

student guide



**HP ProLiant
Systems Technologies**

Rev.3.12A



training

©Copyright 2003 Hewlett-Packard Development Company, L.P.

©Copyright 2003 日本ヒューレット・パッカード株式会社

Microsoft®、Windows®、およびWindows NT®は、米国におけるMicrosoft Corp.の登録商標です。

ここに記載するその他のすべての製品は、それぞれ対応する各社の商標の可能性あります。

日本ヒューレット・パッカード株式会社は、本書の技術的もしくは編集上の間違いや欠落について、一切の責任を負いません。本書の内容は作成時点のものであり、いかなる保証を意味するものでもありません。本書の内容は、将来、予告なしに変更することがあります。日本ヒューレット・パッカード株式会社は、製品保証書で保証する内容以外には、一切の保証はいたしません。本書の内容が、保証期間の延長や保証内容の拡張を意味することは一切ありません。

HP ProLiant Systems Technologies

student guide

Rev.3.12A

2003年8月 第1版

第 1 章 — サーバテクノロジー

目的	1-1
サーバテクノロジー	1-2
HP サーバの設計	1-2
HP サーバのクラス	1-2
汎用サーバ	1-2
スモールビジネスサーバ	1-3
エンタープライズサーバ	1-3
サーバの種類	1-4
サーバ設計時の考慮事項	1-5
ProLiant サーバファミリ	1-6
製品の位置付け	1-7
顧客の環境	1-7
顧客の用途による種類	1-11
世代識別子	1-12
サーバサブシステム	1-13
システムボード	1-14
プロセッサ	1-15
チップセット	1-16
メモリコントローラハブ	1-17
I/O コントローラハブ	1-17
HP 製の ASIC	1-17
バス	1-18
サーババスの種類	1-18
バスコンポーネント	1-19
メモリ	1-21
ストレージコンポーネント	1-22
ドライブアレイ	1-22
ストレージソリューション	1-22
電源サブシステム	1-23
メカニカルコンポーネント	1-25
データフロー	1-26
システムタイミングと速度	1-27
クロック速度	1-27
速度	1-28
設計による性能の向上	1-29
性能を向上させる 3 つの方法	1-29
復習問題	1-30

第 2 章 — プロセッサとマルチプロセッシング

目的	2-1
プロセッサの選択	2-2
プロセッサのしくみ	2-3
命令	2-3
プロセッサのコンポーネント	2-4
命令の処理	2-5
クロック同期	2-6
プロセッサの性能の向上	2-7
パイプラインプロセッサ	2-7
スーパースケーラプロセッサ	2-8
ハイパーパイプライン プロセッサ	2-9
動的実行	2-10
依存問題	2-10
順序外実行	2-11
分岐問題	2-12
分岐予測	2-13
投機的実行	2-14
EPIC	2-15
現在の Intel プロセッサ	2-16
パッケージ形式	2-16
カートリッジ	2-16
Pin Grid Array	2-17
Ball Grid Array	2-18
製造プロセス	2-19
マルチプロセッサ サーバ	2-20
負荷の分散	2-21
メモリとI/Oリソースの共有	2-22
密結合	2-22
疎結合	2-23
その他の要因	2-23
論理プロセッサとハイパースレッディング テクノロジー	2-24
論理プロセッサ	2-24
ハイパースレッディング テクノロジー	2-24
スレッドのスケジューリング	2-25
プロセッサの混用	2-26
テスト	2-26
プロセッサステッピング	2-26
プロセッサの混用に対する Intel のサポート	2-27
プロセッサの混用に対するオペレーティングシステムのサポート	2-27
プロセッサの混用に対する HP のサポート	2-28
復習問題	2-29

第 3 章 — メモリとキャッシュ

目的	3 - 1
メモリとキャッシュ	3 - 2
DRAM (メモリ)	3 - 3
DRAM の欠点	3 - 3
SRAM (キャッシュ)	3 - 4
SRAM の欠点	3 - 4
メモリとキャッシュのコスト比較	3 - 5
メモリ	3 - 6
SIMM と DIMM	3 - 7
メモリのしくみ	3 - 8
メモリチップのしくみ	3 - 9
プロセッサから送信された 1 バイトのデータが格納されるしくみ	3 - 10
メモリ性能の向上	3 - 11
メモリレイテンシ	3 - 11
アクセス時間	3 - 11
サイクル時間	3 - 11
ファストページモード	3 - 12
Extended Data Out DRAM	3 - 12
Synchronous DRAM	3 - 12
Double Data Rate RAM	3 - 13
デュアルインタリーブ メモリ	3 - 14
フォールトトレランス メモリ	3 - 15
ハードエラー	3 - 15
ソフトエラー	3 - 15
パリティ	3 - 16
ECC	3 - 17
アドバンスド ECC	3 - 18
オンラインスペア メモリ	3 - 19
ホットプラグ ミラーメモリ	3 - 20
ホットプラグ RAID メモリ	3 - 22
正しいデータ	3 - 23
不正なデータ	3 - 23
キャッシュとハードディスクドライブ アレイの比較	3 - 24
キャッシュ	3 - 25
キャッシュレベルと配置	3 - 26
キャッシュラインと Tag RAM	3 - 27

キャッシュ性能の向上	3 – 28
L2のキャッシュサイズが性能に及ぼす影響	3 – 28
1 プロセッサシステム	3 – 28
マルチプロセッサシステム	3 – 28
キャッシュアーキテクチャ	3 – 29
ルックアサイド キャッシュ	3 – 29
ルックスルー キャッシュ	3 – 30
キャッシュへの書き込み	3 – 31
ライトスルー キャッシュ	3 – 31
ライトバック キャッシュ	3 – 31
バススヌーピングとバススナッフイング	3 – 32
キャッシュからの読み出し	3 – 33
フルアソシアティブ キャッシュ	3 – 33
ダイレクトマッピング キャッシュ	3 – 33
セットアソシアティブ キャッシュ	3 – 34
復習問題	3 – 35

第 4 章 — バスアーキテクチャ

目的	4 – 1
バスアーキテクチャ	4 – 2
バスの構成要素	4 – 3
バスマルチプレクシング	4 – 4
アドレスバス	4 – 4
アドレスバス幅	4 – 4
データバス	4 – 5
データバス幅	4 – 5
コントロールバス	4 – 6
バスのパフォーマンス	4 – 7
バス幅	4 – 8
バス速度	4 – 8
バスサイクル	4 – 9
最大転送レート	4 – 10
システムパフォーマンスの向上	4 – 11
プロセッサ、バス、およびデバイス速度の高速化	4 – 12
バスの拡張	4 – 13
バーストサイクル	4 – 14
バースト転送レート	4 – 15
並列性の向上	4 – 16
バスマスタ	4 – 16
バスアービトレーション	4 – 16
ISA バス	4 – 17
ISA バステクノロジ	4 – 17
EISA バス	4 – 18
EISA バステクノロジ	4 – 18

PCI バス	4 - 19
完全に独立したバス	4 - 20
プラグ&プレイアーキテクチャ	4 - 20
IRQ 共有	4 - 20
マルチプレクシング	4 - 20
デバイスのロード	4 - 21
PCI バスでの転送	4 - 21
カードスロットのサポート	4 - 21
32/64 ビット 33/66MHz カード	4 - 22
66MHz で動作する 64 ビット PCI	4 - 22
32/64 ビット 33/66MHz カードのルール	4 - 22
複数の PCI バスシステム	4 - 23
PCI バスでのデータ転送	4 - 24
PCI バスのパフォーマンスプラン	4 - 25
PCI バス周波数の調整	4 - 27
PCI バス番号の割り当て	4 - 27
PCI ホットプラグのサポート	4 - 28
PCI-X バス	4 - 29
PCI-X のパフォーマンス	4 - 31
PCI-X の階層型構造	4 - 32
PCI-X システムの柔軟性	4 - 33
PCI と PCI-X の比較	4 - 34
PCI プロトコル	4 - 34
PCI-X のレジスタツリーレジスタ プロトコル	4 - 35
PCI-X のタイミングの違い	4 - 36
プロトコルの拡張	4 - 37
アダプタカードの選択	4 - 39
相互運用性マトリクス	4 - 40
PCI-X での PCI ホットプラグのサポート	4 - 41
ハードウェア関連の影響	4 - 41
ソフトウェア関連の影響	4 - 41
PCI-X 2.0	4 - 42
InfiniBand	4 - 43
InfiniBand テクノロジ	4 - 44
PCI Express	4 - 45
USB	4 - 47
USB コネクタ	4 - 48
シリーズ A	4 - 48
シリーズ B	4 - 48
復習問題	4 - 49

第 5 章 — システムアーキテクチャ

目的	5 - 1
チップセット	5 - 2
チップセット開発におけるパートナーシップ	5 - 2
サーバアーキテクチャの進化	5 - 3
初期のアーキテクチャ	5 - 3
並列バス	5 - 4
システム並列性の拡大	5 - 5
システム並列性の拡大—MIOC	5 - 6
並列 I/O バスによる I/O 並列性の拡大	5 - 7
デュアル メモリコントローラ	5 - 8
クロスバースイッチ	5 - 9
改良型クロスバースイッチ	5 - 9
ServerWorks チップセット	5 - 10
ServerSet I	5 - 11
ServerSet II LE チップセット	5 - 11
ServerSet II HE チップセット	5 - 11
ServerSet III LE チップセット—シングル ノースブリッジ	5 - 11
ServerSet III LE チップセット—デュアルノースブリッジ	5 - 11
ServerSet III LE LP チップセット	5 - 12
ServerSet III HE SL チップセット	5 - 12
ServerSet GC SL チップセット	5 - 12
ServerSet GC LE チップセット	5 - 12
ServerSet GC HE チップセット	5 - 12
ハイリーパラレル システムアーキテクチャ	5 - 13
ProFusion チップセット	5 - 14
ProFusion クロスバースイッチ	5 - 15
F8 チップセットのアーキテクチャ	5 - 16
F8 チップセットの利点	5 - 16
F8 の高い帯域幅容量	5 - 17
復習問題	5 - 19

第 6 章 — ディスクストレージ

目的	6 - 1
ディスクストレージの概要	6 - 2
ATA/IDE	6 - 2
SCSI	6 - 2
ATA/IDE ハードドライブ技術	6 - 3
サーバにおける ATA/IDE	6 - 3
ATA/IDE 標準	6 - 4
ATAPI (ATA Packet Interface)	6 - 5
ATA RAID	6 - 5
新しい技術—シリアル ATA	6 - 6
シリアル ATA II	6 - 6
SCSI とパラレル ATA/IDE	6 - 7
SCSI の概要	6 - 8
SCSI の基礎知識	6 - 9

SCSI コマンドセット.....	6 – 10
SCSI 標準	6 – 11
SCSI-1	6 – 11
SCSI-2	6 – 12
SCSI-3	6 – 13
パラレル SCSI	6 – 18
シリアル SCSI	6 – 19
パラレル SCSI とシリアル SCSI の比較.....	6 – 20
SCSI 標準のまとめ	6 – 21
SCSI の転送モード.....	6 – 22
SCSI プロトコルの互換性	6 – 23
SCSI アドレッシング	6 – 24
SCSI のデバイス ID に関するガイドライン	6 – 26
論理ユニット番号 (LUN)	6 – 27
SCSI の通信方法	6 – 28
SCSI バスのトランザクション	6 – 29
ネゴシエーション	6 – 30
切断と再接続.....	6 – 30
タグ付きコマンドキューイング (TCQ)	6 – 31
SCSI バスフェーズ	6 – 32
SCSI コネクタ	6 – 33
ケーブル	6 – 34
電氣的信号方式.....	6 – 35
シングルエンド SCSI	6 – 36
ディファレンシャル SCSI	6 – 38
ケーブルについての留意事項	6 – 40
ターミネータ	6 – 41
ターミネータ電源.....	6 – 42
HP による SCSI の実装方法	6 – 43
HP による SCSI コントローラチップ	6 – 43
HP による SCSI ストレージソリューション	6 – 44
HP の SCSI アダプタ	6 – 45
HP の SCSI ドライブ	6 – 46
ユニバーサルディスクドライブ	6 – 46
ホットプラグ SCSI ドライブ	6 – 47
iSCSI.....	6 – 48
SCSI のトラブルシューティング	6 – 49
復習問題	6 – 50

第7章 — アレイテクノロジー

目的	7-1
RAID テクノロジーの概要	7-2
RAID の利点	7-2
JBOD	7-3
ドライブアレイ	7-3
アレイのディスク数	7-3
論理ドライブ	7-4
ドライブアレイ機能	7-5
RAID の実装方法	7-6
平均故障間隔	7-8
MTBF の最大化	7-8
RAID レベル	7-9
ストライピングファクタ	7-10
RAID 0 — ディスクストライプ	7-11
RAID 1 — ディスクミラーリング	7-12
ディスクミラーリング	7-12
デュプレキシング (二重化)	7-13
RAID 1 のパフォーマンス	7-14
RAID 1+0 — ストライプミラーセット	7-15
RAID 1+0 のパフォーマンス	7-16
RAID 2 — ECC によるデータセクタ ストライピング	7-17
RAID 3 — パリティチェック	7-17
RAID 4 — データガーディング	7-18
制限事項	7-18
RAID 5 — 分散データガーディング	7-19
RAID 5 のパフォーマンス	7-20
RAID ADG — アドバンスド データガーディング	7-22
RAID ADG のパフォーマンス	7-22
RAID ADG の用途	7-23
Smart アレイコントローラの ADG サポート	7-23
RAID ADG のフォールトトレランスの比較	7-24
RAID レベルのパフォーマンスの比較	7-25
RAID 0	7-25
RAID 1 および RAID 1+0	7-25
RAID 4	7-25
RAID 5	7-26
ADG	7-26
RAID での障害ドライブの復旧時間	7-27
ストライプサイズの最適化	7-28

Smart アレイコントローラ.....	7 - 29
Smart アレイコントローラの命名方法.....	7 - 30
アレイコントローラの比較.....	7 - 31
SMART-2 ファミリー.....	7 - 31
エントリレベル.....	7 - 31
ハイエンド.....	7 - 32
エントリレベルの Smart アレイコントローラ.....	7 - 33
Smart アレイ 532 コントローラ.....	7 - 33
内蔵 Smart アレイコントローラ.....	7 - 33
シングルチャネル RAID (RAID LC2) コントローラ.....	7 - 33
高性能な Smart アレイコントローラ.....	7 - 34
Smart アレイ 5300 コントローラ.....	7 - 34
Smart アレイ 4250ES コントローラ.....	7 - 35
Smart アレイコントローラについての留意事項.....	7 - 36
アレイコントローラテクノロジーとユーティリティ.....	7 - 38
アレイコンフィギュレーションユーティリティ (ACU).....	7 - 39
ACU を使ったドライブアレイ容量の拡張.....	7 - 40
アレイコンフィギュレーション データの位置.....	7 - 40
サポートされているコントローラ.....	7 - 40
アレイコンフィギュレーションユーティリティ (ACU-XE).....	7 - 41
Option ROM Configuration for Arrays (ORCA).....	7 - 42
さまざまな構成方法.....	7 - 42
Array Diagnostics Utility.....	7 - 43
オンラインスペアドライブ.....	7 - 44
オンラインでのドライブアレイ拡張.....	7 - 46
論理ドライブ容量の拡張.....	7 - 47
オンラインでの RAID レベルの移行.....	7 - 48
オンラインでのストライプサイズの移行.....	7 - 49
ハードドライブ障害予測テクノロジー.....	7 - 50
DSR (Dynamic Sector Repair).....	7 - 52
ホットプラグドライブのサポート.....	7 - 53
ホットプラグ可能なドライブの LED.....	7 - 53
Automatic Data Recovery.....	7 - 53
アレイアクセラレータ (読み取り/書き込みキャッシュ).....	7 - 54
アレイのパフォーマンス調整.....	7 - 54
データ保護.....	7 - 55
高度なエラーチェック.....	7 - 55
バッテリーのバックアップ.....	7 - 55
リカバリ ROM.....	7 - 55
復習問題.....	7 - 56

第 8 章 — ファイバチャンネル

目的	8 - 1
ファイバチャンネルの概要.....	8 - 2
ファイバチャンネルの利点	8 - 3
ファイバチャンネルで使われる用語	8 - 5
ファイバチャンネルの構成要素.....	8 - 6
ホストバスアダプタ	8 - 7
GBIC (ギガビット インタフェースコンバータ)	8 - 8
GBIC のしくみ	8 - 9
GBIC-SW	8 - 10
GBIC-LW	8 - 10
GLM (ギガビット リンクモジュール)	8 - 11
ストレージアレイとドライブアレイ.....	8 - 12
ファイバチャンネルアレイコントローラ	8 - 13
ファイバチャンネルハブ	8 - 14
ファイバチャンネルスイッチ	8 - 15
ハブとスイッチの比較	8 - 19
ファイバチャンネルケーブル	8 - 21
光ファイバ	8 - 22
マルチモードステップインデックス ファイバ.....	8 - 25
マルチモードグレーデッドインデックス ファイバ.....	8 - 26
安全上の注意	8 - 27
減衰	8 - 28
マクロバンド.....	8 - 29
マイクロバンド	8 - 29
ファイバチャンネルトポロジ.....	8 - 30
ポイントツーポイントトポロジ	8 - 31
ファイバチャンネル アービトレーテッドループトポロジ	8 - 32
スイッチファブリックトポロジ	8 - 33
新しいテクノロジー	8 - 35
ATM (Asynchronous Transfer Mode)	8 - 35
IP and ARP over Fibre Channel.....	8 - 36
復習問題	8 - 37

第 9 章 — 高度なストレージテクノロジー

目的	9 - 1
データストレージテクノロジー	9 - 2
Direct Attached Storage (DAS)	9 - 3
Network Attached Storage (NAS)	9 - 4
NAS の利点.....	9 - 4
NAS の欠点.....	9 - 4
Storage Area Network (SAN)	9 - 5
SAN の利点.....	9 - 6
NAS と SAN の比較	9 - 6
同機種 SAN	9 - 7
エンタープライズ (異機種混在) SAN	9 - 8
Open SAN	9 - 9
SAN ベースの高可用性 - ビジネスの継続性	9 - 10
SAN アーキテクチャのカテゴリ	9 - 10
HP 標準 SAN トポロジ	9 - 12
カスケードファブリック	9 - 13
メッシュファブリック	9 - 14
リングファブリック	9 - 15
バックボーン SAN ファブリック	9 - 16
Enterprise Network Storage Architecture (ENSA)	9 - 17
Open SAN をサポートするソリューション	9 - 18
ストレージバーチャライゼーション	9 - 18
動的スケールビリティ	9 - 19
簡素化された管理	9 - 19
VersaStor バーチャライゼーション	9 - 20
Virtual Replicator	9 - 21
Enterprise Virtual Array (EVA)	9 - 23
EVA のパフォーマンス	9 - 24
CASA (OpenView continuous access storage appliance)	9 - 25
ENSAextended VersaStor テクノロジ	9 - 26
VersaStor Vector	9 - 26
Virtual SAN Executor	9 - 26
Vraid ソフトウェア	9 - 27
Vsnap	9 - 28
スナップクローン	9 - 28
HSV Element Manager	9 - 28
テープバックアップテクノロジー	9 - 29
セカンダリストレージとバックアップの重要性	9 - 30
Hierarchical Storage Management (HSM)	9 - 31
バックアップ方法	9 - 32
計画とサイズ	9 - 32
速度	9 - 33
バックアップテープのローテーション方式	9 - 34
フルバックアップ	9 - 34
部分バックアップ	9 - 35
Grandfather-Father-Son (GFS) テープローテーション	9 - 36

テープ媒体の平均寿命	9 – 38
テープストリーミング	9 – 39
テープドライブのクリーニング間隔	9 – 40
バックアップ媒体のオフサイトでの保管	9 – 40
テープストレージテクノロジー	9 – 41
能力	9 – 42
ニアラインテクノロジー	9 – 43
クローン化とスナップショット	9 – 44
クローン化	9 – 44
スナップショット	9 – 44
バックアップソリューション	9 – 45
エンタープライズバックアップソリューション (EBS)	9 – 46
EBS の利点	9 – 46
災害復旧計画の準備	9 – 47
災害耐性と復旧	9 – 48
災害耐性	9 – 48
災害復旧	9 – 48
データリпликаーションマネージャ (DRM)	9 – 48
復習問題	9 – 49

第 10 章 — 高可用性

目的	10 – 1
はじめに.....	10 – 2
可用性の要件の比較.....	10 – 2
クリティカルアプリケーションのクラス.....	10 – 3
ミッションクリティカル	10 – 3
ビジネスクリティカル	10 – 3
可用性レベルによるサーバの分類.....	10 – 4
可用性レベル	10 – 5
可用性の割合.....	10 – 5
平均故障間隔	10 – 6
高可用性の設計	10 – 7
可用性ドメイン.....	10 – 7
ダウンタイムの原因.....	10 – 8
ProLiant サーバでの高可用性の実現.....	10 – 9
障害予防	10 – 10
フォールトトレランス.....	10 – 11
ラピッドリカバリ.....	10 – 12
ラピッドリカバリ	10 – 13
クラスタ化.....	10 – 14
クラスタ化の利点	10 – 15
リソースの高可用性.....	10 – 15
ビジネスの成長に対応したスケーラビリティ.....	10 – 16
集中管理.....	10 – 16
負荷分散.....	10 – 17
クラスタ対応アプリケーション	10 – 17
クラスタ要素.....	10 – 18
クラスタテクノロジー.....	10 – 19
用語.....	10 – 19
クラスタモデル.....	10 – 20
シェアードナッシング (Shared nothing)	10 – 20
シェアードディスク (Shared disk)	10 – 21
シェアードエブリッシング (Shared everything)	10 – 21
HP クラスタソリューション	10 – 22
ProLiant クラスタ	10 – 23
DL380 パッケージクラスタ	10 – 23
Smart アレイクラスタ	10 – 24
MSA1000 用の ProLiant Cluster HA/F100、HA/F200.....	10 – 25
Enterprise Virtual Array (EVA) 用の ProLiant Cluster HA/F500.....	10 – 25
HP ハイエンドクラスタソリューション.....	10 – 25
HP クラスタの管理.....	10 – 26
Insight マネージャ 7 の Cluster Monitor.....	10 – 26
復習問題	10 – 27

第 11 章 — ProLiant サーバのユーティリティ

目的	11 – 1
ProLiant Essentials	11 – 2
Foundation パック	11 – 3
SmartStart	11 – 4
SmartStart の機能	11 – 5
SmartStart6.00 の機能	11 – 6
SmartStart 5.5 と SmartStart 6.x の比較	11 – 8
SmartStart 5.5 でサポートされるサーバ	11 – 8
SmartStart 6.x でサポートされるサーバ	11 – 9
サポートしているオペレーティングシステム	11 – 10
ハードウェアの最小要件	11 – 10
SmartStart 5.5 のインストールパス	11 – 11
ユーティリティへのアクセス	11 – 13
システムコンフィギュレーションユーティリティ (SCU)	11 – 14
ROM ベースセットアップユーティリティ (RBSU)	11 – 20
RBSU の起動と終了	11 – 21
アレイコンフィギュレーションユーティリティ (ACU)	11 – 22
アレイコンフィギュレーションユーティリティ XE (ACU-XE)	11 – 24
Array Diagnostics Utility (ADU)	11 – 25
システムパーティションアップグレードユーティリティ	11 – 26
ディスクビルダ	11 – 27
System Erase ユーティリティ	11 – 28
Fibre Channel Diagnostics と Fibre Channel Fault Isolation Utility	11 – 28
ROMPaq Utility	11 – 29
ProLiant Support Pack (PSP)	11 – 30
リモート展開ユーティリティ (RDU)	11 – 30
リモート展開コンソールユーティリティ	11 – 30
サーバの展開	11 – 31
SmartStart Scripting Toolkit	11 – 32
Rapid Deployment パック (RDP)	11 – 33
Workload Management パック	11 – 34
Resource Partitioning Manager	11 – 34
Management CD	11 – 35
Insight Management Suite	11 – 36
Insight マネージャ 7	11 – 37
機能	11 – 38
マネジメントエージェント	11 – 40
ActiveUpdate	11 – 41
Survey ユーティリティ	11 – 43
リモート Insight Lights-Out テクノロジー	11 – 44
RILOE の機能	11 – 45
内蔵 Lights-Out (iLO)	11 – 46
iLO Standard	11 – 46
iLO Advanced	11 – 48
iLO Advanced 機能の有効化	11 – 48
復習問題	11 – 49

第 12 章 — ラック/電源テクノロジー

目的	12 - 1
HP ラック/電源テクノロジー	12 - 2
ラック 10000 シリーズ	12 - 3
オプション	12 - 4
rack builder online.....	12 - 5
追加機能	12 - 6
HP ラックの保守に関する注意事項.....	12 - 7
コンソールマネジメントコントローラ (CMC)	12 - 8
インテリジェント ラックマネージャ Lite	12 - 9
タワーUPS モデル	12 - 10
ラックマウント UPS モデル	12 - 12
すべての UPS の保守に関する注意事項.....	12 - 14
パワーディストリビューションユニット (PDU)	12 - 15
サーバコンソールスイッチ	12 - 16
PCI カード KVM スイッチ	12 - 19
ラックアクセサリ.....	12 - 20
ケーブルマネジメントアーム.....	12 - 22
パワーマネジメントソフトウェア	12 - 23
マルチサーバ UPS カード	12 - 23
復習問題	12 - 24

第 13 章 — サポートサービス

目的	13 - 1
ドキュメント、オンラインなどのサポートサービス.....	13 - 2
シリアル番号	13 - 3
コンフィギュレーションコード	13 - 3
シリアル番号の詳細 (1998 年以前)	13 - 4
シリアル番号の詳細 (1998 年から 2003 年 2 月まで)	13 - 5
シリアル番号の詳細 (2003 年 3 月以降)	13 - 6
パーツ番号	13 - 7
Maintenance and Service Guides (MSG)	13 - 8
アクセスラベル	13 - 9
HP サポートリソース Web サイト.....	13 - 10
ソフトウェアとドライバ.....	13 - 11
Reference Library	13 - 12
ActiveAnswers	13 - 13
HP の標準保証	13 - 14
復習問題	13 - 15

付録 A — テクノロジに関する付録

第 2 章のトピック	A - 1
RISC 命令と CISC 命令	A - 1
命令セットと拡張	A - 2
MMX	A - 2
SSE	A - 2
SSE2	A - 2
パイプラインハザードの例	A - 3
分岐の問題	A - 4
依存の問題	A - 4
パイプラインハザードの影響	A - 4
VLIW	A - 5
ProLiant サーバで採用されているプロセッサ	A - 6
Pentium III プロセッサ	A - 6
Pentium III Xeon プロセッサ	A - 7
Pentium 4、Xeon、XeonMP プロセッサ	A - 7
Windows ベースのサーバと論理プロセッサ	A - 8
第 3 章のトピック	A - 9
セルのデータ格納方法	A - 9
セルのデータ読み取り方法	A - 9
メモリコントローラのデータビットの格納方法	A - 10
メモリコントローラのデータビットの読み取り方法	A - 11
データのリフレッシュ	A - 12
メモリチップサイズ	A - 13
CAS レイテンシ	A - 14
第 4 章のトピック	A - 15
AGP	A - 15
サポートするオペレーティング システム	A - 15
ダイレクト メモリ アクセス	A - 16
第 6 章のトピック	A - 17
ATA 規格	A - 17
ATA-1	A - 17
ATA-2	A - 17
ATA-3	A - 17
ATA-4 (Ultra-ATA/33)	A - 18
ATA-5 (Ultra-ATA/66)	A - 18
ATA-6 (Ultra-ATA/100)	A - 18
ATA-7 (Ultra-ATA/133)	A - 18
ATA の動作モードと速度	A - 19
PIO (Programmed I/O: プログラム I/O)	A - 19
DMA (Direct Memory Access: ダイレクトメモリアクセス) モード	A - 20
ATA 変換	A - 22
Standard CHS (標準シリンダ/ヘッド/セクタ)	A - 22
Extended CHS (拡張シリンダ/ヘッド/セクタ)	A - 22
LBA (Logical Block Addressing: 論理ブロックアドレッシング)	A - 22
ドライブ 0 とドライブ 1 の関係	A - 23
ケーブルセレクト	A - 23
ドライブセレクト	A - 23

付録 B — 復習問題の解答

第 1 章 – サーバテクノロジー	B – 1
第 2 章 – プロセッサとマルチプロセッシング	B – 2
第 3 章 – メモリとキャッシュ	B – 4
第 4 章 – バスアーキテクチャ	B – 7
第 5 章 – システムアーキテクチャ	B – 8
第 6 章 – ディスクストレージ	B – 9
第 7 章 – アレイテクノロジー	B – 10
第 8 章 – ファイバチャネル	B – 11
第 9 章 – 高度なストレージテクノロジー	B – 12
第 10 章 – 高可用性	B – 14
第 11 章 – ProLiant サーバのユーティリティ	B – 15
第 12 章 – ラック/電源テクノロジー	B – 17
第 13 章 – サポートサービス	B – 17

目的

この章では、次の内容について学習します。

- HP のサーバ設計テクノロジーおよび考慮事項
- ProLiant サーバファミリ
- ProLiant サーバ用の構成およびセットアップツール
- 次のサーバサブシステムの機能
 - システムボード
 - プロセッサ
 - チップセット
 - バス
 - メモリ
 - ストレージコンポーネント
 - 電源サブシステム
 - メカニカルコンポーネント
- サーバの性能を向上させる3つの方法

サーバテクノロジー

HP の革新的な技術は、自社サーバの品質、信頼性、保守性、性能を向上させると同時に、TCOも削減します。HP は後に業界標準として採用されることになった数々の新しいテクノロジーを開発してきました。ProLiant サーバに代表されるこれらのテクノロジーは、企業が競争の激しい環境に適応し、限られたリソースを活用し、顧客のニーズに応えることを可能にします。

HP サーバの設計

HP サーバと他社のサーバとの違いを理解するためには、サーバの基本的なアーキテクチャ設計を理解する必要があります。HP サーバは次のような多くの要素を考慮に入れて設計されています。

- 変化する顧客のニーズ
- 新しいテクノロジー
- 新しい標準

HP サーバのクラス

サーバはさまざまなタスクを実行しますが、それには特定のニーズに応じたサーバの構成が必要になります。サーバ設計の第一歩は、明確な顧客の要求セットを定義することです。サーバは、市場の要求に応え、投資を最大限に回収できるものでなければなりません。

HP は、業界のあらゆる要求に応えることができる広範な製品を提供しています。

汎用サーバ

汎用サーバとは、さまざまな機能を実装できるサーバプラットフォームです。企業規模にかかわらず、汎用サーバはどのような環境でも使用できます。また、部門内での使用から企業全体のクラスタアプリケーションの管理まで、広範な目的で使用されます。

スモールビジネスサーバ

スモールビジネスサーバプラットフォームには、使い易さや構築のし易さを向上させるための機能が組み込まれています。このプラットフォームは、限られた環境におけるすべての典型的なサーバタスクを処理するための多様な特徴を持ち、ファイル、印刷、およびクライアント/サーバアプリケーションの要求を処理します。また、業務にただちに影響するダウンタイムから迅速に復旧するための機能も備えています。

エンタープライズサーバ

エンタープライズサーバはどのような環境でも使用できますが、一般的には大企業の基幹として、次のような特定の用途で使用されます。

- 電子メールの処理と配信
- 財務データの格納
- データベースの管理

このサーバは企業の業務にとって必須であるため、高い可用性を保証する必要があります。したがって、予期せぬダウンタイムをなくすためのソフトウェアおよびハードウェアコンポーネントが必要になります。

サーバの種類

ビジネスコンピューティングには特定の種類のサーバが必要になります。HP のエンジニアは、顧客およびソフトウェアパートナーと緊密に協力しながら、サーバの設計、統合、テストを行っています。このため、HP は、ビジネスコンピューティングのニーズに的確に応えることができる総合的なサーバおよびソリューションを提供することができます。

サーバの種類およびその用途には次のようなものがあります。

- **アプリケーションサーバ** — クライアントにアプリケーションへのアクセスを提供する。
- **ブリッジサーバ** — ネットワークトラフィックをフィルター処理し、転送する。
- **データベースサーバ** — データの格納、取り出し、管理を行う。
- **ドメインネームサーバ** — ホスト名を IP アドレスに解決する。
- **DHCP (Dynamic Host Configuration Protocol) サーバ** — ネットワーク上のクライアントに IP アドレスを割り当てる。
- **ファックスサーバ** — クライアントワークステーションからのファックスの送受信と管理を行う。
- **ファイルサーバ** — クライアント用のファイルを格納する。
- **ファイアウォールサーバ** — 受信および送信ネットワークトラフィックを監視し、不正なアクセスを防ぐ。
- **FTP (File transfer protocol) サーバ** — ファイルをリモートからダウンロード、アップロードする。
- **ゲートウェイサーバ** — 異種ネットワークを接続する。
- **メールサーバ** — 受信および送信メールを管理する。
- **プリントサーバ** — クライアントからの印刷ジョブを管理する。
- **プロキシサーバ** — 送信されるネットワーク要求をフィルター処理する。
- **リモートアクセスサーバ** — ユーザがリモートからネットワークにアクセスすることを可能にする。
- **ルータサーバ** — ネットワーク間のトラフィックを管理する。
- **Web サーバ** — Web ページを配信する。
- **WINS (Windows Internet Name Service) サーバ** — ネットワーク名を解決する。

▲ 注記

上記のリストはすべてのサーバおよびその機能を網羅しているわけではありません。このリストは、さまざまな種類のサーバが市場に出回っていることを示しています。

サーバ設計時の考慮事項

ProLiant サーバの設計は、性能およびその価値において業界をリードしています。HP は複数のサブシステムで ProLiant サーバを構築することにより、高い性能を実現しながら価格は抑えた、バランスの取れたコンピュータアーキテクチャを提供しています。

- **TCO** — あらゆるタイプの顧客に適した車種がないように、あらゆるタイプのビジネスに適した単一のサーバモデルまたは構成は存在しません。HP は、1 つのサイズ、価格、または性能レベルのサーバではすべての顧客の要求を満たせないため、複数のサーバを提供しています。
- **性能** — サーバシステムの性能は、バランスの取れた設計であるかどうか、つまりコンピュータの帯域幅がメモリの帯域幅と一致し、I/O 帯域幅がソフトウェアの要求に一致しているかどうかにかかっています。命令やデータ転送に対するプロセッサ要求よりもメモリが大幅に遅い場合、プロセッサはメモリ不足になり、性能をフルに活用することはできません。逆に、メモリシステムが高速でプロセッサが低速の場合は、プロセッサがボトルネックとなります。
- **信頼性** — サーバシステムは途切れることなく動作し続けることができるように、フォールトトレランス機能および故障した部品をスムーズに交換するための機能を備えている必要があります。高可用性機能の実装方法は、サーバがどのように使用されているかによって異なりますが、どのようなサーバにとっても非常に重要な機能です。また、要求される信頼性および高可用性のレベルは、サーバのクラスごとに異なります。
- **容量** — 多くのサーバアプリケーションでは、スケーラビリティとサービスは非常に重要です。RAM の拡張機能やストレージ容量など、コンポーネントに関する要件は、サーバのスケーラビリティの問題を解決するものとみなされています。
- **使用、保守の容易性** — サーバの TCO を大きく左右する 2 つの要因は、使用の容易性と保守の容易性です。これらの要因に影響する設計基準には、次のものがあります。
 - メカニカルコンポーネント
 - 放熱設計
 - 点検のし易さ
- **セキュリティ** — 企業によっては、ユーザデータおよびシステムコンポーネントへのアクセスを保護しなければならない場合があります。システム設計者は、セキュリティ要件およびハードウェア設計上の推奨事項を考慮する際には、常にコストと性能を秤にかける必要があります。

ProLiant サーバファミリ



ProLiant 製品ライン

HP は、顧客およびソフトウェアパートナーと緊密に協力しながら、サーバの設計、統合、テストを行っています。豊富な専門技術をもとに、HP は次のような総合的なサーバ製品を設計、開発しています。

- フロントエンドサーバミッドティアサーバ
- バックエンドサーバ

製品の位置付け

ProLiant ファミリの位置付けは、次の 2 つの基準に基づいて行われています。

- **顧客の環境** — モデル名の先頭の ML、DL、または BL で示されます。
- **顧客の用途** — 300/500/700 シリーズ番号で示されます。

顧客の環境

顧客の環境に応じた ProLiant サーバファミリの位置付けは、モデル名の先頭の ML、DL、BL で示されています。

- **ML サーバ** — システム内部の拡張性が最大化されています。
- **DL サーバ** — ラックマウント環境用に、集積度が最適化されています。
- **BL サーバ** — モジュール化された超高密度ブレードサーバ。

ML ライン



ProLiant ML530 Generation 3

ProLiant ML ラインでは、システムを最大限に内部拡張できます。ML シリーズは、サーバシャーシ内の拡張スロット数を増やし、ドライブ拡張機能を向上させることにより、柔軟な拡張を可能にしています。また、これらのサーバは最大限のアップタイムを保証するための高可用性機能も備えています。このため、顧客は幅広い価格帯と性能レベルから柔軟にサーバを選択することができます。

ML サーバはラック型とタワー型の両方のモデルで提供されています。柔軟性が高く、さまざまな構成が可能な ML ラインの設計は、リモート、インフラストラクチャ、データセンター環境など、広範な用途モデルをサポートしています。

これらのサーバは支社やリモートオフィスに最適で、サーバおよびストレージソリューションをすべて 1 つの製品で提供します。

DL ライン



ProLiant DL360 G3

ProLiant DL ラインは、設置スペースに制約がある環境やラックマウント環境向けに集積度が最適化されたサーバです。DL シリーズは内蔵コンポーネントの活用により、サーバの機能セットを増やししながら、シャーシサイズを抑えています。

データセンタや外部ストレージ環境向けのこれらのサーバは、マルチサーバ環境に最適です。DL ラインは、ラック最適化型でのみ提供されており、次のような特長があります。

- 小型の省スペース設計で最大限のコンピューティング能力を実現
- 多くの内蔵オプションを統合
- 効率的なクラスタリング機能

BL ライン



ProLiant BL e-Class ブレードエンクロージャ



ProLiant BL p-Class システム

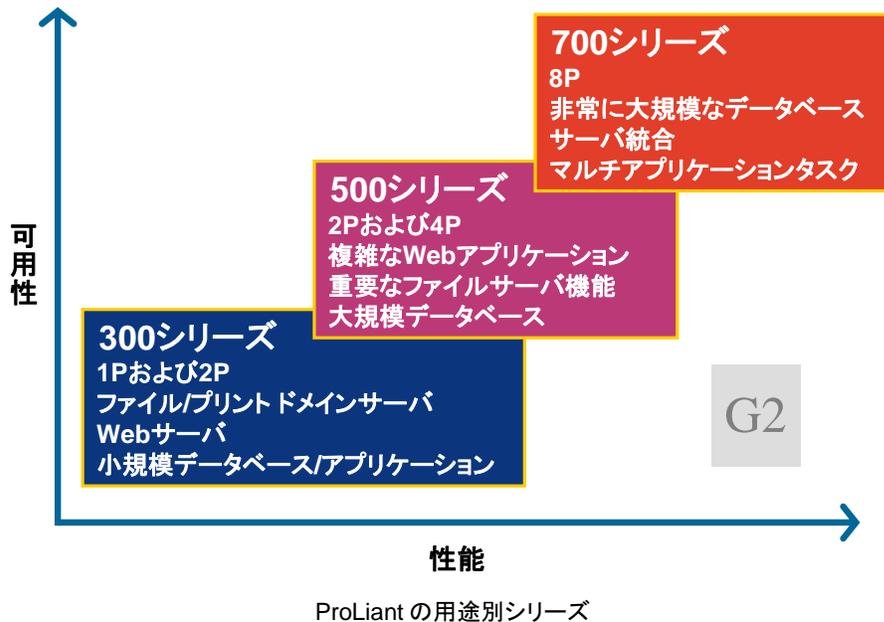
ProLiant BL ラインは、消費電力を抑えたモジュール式のサーバです。ProLiant のブレードアーキテクチャでは道具を使わずに個々のモジュールを取り付けられるため、サーバおよびインフラストラクチャを迅速に配備することができ、保守も容易です。

BL ラインは、企業やサービスプロバイダが、変化する市場の要求に適応し、限られた人的および技術的リソースを活用し、効率を最大限に高め、動的なビジネス環境において顧客の要求に迅速に応えることを支援する目的で設計されています。

ProLiant BL ラインには 2 つの製品クラスがあります。

- **e-Class** — エンタープライズクラスのチップセット、超低電圧プロセッサ、およびその他の省電力コンポーネントを超高密度設計に統合することにより消費電力を抑えた、フロントエンド向けのサーバです。標準ラックに最大 280 台 (42U ラックの場合) の低消費電力エッジサーバブレードを搭載でき、データセンタのスペースを最大限に活用することを可能にします。
- **p-Class** — 複数階層環境向けの高性能、高可用性サーバブレードを提供します。p-Class インフラストラクチャの電源は集中化および二重化されており、またその他の優れた管理機能も備えています。e-Class と比較した場合、p-Class はより低い密度でより高い性能を提供します。

顧客の用途による種類



各 ProLiant サーバラインは、3つのシリーズに分かれています。各シリーズは、性能レベルとサーバの可用性によって定義されます。

300 シリーズ

300 シリーズのサーバは、次のような用途に最適です。

- ファイル/プリントおよびドメインサーバ
- Web サーバ機能
- 小規模のデータベースやアプリケーション
- インフラストラクチャ アプリケーション

500 シリーズ

500 シリーズのサーバは、さらに高い可用性および性能を提供します。

- 複雑な Web アプリケーション
- 大規模データベース
- 重要なファイルサーバ アプリケーション

700 シリーズ

700 シリーズのサーバは、最大限の性能と可用性を提供し、次のような用途に適しています。

- 非常に大規模なデータベース
- サーバ統合ソリューション
- マルチアプリケーション環境

世代識別子



ProLiant サーバの世代が代わる際には、サーバの世代を識別するための手段が必要になります。技術資料、オプション、部品などを特定するために、第 2 世代以降のサーバには、1 cm² のラベルが貼られます。

このラベルは、次に示す一定の個所に貼られています。

- **ラック型** — Intel ロゴの反対側にある左のラックねじ内
- **タワー型** — 左上シャーシのドア (ベゼル) の裏

世代によって仕様が異なる場合には、技術文書内でも世代識別子が使用されます。データベースフィールド内で使用可能なスペースなどの制約の有無に応じて、世代識別子は次のような形式で使用されます。

- ProLiant DL360 Generation 2 サーバ
- ProLiant DL360 G2

世代識別子は、カタログや写真などの特定のマーケティング資料では使用されていない場合もあります。これまでに作成された文書に付与し直すことはありませんが、今後登場するサーバの世代については、文書内でも世代識別子を使用する予定です。

サーバサブシステム

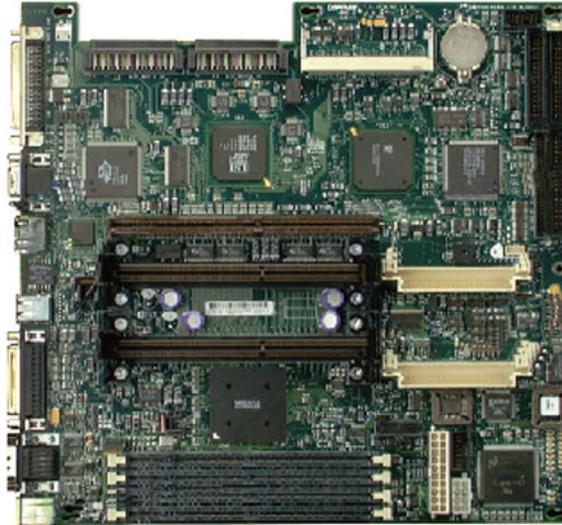


サーバを構築するには、要求されているサーバ設計要件を満たし、最高レベルの品質を提供するアーキテクチャにシステムコンポーネントを統合する必要があります。HPは、複数のサブシステムを使って各 ProLiant サーバを設計することにより、高い性能を実現しながら価格を抑えた、バランスの取れたコンピュータアーキテクチャを提供しています。HP サーバのアーキテクチャ設計は、設計、製造プロセスを通して厳しくテストされています。

このコースでは、次のサブシステムについて解説します。

- システムボード(マザーボード、バックプレーンボード)
- プロセッサ(プロセッサとマイクロプロセッサ)
- チップセット(ASIC)
- バス(I/O 相互接続)
- メモリ
- ストレージコンポーネント
- 電源
- メカニカル(サーバシャーシ、コンポーネントシャーシ)
- ソフトウェア

システムボード

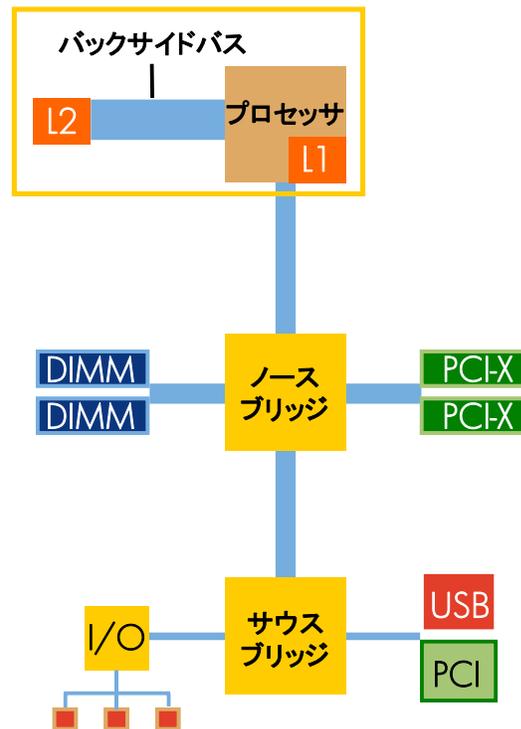


システムボードは、処理コンポーネントの大半が搭載された集積回路ボードです。システムボードには、次の3種類があります。

- **アクティブ** — 従来型のシステムボード(マザーボード)。アクティブボードは、アクティブなチップセットを内蔵しています。
- **パッシブ** — このボードはチップセットを内蔵していませんが、複数の集積回路コンポーネントボードを接続する相互接続ボードとして機能します。
- **モジュール式** — このシステムボードは、アクティブとパッシブの混合型です。いくつかのチップセットを内蔵しており、アップグレード可能な集積回路コンポーネントボード用の相互接続ボードとして機能します。

ProLiant サーバファミリでは、アクティブとモジュール式の両方のシステムボードが採用されています。アクティブボードを使用した設計は、安価なソリューションとして、エントリーレベルのサーバおよび集積度が最適化されたサーバに適しています。一方、モジュール式ボードを使用した設計では、サーバ内で最大限のスケラビリティとコンポーネントのアップグレードが可能です。

プロセッサ



プロセッサはサーバの中枢部であり、サーバコンポーネント間のすべてのアクティビティを制御します。プロセッサは、メモリやバスなど、1つの集積回路上に実装された、連携する複数のコンポーネントから構成されています。

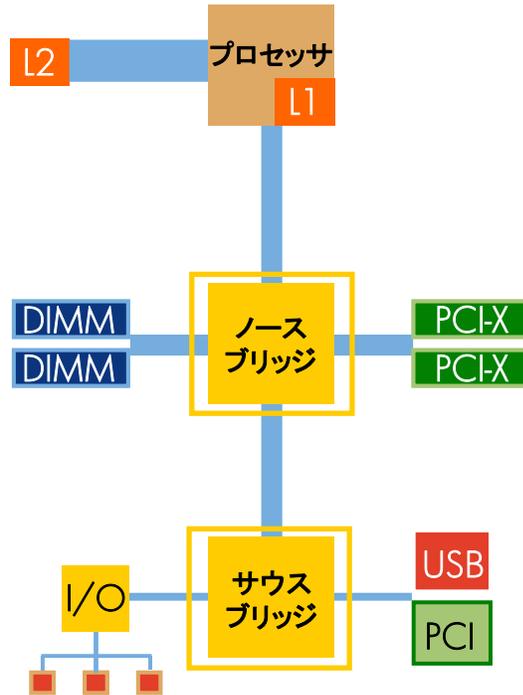
HPのエンジニアはIntelと共同で開発を進めることにより、互換性を保証しています。

現在、業界で最も広い互換性を持っているのは、32ビットプロセッサです。現在の32ビットマイクロプロセッサアーキテクチャは64ビットのデータバスを持ちますが、プロセッサ内のレジスタおよび実行ユニットがアクセスできるアドレスおよび操作できるデータの幅は32ビットのみです。

ほとんどのエンタープライズアプリケーションにとって、64ビットコンピューティングの最大かつ直接的な利点は、より大容量のシステムメモリプールを扱えるようになることです。この能力によって、アプリケーションはより大きなデータセットを処理しながら、実行速度を高速に保つことができます。

64ビットのプロセッサは、64ビット幅のアドレスにアクセスし、64ビット幅のデータを操作することができます。このようなコンピューティング環境では、64ビットのプロセッサに加え、64ビットのオペレーティングシステムが必要となります。

チップセット



チップセットはサーバの中枢神経系であり、すべてのサーバコンポーネント間の通信とデータ交換を制御し、サーバの全般的な機能を決定します。

チップセットは、多くの基本的なサーバロジック機能が 1 つまたは複数のチップに統合された、特定用途向け集積回路 (ASIC) です。ASIC には次のような機能が統合されています。

- システムコントローラ
- メモリコントローラ
- I/O コントローラ
- プログラマブル割り込みコントローラ

チップセットでは、これらの機能を少数の ASIC に統合することにより、信号経路が短縮され、消費電力も抑えられます。この結果、信号回路はより高速に動作し、システム性能が向上します。このほかにも、次のような利点があります。

- システムボード上で占有するスペースが少ない
- 製造コストを抑えられる
- 信頼性が向上する

▲ 注記

ASIC チップセットは、「ブリッジ」とも呼ばれます。

主要なチップセットコンポーネントには次のものがあります。

- メモリコントローラハブ — ノースブリッジ
- I/Oコントローラハブ — サウスブリッジ

ブリッジは、複数のバスを接続するデバイスです。「ノース」、「サウス」という名称は、ブロック図でデバイスが PCI バスの北(ノース)と南(サウス)のどちらにあるのかを指しています。サーバアーキテクチャ設計によっては、各タイプのブリッジが複数存在する場合があります。

メモリコントローラハブ

メモリコントローラハブは、システムコントローラおよびメモリコントローラとして機能します。このハブは、システムタイミングを管理し、高速メモリ調整を行うことにより、バス間を移動するデータの交通整理を行います。ハブは 4 方向のトラフィックを処理する必要があり、すべてのトラフィックはメモリへの経路に簡単にアクセスできなければなりません。また、各デバイスがあたかもメモリへの専用アクセスを持つかのように見せるために、ハブは並列アクセスも提供する必要があります。

I/O コントローラハブ

I/O コントローラハブは、一連の I/O バスとポートを通して、サーバが周辺機器と通信することを可能にします。これらの I/O バスとポートは、プロセッサやメモリよりも大幅に遅い速度で動作します。ハブは、PCI、IDE、USB、ドライブコントローラ、シリアルポート、マウス、キーボードなど、低速の周辺機器をすべて統合します。

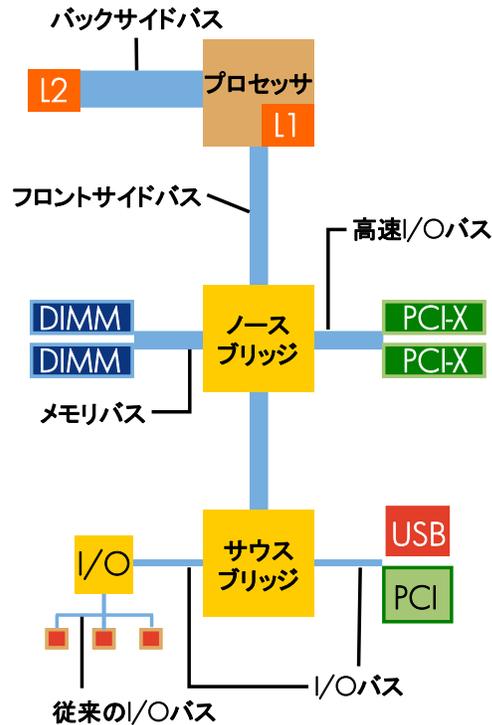
HP 製の ASIC

ASIC チップの開発に必要なリソースと技術に投資できる企業は多くありませんが、HP は、テクノロジーの分野でリーダーシップを取るためには、ASIC チップの開発は重要だと考えています。HP は自社で設計行っているため、他社の設計を採用するベンダよりも、顧客が求める要求の実現に力を注ぐことができます。また、次のような革新的技術で、新たな業界標準を築くこともできます。

- PCI-X
- PCI ホットプラグ
- サーバ管理機能

このような能力により、HP は新しい標準をいち早く市場に導入することができます。

バス



バスは、バイナリデータを表す電気信号のパターンが通る電気回路の集合です。バス(または I/O サブシステム)は、サーバのすべてのコンポーネントをつなぐ通路または回路と考えることができます。

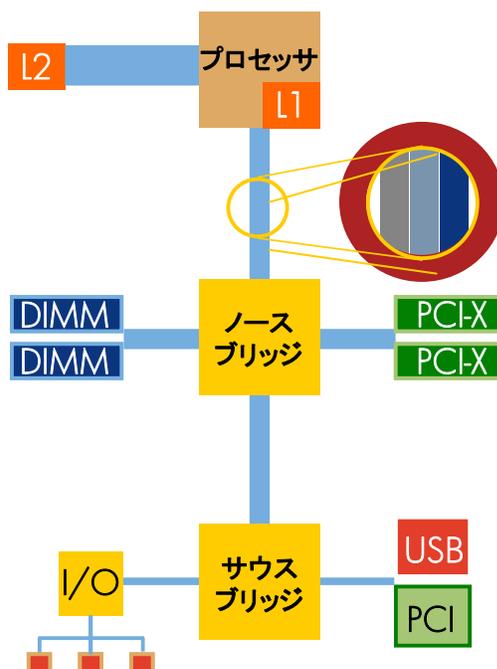
I/O バスサブシステムの性能が、高速なプロセッサやメモリによる最適化と足並みを揃えることができれば、潜在的な性能向上を達成することはできません。

サーババスの種類

サーバ内では多くの種類のバスが連携することにより、最小限のボトルネックで高い性能を実現します。サーバ内に存在する基本的なバスの種類には次のものがあります。

- **バックサイドバス** — 1 つまたは複数のプロセッサをオンボード キャッシュメモリに接続するバス
- **フロントサイドバス** — プロセッサをシステムコントローラ チップセットに接続するバス
- **メモリバス** — メインメモリをメモリコントローラに接続するバス
- **ローカル I/O バス** — 高速の内部および外部 I/O デバイスを I/O コントローラハブに接続するバス
- **拡張 I/O バス** — I/O ポート、拡張 I/O コントローラ、拡張スロットを I/O コントローラハブに接続するバス

バスコンポーネント



I/O サブシステムは、次の共通電気回路の集合です。

- アドレスバス
- データバス
- コントロールバス

アドレスバス

プロセッサがデータを移動するには、アドレスバスによって、データが置かれている可能性のあるターゲットデバイス内の位置が特定される必要があります。ターゲット位置とは、メモリサブシステム、システム ROM BIOS、または I/O バス上の周辺機器のいずれかです。

Intel ベースのシステムでは、アドレスは、1 バイトのデータが置かれている、メモリまたは I/O デバイス内の特定の位置を指します。デバイス内のデータにアクセスするには、1 と 0 の特定の組み合わせから構成されるそのアドレスがアドレスバスに渡されなければなりません。

データバス

データバスは、プロセッサとの間、またはメモリとの間で転送されるデータが通過する内部通路です。

2つのデバイス間を移動するデータは、データバスを経由します。この場合、データとはマイクロプロセッサに対する命令またはマイクロプロセッサが送信する情報です。この情報は、メモリまたは I/O サブシステムとの間でやり取りされます。

コントロールバス

コントロールバスは、コントロール信号の集まりであって、実際にはバスではありません。信号は、取り扱うコンポーネントの動作を決定します。

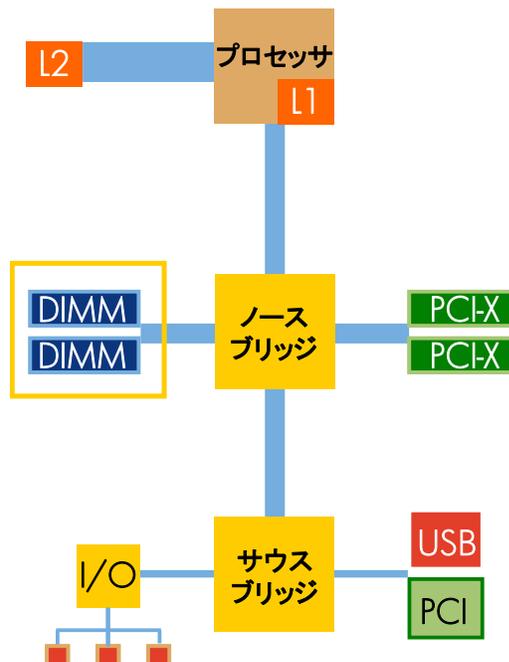
コントロールバスは、バスサイクルのタイプを識別するためと、サイクルの完了を示すために使用されます。また、次のことを決定するために必要なコントロール信号を運ぶためにも使用されます。

- デバイスはデータを読み出すのか、書き込むのか
- 実行中の操作の対象はメモリなのか、I/O なのか

アーキテクチャにかかわらず、どのようなコンピュータでも、主要なサブシステム間の通信は、システムバスまたはホストバス経由でデータを読み書きすることにより行われます。16ビットのバスは 16本のワイヤを使って一度に 2バイトずつ転送し、32ビットのバスは 32本のワイヤを使って同時に 4バイトずつ転送します。

サーバ内の I/O バスは、あらゆる人が通る公共の道路に例えることができます。データの事故を防ぐために、すべてのデータ送信に対して、一連の規則が適用されます。これらの規則は、チップセットシステムおよびメモリコントローラによって適用されます。

メモリ



メモリは、次に示すサーバ内の複数の個所で使用されます。

- メインシステム
- キャッシュ
- ビデオ
- ドライブ

一般的には、このメモリは **RAM** と呼ばれます。**RAM** は、その中のデータをプロセッサが読み取ることができ、処理を通して変更でき、格納するために書き戻すことが可能なメモリです。

RAM には次の基本的な種類があります。

- **DRAM** — メインメモリサブシステムで最も一般的に使用されている **RAM**。このメモリテクノロジーは安価ですが、プロセッサの要件と比較すると、非常に低速です。
- **SRAM** — 非常に高速な、揮発性メモリデバイス。ほとんどの場合、キャッシュメモリで使用され、多くのスペースを占有し、大量の熱を放出します。

ProLiant サーバで採用されているような、綿密に設計されたメモリサブシステムは、非常に高い信頼性を持ちます。HP のエンジニアは、最高の品質を保証すべく、すべての ProLiant サーバで使用されているメモリサブシステムを設計し、入念にテストしています。

ストレージコンポーネント

ディスクドライブは、データ、アプリケーション、オペレーティングシステムを格納し、アクセスするための、比較的高速で信頼性が高い、安価な手段です。サーバは通常、次のようなストレージコンポーネントから構成されています。

- リムーバブルストレージ
 - CD-ROMドライブ
 - ディスケットドライブ
 - テープバックアップドライブ
- 固定ストレージ
 - IDE ディスクドライブ
 - SCSI ディスクドライブ

ドライブアレイ

HP は、ドライブアレイストレージテクノロジーの分野をリードしています。**RAID** は、サーバ内でのディスクアレイの使用に関する業界標準です。**RAID** は、1つの大きなディスクのように機能する複数のディスクの配列で、マルチディスクシステムよりも高い性能とフォールトトレランスを実現します。

ストレージソリューション

ストレージに対するニーズがサーバだけでは満たせなくなり、またストレージデバイスの種類が増加するのに伴って、サーバとストレージデバイス間的高速な接続の必要性が高まりました。このようなニーズに応えるためのストレージソリューションには、次のものがあります。

- **Direct-attached storage (DAS)** — ディスクドライブ、テープドライブ、外部ストレージシステムなど、サーバに直接接続、内部接続、または外部接続されたストレージデバイス。
- **Network attached storage (NAS)** — 既存の LAN に直接接続される、独立型の専用インテリジェントサーバ。データは、業界標準のネットワークプロトコルを介し、業界標準のファイル共有プロトコルを使って、クライアントとの間でやり取りされます。
- **Storage Area Network (SAN)** — 共通のストレージデバイスまたはストレージプールへの物理的な直接アクセスを提供する、集中管理され、安全な、専用の情報インフラストラクチャネットワーク。
- **エンタープライズネットワークストレージアーキテクチャ(ENSA)** — ストレージに対する企業のニーズに応えるために、DAS、NAS、SANを1つに統合するアーキテクチャ。

電源サブシステム



パワーサプライ

サーバのコンポーネントには、厳密に制御、調整された電力が供給される必要があります。標準構成でのサーバの電力負荷だけでなく、追加またはアップグレードされたコンポーネントをサポートするだけの電力が供給されなければなりません。

主な ProLiant サーバには、障害防止およびフォールトトレランス機能を通して信頼性を保証する、次のような電源システム設計が組み込まれています。

- インテリジェントなパワーサプライとファン
- 二重化されたパワーサプライとファン



ファン

電源サブシステムには、電源、放熱、気流のコンポーネントが含まれます。

内部電源コンポーネントには次のものがあります。

- パワーサプライ
- Voltage Regulator Module (VRM) または Processor Power Module (PPM)
- ファン
- システムボード サーミスタ

外部電源コンポーネントには次のものがあります。

- 無停電電源装置 (UPS) とケーブル
- パワーディストリビューションユニット (PDU)
- 冗長化された AC 電源

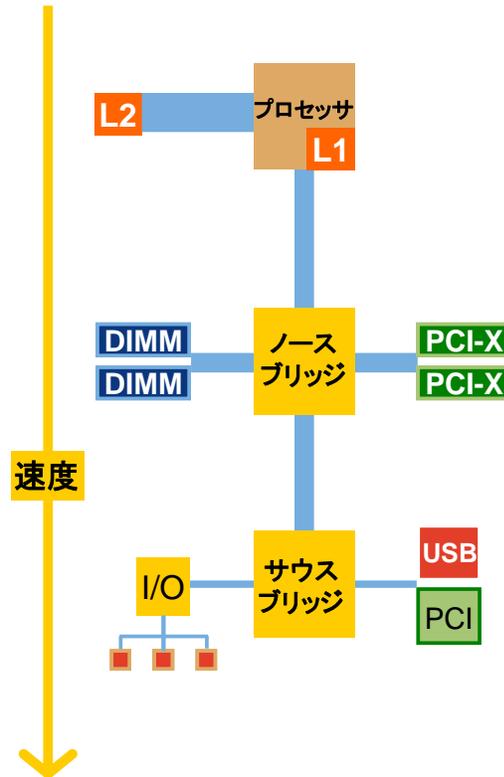
メカニカルコンポーネント

サーバの TCO は、使い易さと保守の容易さという 2 つの要因によって左右されます。サーバのメカニカルコンポーネントは、これらの要因に大きな影響を及ぼします。ProLiant サーバのメカニカルコンポーネントは、HP 社内で開発されています。自社でメカニカルコンポーネントを設計することにより、HP はどのような新しいサーバの設計でも生じる、メカニカルコンポーネントに関する主要な課題を克服することができます。

サーバのメカニカルコンポーネントを適切に設計することは、次のような理由から非常に重要です。

- シャーシ構造はサーバの機構の基盤であり、機械的な強度と保全性を決定する
- シャーシの材質が、電源およびその他のサブアセンブリのアースの役割を果たす
- 熱によってサーバの信頼性や可用性が低下することがないように、シャーシの設計によって効果的なエアフローと冷却を確保する必要がある
- 厳しい電磁放出基準 (EMC) にも準拠するように、エンクロージャの設計によって電子ノイズの漏れを最小限に抑える必要がある
- サーバの保守のし易さは、メカニカルコンポーネントの設計によって大きく左右される

データフロー

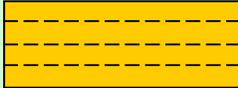
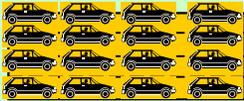


サーバ設計における目標は、システム全体で一定のデータフローを確保しながら、ボトルネックを最小限に抑えるような方法ですべてのコンポーネントを接続するということです。

サーバ内のあらゆるサブシステムコンポーネントは、サーバのデータフローや全般的な性能に影響を及ぼします。つまり、サーバの設計とユーザによるサーバの構成方法によって、サブシステムがサーバの全般的な性能を向上させるか、全般的な性能を制約するボトルネックとなるかが決まります。

サーバ内のデータフローは、高速なサブシステムから低速なサブシステムへという順序で示された、階層的な論理ブロック図で表すことができます。データはシステム内で、高速なサブシステムから低速なサブシステムへと、またその逆方向へと、効率的に流れなければなりません。高速なサブシステム(プロセッサとメモリ)が効率的に動作し続けるためには、コントローラ、メモリバッファ、パラレルバス、および並列読み取り/書き込みプロセスに依存します。

システムタイミングと速度

	バス幅 — 高速道路の車線数
	クロック速度 — すべての車の速度
	容量(ピークスループット) 1つの地点から別の地点までの間を走行可能な車の(理論上の)最大数
	性能(持続スループット) 1つの地点から別の地点までの間を実際に走行している車の数
	待ち時間 — 一時停止標識、交通信号

データフローを理解するには、まずその基本的なタイミング規則を理解する必要があります。すべてのサーバコンポーネントの動作と通信は、システムクロックによって生成される周波数によって同期します。システムは次のような複数のクロックを使用します。

- システムクロック
- バスクロック
- リアルタイムクロック

クロック速度

システムのクロック速度は、発振器回路(通常はシステムチップセット上にあります)によって生成されます。発振器は、システムクロック信号として使用される周波数を発信します。このクロック信号はクロックサイクル(クロックチックとも呼ばれます)で表されます。これらのサイクルは、メガヘルツ単位で表します。

速度

- **クロックサイクル** — 1クロックサイクルは、サーバシステムで最も小さい時間の単位です。どのようなアクションでも1つ以上のサイクルを必要とし、通常は複数のサイクルを使用します。
- **ウェイトステート** — ウェイトステートとは、その間何も行われぬサイクル(クロックチック)のことです。プロセッサが、転送を行うためにコントローラとの同期を待っている状態です。プロセッサは、ウェイトステートを追加することにより、より低速のコンポーネントと同期することができます。ウェイトステートが少ないほど性能は向上します。
- **フロー制御** — コンポーネント間に速度差がある場合に、コンポーネントを同期させるためにウェイトステート(クロックサイクル)を挿入することです。追加されたウェイトステートは、レイテンシ(Latency)と呼ばれます。
- **レイテンシ(Latency)** — システムがデータ転送を準備するために使用する時間。ウェイトステートの数で表します。どのようなサーバコンポーネントでも、ある程度のレイテンシが発生します。レイテンシの例として、速度が異なるコントローラと同期するためにプロセッサによって追加されるウェイトステートや、メモリがマトリクス内のデータのアドレスを見つけるために追加されるウェイトステートなどがあります。
- **バスサイクル** — バス上の1クロックサイクル。バスの速度はバスの種類によって異なります。バスのクロック速度は、チップセットコントローラによって制御されます。
- **バス幅** — 1クロックサイクル内に何ビットがチップに転送されるかを決定する、システム設計要因です。
- **ピークスループット** — コンポーネント間で転送可能な最大データ量。ピークスループットは、性能ではなく容量を表します。
- **持続スループット** — コンポーネント間で実際に転送されるデータ量。持続スループットは実際の性能を表します。
- **バースト転送** — すべてのクロックサイクルでデータを転送します。(非バースト転送では、データは1つおきのサイクルで転送されます)

設計による性能の向上

サーバ性能の向上は主に、設計をより高速かつ効率的にするためにサーバアーキテクチャに適用される新たなテクノロジーによってもたらされます。

また、性能の向上は、適切なサーバの選択と適用によっても実現することができます。サーバ上で実行する予定のアプリケーションに適したサーバ設計を選択することは、システムの全般的な性能を向上させるだけでなく、TCOの削減にもつながります。

サーバアーキテクチャのサブシステムがどのように連携するかを理解することは、特定の用途でのニーズに応え、最大限の性能を引き出すために適切なサーバを選択し、正しく構成する上で役立ちます。

性能を向上させる 3つの方法

どのようなサブシステムでも、次の3つの方法でサーバの性能を向上させることができます。

- **速度を上げる** — システムのパフォーマンスの引き上げ
 - クロック速度を上げる
 - データパスを広げる
 - プロセッサ/メモリ間および I/O の速度を上げる
- **並列処理を行う** — システム全体の並列性
 - パラレルバス
 - バスマスタ
- **レイテンシを減らす** — 待ち時間の原因をなくす
 - データバッファ(キャッシュ)
 - バーストサイクル

復習問題

1. 汎用サーバはどのような目的で使用されますか。
.....
.....
2. 各用語とその説明を対応付けてください。
 - a. ML サーバ 消費電力を抑えた、モジュール式の超高密度ブレードサーバ
 - b. BL サーバ ラックマウント環境用に集積度が最適化されたサーバ
 - c. DL サーバ システム内部の拡張性が最大化されたサーバ
3. すべてのサーバコンポーネント間の通信とデータ交換を制御し、サーバの全般的な機能を決定するのは、サーバサブシステムのどの部分ですか。
.....
4. 高速なサブシステム(プロセッサとメモリ)の効率的な動作を可能にするのは、システム内の次のどのコンポーネントですか(4つ選択してください)。
 - a. コントローラ
 - b. NIC
 - c. メモリバッファ
 - d. パラレルバス
 - e. ハードディスクドライブの速度
 - f. 並列読み取り/書き込みプロセス
 - g. システム相互接続

5. サーバ性能を向上させる 3 つの方法を挙げてください。

.....
.....
.....

6. システムが使用する 3 つのクロックを挙げてください。

.....
.....
.....

7. HP サーバの設計に影響を及ぼす 3 つの要因を挙げてください。

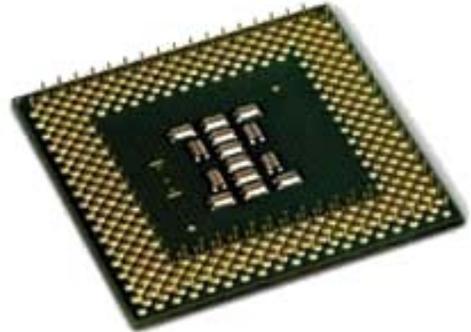
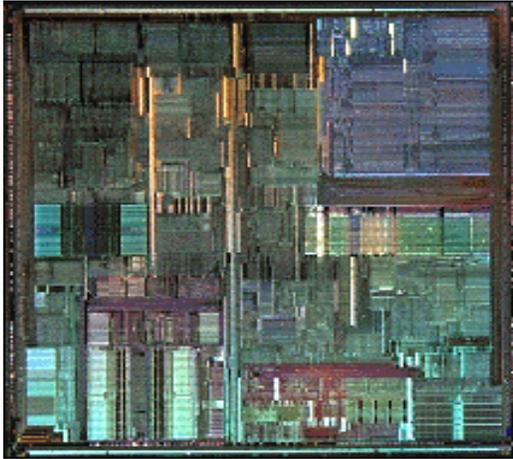
.....
.....
.....

目的

この章では、次の内容について学習します。

- 命令の定義、プロセッサのコンポーネントおよびその機能、プロセッサによる入力の処理方法、クロック同期のしくみ
- パイプライン、スーパースケーラ、ハイパーパイプライン、動的実行、EPIC など、プロセッサの性能を向上させた各種テクノロジー
- Intel プロセッサの機能、現在の ProLiant サーバで使用されているプロセッサ、Pentium III、Pentium III Xeon、Itanium の各 Intel プロセッサの主要な特性
- マルチプロセッシングの定義、アシンメトリック(非対称型)マルチプロセッシングとシンメトリック(対称型)マルチプロセッシングの特性と違い、密結合アーキテクチャと疎結合アーキテクチャの意味
- プロセッサの混用に関する規則と、ProLiant サーバでプロセッサの混用がサポートされる条件

プロセッサの選択



プロセッサの速度が、特定の用途で使用するために考慮すべき唯一の要因であると誤解している人は少なくありません。このような人は、たとえば 1GHz のプロセッサを搭載したサーバは、800MHz のプロセッサを搭載したサーバよりも常に高い性能を発揮すると思いがちです。しかし、必ずしもそうではありません。特定の条件でプロセッサがどの程度の性能を発揮するかは、プロセッサの速度以外の多くの要因によって左右されます。

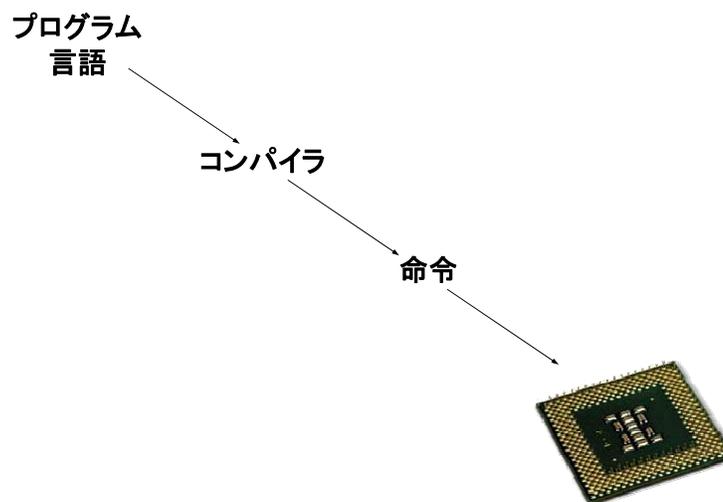
特定のタスクを行うために適切なサーバを選択するには、次のことを理解する必要があります。

- プロセッサのしくみ
- プロセッサの速度を向上させた技術革新
- プロセッサの種類による違い
- マルチプロセッサ サーバとシングルプロセッサ サーバの違い

プロセッサのしくみ

簡単に言えば、プロセッサは、命令を実行するコンポーネントの集合です。

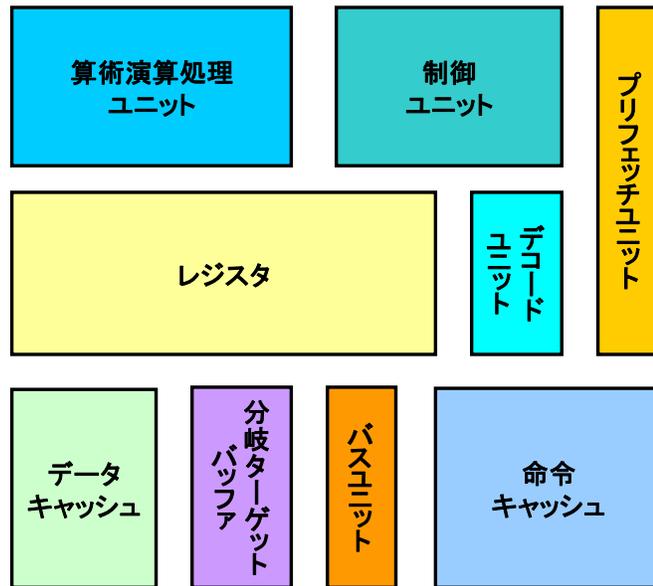
命令



命令とは、コンピュータプログラムからプロセッサに渡される指示(加算、減算、比較など)です。命令はバイナリコードで書かれています。つまり、1と0で表されています。バイナリコードは、互いに通信する必要がある、プロセッサ内の2つのデバイス間のゲートを開いたり閉じたりします。このプロセスにより、1つのデバイスから別のデバイスへと情報が移動します。

プログラマは通常、高水準プログラミング言語を使って新しいアプリケーションを開発します。高水準プログラミング言語のコマンドは、複数の命令に対応しています。プログラミングが完了すると、コンパイラを使って、高水準プログラムを命令に変換します。

プロセッサのコンポーネント



プロセッサのコンポーネント

プロセッサは次のようなコンポーネントから構成されます。

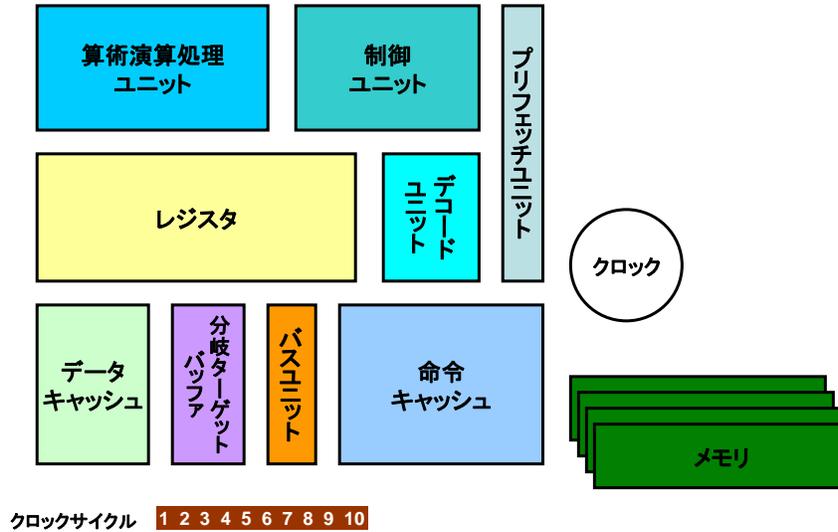
- **プリフェッチユニット** — プロセッサがその後必要とする命令やオペランドを置いておく場所
- **デコードユニット** — 命令をその構成要素に分解するコンポーネント
- **実行ユニット** — 加算や減算など、実際のデータ処理を行うコンポーネント
 - 算術演算処理ユニット (ALU)
 - 浮動小数点演算ユニット (FPU)
- **制御ユニット** — 実行ユニットのスケジューラとして機能するコンポーネント
- **レジスタ** — 制御ユニットと実行ユニットによって、一時的にデータを格納するために使用される少数のメモリ位置
- **L1 キャッシュ** — 最近使用された命令およびデータを格納する、小容量の高速のメモリ領域

▲ 注記

一部のプロセッサはハーバードアーキテクチャをもとに設計されています。上図に示すように、このアーキテクチャでは、L1 キャッシュは命令キャッシュ (I キャッシュ) とデータキャッシュ (D キャッシュ) に分かれています。

- **分岐ターゲットバッファ** — 分岐予測を行うために、最近辿った分岐を格納するレジスタ
- **バスインタフェースユニット** — アドレスバスおよびデータバスへのアクセスを制御するコンポーネント

命令の処理



一般には、プロセッサは入力を受け取ると、次の5つのステージを実行します。

- **ステージ 1: 命令をフェッチする** — プリフェッチユニットが、入力を処理するための命令を探します。まず、L1 キャッシュ(または L1 の Iキャッシュ部分)内で命令を探します。そこで命令が見つからなければ、プリフェッチユニットはメモリ階層のそれ以降の層を検索します。
適切な命令が見つかり、その命令は L1キャッシュに格納されます。その後、プリフェッチユニットがキャッシュから命令をフェッチし、デコードユニットに送ります。
- **ステージ 2: 命令をデコードする** — デコードユニットが、インデックス、データ、オペレーションコードなどの基本要素に命令を分解し、そのデコード結果を制御ユニットに送ります。
- **ステージ 3: 命令を実行する** — 制御ユニットが、デコードされた命令を ALU などの実行ユニットに送ります。実行ユニットが、加算、減算、乗算、除算などのコマンドを実行します。
- **ステージ 4: データを転送する** — 命令がコマンドである場合、コマンドを実行するために必要なデータが、レジスタから実行ユニットに転送されます。
- **ステージ 5: データを書き込む** — 必要であれば、実行ユニットが実行結果をレジスタとデータキャッシュに書き込みます。

クロック同期

	クロック サイクル 1	クロック サイクル 2	クロック サイクル 3	クロック サイクル 4	クロック サイクル 5	クロック サイクル 6	クロック サイクル 7	クロック サイクル 8	クロック サイクル 9	クロック サイクル 10
命令 1	命令を フェッチ する	命令を デコード する	命令を 実行する	データを フェッチ する	データを 書き込む					
命令 2						命令を フェッチ する	命令を デコード する	命令を 実行する	データを フェッチ する	データを 書き込む

プロセッサの各ステージはシステムクロックと同期します。各ステージの完了に 1 クロックサイクルかかるのであれば、1 つの命令は 5 クロックサイクルで処理されます。5 クロックサイクルが経過すると、次の命令を処理できます。

プロセッサの性能の向上

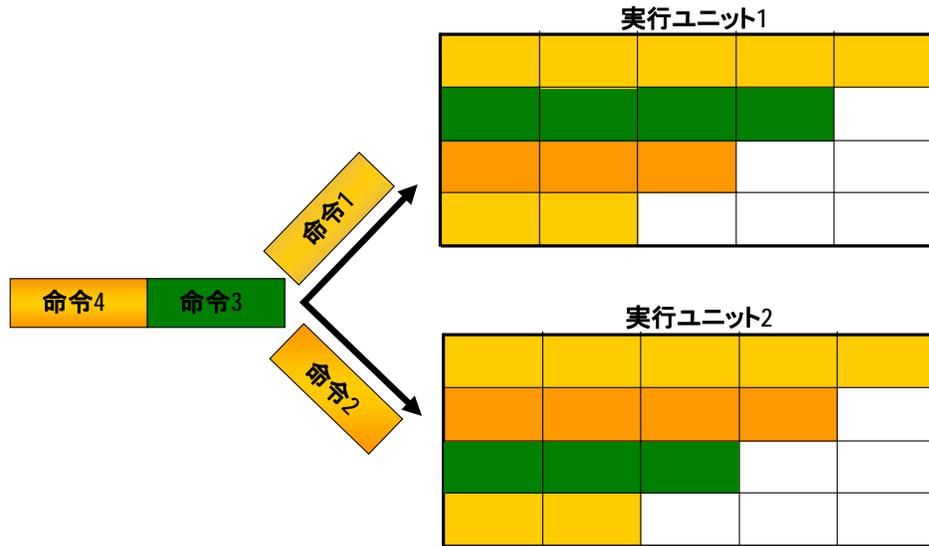
エンジニアは長年にわたって、プロセッサの性能を向上させる方法を模索してきました。性能上のボトルネックをなくす 1 つの方法は、1 クロックサイクルでプロセッサにより多くの処理をさせることです。

パイプラインプロセッサ

フェッチ	デコード	実行	転送	書き込み
		命令 1		
		命令 2		
		命令 3		
		命令 4		

1989 年 4 月、Intel は最初のパイプラインプロセッサである i486 を発表しました。パイプラインプロセッサは、1 つの命令の完了を待たずに、別の命令を開始します。つまり、プリフェッチユニットはステージ 1 を完了し、命令をデコードユニットに送ると、ただちに次の命令の処理を開始します。また、デコードユニットも、1 つの命令のデコードを完了すると、すぐに次の命令の処理に移ります。他のユニットも同様です。

スーパースケーラプロセッサ



1993年に発表された Pentium プロセッサでは、プロセッサに実行ユニットがもう 1 つ追加されました。この技術革新は、2 つのパイプラインを持つことを可能にしました。両方のパイプラインが同時に動作すると、プロセッサは 1 つのクロックサイクルで 2 つの命令を実行することができます。これを、「2 命令の同時実行」と言います。

1 つのクロックサイクルで複数の命令を実行できるプロセッサを、スーパースケーラプロセッサと言います。Pentium II と Pentium III プロセッサは、スーパースケーラプロセッサです。どちらにも、5 つの実行ユニットがあります。

ハイパーパイプライン プロセッサ

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
								命令	1										
								命令	2										
								命令	3										
								命令	4										

実行ユニットを追加することによって、1つのクロックサイクルでプロセッサが処理できる命令の数を増やすことができますが、もう1つ、パイプライン内のステージの数を増やすという方法もあります。

一見、パイプライン内のステージ数を増やすことは理にかなわないように思えるかもしれませんが。たとえば、完了するのに5サイクルしかかからなかった命令が、20サイクルもかかるようになります。ただし1つの利点は、作業をより小さなステップに分解すると、各ステップでプロセッサが行わなければならない作業が少なくなるということです。これにより、システムクロック速度を大幅に引き上げることができ、同じ時間でより多くの命令を実行できるようになります。

例

ハイパーパイプライン プロセッサは、20サイクルで11個の命令を実行できます。パイプラインがなければ、20サイクルで2つの命令しか実行できません。つまり、ハイパーパイプライン プロセッサは、5.5倍の速度を持つことになります。

動的実行

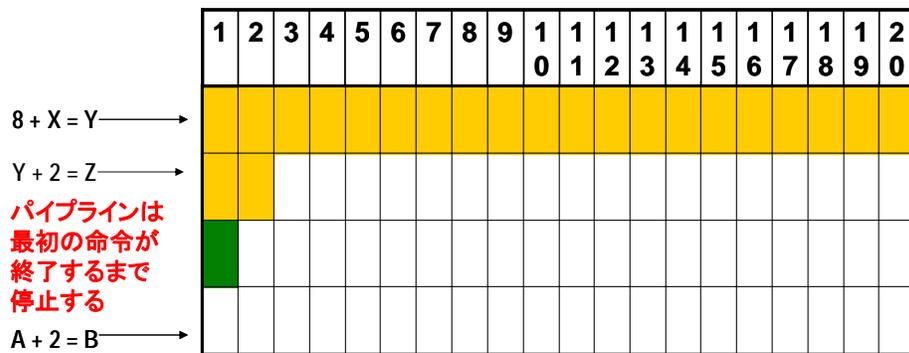
パイプラインはプロセッサの性能を向上させますが、特にハイパーパイプライン プロセッサでは、問題を発生させることがあります。この種の問題を「パイプライン障害」と呼びます。

パイプライン障害が発生すると、障害が取り除かれるまで、パイプラインはブロックされます。これを「パイプラインストール」と言います。パイプラインが深いほど、パイプラインストールによる被害は大きくなります。20 ステージのパイプラインでは、正しい分岐が処理されるまでに、最長で 19 サイクル待たなければなりません。10 ステージのパイプラインでは、同じような状況でも、最長で 9 サイクルしか待つ必要はありません。

1995 年に、Intel は P6 アーキテクチャに基づく最初のプロセッサである Pentium Pro を発表しました。P6 アーキテクチャには、集合的に「動的実行」と呼ばれる一連のテクノロジーが採用されています。順序外実行、分岐予測、投機的実行から構成される動的実行は、パイプライン障害を解決するために開発されたテクノロジーです。

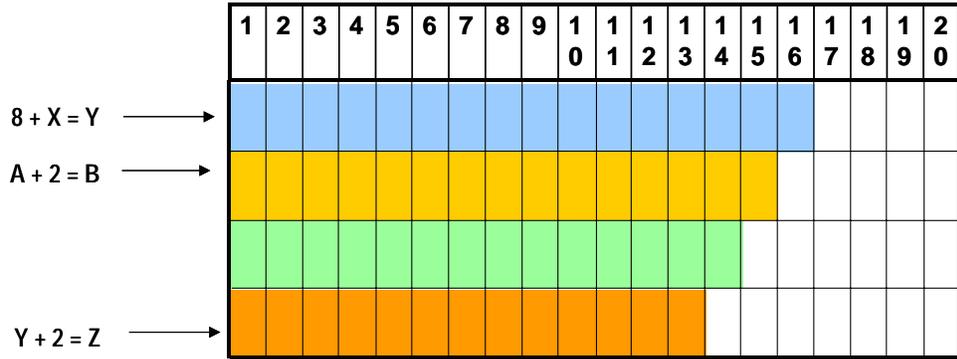
依存問題

命令2は命令1に依存する



パイプライン障害の 1 つは、依存問題です。依存問題は、ある命令の操作の完了に別の命令からのデータが必要な場合に発生します。2 つ目の命令でデータが必要となる前に、1 つ目の命令が完了していないと、パイプラインは 1 つ目の命令が完了するまで待たなければなりません。

順序外実行



依存問題に対処するために、プロセッサは順序外実行を行います。つまり、1つ目の命令が完了するまでパイプラインを保留にする代わりに、プロセッサは1つ目の命令に依存しないその他の命令を処理します。これらの命令の実行結果は、後で必要になるまで格納されます。1つ目の命令の処理が完了した時点で、その命令に依存している命令がパイプラインに取り込まれます。この結果、無駄になるサイクルの数が少なくなります。

例

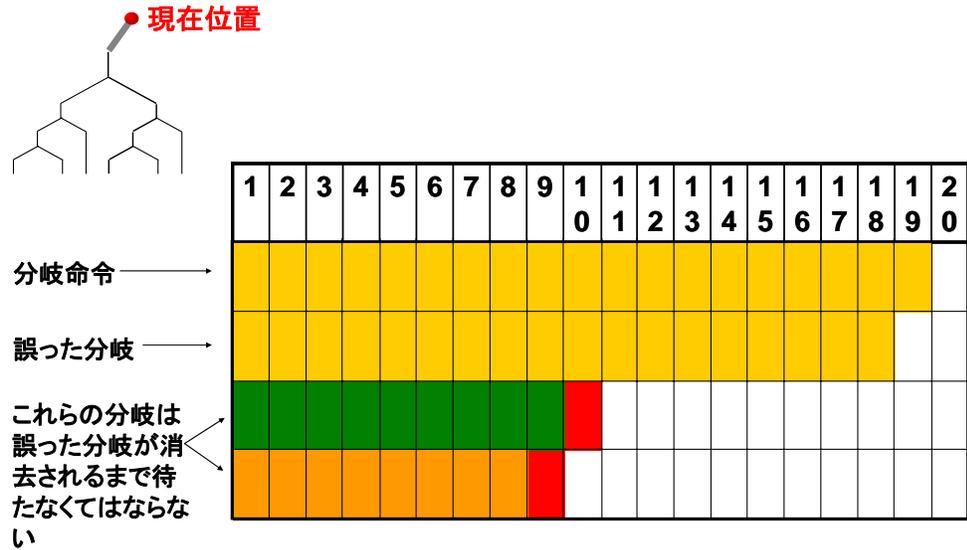
命令 1: $8 + X = Y$

命令 2: $Y + 2 = Z$

命令 3: $A + 2 = B$

命令 2 は命令 1 に依存しています。順序外実行を行えるプロセッサは、命令 3 は命令 1 に依存しないため、命令 2 の前に命令 3 をパイプラインに取り込むことができます。命令 1 が完了すると(その時点で命令 3 はパイプライン内の先のステップに進んでいます)、命令 2 がパイプラインに取り込まれます。

分岐問題

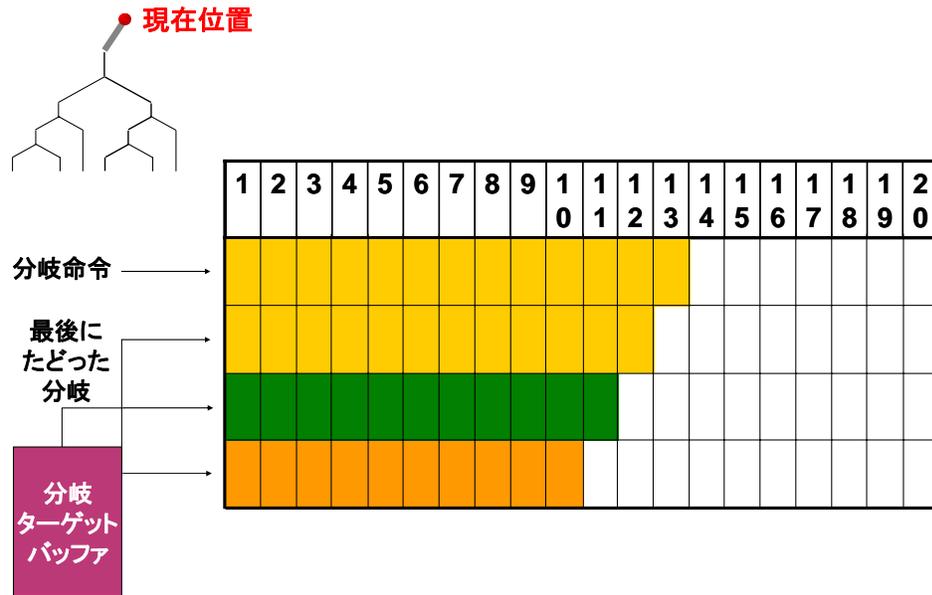


もう1つのタイプのパイプライン障害は、分岐問題です。分岐命令は条件を調べ、条件が真であれば命令の一方の分岐を実行し、条件が偽であれば他方の分岐を実行します。

分岐命令の完了には最長で20サイクルかかりますが、プリフェッチユニットが命令をパイプラインに取り込み続けます。このため、分岐命令が完了する前に、誤った分岐からの命令がパイプラインに送り込まれることがあります。

この状態が発生すると、誤った分岐からの命令がパイプラインから消去されるまで、パイプライン内のすべての命令が停止します。

分岐予測



分岐予測を行うことにより、分岐問題の発生を抑えることができます。分岐予測では、分岐命令が初めて実行される際に、その命令のアドレスとその後に実行された命令（正しい分岐）が分岐ターゲットバッファ(BTB)に格納されます。プロセッサはこの情報を使用して、その命令を次回実行する際にどちらに分岐するかを予測します。分岐予測の正解率は、90%を超えます。

1つ目の命令が再び使われると、プロセッサは前回採用された分岐をプリフェッチするため、待ち時間を最小限に抑えることができ、また誤った命令を消去する頻度も減ります。分岐予測が正しいと、分岐を実行してもパイプラインストールは発生しません。

投機的実行

動的実行の3つ目の構成要素は、投機的実行です。これは、分岐予測と順序外実行を組み合わせたものです。

投機的実行を行えるプロセッサは、分岐命令が辿る経路を予測し、その経路上の非条件命令をプリフェッチし、実行を開始します(分岐予測では命令をプリフェッチするだけで、実行はしません)。

通常のように、命令の実行結果を汎用レジスタに書き込む代わりに、プロセッサは投機的実行の結果を一時レジスタに格納します。分岐命令が解決された時点で、事前に計算した結果を一時レジスタから取り出し、発行された順序で汎用レジスタに書き込むことができます。

EPIC

2001年5月、IntelはExplicitly Parallel Instruction Computing (EPIC)モデルに基づくItaniumプロセッサを発表しました。

このアーキテクチャでは、コンパイラによって3つの命令が1つのグループにまとめられます。各グループには、テンプレートと呼ばれる少数のビットが含まれています。テンプレートは、命令が依存関係にあるかどうかを実行ユニットに知らせます。コンパイラが実行ユニットに対して、どの命令を並列に処理できるかを指示するため、Intelではこのアーキテクチャを「explicitly parallel」(明示的並列)と呼んでいます。Itaniumプロセッサは、順序外実行を行うプロセッサのように、自身で処理順序を判断する必要はありません。

EPICアーキテクチャでは、分岐も新たな方法で処理されます。各命令は、1ビットの述部を持ちます。一方の分岐内の命令は述部が真に設定され、他方の分岐内の命令は述部が偽に設定されます。

分岐命令がプロセッサに送信され、その後に両方の分岐が続きます。プロセッサは、その分岐命令が真か偽かを判断します。真の場合は、真の述部を持つ命令のみが処理され、偽の場合は、偽の述部を持つ命令のみが処理されます。このため、予測を誤った分岐をパイプラインから消去する必要はありません。

現在の Intel プロセッサ

ProLiant サーバにはすべて Intel プロセッサが採用されていますが、プロセッサにはさまざまな種類があります。たとえば、速度、フォームファクタ、キャッシュサイズ、サポートされている命令セットとチップセットは、プロセッサの種類ごとに異なります。システム設計者は、特定のサーバでどのプロセッサを使用するかを決定する際には、これらの要因をすべて考慮に入れる必要があります。また、顧客の用途に合ったサーバを選択する際にも、これらの要因を考慮しなければなりません。

パッケージ形式

一般に、ProLiant サーバで使用されている Intel プロセッサは、カートリッジ、pin grid array (PGA)、ball grid array (BGA) の 3 つのうちのいずれかのパッケージ形式を持ちます。

カートリッジ

SECC 2



SECC2 パッケージの Pentium III プロセッサ

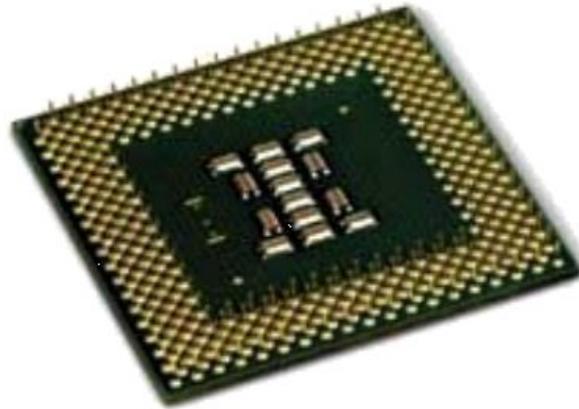
カートリッジプロセッサは、プラスチック製の箱に入っており、メモ리카ードのコネクタに似たゴールドコネクタを、マザーボードまたはプロセッサボード上の特殊なコネクタに挿入します。カートリッジプロセッサは、Single Edge Contact Cartridge (SECC 2) などのカートリッジのタイプ、または SC242 や SC330 などのコネクタのタイプで呼ばれます。

コネクタのタイプは、Slot (Slot 1、Slot 2 など) と呼ばれる場合もあります。これは古い呼び方です。1 基または 2 基での使用が可能な Pentium III プロセッサには、Slot 1 タイプのカートリッジが使用されています。4 プロセッサおよび 8 プロセッサ サーバで使用可能な Pentium III Xeon プロセッサには、Slot 2 タイプのカートリッジが使用されています。

Pin Grid Array

pin grid array (PGA) プロセッサの底面には、多数の小さなピンがあります。これらのピンは、マザーボード上の特殊なコネクタに挿入されます。破損を防ぐために、レバーを使ってプロセッサを所定の位置にロックし、取り外す際にもレバーを使って持ち上げます。

FC-PGA



Flip Chip-Pin Grid Array (FC-PGA) プロセッサ (底面)

flip chip-pin grid array (FC-PGA) パッケージでは、発熱するコンポーネントが上面に来るように、プロセッサダイ上のコンポーネントが上下逆さに付いています。ダイは基板に取り付けられ、その上にヒートシンクが設置されています。ヒートシンクは、熱を逃がすことによりプロセッサを冷却するコンポーネントです。FC-PGA では、発熱するコンポーネントがヒートシンクに面しているため、放熱を効率的に行うことができます。

FC-PGA2



ヒートシンクに取り付けられた FC-PGA2 プロセッサ (底面)



FC-PGA2 プロセッサのヒートスプレッダ (上面)

FC-PGA2 プロセッサは、ヒートスプレッダを内蔵しています。ヒートシンクは、FC-PGA パッケージのようにダイではなく、ヒートスプレッダに接触します。

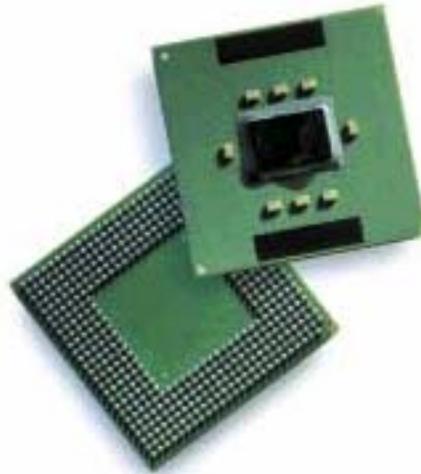
マイクロ PGA



μ Pin Grid Array (マイクロ PGA)プロセッサ

マイクロ PGA は、FC-PGA よりもピンピッチを狭めています。これにより、パッケージのサイズを小さくするだけでなく、電気的な特性を高める効果があります。

Ball Grid Array



Flip Chip-Ball Grid Array (FC-BGA)プロセッサ

ball grid array (BGA)プロセッサは PGA プロセッサと似ていますが、主要な違いは、ソケットに挿入する金属製のリードの代わりに、半田ボールが使用されている点です。BGA プロセッサは、e-Class サーバのブレードなど、プロセッサがマザーボードに固定されたときにスペースに制約のある環境で使用されています。BGA プロセッサには、FC-BGA と呼ばれる、上下逆さの設計もあります。

製造プロセス

チップは、ウェーハと呼ばれる大きな円盤状のシリコンの板上で製造されます。機械によって、1枚のウェーハに多数のチップの回路線がエッチングされます。その後、ウェーハは個々のチップに裁断されます。

1枚のウェーハにエッチング可能なチップの数を決定する要因の1つは、回路線の幅です。回路線が細いほど、1枚のウェーハにエッチングできるチップの数は多くなります。また、回路線が細いほど、プロセスが発する熱量は少なくなります。どちらの要因も、チップの製造コストを低減します。また、放熱量の少ない、細い回路線は、プロセスの速度を向上させるだけでなく、少ない電力で動作します。

回路線幅はミクロン単位で表されます。この数年間で、Intelは0.25 μ テクノロジー、0.18 μ テクノロジー、そして最近では、0.13 μ テクノロジーを発表しました。

マルチプロセッサ サーバ



ProLiant DL580 G2とProLiant DL760

より高いコンピューティング能力に対する要求に応える1つの方法は、サーバに複数のプロセッサを搭載することです。複数のプロセッサが搭載されたサーバを「マルチプロセッササーバ」と言います。

「ウェイ」という用語は、サーバ内のプロセッサの数を示すために使用されています。たとえば、ProLiant DL580 G2などの4ウェイサーバには、4つのプロセッサを搭載することが可能です。また、ProLiant DL760などの8ウェイサーバには、8つのプロセッサを搭載することが可能です。

マルチプロセッサシステムを設計する際には、エンジニアは次のことを決定する必要があります。

- プロセッサ間でどのように負荷を分散させるか
- プロセッサ間でどのようにメモリとI/Oリソースを共有させるか

負荷の分散

- アシンメトリック (非対称型) マルチプロセッシング (AMP) — 特定のスレッドを特定のプロセッサに割り当てる



- シンメトリック (対称型) マルチプロセッシング (SMP) — 次のスレッドを次に使用可能なプロセッサで実行する



マルチプロセッサ サーバでは、次のいずれかの方法で、プロセッサ間で負荷を分散させることができます。

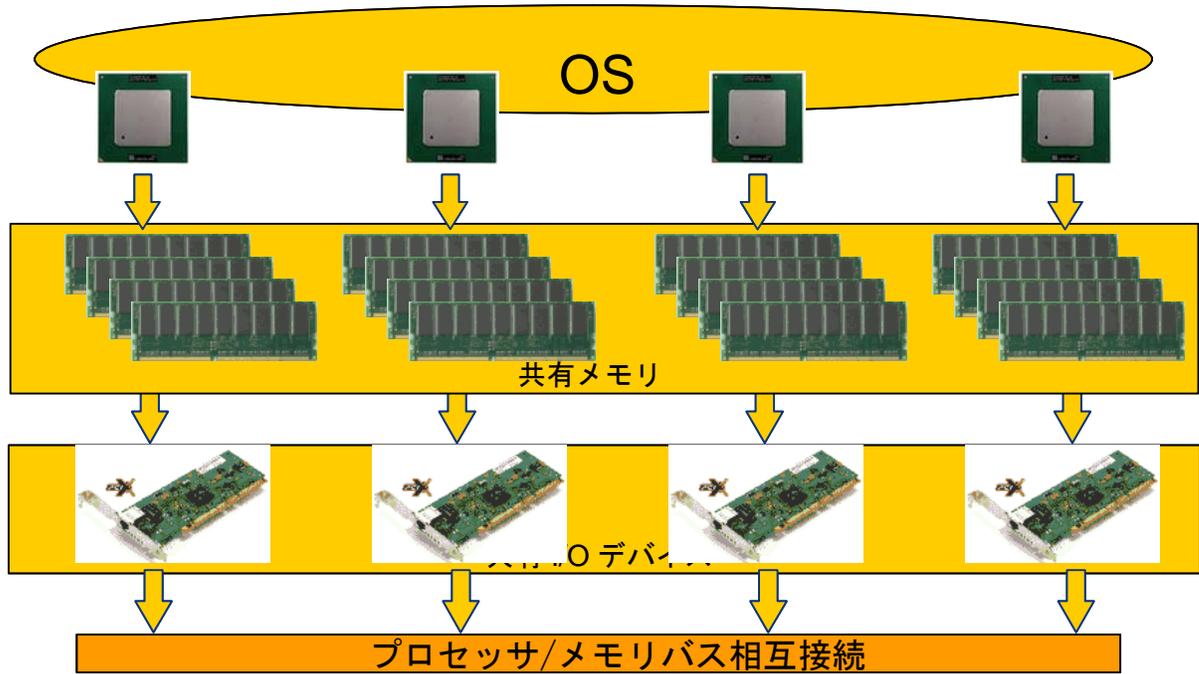
- **アシンメトリック (非対称型) マルチプロセッシング (AMP)** — 特定のタスクを特定のプロセッサに割り当てます。アプリケーションの種類によっては、負荷は各プロセッサに均等に配分されないため、最も効率的な方法とは言えません。また、AMP システムにプロセッサを追加しても、必ずしもコンピューティング能力は向上しません。ただし、状況によっては、1つのプロセッサで1つのプロセスを処理する方が望ましい場合があります。
- **シンメトリック (対称型) マルチプロセッシング (SMP)** — 次のタスクを、次に使用可能なプロセッサで実行します。この方法では、非常に効率の良い負荷分散を行えますが、オペレーティングシステムが SMP をサポートしている必要があります。プロセッサを追加すると、通常はコンピューティング能力が向上します。

メモリとI/Oリソースの共有

メモリとI/Oリソースを共有するためのマルチプロセッサ システムを設計する際には、2つのアーキテクチャを使用できます。

密結合

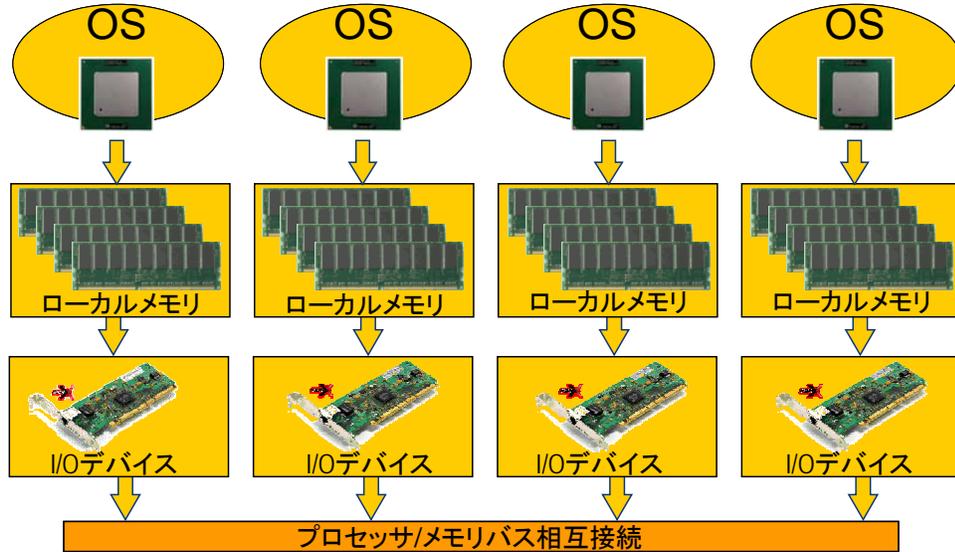
すべてのプロセッサはすべてのメモリと I/O リソースを共有する



密結合アーキテクチャでは、すべてのプロセッサがすべてのメモリとI/Oリソースを共有します。すべてのリソースはオペレーティングシステムによって管理されます。メモリに障害が発生すると、すべてのプロセッサが影響を受けます。このアーキテクチャには、すべてのプロセッサと共有コンポーネントに共通のシステムバスが必要になります。

疎結合

各プロセッサにメモリとI/Oリソースが割り当てられ、
独立したコンピュータのように機能する



疎結合アーキテクチャでは、各プロセッサにメモリとI/Oリソースが割り当てられ、独立したコンピュータのように機能します。各プロセッサは、他のプロセッサとリソースを共有することはありませんが、互いに通信し、連携します。メモリに障害が発生した場合には、そのメモリを使用するプロセッサのみが影響を受けます。

その他の要因

マルチプロセッササーバの性能に影響を及ぼす要因は、このほかにもいくつかあります。たとえば、オペレーティングシステムがマルチプロセッシング環境を処理できなければなりません。マルチプロセッシングの処理方法は、オペレーティングシステムごとに異なります。オペレーティングシステムによってすべてのI/O割り込みが1つのプロセッサに送られる場合もあれば、複数のプロセッサ間でI/O割り込みを共有する場合があります。マルチプロセッシングソリューションを設計する際には、そのソリューションが顧客の要求を満たせるように、綿密に計画を立てる必要があります。

論理プロセッサとハイパースレッディング テクノロジ

論理プロセッサ



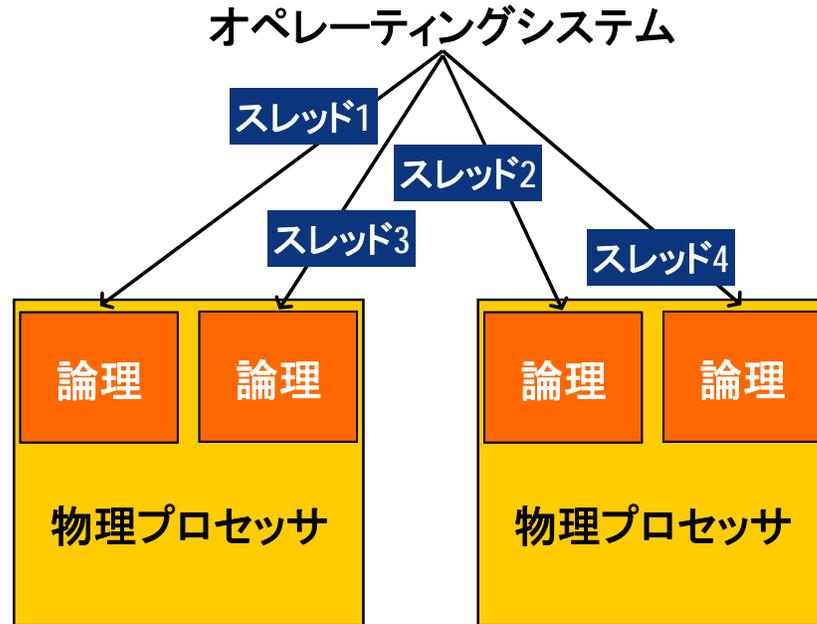
Intel Xeon や一部の Pentium 4 プロセッサには、1つのコア上に2つのレジスタのセット(アーキテクチャステート)があります。これにより、各物理プロセッサは、2つの論理プロセッサとして機能することができます。コア内の実行ユニットおよびその他のリソースは2つのアーキテクチャステートによって共有されるため、実行ユニットはより頻繁に活用されるようになります。この結果、ハイパースレッディング テクノロジを使用しているプロセッサは、このテクノロジーを使用していないプロセッサと比較して、約30%の性能向上を達成することができます。

ハイパースレッディング テクノロジ

ハイパースレッディング テクノロジでは、2つの論理プロセッサが、共有ハードウェアリソースを使って、異なるタスク(スレッド)を同時に実行できます。ソフトウェアまたはアーキテクチャの観点から言えば、オペレーティングシステムやユーザプログラムは、2基の物理プロセッサを使用しているかのように、両方の論理プロセッサに対してスレッドをスケジュールすることができます。マイクロアーキテクチャの観点から言えば、両方の論理プロセッサからの命令が、共有実行リソース上で同時に持続し、実行されます。

この結果、より多くのトランザクションが処理されるため、多くのインターネットアプリケーションや e-business アプリケーションで応答時間の短縮が見られます。さらに、Xeon プロセッサはより多くのユーザをサポートできます。

スレッドのスケジューリング



論理プロセッサを認識するオペレーティングシステムは、論理プロセッサを物理プロセッサと同様に管理します。つまり、論理プロセッサに対してタスク(スレッド)をスケジュールします。複数の Xeon プロセッサが搭載されていれば、オペレーティングシステムは同時に、個々の物理プロセッサと、各物理プロセッサ上の個々の論理プロセッサに対してスレッドをスケジュールできます。

プロセッサをカウントし、識別する方法により、オペレーティングシステムは常に、異なる物理プロセッサ上にある複数の論理プロセッサにスレッドをスケジュールしてから、同じ物理プロセッサに複数のスレッドをスケジュールします。このような最適化が行われるため、ソフトウェアスレッドは、可能であれば異なる物理実行リソースを使用することができます。

2つ目の論理プロセッサが不要な場合は、オペレーティングシステムは HALT 命令を発行することによってその論理プロセッサをオフにします。これにより、2つ目の論理プロセッサが処理すべきスレッドの有無を繰り返しチェックすることがなくなり、実行リソースを節約することができます。

▲ 注記

ハイパースレッディング テクノロジは、ROM ベースセットアップユーティリティ(RBSU)を使ってオフにすることができます。これは、エンタープライズアプリケーションの使用時に性能向上が見られるかどうかを検証する場合に必要になります。マルチスレッディング用に設計されたエンタープライズアプリケーションであれば、ハイパースレッディング テクノロジによって性能が向上するはずですが。

プロセッサの混用

Intel プロセッサラインには、コア周波数、L2 キャッシュのサイズ、ステッピングの異なるプロセッサから構成されるプロセッサファミリーが複数あります。顧客は既存の投資を活用することを望んでいますが、同時に処理能力を拡大し、サーバコンポーネントを配備し直す必要にも迫られています。このため、デュアルプロセッサ システムやマルチプロセッサ システムでプロセッサを混用できることは、顧客にとっては大きなメリットとなります。

製造元による保証およびサポートを受けるためには、ユーザは製造元のガイドラインに従ってコンピュータのハードウェアおよびソフトウェアを構成する必要があります。サーバ内で使用されているプロセッサについては、プロセッサの製造元、サーバの製造元、そしてオペレーティングシステムのベンダといった複数のベンダのガイドライン間での調整が必要になります。Intel プロセッサの混用がベンダによってサポートされているかどうかは、テストの計画がベンダごとに異なり、ベンダによってはサポート方針の表現があいまいなため、明確ではありません。

テスト

新しいプロセッサが次々と発表され、また競争に打ち勝つためには新製品をいち早く市場に投入しなければならないため、どのベンダも、プロセッサのあらゆる組み合わせをテストし、潜在的な問題をすべて発見するために十分な時間やリソースを割くことはできません。Intel は、自社のプロセッサに関する独自のテスト計画を実行しており、発見した不具合や回避策を自社の Web サイトで公開しています。

Microsoft は、自社の Windows Hardware Quality Labs (WHQL) による互換性テストおよびハードウェア用の Microsoft Windows Logo Program による認定を受けるためには、システムが、定義されているプロセッサの構成制限を順守している必要があるとしています。

その他のオペレーティングシステム ベンダも、マルチプロセッサ サーバに対する独自の認定手続きを規定しています。また、必要であれば、HP、他の製造元、およびユーザは、プロセッサを混用した構成を独自に検証することもできます。

プロセッサステッピング

プロセッサステッピングとは、同じプロセッサモデル (Intel Pentium III Xeon プロセッサなど) の異なるバージョンであり、新しいステッピングには、通常は性能や歩留まりを向上させるためにわずかな改良が加えられています。ステッピングが変わった場合には、システム ROM に変更を加える必要があります。Intel は、各プロセッサステッピングに対して、システム ROM に適用するマイクロコードパッチを提供しています。

各 Intel プロセッサステッピングには固有のプロセッサ ID が割り当てられており、対応するマイクロコードパッチには同じプロセッサ ID が含まれています。マイクロコードパッチは、システム ROM 内のテーブルに格納されています。HP およびその他のサーバベンダは、ROM を最新に保つために、Intel のパッチが新たにリリースされるたびに ROM に追加する必要があります。

現在のシステム ROM にそのプロセッサ用の適切なマイクロコードが含まれない ProLiant サーバにプロセッサがインストールされた場合、ユーザが誤ったマイクロコードでプロセッサを動作させないように、ROM が次のエラーメッセージを表示します。

UNSUPPORTED PROCESSOR DETECTED. SYSTEM HALTED.

例

顧客が 2 基のプロセッサを搭載した 4 ウェイ ProLiant サーバを、2 基のプロセッサを追加することによりアップグレードしようとしています。新しいプロセッサのステッピングは、既存のプロセッサとは異なる可能性があります。この場合、現在のシステム ROM には、新しいプロセッサに対する適切なマイクロコードパッチが含まれていないかもしれません。システム ROM に最新のマイクロコードを追加するには、新しいプロセッサをインストールする前に、アップデートされたシステム ROMPaq を使って、サーバ上の ROM をフラッシュする必要があります。

一般にプロセッサチップの上部はヒートシンクで覆われてしまうため、プロセッサのインストール後にステッピングを確認することは容易ではありません。このような場合、オペレーティングシステム付属のユーティリティを使って調べることができます。たとえば、Windows NT では、システム管理者は Windows NT 診断プログラムを使って、インストールされているプロセッサのステッピングレベルを確認できます。または、Survey ユーティリティを使って、インストールされているプロセッサのステッピングを表示することもできます。

INTERNET Survey ユーティリティは、SmartStart CD に含まれています。また、特定の ProLiant サーバに対するユーティリティを、HP の Web サイトの Software & Drivers Support ページ (<http://www.hp.com/support/files/>) からダウンロードすることもできます。

プロセッサの混用に対する Intel のサポート

本書の執筆時に Intel の Web サイトに掲載されている情報によれば、Intel は次の条件下でのみ、ステッピングの異なるプロセッサの混用をサポートしています。

- システム内のすべてのプロセッサが、同じファミリ番号とモデル番号 (CPU ID 命令で確認可能) を持っている
- システム内のすべてのプロセッサが、同じ周波数 (すべてのプロセッサで共通にサポートされている最高周波数) で動作する
- システム内のすべてのプロセッサが、同じキャッシュサイズを持っている
- 最も少ない機能セットを持つプロセッサが、ブートストラップ プロセッサである

プロセッサの混用に対するオペレーティングシステムのサポート

Intel と同様、オペレーティングシステムのベンダ各社も、システム内のすべてのプロセッサが同じファミリ番号とモデル番号を持ち、同じ周波数で動作し、同じキャッシュサイズを持っている場合に限り、プロセッサの混用をサポートしています。また、最も少ない機能セットを持つプロセッサが、ブートストラップ プロセッサでなければならない点も同様です。オペレーティングシステムのベンダは、多くの場合、プロセッサステッピングの混用のサポートに関して明確な表明を行っていません。ただし、通常のオペレーティングシステムでは、ステッピングの異なる複数のプロセッサを含むサーバの動作を一般に禁じていません。

プロセッサの混用に対する HP のサポート

HP は、ProLiant サーバでのプロセッサの混用を、以下の一般的な条件下でサポートしています。

- サーバ内のすべてのプロセッサが、同じファミリーおよびモデルのものでなければなりません。つまり、Pentium II Xeon プロセッサと Pentium III Xeon プロセッサを混用することはできません。
- 異なるキャッシュサイズを持つプロセッサ (Pentium II Xeon と Pentium III Xeon プロセッサのみ) を使用する場合、同じキャッシュサイズを持つプロセッサ同士をペアでインストールする必要があります。最初のペアを Slot 1 と Slot 2 に、2 つ目のペアを Slot 3 と Slot 4 に、3 つ目のペアを Slot 5 と Slot 6 に、そして 4 つ目のペアを Slot 7 と Slot 8 にそれぞれインストールします。
- 最も低いステッピング番号のプロセッサを、ブートストラップ プロセッサとしてインストールします。

ステッピングの異なるプロセッサを含む ProLiant サーバを正しく動作させるためには、ユーザはプロセッサをインストールする前にシステム ROM をフラッシュする必要があります。最新のシステム ROM は、HP の Web サイトの Support & Drivers Support ページ (<http://www.hp.com/support/files/>) からダウンロードできます。

復習問題

1. 命令とは何ですか。

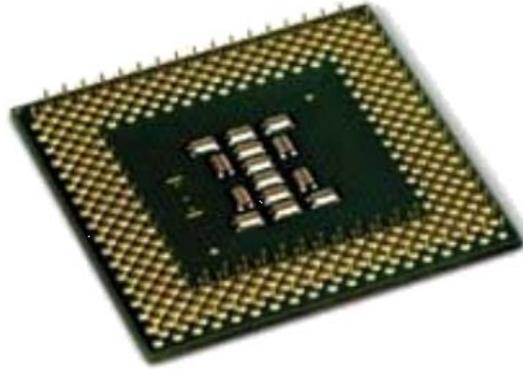
2. プロセッサの各コンポーネントとその機能を対応付けてください。

a. プリフェッチユニット	最近使用された命令およびデータを格納する、小容量の高速メモリ領域
b. デコードユニット	アドレスバスおよびデータバスへのアクセスを制御するコンポーネント
c. 実行ユニット	命令をその構成要素に分解するコンポーネント
d. 制御ユニット	制御ユニットと実行ユニットによって、一時的にデータを格納するために使用される少数のメモリ位置
e. レジスタ	分岐予測を行うために、最近辿った分岐を格納するレジスタ
f. L1 キャッシュ	加算や減算など、実際のデータ処理を行うコンポーネント
g. 分岐ターゲットバッファ	プロセッサがその後必要とする命令やオペランドを置いておく場所
h. バスインタフェースユニット	実行ユニットのスケジューラとして機能するコンポーネント

3. プロセッサによる入力の処理方法を示す次の手順を、正しい順序に並べ直してください。
 - a. 命令を実行する
 - b. データを書き込む
 - c. 命令をフェッチする
 - d. データを転送する
 - e. 命令をデコードする

4. 各テクノロジーとその説明を対応付けてください。
- a. パイプライン 1つの命令の完了を待たずに別の命令を開始するプロセッサ
 - b. スーパースケーラ 1つの命令を実行するために使用するステップの数を増やしたプロセッサ
 - c. ハイパーパイプライン 1つのクロックサイクルで複数の命令を実行できるプロセッサ
 - d. 分岐予測 コンパイラが実行ユニットに対して、どの命令を並列に処理できるかを指示するプロセッサ
 - e. 順序外実行 ある分岐命令を初めて実行する際に、その命令のアドレスと正しい分岐のアドレスを分岐ターゲットバッファに格納するテクノロジー
 - f. EPIC 別の命令に依存しない命令を先に処理できるプロセッサテクノロジー

5. 次の図に示されているのは、どのタイプのパッケージですか。



- a. PAC418
 - b. SECC 2
 - c. FC-PGA
 - d. BGA
6. 次の図に示されているのは、どのタイプのパッケージですか。



- a. PAC418
 - b. SECC 2
 - c. FC-PGA2
 - d. FC-BGA
7. 回路線の幅に関する次の説明のうち、正しいものはどれですか(該当するものをすべて選択してください)。
- a. 回路線が細いほど、1枚のウェーハにエッチングできるチップの数が増える
 - b. 回路線が細いほど、プロセッサの発熱量が少なくなる
 - c. 回路線が細いほど、プロセッサの速度が向上する
 - d. 回路線が細いほど、プロセッサの動作に必要な電力が少なくなる
 - e. 回路線が細いほど、 α 粒子の影響を受けにくい

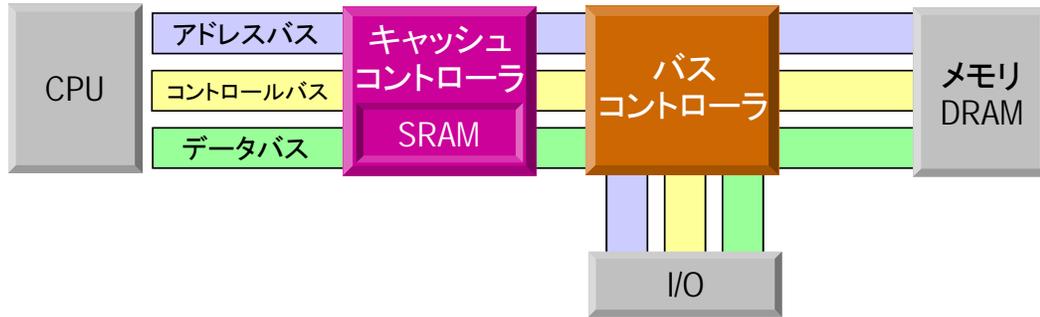
8. マルチプロセッサ サーバとは何ですか。
-
-
9. 各テクノロジーとその説明を対応付けてください。
- | | | | |
|----|--------------------------|-------|--------------------------------------|
| a. | アシンメトリック(非対称型)マルチプロセッシング | | すべてのプロセッサがメモリを共有する |
| b. | シンメトリック(対称型)マルチプロセッシング | | 次のタスクを、次に使用可能なプロセッサで実行する |
| c. | 疎結合 | | 特定のタスクを特定のプロセッサに割り当てる |
| d. | 密結合 | | 各プロセッサにメモリが割り当てられ、独立したコンピュータのように機能する |
10. プロセッサを混用する場合、各プロセッサのコア周波数は最も速いプロセッサの周波数に合わせて設定する必要があります。
- 正
- 誤
11. キャッシュサイズの異なる Xeon プロセッサを混用する場合、同じキャッシュサイズを持つプロセッサ同士をペアでインストールする必要があります。
- 正
- 誤
12. 最も高いステッピング番号を持つプロセッサを、ブートストラップ プロセッサとしてインストールする必要があります。
- 正
- 誤

目的

この章では、次の内容について学習します。

- メモリ(DRAM)とキャッシュ(SRAM)の特性および違い
- メモリへのデータの格納方法および読み出し方法
- メモリレイテンシと、メモリ性能を向上させる次の DRAM テクノロジ
 - FPM と EDO RAM
 - SDRAM と DDR RAM
 - デュアルインタリーブ メモリ
- ハードメモリエラーとソフトメモリエラーの違い、および次のフォールトトレランス メモリテクノロジの詳細
 - パリティ
 - ECC
 - アドバンスド ECC
 - オンラインスペア メモリ
 - ホットプラグ ミラーメモリ
 - ホットプラグ RAID メモリ
- キャッシュメモリの機能、キャッシュレベルと配置、キャッシュラインと Tag RAM の定義
- 次のキャッシュテクノロジの詳細と、各テクノロジが性能に及ぼす影響
 - ルックアサイドとルックスルー
 - ライトスルーとライトバック
 - バスヌーピングとバススナッフイング
 - フルアソシアティブ、ダイレクトマッピング、セットアソシアティブ

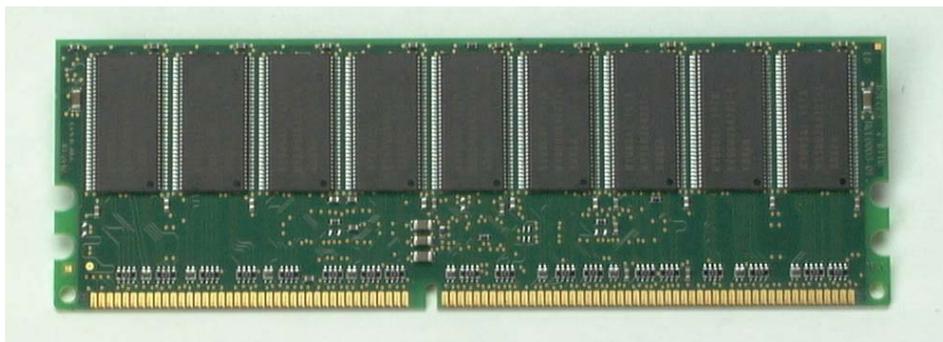
メモリとキャッシュ



基本的なコンピュータシステム

サーバの性能を向上させる 1 つの方法は、すべてのクロックサイクルで、プロセッサのパイプラインに命令を送ることです。プロセッサの速度や容量は技術革新によって向上しましたが、システム設計者には、いかにしてパイプラインをいっばいにし続けるかという課題が残されました。

DRAM(メモリ)



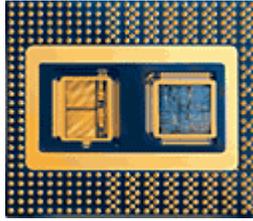
メモリモジュール

メモリモジュールでは、DRAM テクノロジを使用して、プロセッサに対するデータの格納と読み出しを行います。DRAM モジュールでは、データのビットがコンデンサに格納されます。コンデンサとは、電荷を蓄えることができる電子素子です。電荷を帯びたコンデンサはバイナリの 1 に相当し、電荷のないコンデンサはバイナリの 0 に相当します。

DRAM の欠点

DRAM (dynamic RAM) という名前は、コンデンサがいつまでも負荷を蓄えておくことができないことを表しています。放っておけば、コンデンサの負荷は数ミリ秒でなくなります。データを保持するためには、コンデンサは 1 秒間に何千回も再充電する必要があります。再充電中は、そのメモリチップを使用してプロセッサとの間でデータを送受信することはできません。

SRAM(キャッシュ)



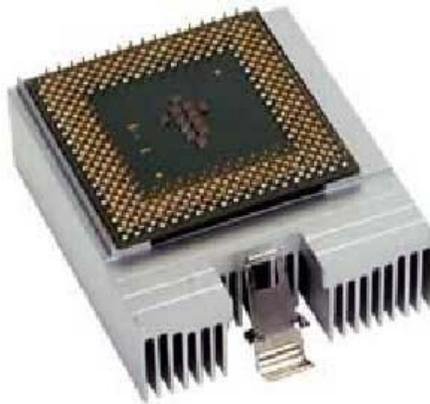
512KB のキャッシュ(左側)を内蔵した Pentium Pro プロセッサ

DRAM の欠点を克服するためにキャッシュが開発されました。キャッシュでは、フリップフロップ回路(ラッチ)として配置された一連のトランジスタを使用してデータを格納します。ラッチとは、2つの状態が交互に切り替わる電子回路です。電流を加えると、ラッチはもう一方の状態に切り替わります。

キャッシュが SRAM (static RAM) と呼ばれるのは、ラッチはコンデンサのようにリフレッシュする必要がないためです。このため、キャッシュはメモリよりも高速で動作することができます。

エンジニアはこれまで、プロセッサとメモリとの間に数層のキャッシュを追加してきました。また、キャッシュをマザーボードからプロセッサ自体に移す方法も開発しました。たとえば、Itanium プロセッサには、3レベルのキャッシュが内蔵されています。

SRAM の欠点

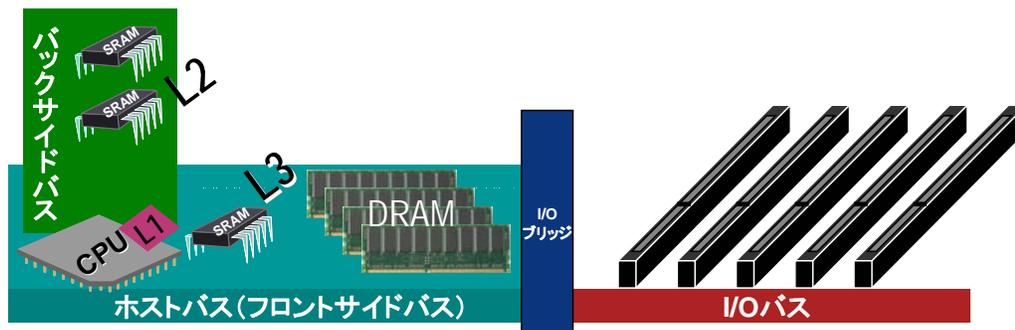


ヒートシンク上に設置された FC-PGA プロセッサ

高速なキャッシュにも欠点があります。まず、SRAM は DRAM よりも高価です。DRAM では、1つのビットは1つのコンデンサに格納されますが、SRAM では、1つのビットを格納するのに6~8個のトランジスタが必要になります。トランジスタの数が多いほど、チップの製造コストは高くなります。

2つ目に、ラッチ内のトランジスタはキャッシュを非常に高温にします。このため、熱を逃がすヒートシンクなどの特別な冷却メカニズムを開発する必要があります。これも、SRAM のコストを引き上げる要因です。

メモリとキャッシュのコスト比較

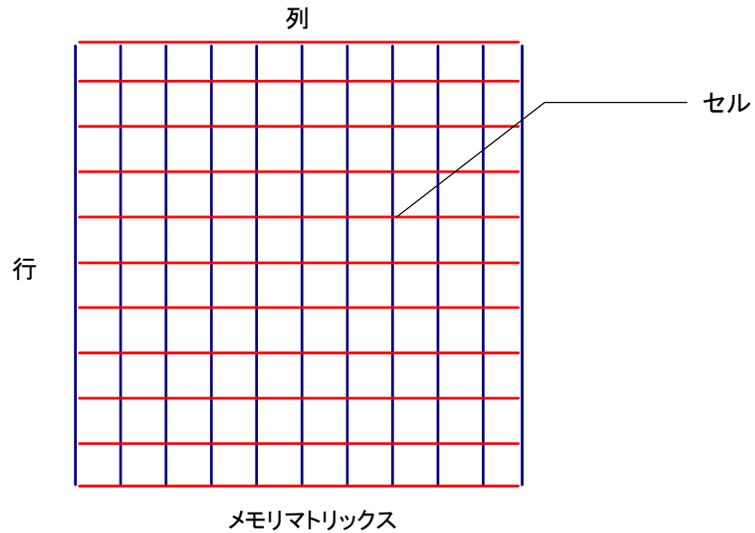


L1のコスト = \$\$\$、サイズ = 小
L2のコスト = \$\$、サイズ = 中
メインメモリのコスト = \$、サイズ = 非常に大きい

顧客のニーズに最適なサーバを選択する際には、データをどれだけ迅速にプロセッサに送る必要があり、それにどれくらいのコストがかかるのかを天秤にかけなければなりません。

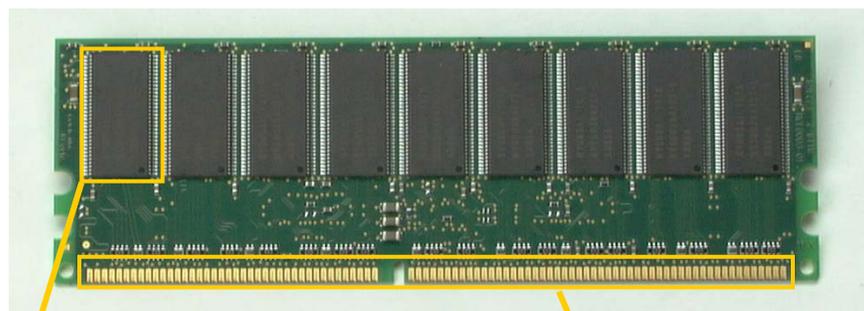
この章では、メモリとキャッシュの両方のしくみについて学習します。また、メモリとキャッシュの性能を向上させた技術革新およびフォールトトレランス機能についても学びます。

メモリ



DRAM モジュールでは、各データビットはそれぞれ 1 つのコンデンサに格納されます。DRAM モジュール上のコンデンサは、メモリマトリックスまたはメモリチップと呼ばれるグリッド状に配列されています。グリッドの行と列は、チップにエッチングされた導電線です。行と列の交点をセルと言います。行アドレスと列アドレスによって識別される各セルには、トランジスタとコンデンサがそれぞれ 1 つずつ含まれています。

SIMM と DIMM



メモリチップは、このプラスチック製のキャップの下にある

金メッキピンコネクタ

DIMM

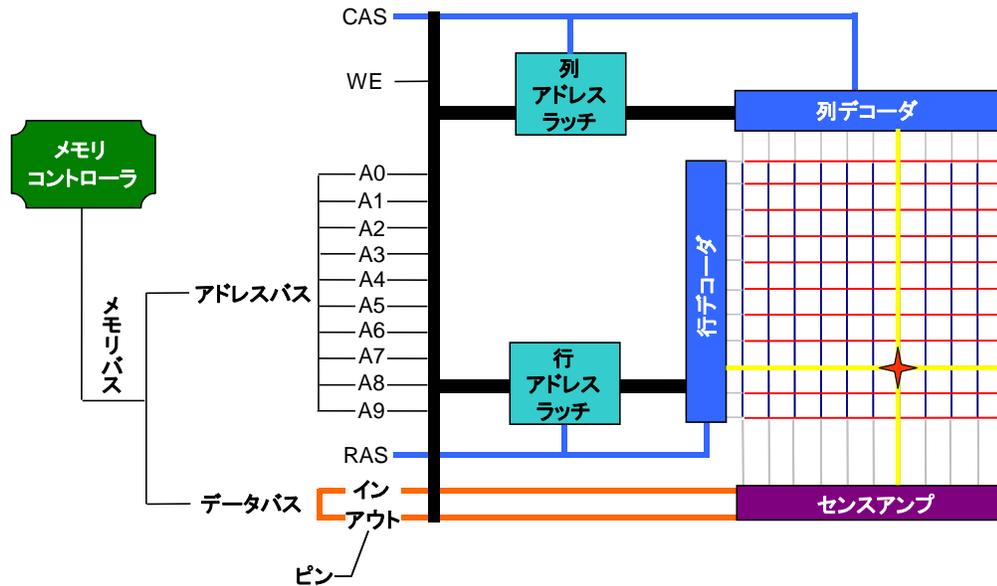
当初は、メモリチップはコンピュータのマザーボードに直接取り付けられていました。大容量のメモリに対する需要が高まり、マザーボード上のスペースが少なくなるにつれて、メモリチップはメモリモジュールと呼ばれるカードに取り付けられるようになりました。

メモリチップの電気線は、メモリチップからメモリモジュールの一番下にある金メッキピンへと配線されています。これらのピンは、マザーボードまたはメモリボード上のスロットに挿入されます。

最初に開発されたメモリモジュールの1つが、シングルインラインメモリモジュール(SIMM)でした。メモリチップはSIMMの片側または両側に取り付けられていますが、両側の場合も、すべての電気信号が同じ金メッキピンのセットに配線されます。SIMMには、30ピンと72ピンという2つのフォームファクタがありますが、一度に処理できるデータは32ビットのみです。現在、ほとんどのサーバは一度に64ビットのデータを処理するように設計されているため、SIMMを使用する場合は、2つ1組で設置する必要があります。

現在のほとんどのサーバでは、デュアルインラインメモリモジュール(DIMM)が使用されています。DIMMは両側にメモリチップを搭載しており、両側にそれぞれ72本のピン(合わせて144本)があります。168ピンのDIMMもあります。残りのピンは、エラー検出および修正(ECC)機能のためのものです。

メモリのしくみ



メモリコントローラとメモリモジュールの金メッキピンとの間の経路をメモリバスと言います。メモリバスは、アドレスバスとデータバスから構成されます。一般的なメモリモジュールには、アドレスピンとデータピンのほかに、メモリコントローラとの通信に使用されるピンがあります。行アドレスストロブ (RAS) および列アドレスストロブ (CAS) は、メモリが設定するセルのアドレスを伝えるために使用されるピンです。これらのピンには通常、高い電圧がかかっています。メモリコントローラは、RAS と CAS の電圧を変えることにより、メモリチップと通信します。

メモリコントローラは、ライトイネーブル (WE) ピンを使って、操作が読み出しと書き込みのどちらであるのかを示します。

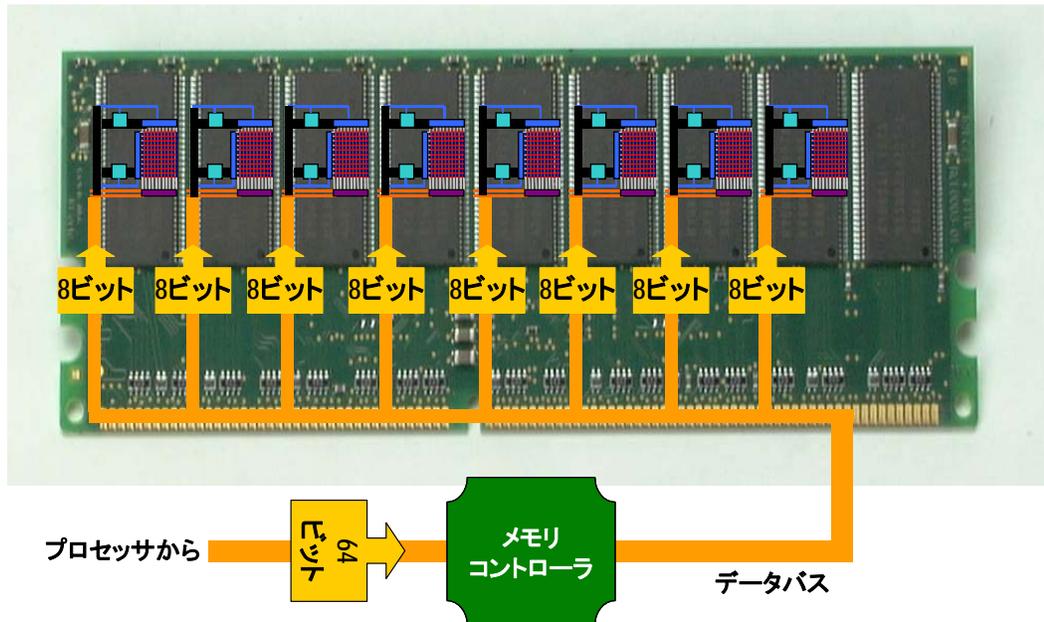
メモリチップのしくみ

データの読み出しまたは書き込みを行うには、メモリコントローラはアドレスピンを通して、行アドレスを RAS に、列アドレスを CAS にそれぞれ送信します。

書き込み操作では、指定されたアドレスの交点にあるセルに、データインピン経由でデータが送られます。

読み出し操作では、センスアンプによって、交点のセルの電荷が測定されます。電荷を帯びている場合は、センスアンプはデータアウトピンを 1 にします。電荷を帯びていない場合は、データアウトピンを 0 にします。

プロセッサから送信された1バイトのデータが格納されるしくみ



メモリセルは1ビットのデータしか格納できませんが、プロセッサはバイト単位でデータを操作するように設計されています。データをメモリに格納する必要がある場合、プロセッサは1バイトのデータとアドレスをシステムバス経由でメモリコントローラに送ります。

メモリコントローラは1バイトのデータを8ビットに分解し、プロセッサから送信されたアドレスを使って行アドレスと列アドレスを決定します。メモリコントローラは8つのビットをそれぞれメモリモジュール上の異なるメモリチップに送信します。8つのビットは、異なるチップ上で同じアドレスを持つことになります。

プロセッサがデータを要求すると、メモリコントローラは同じ要求を各チップに送り、データのビットを受け取ると、バイト単位に組み立て直します。

メモリ性能の向上

プロセッサの速度は飛躍的に向上しましたが、メモリの速度はそれほど向上していません。メモリ性能を向上させる手段の1つは、メモリレイテンシを短縮することです。

メモリレイテンシ

メモリレイテンシとは、DRAM モジュールがデータをメモリコントローラに戻すためにかかる時間です。主に次の2つの要因がメモリレイテンシに影響を及ぼします。

- **アクセス時間** — 行が RAS によって有効化されてからデータがデータバス上に読み込まれるまでにかかる時間
- **サイクル時間** — 2つの読み出し操作間にかかる時間

アクセス時間

読み出し操作の開始時に、メモリコントローラはRAS電圧を下げます。これにより、行アドレスラッチ上に行アドレスが設定されます。RASラインが規定の電圧レベルまで下がるのにかかる時間を t_{RAC} と言います。 t_{RAC} が長いほど、アクセス時間は長くなります。

その後、メモリコントローラはCAS電圧を下げます。これにより、列アドレスラッチ上に列アドレスが設定されます。CASラインが規定の電圧レベルまで下がるのにかかる時間を t_{CAC} と言います。 t_{RAC} と同様、 t_{CAC} が長いほど、アクセス時間は長くなります。

サイクル時間

読み出し操作の終了時に、メモリコントローラは操作の終わりを示すために、RASラインおよびCASライン上の電圧を上げます。どちらかのラインの電圧を規定のレベルまで引き上げるのにかかる時間を「プリチャージ遅延」と言います。両方のラインの電圧が規定レベルに達するまで、次の操作は実行できません。プリチャージ遅延が長いほど、サイクル時間は長くなります。

ファストページモード (Fast Page Mode)

レイテンシを解消しようと試行錯誤を繰り返す中で、エンジニアは、プロセッサがデータを要求する場合、その次の要求は通常、同じ行 (ページ) 内のデータに対するものであることを学びました。プロセッサへのデータの送信を速くするために、ファストページモード (FPM) と呼ばれる新たな手法が開発されました。FPM では、プロセッサがデータを要求すると、メモリコントローラは、そのデータのほかに、各メモリチップからさらに 3 つのデータ列を要求します。

メモリコントローラはデータを、1 バイトの塊 4 つにまとめます。この 4 つの塊をワードと言います。メモリコントローラはワードをキャッシュに送信 (バースト) します。キャッシュは受け取ったワードをキャッシュラインに格納し、要求されたデータバイトをプロセッサに送信します。

プロセッサからの次の要求が同じページ上にあるデータに対するものであれば、要求されたデータをキャッシュラインからただちに送信することができます。これを「ページヒット」と言います。要求されたデータが同じページになければ、データはメモリから取り出されます。これを「ページミス」と言います。ページミスが発生すると、データの取得にかかるサイクル数は 2 倍以上になります。

FPM は、データの最後の 3 ビットに対する RAS のサイクル時間をなくすことにより、レイテンシを短縮します。

Extended Data Out DRAM

FPM の欠点の 1 つは、1 つの列からのデータがデータアウトピンに送信されないかぎり、次の列を有効化できないという点です。1996 年、Extended Data Out (EDO) DRAM の発表により、メモリ設計者はこの欠点を克服しました。

EDO は、データがデータアウトピン上にある間に次の列を有効化することにより、レイテンシを短縮します。DRAM のタイミングが変わることにより、メモリモジュールからメモリコントローラにデータが到達するのにかかる時間が 20~30% 短縮されます。

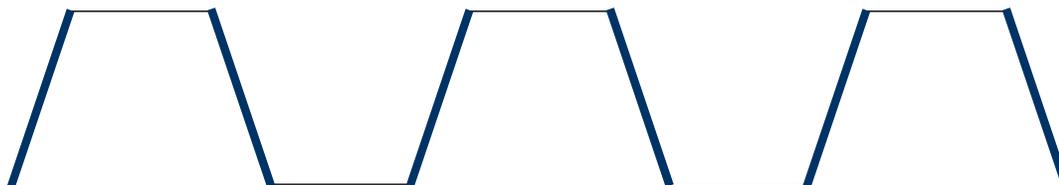
Synchronous DRAM

FPM と EDO はどちらもシステムクロックに従って動作しないため、非同期テクノロジーと呼ばれます。どちらも、メモリからのデータの読み出しおよび書き込みとシステムクロックとの調整を取る独自のタイミングメカニズムを持っています。

1997 年に、synchronous DRAM (SDRAM) が発表されました。SDRAM は、メモリバスと同じクロックを使用します。このテクノロジーの開発により、特別なタイミングメカニズムは不要になりました。

SDRAM の DIMM は、チップのバンクを複数含むという点でも asynchronous (非同期) RAM と異なります。メモリコントローラは、1 つのバンクでデータを受信しながら、別のバンク内の行をプリチャージすることができます。このプロセスにより、データが使用可能になるまでコントローラが待機しなければならない時間が短縮されます。

Double Data Rate (DDR) RAM

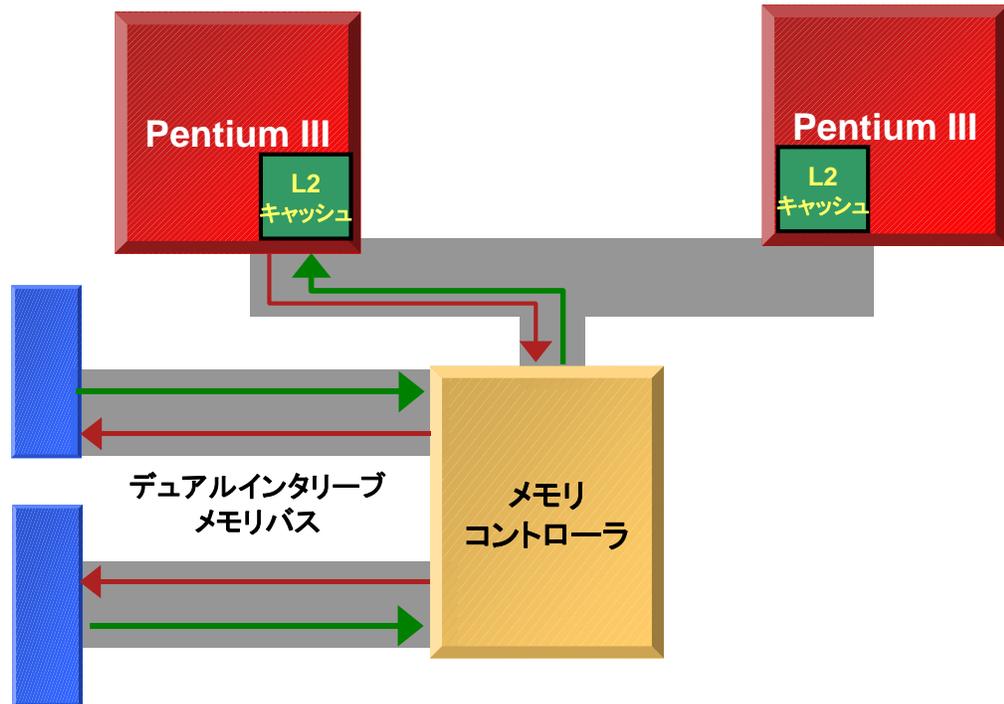


データは各クロックサイクルの立ち上がりエッジまたは立ち下がりエッジで転送される

クロックサイクルは、立ち上がりエッジと立ち下がりエッジ、およびエッジ間にある水平域から構成される波形として表すことができます。

SDRAM DIMM では、データはクロックサイクルの立ち上がりエッジでのみ転送されます。2002 年、HP は Double Data Rate (DDR) RAM を搭載した ProLiant ML530 G2 と DL580 G2 を発表しました。DDR RAM では、データは各クロックサイクルの立ち上がりエッジと立ち下がりエッジの両方で転送されます。つまり、1 つのクロックサイクルで、プロセッサとの間で 2 倍のデータを送受信することが可能です。

デュアルインタリーブ メモリ



デュアルインタリーブ メモリテクノロジーは、1回のメモリアクセスで取得できるデータ量を、64ビットから128ビットへと2倍に増やします。この結果、レイテンシは大幅に短縮されます。

データがメモリに書き込まれるとき、メモリコントローラは、2つのバンクにあるDIMMにデータを分配(インタリーブ)します。プロセッサがメモリコントローラに要求を送信すると、メモリコントローラは両方のDIMMに同時に要求を送信します。両方のDIMMから、要求されたアドレスにあるデータと、それに続く連続したアドレスにあるデータが返されます。

連続したアドレスからのデータは、将来必要になることを見越してプロセッサのL2キャッシュに送られ、プロセッサのキャッシュラインに格納されます。それ以降の要求に対し、プロセッサは高速なL2キャッシュを使用し、必要なデータが見つからなくなった時点で、新たな要求をメモリコントローラに送ります。

デュアルインタリーブ メモリを使用することにより、L2キャッシュはプロセッサからのデータ要求の98%以上に応えることができます。これには2つの利点があります。

- プロセッサとL2キャッシュ間のバスは、プロセッサの速度で動作します。一方、L2キャッシュとメモリ間のバスは、それよりもかなり遅い速度で動作します。デュアルインタリーブ メモリでは、L2キャッシュがより速く充填されるため、プロセッサはキャッシュを迅速に充填できます。
- ほとんどの要求がL2キャッシュから充填されるため、メモリコントローラに送られる要求が少なくなります。この結果、2基目のプロセッサがメモリコントローラへのアクセスを待つことが少なくなります。

フォールトトレランス メモリ

メモリは電子記憶装置であるため、元々格納された情報とは異なる情報を返す可能性があります。

一般的にメモリシステムでは、ハードエラーとソフトエラーという 2 種類のエラーが発生することがあります。

ハードエラー

ハードエラーは、メモリモジュールの欠陥や接触不良などの物理的な問題が存在する場合に発生します。問題を修復しないかぎり、ハードエラーは発生し続けます。

HP は、顧客が遭遇するハードエラーの数を減らすために、自社のサーバに搭載するすべてのメモリモジュールをテストしています。

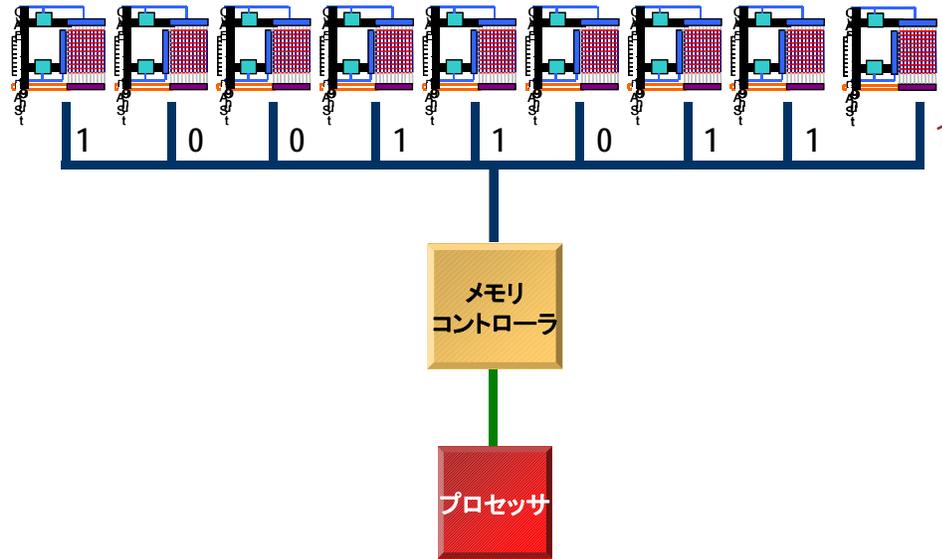
ソフトエラー

ソフトエラーは、メモリセル付近の電気障害によってコンデンサの電荷が変わることにより不規則に発生します。ソフトエラーはメモリデバイスの問題によるものではないため、データが修正されると、同じエラーは再発しません。

ソフトエラーは、ハードエラーよりも頻繁に発生します。調査によれば、メモリの容量が増えると、ソフトエラーの数も増えます。それぞれ 64MB のメモリを搭載した 10,000 台のコンピュータで年間 460 件のソフトエラーが発生するという調査結果に基づけば、1 台のコンピュータに搭載された 4GB のメモリに付き、年間で約 3 件のソフトエラーが発生することになります。

HP は、ソフトエラーを検出し、多くの場合、修正するための多くのテクノロジーを使ってきました。

パリティ

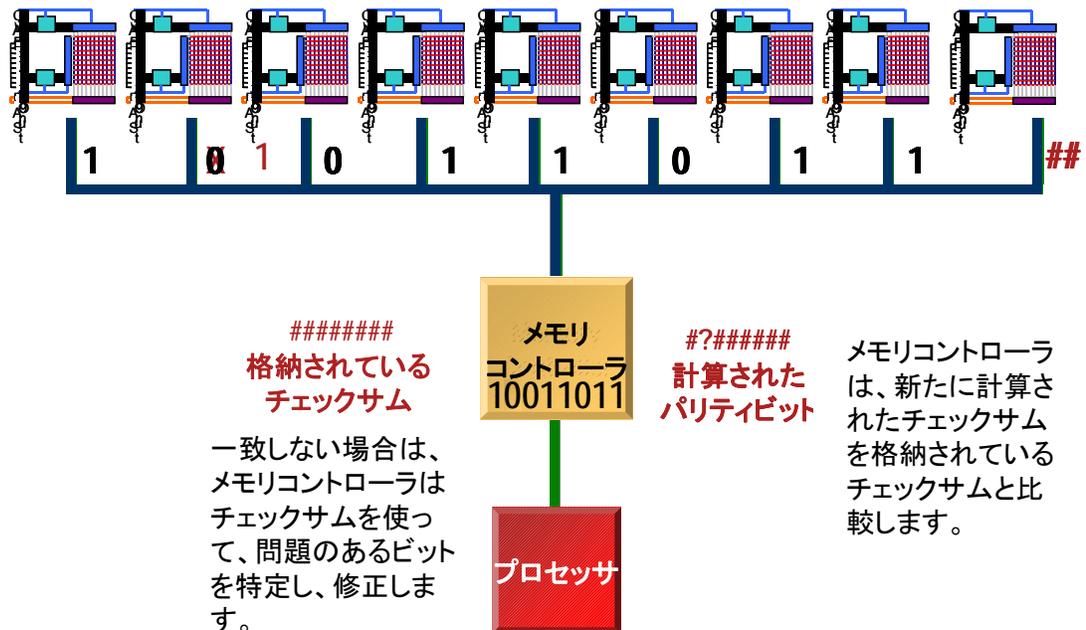


HP が最初に使用したエラー検出テクノロジーは、パリティチェックでした。これは元々、デスクトップコンピュータ用に開発されたテクノロジーです。パリティチェックを使用するシステムは、データをメモリに書き込む際に各バイトに1つのパリティビットを追加します。システムは、バイト内の1の数を数え、偶数であればパリティビットを0に、奇数であればパリティビットを1に設定します。

システムがメモリからバイトを読み出すとき、メモリコントローラはパリティビットを再計算し、格納されているパリティビットと比較します。バイト内の1つのビットが変更されていれば、新しいパリティ値は格納されているパリティ値と一致しません。これにより、コントローラはソフトウェアが発生したことを認識し、データ破壊を防ぐためにシステムをシャットダウンします。エラーはログには記録されません。

パリティチェックでは多くのエラーが検出されますが、欠点もあります。パリティチェックでは、奇数個の1ビットエラーしか検出できません。同じバイト内で2件のソフトウェアが発生した場合、バイトは変更されているにもかかわらず、パリティビットは正しくなります。また、パリティチェックではどのビットが不正かわからないため、そのビットを修正することはできません。

ECC



ECC メモリは、1992 年に Compaq Systempro XL サーバに初めて採用され、すべての ProLiant サーバで使用されています。ECC メモリは、1 バイトあたり 2 ビットまでのエラーを検出できます。さらに、1 ビットのエラーを修正することが可能です。

ECC システムはパリティシステムに似ていますが、1 つ大きな違いがあります。メモリに書き込まれた各データバイトに対してパリティビットを計算する代わりに、ECC システムでは、64 ビットのデータごとに 72 ビットのシンδροームを計算します。シンδροームは、64 ビットのデータと 8 つのチェックビットから構成されます。

読み出し操作では、メモリコントローラがシンδροームを再計算し、元のシンδροームと比較します。2 つのシンδροームが一致しない場合、エラーが発生したとわかります。システムはシンδροームを使ってどのビットが変更されているかを判断し、要求元のデバイスにデータを送る前に、エラーを修正します。

ソフトエラーにはマルチビットエラーもあります。従来の ECC ではこの種のエラーを修正することはできず、ECC システムが機能なくなる原因となります。マルチビットエラーが発生する確率は、メモリ容量の増加とともに大きくなります。サーバの世代が変わるごとにメモリ容量も増加し、それに伴ってシステム障害の危険性も高まります。

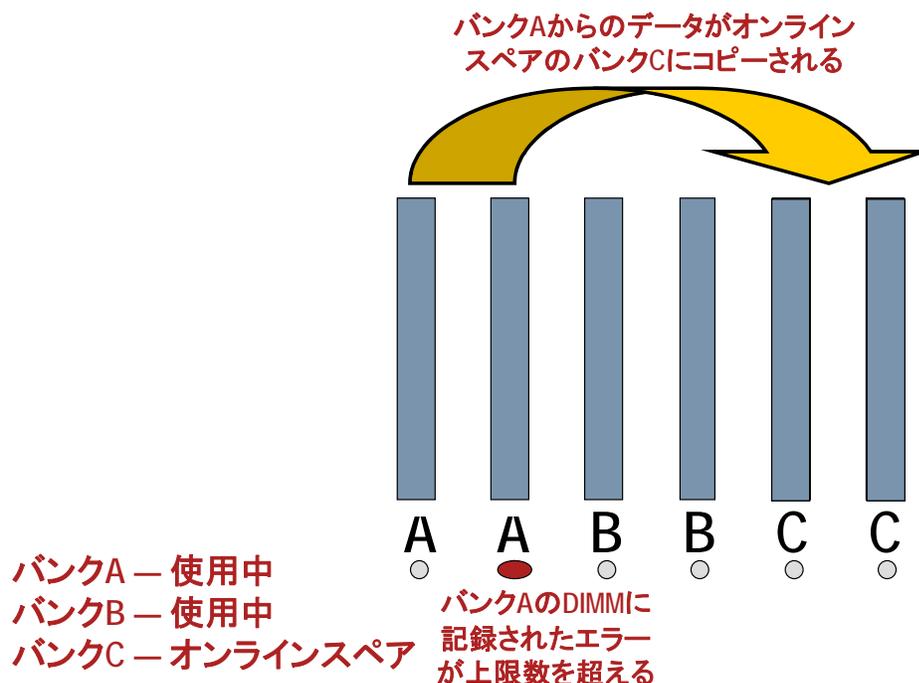
アドバンスト ECC

メモリ保護機能を向上させるために、Compaq は 1996 年にアドバンスト ECC テクノロジーを発表しました。HP およびその他のサーバメーカーは、現在の製品ラインでもこのソリューションを使用しています。

通常の ECC デバイスは、DIMM からの読み出し時に 1 ビットエラーを修正できます。一方、アドバンスト ECC は、1 つの DRAM チップ内で発生したマルチビットエラーを修正できるため、DRAM チップ障害を完全に修正できます。

アドバンスト ECC はメモリを障害から保護しますが、マルチビットエラーを確実に修正できるのは、そのエラーが 1 つの DRAM チップ内で発生した場合のみです。フェイルオーバ機能やホットプラグ機能は提供しません。

オンラインスペア メモリ



以前のサーバでは、メモリモジュールで大量の修正可能な1ビットエラーが発生した場合、システムは障害前の警告を発していました。このような状態が発生した場合に取るべき手段は、すぐにサーバの電源を切り、問題のあるDIMMを交換し、システムを再起動するというものでした。

オンラインスペアメモリでは、問題のあるDIMMを含むメモリバンクが自動的にスペア(予備)のDIMMバンクにフェイルオーバーします(切り替わります)。

1組のDIMMがスペアバンクとして機能します。他のスロットには、最大2つのメモリバンクを設置できます。

あるバンクのDIMMが規定の1ビットエラーしきい値に達すると、システム管理ドライバの指示によって、ROMが問題のあるバンクの内容をスペアバンクに128KB単位でコピーし始めます。この間、読み出しアクセスは問題のあるバンクに対して行われます。一方、データの書き込みは両方のバンクに対して行われます。

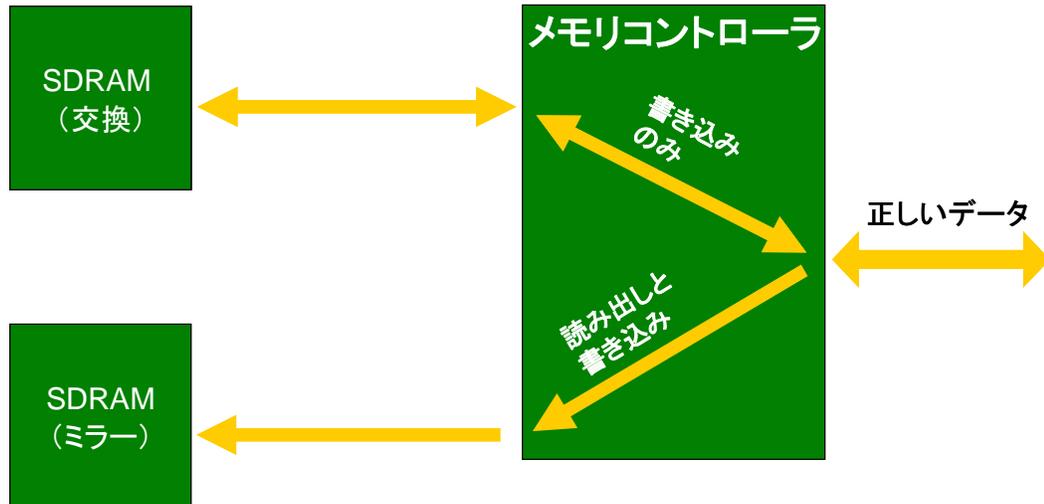
注

▲ スペアバンクは、問題のあるバンク内のすべての情報を格納できなければならないため、スペアバンク内のDIMMは、他のバンクと同じサイズかそれ以上でなければなりません。

メモリのコピーが完了すると、ROMはバンクを切り替えるようにシステム管理ドライバに伝えます。これ以降の読み出しおよび書き込みは問題のあるバンクではなく、スペアバンクに対して行われるようになります。

スケジュールされたシャットダウン時に、問題のあるDIMMを正常なDIMMと交換します。サーバが再起動すると、メモリバンクは元の機能に戻ります。

ホットプラグ ミラーメモリ



ホットプラグ ミラーメモリは、オンラインスペア メモリよりも高い可用性を実現する、フォールトトレランス メモリオプションです。ホットプラグ ミラーメモリは、1ビットエラーとマルチビットエラーの両方からメモリを保護します。このメモリは、問題のある DIMM を交換するためにサービスを中断することが困難な顧客のために設計されたものです。

ホットプラグ ミラーメモリを使用するには、サーバには複数の DIMM バンクを持つ同一のメモリボードが 2 つ必要です。メモリコントローラは、同じデータを、両方のメモリボード上の同様に構成された DIMM バンクに書き込みます。ただし、データの読み出しは、1 つのグループからのみ行います。

いずれかの DIMM バンクでマルチビットエラーが発生した場合、システムは次の処置を行います。

1. もう一方のメモリボード上のミラー化されたバンクから、正しいデータを読み出し直します。
2. それ以降のすべての読み出しを、もう一方のメモリボードから行います。
3. メモリボードの LED、フロントパネルの診断 LED、および Insight マネージャ 7 を通して、DIMM の障害を通知します。

両方のバンクのセットでメモリエラーの有無を監視するため、エラーが発生しなくても、システムは読み出し元のバンクのセットを定期的に切り替えます。

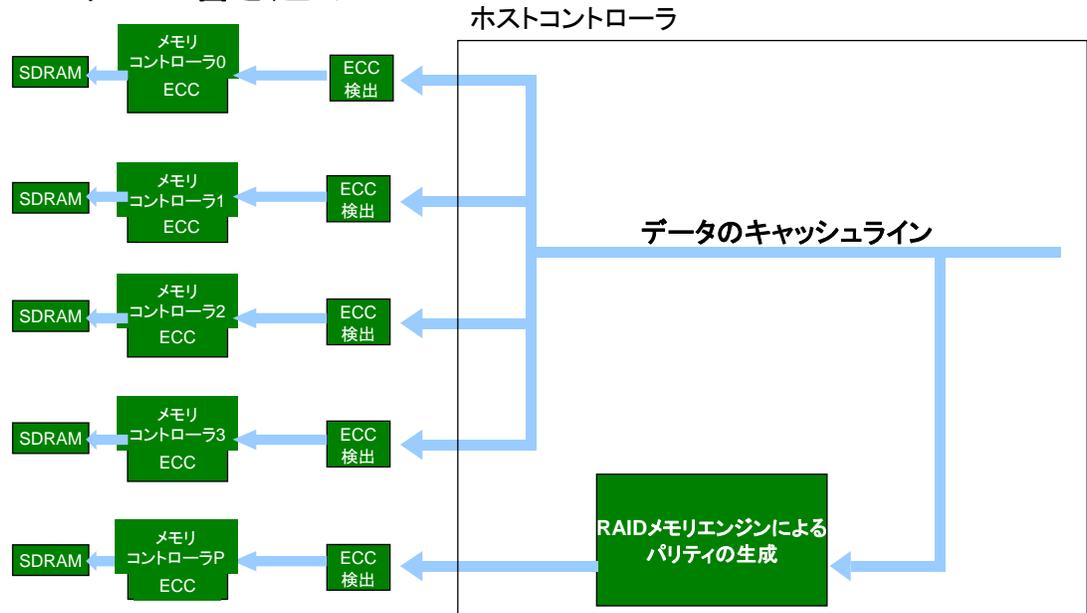
**重要**

DIMM が、HP によって規定されている修正可能な 1 ビットエラーのしきい値を超えた場合、システムはスペアバンクにフェイルオーバーせずに、Insight マネージャ 7 を通してその状態を通知します。

ホットプラグ ミラーメモリには、ホットプラグ交換機能も備わっています。つまり、サーバをシャットダウンせずに、障害が発生したメモリボードまたは問題のあるバンクを交換できます。また、DIMM バンクを交換した後、サーバが稼動中でも、メモリボードをインストールし直すことができます。ボードが再インストールされると、システムは自動的にミラー化状態に戻ります。また、スケジュールされたシャットダウン時に、問題のある DIMM を交換することもできます。

ホットプラグ RAID メモリ

メモリへの書き込み



もともと RAID という用語は、フォールトトレランス機能を備えたハードディスクドライブ テクノロジーを表すために使用されていたものです。ストレージ用の RAID については、この後の章で詳しく学習します。現在は、同じ RAID 理論が DIMM にも適用されています。HP では、これをホットプラグ RAID メモリと呼んでいます。

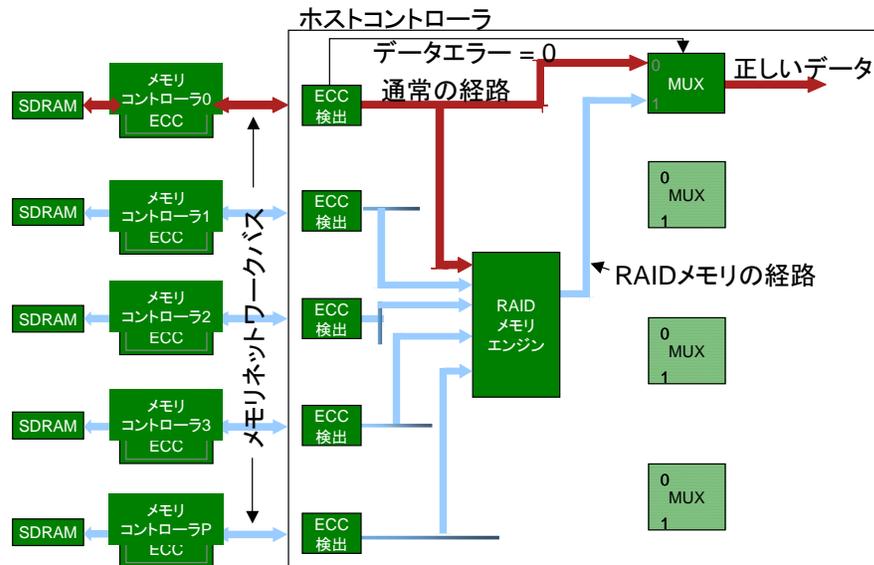
ホットプラグ RAID メモリは、メモリサブシステムが完全な DIMM 障害に耐え、通常どおり動作することを可能にします。

ホットプラグ RAID メモリテクノロジーを使用するサーバは、5 つのメモリコントローラを使って 5 つの SDRAM DIMM を制御します。これらのメモリコントローラは、HP が設計した次世代チップセットに搭載されており、それぞれが特定用途向け集積回路 (ASIC) です。

メモリへのデータの書き込みでは、データの 1 つのキャッシュラインが 4 つのブロックに分割されます。メモリコントローラの内の 4 つが、関連付けられた DIMM にそれぞれ 1 ブロックのデータを書き込みます。RAID エンジンがパリティ情報を計算し、5 つ目のメモリコントローラがパリティ情報を 5 つ目の DIMM に格納します。

正しいデータ

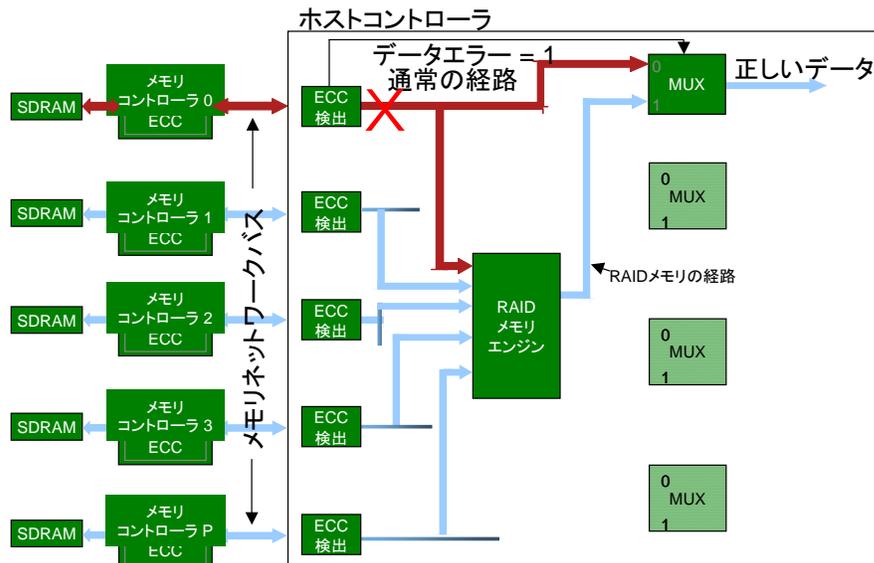
メモリからの読み出し—正しいデータ



メモリエラーが発生しなかった場合、データは通常の経路でプロセッサに送られます。

不正なデータ

メモリからの読み出し—不正なデータ



メモリエラーが発生した場合、またはいずれかの DIMM が取り外された場合、不正なデータまたは足りないデータは、残りの 4 つの DIMM 上にあるデータから計算し直すことができます。

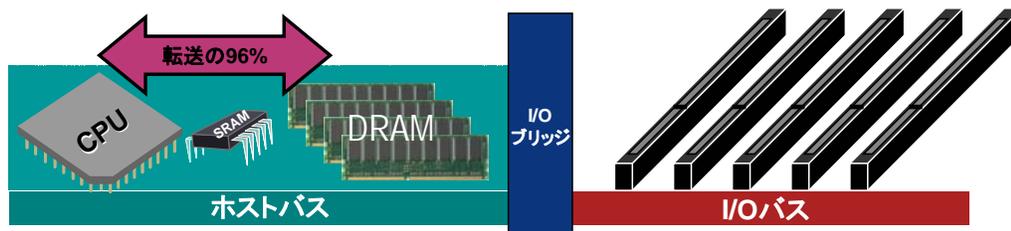
キャッシュとハードディスクドライブ アレイの比較

概念上、ホットプラグ RAID メモリはハードディスクドライブ アレイに似ていますが、いくつかの相違点があります。

ホットプラグ RAID メモリでは、ハードディスクドライブ アレイのようにシーク時間による機械的な遅延や回転遅延は発生しません。ストレージサブシステム アレイは、1 つのバスを使って、複数のドライブに順番にストライプを書き込みます。一方、ホットプラグ RAID メモリは 5 つの並行接続を使用するため、データは複数の DIMM に同時に書き込まれます。

また、ホットプラグ RAID メモリでは、通常のストレージサブシステムの RAID 実装で発生する書き込みボトルネックは発生しません。ストレージアレイでは、一般に RAID コントローラは既存のパリティの読み取り操作を行ってからでないと書き込み操作を終了することはできません。このため、専用のパリティドライブが使用されている場合、ボトルネックが発生します。これに対し、ホットプラグ RAID メモリはデータのキャッシュライン全体を一度に操作するため、書き込み操作の前に既存のパリティを読み取る必要はなく、このようなボトルネックは発生しません。

キャッシュ

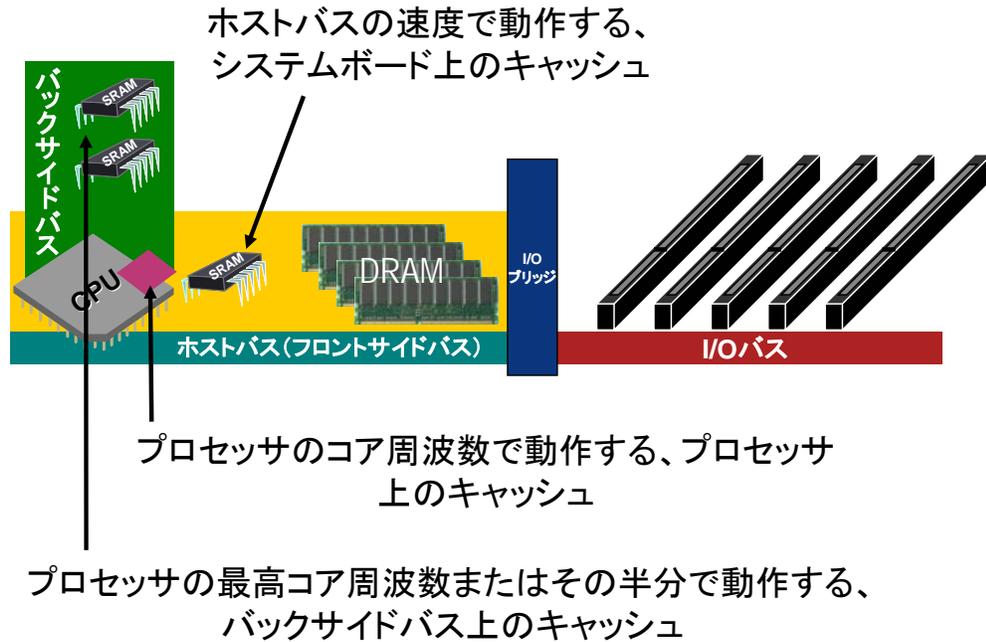


ハードディスクドライブやテープからのデータの取り出しと比べれば、メモリはかなり高速ですが、プロセッサのパイプラインを常に一杯にしておくことはできません。このためキャッシュが使用されます。

メモリと同様、キャッシュではデータはグリッド状に格納されます。このグリッドは、コンデンサではなくラッチで形成されています。ラッチはリフレッシュする必要がないため、キャッシュはメモリよりもかなり高速でデータの読み書きを行うことができます。

キャッシュは、データの局所性に基づいて機能します。つまり、プロセッサがデータを要求した場合、その次の要求が最初のデータの近くにあるデータに対するものである可能性は非常に高くなります。このため、プロセッサがメモリに対してデータを要求すると、メモリは要求されたデータに加えて、近くのデータも送り返します。送られたすべてのデータがキャッシュに格納されます。このようにすることにより、それ以降のデータ要求を、メモリからではなく、より高速なキャッシュから充填することが可能になります。

キャッシュレベルと配置



もともとキャッシュはマザーボード上の、プロセッサとメモリの間に置かれていました。プロセッサには、システムバスを通して接続されていました。その後、キャッシュはプロセッサ自体に組み込まれるようになりました。現在のプロセッサは通常、複数のキャッシュを持っています。一次キャッシュはL1、二次キャッシュはL2というように呼ばれます。

L1は最も高速なキャッシュですが、格納できるデータ量は最も少なくなります。直接プロセッサに接続されているため、プロセッサと同じ速度で動作します。最高10nsでデータを配信できます。

L2キャッシュも当初はマザーボード上に置かれており、システムバス経由でプロセッサに接続されていました。つまり、L1キャッシュよりは低速でしたが、メモリよりは高速でした。現在、ProLiantサーバで使用されているすべてのプロセッサでは、プロセッサの一部としてL2キャッシュが組み込まれています。L2キャッシュはバックサイドバスを通してプロセッサに接続されています。バックサイドバスの速度は、プロセッサに依存し、プロセッサの速度からシステムバスの速度までの間の速度で動作することができます。

L2キャッシュがプロセッサパッケージに移動すると、マザーボード上のキャッシュはL3キャッシュと呼ばれるようになりました。L3キャッシュはシステムバスの速度(メモリバスよりも高速)で動作します。現在のItaniumプロセッサでは、L3キャッシュもプロセッサ上に置かれています。

キャッシュレベルが上がるごとに、キャッシュの容量は大きくなります。L2には、L1の内容すべてが格納されており、さらに追加のデータを格納できます。L3には、L2の内容すべてが格納されており、さらに追加のデータを格納できます。L2とL3は、20~30nsでデータを配信できます。これに対し、メインメモリはメモリバスの速度で動作するため、約60nsでデータを配信します。一方、ハードディスクドライブは、データをプロセッサに返すのに、最高12msかかります。

キャッシュラインと Tag RAM

一般的に DRAM では、簡単にリフレッシュできるように行の数が少ないのに対し、SRAM にはこのような制限はありません。この結果、キャッシュは通常、多数の行と少数の列を持つように設計されます。たとえば、512KB のキャッシュは、16,384 本の行 (キャッシュライン) と 32 本の列を持ちます ($16,000 \times 32 = 512,000$)。

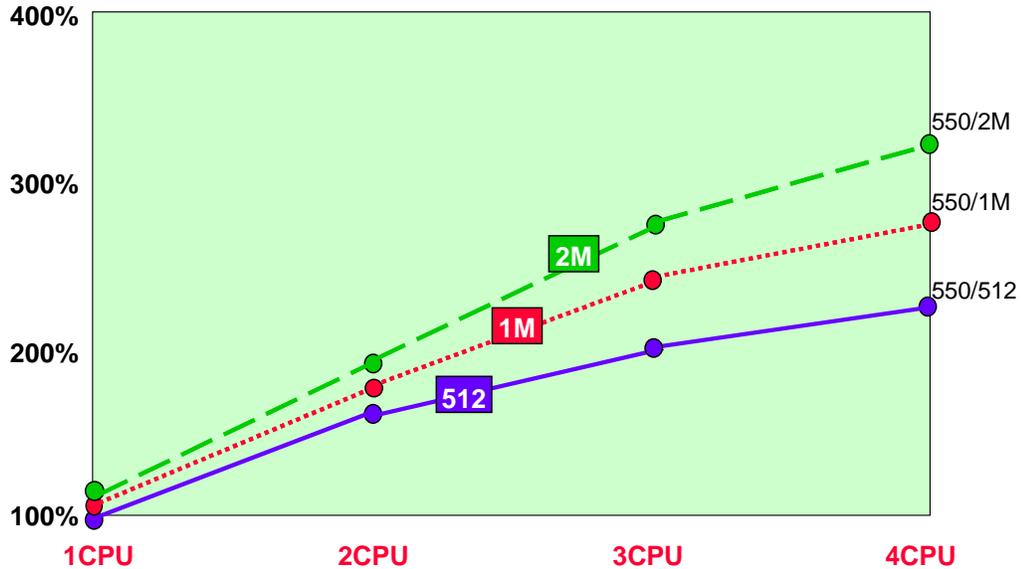
キャッシュ内の各行をキャッシュラインと言います。最新のプロセッサでは、プロセッサがデータを要求すると、メインメモリは、要求された 1 バイトのデータのほか、その後の 31 バイトをプリフェッチします。この 32 バイトの塊がキャッシュラインに格納されます。これをキャッシュラインフィルと言います。プロセッサが追加の 31 バイトに含まれるデータを要求した場合 (これはかなりの確率で発生します)、迅速にデータを取得することができます。

Tag RAM と呼ばれる小さな SRAM チップに、各キャッシュライン内のデータに対するメモリアドレスが格納されます。プロセッサがデータを要求すると、キャッシュコントローラは要求内のアドレスと Tag RAM 内のアドレスとを比較します。アドレスが見つかり、キャッシュコントローラは関連するデータをプロセッサに返します。これを「キャッシュヒット」と言います。コントローラがアドレスを見つけれない場合は「キャッシュミス」となります。

キャッシュ性能の向上

エンジニアは各クロックサイクルの効率を向上させる方法を常に模索しています。これまで、システムがデータの読み書きに要する時間を短縮するために、キャッシュとメモリの設計にいくつもの改良が加えられてきました。

L2 のキャッシュサイズが性能に及ぼす影響



1 プロセッサシステム

1 基のプロセッサを搭載したシステムでは、L2 キャッシュのサイズは、複数のプロセッサを搭載したシステムのように重要ではありません。1 プロセッサシステムでは、通常は 256KB で十分です。プロセッサからのメモリアクセスのほかは、バスマスタ PCI カードが、システムバスで競合するのみです。キャッシュサイズを 256KB から 512KB へと 2 倍に増やすと、性能は 3～5% 向上します。キャッシュサイズが大きいくほどヒット率は高くなりますが、1 プロセッサシステムでは、コストに見合った効果は得られません。

マルチプロセッサシステム

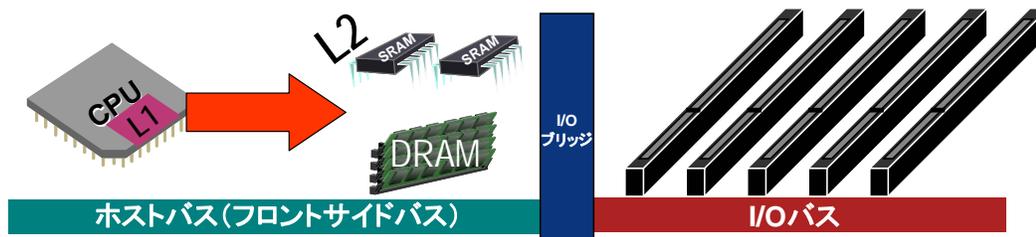
マルチプロセッサシステムでは、1 本のプロセッサバスにつき、最大 4 基のプロセッサがメモリと I/O アクセス権を競合します。システムバスは大きなボトルネックとなり、トラフィックが増えると性能は低下します。このため、プロセッサを追加しても、それに比例して性能は向上しません。

4 基のプロセッサが 1 本のバスを共用するため、メインメモリからのデータを要求するプロセッサは、システムバスにアクセスできるようになるまで待たなければならない可能性が高くなります。ただし、プロセッサに大容量のキャッシュがあれば、必要なデータのほとんどに簡単かつ迅速にアクセスすることができます。これにより、性能は大幅に向上します。

キャッシュアーキテクチャ

プロセッサからの要求をキャッシュが充填する速度は、メインメモリからのデータがキャッシュ内のどの位置に格納されているかによってある程度決まります。

ルックアサイド キャッシュ

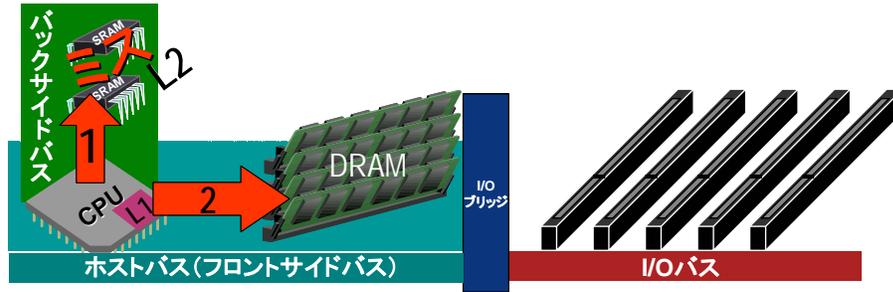


キャッシュを持たなかった初期のプロセッサは、メモリに対する読み書き要求をシステムバスに送信していました。このため、要求はメモリのみに対するものであるのに、バスに接続されているすべてのデバイスが要求を受け取り、その要求がそのデバイスに対するものであるかどうかを判別する必要がありました。一度にシステムバスを使用できるのは1台のデバイスのみであるため、メモリ要求はシステム性能を低下させました。

キャッシュが実装されてからも、プロセッサは要求をシステムバスに送信していました。つまり、キャッシュとメモリの両方が要求を受け取っていました。キャッシュヒットがあった場合は、キャッシュコントローラが、システムバス上のその他のデバイスへの要求の送信を終了しました。キャッシュミスとなった場合は、要求は通常どおり、他のデバイスにも送られました。このタイプのアーキテクチャをルックアサイドキャッシュと言います。

ルックアサイドキャッシュの利点は、キャッシュミスが発生した場合、メモリにはすでに通知されているため、要求をすばやく満たせることです。ただし、キャッシュヒットが発生した場合は、メモリは列と行のプリチャージに無駄な時間を費やしたことになり、またシステムバス上の他のデバイスもバスを使用するためにその間待機していたこととなります。

ルックスルー キャッシュ



ルックスルー キャッシュは、システムバスが受け取る要求の数を減らします。このタイプのシステムでプロセッサが読み出し要求を送信すると、要求はまずキャッシュに送られます。キャッシュヒットがあれば、要求がシステムバスに送られることはなく、他のデバイスがバスを使用することができます。キャッシュミスの場合は、キャッシュコントローラが要求をシステムバスへ送ります。

ルックスルー キャッシュの利点は、システムバスに送られるメモリ要求の数が少ないことです。一方、ルックスルー キャッシュの欠点は、キャッシュによってキャッシュミスであると判断されてからでないと、メインメモリがデータの検索を開始しないことです。この遅延は通常「ルックアップペナルティ」と呼ばれています。

ルックスルー キャッシュは、Pentium プロセッサ以降のすべての Intel プロセッサで採用されています。

キャッシュへの書き込み

システムがキャッシュおよびメモリに書き込むために使用する方法は、システムの速度に影響します。

ライトスルー キャッシュ

プロセッサはデータを書き込むたびに、L1 キャッシュを更新します。これによって L2 キャッシュが更新され、また L2 キャッシュの更新によって L3 キャッシュが更新され (L3 キャッシュがある場合)、L3 キャッシュの更新によってメインメモリが更新されます。この更新プロセスの完了には多数のクロックサイクルを要します。プロセッサが同じデータをすぐに必要としても、そのデータがメインメモリ内で更新されるまで待たなければなりません。

データを再度使用する前に、システムはすべてのメモリレベルにデータを書き込む必要があるため、このタイプのキャッシュアーキテクチャはライトスルー キャッシュと呼ばれます。

ライトバック キャッシュ

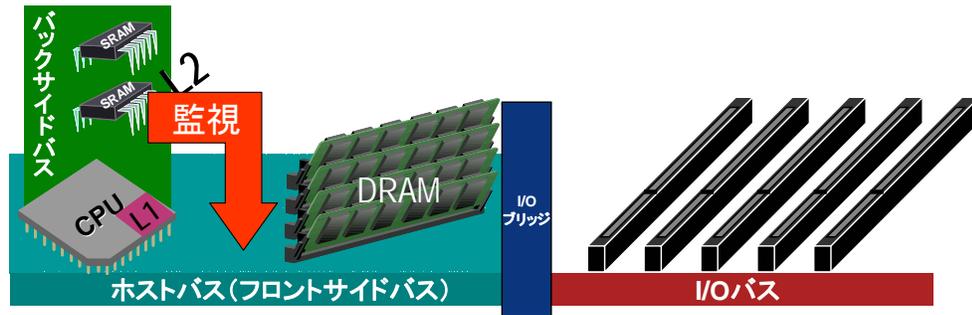
プロセッサが書き込んだばかりのデータを取り出すのに待機しなければならない時間を短縮するために、ライトバック キャッシュが開発されました。

ライトバック キャッシュでは、プロセッサは L1 キャッシュを更新しますが、メインメモリはすぐには更新されません。代わりに「ダーティビット」と呼ばれる、キャッシュラインに添付される特別なステータスビットに、データがメモリにまだ書き込まれていないことを示すフラグが付けられます。これにより、メモリバスは他の要求を受け付けることが可能になります。

プロセッサがデータを再度要求すると、キャッシュコントローラはステータスビットをチェックし、キャッシュライン内のデータがメモリ内のデータと同じであるかどうかを確認します。違っている場合は、データはプロセッサに返される前にメモリに書き込まれます。

プロセッサがデータを書き込む必要がある場合は、キャッシュコントローラはステータスビットをチェックし、最後に使用されてから最も経過時間が長いキャッシュラインがメモリに書き込まれているかどうかを確認します。まだ書き込まれていなければ、キャッシュコントローラはメインメモリへの書き込みを開始してから、プロセッサの書き込み要求でキャッシュラインを上書きします。

バスヌーピングとバススナッピング



サーバ内のデバイスでメモリへの読み書きを試みるのは1基のプロセッサだけではありません。マルチプロセッシングシステムでは、他のプロセッサもメモリへのアクセスを試みます。新しいサーバの多くには「バスマスタ」と呼ばれるデバイスが含まれています。バスマスタはプロセッサを経由せずにメモリにアクセスすることができます。

ライトバックシステムでは、これらのデバイスの内の1つがメモリからデータを要求した場合に、内容が古い可能性があります。キャッシュには、まだメモリに書き込まれていないダーティデータが含まれているかもしれません。この場合、メモリにデータを要求しているデバイスは不正なデータを受け取る可能性があります。

このような問題が発生するのを防ぐために、キャッシュコントローラはバスマスタからのメモリ要求がないか、システムバスのトラフィックを監視するように設計されています。これを「バスヌーピング」と言います。

バスマスタがメモリからデータを読み出そうとしている場合、キャッシュコントローラは、より新しいデータがキャッシュ内にはないか確認します。新しいデータがあった場合は、キャッシュコントローラは、最新のデータでメモリを更新するまでバスマスタの要求を保留します。

バスマスタがメモリにデータを書き込もうとしている場合、キャッシュコントローラは書き込まれようとしているデータをキャプチャしてキャッシュに書き込みます。このデータのキャプチャ方法を「バススナッピング」と言います。

キャッシュからの読み出し

キャッシュがプロセッサからの要求を満たす速度は、メインメモリからのデータがキャッシュ内のどの位置に格納されているかによってある程度決まります。

フルアソシアティブ キャッシュ

メインメモリからのデータをどのキャッシュラインにも格納できるようにキャッシュを設計することができます。このようなキャッシュを「フルアソシアティブ キャッシュ」と言います。

フルアソシアティブ キャッシュでは、メモリからキャッシュへのデータの移動は迅速に行われます。最も古いデータを持つキャッシュラインが置き換えられます。ただし、後でこのデータを見つける際には時間がかかります。512KB のキャッシュでは、キャッシュコントローラは必要なデータを見つけるまでに、16,384 本のキャッシュラインをすべて検索しなければならない可能性もあります。

ダイレクトマッピング キャッシュ

ダイレクトマッピング キャッシュと呼ばれる別のキャッシュアーキテクチャでは、各キャッシュラインにアドレスのグループを割り当てます。これにより、キャッシュコントローラは特定のデータがキャッシュにあるかどうかをすばやく判断することができます。

コントローラは、Tag RAM 内の 1 つのアドレス (要求されたアドレスが割り当てられているアドレス) だけしか調べる必要がありません。要求されたアドレスが Tag RAM 内のアドレスと一致する場合はキャッシュヒットになります。一致しない場合も、キャッシュコントローラは他の Tag RAM アドレスを調べる必要はなく、次のステップに進むことができます。

ダイレクトマッピング キャッシュではキャッシュスラッシングが発生することがあります。キャッシュスラッシングとは、キャッシュミスが多発し、メモリへのアクセスとキャッシュの更新が頻発する状況を表します。

例

プロセッサが 1 つのアドレス (D1) にあるデータを要求します。メインメモリが D1 をプロセッサとキャッシュに送ります。D1 はキャッシュライン 1 (CL1) に割り当てられます。

プロセッサは直ちに別のデータ (D2) を要求します。D2 もたまたま CL1 に割り当てられています。キャッシュコントローラは CL1 内で D2 を探しますが、D1 を含む塊が CL1 にあるため、キャッシュミスになります。

キャッシュコントローラは、キャッシュヒットになるまで、それ以降のメモリレベルに順に要求を送ります。この時点で、D2 が CL1 内の D1 を置き換えます。

プロセッサはループを実行しているため、再度 D1 を要求します。今度は D2 が CL1 を占有しているため、またキャッシュミスになります。このプロセスはループが終了するまで続きます。この間、キャッシュヒット率は 0% になります。

セットアソシアティブ キャッシュ

現在使用されているほとんどのキャッシュアーキテクチャは、ダイレクトマッピングキャッシュとフルアソシアティブ キャッシュの間を取ったもので、セットアソシアティブ キャッシュと呼ばれます。

セットアソシアティブ キャッシュでは、メモリアドレスのグループを、セットと呼ばれるキャッシュラインの特定のグループに割り当てます。キャッシュによって、各セットは2本、4本、または8本のラインから構成されます。2ウェイセットアソシアティブ キャッシュでは、1つのメモリアドレスグループからのデータで、2本のキャッシュラインのどちらかを充填できます。前述の例の場合、D1でCL1を充填し、D2でCL2を充填することができます。

ダイレクトマッピング キャッシュと比較すると、Tag RAM内で2つのアドレスを調べなければならぬため、キャッシュコントローラがデータを見つけるのにかかる時間は多少長くなりますが、キャッシュヒットの数が増えるという利点があります。一般的に、各セット内に含まれるキャッシュラインの数が多いほど、キャッシュヒットの確率は高くなります。ただし、キャッシュラインの数が多いほど、キャッシュコントローラが一致の有無を調べなければならないキャッシュラインの数は多くなります。

復習問題

最初の 5 つの質問については、DRAM、SRAM、その両方、のどれが該当するのか答えてください。

1. データを格納するためにコンデンサを使用するのはどのテクノロジーですか。
.....
2. どちらのテクノロジーの方が高速ですか。
.....
3. キャッシュで使用されているのはどのテクノロジーですか。
.....
4. データをグリッド状に格納するのはどのテクノロジーですか。
.....
5. DIMM で使用されているのはどのテクノロジーですか。
.....
6. セル、CAS、メモリコントローラ、RAS の各用語を使って、次の文を完成させてください。
データを読み書きするために、_____がアドレスピン経由で行アドレスを_____に送信し、列アドレスを_____に送信します。データはアドレスの交点にある_____に送られます。
7. 各 DRAM テクノロジーとその説明を対応付けてください。
 - a. FPM メモリコントローラは、データがデータアウトピン上にある間に次の列を有効化することにより、レイテンシを短縮します。
 - b. EDO この DRAM テクノロジーはメモリバスと同じクロックを使用するため、特別なタイミングメカニズムを必要としません。
 - c. SDRAM RAS が有効化されると(電圧が下げられると)、メモリコントローラは、順に電圧を下げることにより、次の 3 つの列のデータを要求します。

8. DDR RAM で初めて導入された革新的技術とは何ですか。
 - a. 各クロックサイクルの立ち上がりエッジと立ち下がりエッジの両方でデータを転送する
 - b. 2つのバンクの DIMM にデータを分配する
 - c. メモリに書き込む際に各バイトにパリティビットを追加する
 - d. 1ビットエラーを修正する
9. 1回のメモリアクセスで取得できるデータ量を 64 ビットから 128 ビットへと 2 倍に増やすメモリテクノロジーは、次のどれですか。
 - a. DDR RAM
 - b. オンラインスペア メモリ
 - c. ホットプラグ RAID メモリ
 - d. デュアルインタリーブ メモリ
10. α 粒子が DIMM に当たると、メモリセル内で電気障害が発生します。これはハードエラーとソフトエラーのどちらですか。

.....
11. 欠陥のあるデバイスを修理しないと再発するのは、どのタイプのメモリエラーですか。

.....

12. 各フォールトトレランス テクノロジとその説明を対応付けてください。
- a. パリティ このテクノロジはチェックサムを使用してエラーを分析し、どのバイトが壊れているかを判断し、修正します。
 - b. ECC 4つのメモリコントローラがそれぞれ4つの DIMM の内の1つに1ブロックのデータを書き込みます。5つ目のメモリコントローラがパリティ情報を5つ目の DIMM に格納します。
 - c. アドバンスド ECC 問題のある DIMM を含むメモリバンクが自動的にスペア(予備)の DIMM バンクにフェイルオーバーします。
 - d. オンラインスペア メモリ メモリコントローラが、2つのメモリボード上の同様に構成された DIMM バンクに同じデータを書き込みます。
 - e. ホットプラグ ミラーメモリ メモリコントローラがバイトをメモリに書き込む際に、バイト内の1の数に基づいて各バイトに1ビットを追加します。
 - f. ホットプラグ RAID メモリ このテクノロジは、1つの DRAM チップ内で発生したマルチビットエラーを訂正します。
13. サーバのキャッシュの利点とは何ですか。
- a. メモリよりも速く、プロセッサからのデータ要求を満たす
 - b. ハードディスクドライブ上に格納できるデータ量を2倍に増やす
 - c. 命令をデコードすることにより、プロセッサの処理速度を向上させる
 - d. メモリバスのクロック速度を向上させる
14. L1 キャッシュは、L2 キャッシュや L3 キャッシュよりも多くのデータを格納できます。
- 正
 - 誤
15. L2 キャッシュをプロセッサに接続するのはどのバスですか。
- a. フロントサイドバス
 - b. バックサイドバス
 - c. システムバス
 - d. PCI バス

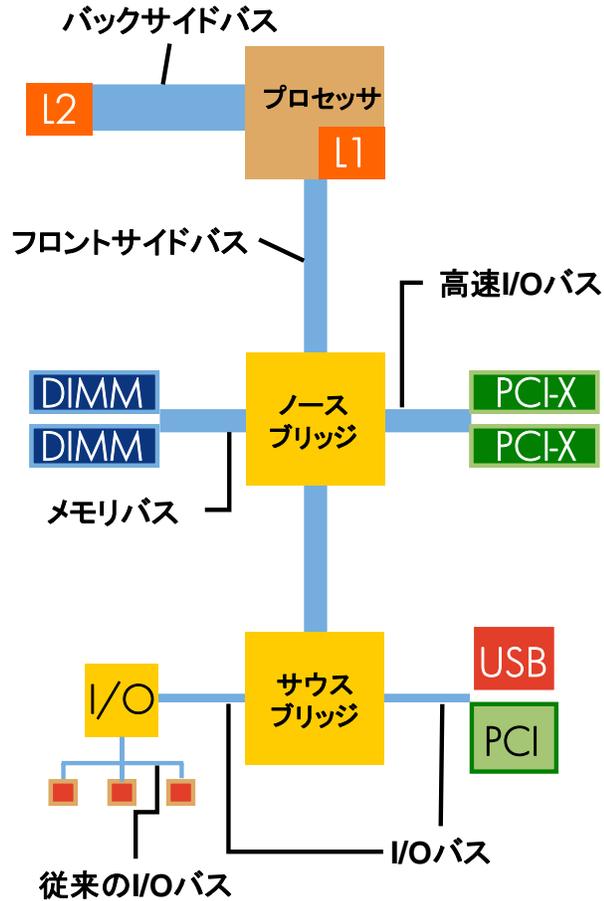
16. マザーボード上のキャッシュは、プロセッサの最高速度と同じ速度で動作します。
- 正
- 誤
17. キャッシュ内の各行を何と言いますか。
-
18. Tag RAM の機能とは何ですか。
-
19. 各キャッシュ実装形式をその定義と対応付けてください。
- | | | |
|------------------|-------|---|
| a. ルックアサイド | | キャッシュコントローラが、バスマスタによるメモリ要求がないか、システムバスのトラフィックを監視します。 |
| b. ルックスルー | | バスマスタがメモリにデータを書き込む際に、キャッシュコントローラが書き込まれようとしているデータをキャプチャしてキャッシュに書き込みます。 |
| c. フルアソシアティブ | | キャッシュラインに添付されたビットに、データがまだメモリに書き込まれていないことを示すフラグが付けられます。 |
| d. ダイレクトマッピング | | メインメモリからのデータをどのキャッシュラインにも格納できます。 |
| e. セットアソシアティブ | | データを再度使用する前に、システムはすべてのメモリレベルにデータを書き込む必要があります。 |
| f. ライトスルー | | キャッシュとメモリの両方がメモリ要求を受け取ります。キャッシュヒットがあれば、キャッシュコントローラがその他のデバイスへの要求を取り消します。 |
| g. ライトバック | | 各キャッシュラインにメモリアドレスのグループが割り当てられます。 |
| h. バススヌーピング | | キャッシュヒットがあった場合、要求はシステムバスに送られません。 |
| i. バススナッフイング | | キャッシュラインの特定のグループにメモリアドレスのグループが割り当てられます。 |

目的

この章では、次の内容について学習します。

- バスアーキテクチャについて
- バスの構成要素
- バスパフォーマンスを向上させる方法
- 次のバスの速度、転送レートおよび長所の比較
 - ISA
 - EISA
 - PCI
 - PCI-X
 - PCI-X 2.0
 - InfiniBand
 - PCI Express
 - Universal Serial Bus (USB)

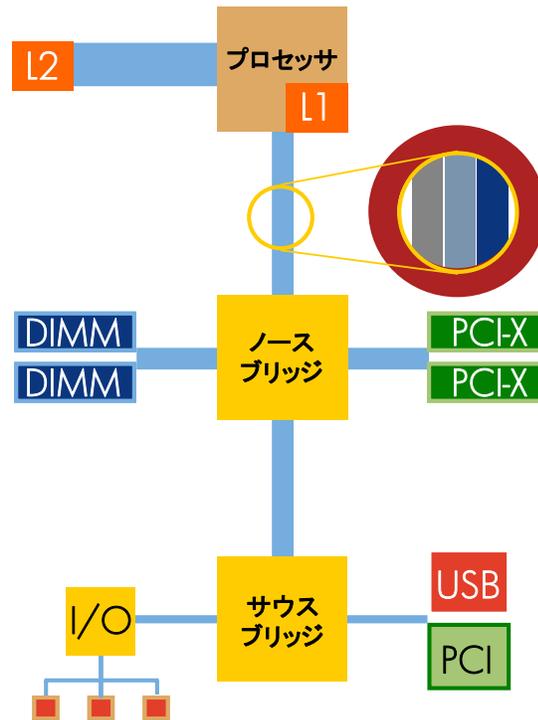
バスアーキテクチャ



サーバ内部のコンピューティング構成要素は、幅や速度が異なるバスの集まりによって接続されています。これらの内部システムバスは、以下の構成要素間でデータを並列に転送するための通路を提供します。

- チップセットコントローラ
- プロセッサ
- キャッシュ
- メインメモリ
- I/Oバス

バスの構成要素



主なバスの構成要素には次の3つがあります。

- アドレスバス
- データバス
- コントロールバス

バスは以上3つの要素で構成されており、これらは物理的には分かれていません。サーバのシステムボードを見ても、3つのバス構成要素は別々ではありません。ICボード内に内蔵された電気回線が見えるだけです。

バスマルチプレクシング

これら 3 種類のバス構成要素の用途は、バス規格により異なります。

例

あるバス規格では、アドレスバスとデータバスは物理的に別々の電気バスラインを使い、別のバス規格では両方に同じ電気バスラインを使います。

バスがアドレス指定用とデータ用に同じ回線を使うことを、マルチプレクシングと呼びます。マルチプレクシングでは、バスラインやバス接続のセットに対して 1 つ以上の機能が与えられます。たとえば、システムコントローラが 32 本のバスラインでアドレスを転送するために、コントロール信号でバスをアサートし、その後同じ 32 本のバスラインをデータ転送に使うために、新しいコントロール信号を再度アサートすることがあります。

アドレスバス

アドレスバスは、次の目的で使われます。

- データバス上で伝達されるデータの発信元アドレスと送信先アドレスを転送する。デバイスまたはメモリ内のデータにアクセスするのに、そのアドレスはプロセッサ、コントローラ、またはバスマスタなどのイニシエータによって、必ずアドレスバスにセットされる。
- マイクロプロセッサ、バスマスタまたは DMA コントローラが生成したメモリの場所を特定する。デバイスがデータを転送するためには、アドレスバスは対象デバイス内のデータが存在する場所を見つける必要がある。

アドレスバス幅

アドレスバスの使用回線(ビット)数によって、アドレス空間が決まります。アドレス指定が可能なメモリサイズは、2(各回線またはビットが 0 か 1 かを表すバイナリ)の n 乗(2^n)で求められます。指数(n)はアドレスバスの回線数で、アドレスバスのビットサイズと同値です。

例

Intel Pentium プロセッサは、32 本のアドレスラインを持ち、 2^{32} 個のアドレス(4GBのアドレス空間)を提供します。

次の表に、Intel プロセッサファミリのアドレスバス幅とアドレス指定可能な最大メモリサイズを示します。

プロセッサ	アドレスライン(ビット)	アドレス指定可能な理論上の最大メモリ
8086	20	1MB
80286	24	16MB
80386 と 80486	32	4GB
Pentium	32	4GB
Pentium Pro	36	64GB
Pentium II、III と 4	36	64GB
Pentium II Xeon と III Xeon	36	64GB
Itanium	64	2 ⁶⁴ GB

注記

最大アドレス数を提供するプロセッサの性能は、チップセットおよびオペレーティングシステムにより制限される場合があります。

データバス

データバスは、プロセッサとメモリが送信または受信するデータを転送する内部通路です。データは、データバスを通過して2つのデバイス間を移動します。データの種類としては、マイクロプロセッサに対する命令や、マイクロプロセッサが伝達している情報があります。この情報の送信元または受信先は、メモリまたは I/O サブシステムです。

データバス幅

データバスの幅と速度は、システムのパフォーマンスに直接影響し、特にスループットを大きく左右します。バス幅は、バスを構成している回線またはワイヤ数で表されます。データバスの各回線または各ワイヤは、1ビットの情報を伝達できます。

バス幅は、マイクロプロセッサなどのデバイス上のデータピン数と同じです。データバスの各回線は、1ビットのデータを運び、8ビットで1バイトが構成されることから、8の倍数の回線数を持ちます。したがって、データバス幅は、バスが一度に転送できるバイト数を示します。32ビットバスでは一度に4バイト、64ビットバスでは1サイクルに8バイトを転送できます。

プロセッサ	データバス幅
8086 と 80286	16ビット
80386 と 80486	32ビット
Pentium と Pentium Pro	64ビット
Pentium II、III と 4	64ビット
Pentium II Xeon と III Xeon	64ビット
Itanium	64ビット

コントロールバス

コントロールバスは、コントロール信号の集まりで、実際のバスではありません。信号は、アドレス指定された構成要素の動作を決定します。

コントロールバスは、次の目的で使われます。

- デバイスの調整 — バスを制御するデバイスを特定し、衝突を防止します。
- データフローの方向 — バスサイクルの方向 (読み出しか、書き込みか) を特定し、サイクルが完了したことを示します。
- メモリアドレス指定型 — 実行中の操作の対象はメモリか I/O かを特定します。

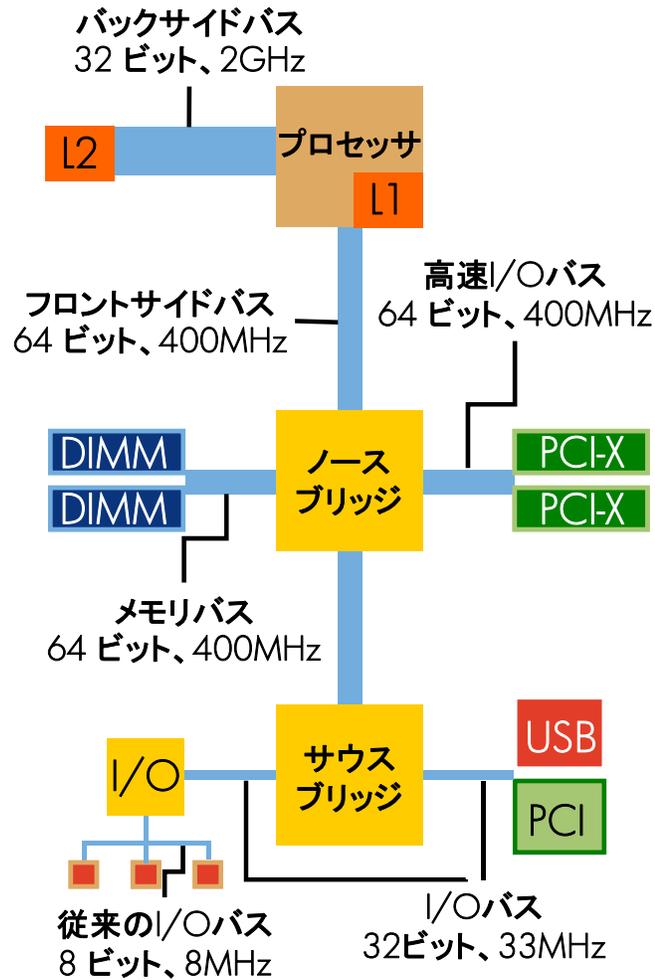
プロセッサまたは他のデバイスが生成するコントロール信号の表記方法には次のようなものがあります。

- W/R — 書き込み/読み出し
- IRQ — 割り込み要求
- BCLK — バスクロック
- DRQ — DMA 要求

多くの信号は、ポイントツーポイントです。ポイントツーポイント信号方式は、専用回線の両端にデバイスが 1 台のみ存在する場合に使われます。

- 転送は、バスクロックのビートに同期して行われる
- バスクロックの速度を増加すると、ある時間内の転送データ量も増加する

バスのパフォーマンス



主要なサブシステム間の通信は、システム内のバスを通してデータを読み出すか書き込むことによって発生します。バスは、サブシステム間で効率的にデータを転送し、ボトルネックを避けるか最小限に抑えるために、システム全体で適用される明確な規格をもとに設計されます。

プロセッサの要求に応えるために 3 クロックサイクル以上を必要とするデバイスが存在すると、プロセッサがそのデバイスにアクセスしている間、システムの速度が低下します。

バスの各規格の長所を判断するためには、基本的なパフォーマンスの評価基準を比較します。

- バス幅
- バス速度

バス幅

バス内の電気回線の本数によって、バス幅が決まります。ちょうど道路の幅がその道路を通過できる交通量を制限するように、バス幅は一度に転送されるデータ量を制限します。

バス幅の測定単位はビットです。たとえば、16ビット幅のバスは、16本のワイヤを使って一度に2バイトのデータを転送し、32ビット幅のバスは、32本のワイヤで一度に4バイトのデータを転送します。このようにバス幅は構成要素間の最大データ転送量を制限します。

データ転送を開始するのはバスではありません。プロセッサやチップセットコントローラがデータ要求を開始し、バス上でデータを転送します。

バス速度

1秒間に発生するバスサイクル数をバス速度と呼びます。バス速度の測定単位はヘルツ(Hz)です。バス速度は周波数とも呼ばれます。

バス幅がバスによって違うように、同一システム内でもバスによってバス速度が違う場合があります。サーバ内の各サブシステムはそれぞれ独自の動作周波数を持ち、タイミングルールをベースに通信を行います。

各構成要素間で通信を行い、データを転送するには、構成要素間のタイミングがシステムのハートビートと同期している必要があります。またバスもこのタイミングルールを守り、バスクロックと同期できなければいけません。バスクロックはチップセットに制御され、システムクロック周波数の約数になります。

バスサイクル

プロセッサ、チップセット、またはバスマスタデバイスなどの構成要素が、バスを使って通信、またはデータを転送する場合、その転送実行プロセスはバスサイクルと呼ばれます。

バス上では、情報を転送するために次のような活動が行われます。

- 一連の命令が、コントロールバスとアドレスバスから転送される
- データがデータバス上を移動する

バスサイクルには、次の基本的な種類があります。

- メモリリード
- メモリライト
- I/Oリード
- I/Oライト

すべてのデータ転送は、1つ以上のバスサイクルの結果として発生します。メモリまたは I/O デバイス内のコードやデータが必要になると、プロセッサまたはバスはそれを得るためにバスサイクルを実行する必要があります。

最大転送レート

$$\frac{\text{速度(MHz)} \times \text{幅(バイト)}}{\text{転送毎のクロックサイクル}}$$

例1: 1ウェイトステート

$$400\text{MHz} \times 4\text{バイト}$$

3サイクル

$$533\text{MB/s}$$

例2: ウェイトステートなし

$$400\text{MHz} \times 4\text{バイト}$$

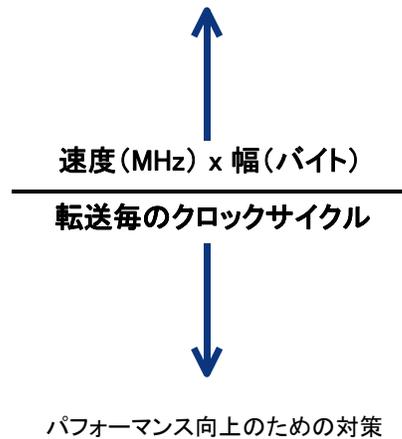
2サイクル

$$800\text{MB/s}$$

バスがある一定時間内に送信できるデータ量を、*最大転送レート*と呼びます。最大転送レートは、サーバのパフォーマンスを計測する1つの方法です。

転送レートは、1秒間に転送されるバイト数で示されます。最大転送レートは、「MTR = 周波数(または速度) x 幅/バスサイクル毎のクロックサイクル数」で求められます。

システムパフォーマンスの向上



システムパフォーマンスを向上させるためには、次の4つの基本的な対策が考えられます。

- バスサイクルの実行速度を高速化する。これは普通、プロセッサのクロックレートを増加することを意味する。デバイス、特にシステムメモリが、プロセッサと通信できる速度を増加する。具体的には、高速メモリの実装やキャッシュの追加を行う
- データバス幅を広げて、1バスサイクルで転送する情報量を増加する
- バーストサイクルなどのバスサイクルを変更する
- デュアル インディペンデント バスやマルチプロセッシングなどの並列プロセスを追加する

プロセッサ、バス、およびデバイス速度の高速化

速度(MHz) x 幅(バイト)

転送毎のクロックサイクル

533MHz x 4バイト

2クロックサイクル

1.066GB/s

フロントサイドバス速度の増加によるシステム速度の高速化

システム内のデバイス速度を高速化することによって、システムパフォーマンスを向上できます。新しいプロセッサは一般に、従来に比べ高速なバス速度をサポートしています。速いプロセッサを遅いバスに配置してもパフォーマンスは向上しません。

バスの拡張

$$\text{速度(MHz)} \times \text{幅(バイト)}$$

$$\text{転送毎のクロックサイクル}$$
$$533\text{MHz} \times 8\text{バイト}$$

$$2\text{クロックサイクル}$$
$$2.132\text{GB/s}$$

バスの拡張によるシステム速度の高速化

バスの拡張によっても、システムパフォーマンスを向上できます。バス幅が拡張されると、データが移動できる電気の通路が増えることとなります。

バーストサイクル

$$\frac{\text{速度 (MHz)} \times [\text{幅 (バイト)} \times \text{バースト毎の転送}]}{\text{転送毎のクロックサイクル}}$$

$$\frac{533\text{MHz} \times [8\text{バイト} \times 4\text{転送}]}{4\text{クロックサイクル}}$$

4.264GB/s

注記: 転送レートの計算では、アドレス時間は普通考慮されない

80486 プロセッサファミリ以降、すべての Intel プロセッサは、2 データサイクル以上を必要とするデータ要求に応えるために、バーストサイクルをサポートしています。バースト転送では、データは、非バースト転送のような 2 クロックサイクル毎ではなく、1 クロックサイクル毎に転送されます。

1 回のバースト転送には、最初のデータ (A1 + D1) を転送する際に最低 2 クロックサイクルが必要です。その後、最大 3 つのデータを転送することができます (D2 + D3 + D4)。1 バーストサイクル中に達成されるスループットは、様々な要因によって異なります。

デフォルトでは、すべてのキャッシュラインフィルは、バーストサイクルで実行されます。キャッシュコントローラは、1 クロックサイクルにつき 16 ビットアドレスを 4 個、または 32 ビットアドレスを 2 個バーストできます。データバス幅は 64 ビットなので、32 バイトアドレスをバーストするのに、最低 4 クロックサイクル (8 バイト/サイクル x 4 サイクル = 32 バイト) が必要です。

0 ウェイトステート バーストサイクルでは、最初のアドレスにアクセスした後、一連のデータを連続して転送できます。データが転送されている間に、残りのアドレスが 2 次キャッシュコントローラにより算出され、自動的に続けて転送されます。

バースト転送レート

合計バーストサイクルで転送された合計量を出し、それを合計データ時間で割ると、バースト転送レートが算出されます。



注記

この式で算出された転送レートでは、アドレス時間は考慮されていません。

並列性の向上

バスマスタ

バスマスタは、バスに接続されたデバイスで、プロセッサを通さずにバス上の他のデバイスと直接通信を行います。

バス マスタリングは、プロセッサがバスの制御を I/O デバイスへ実際に渡すときに使われるプロトコルです。このプロトコルを用いると、I/O デバイスはデータを直接メモリに転送できます。プロセッサがバス上の仲介装置とならないので、転送時のプロセッサの使用が最小限に抑えられ、高速データ転送が実現します。

バスマスタデバイスは、パフォーマンスの向上に役立ちます。プロセッサは、転送時に必要とされない分、他の要求を処理できるからです。

複数のバスマスタ デバイスが同時にバスを要求した場合、I/O コントローラがバスマスタ デバイス間の調停を行い、ロックアップを防止します。

次の構成要素は、バスマスタとして機能する場合があります。

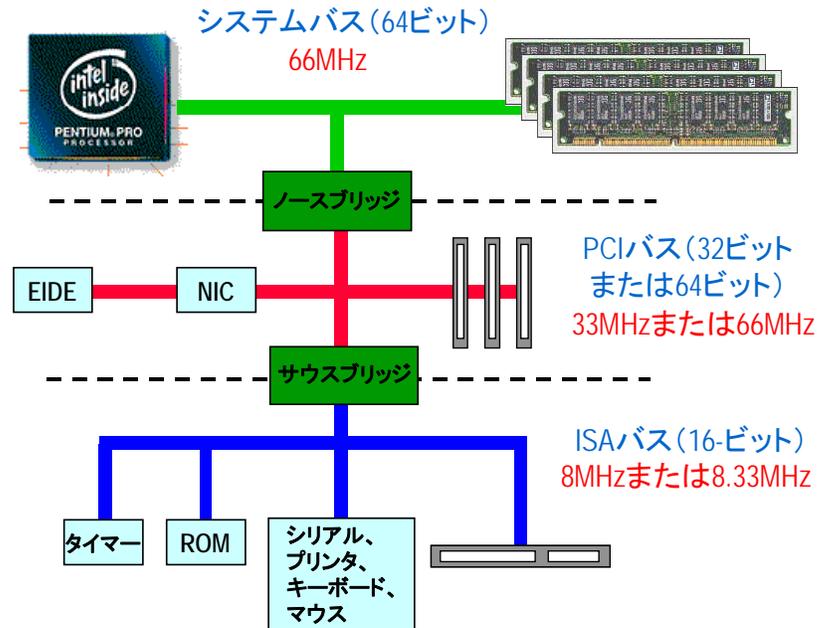
- プロセッサ
- DMA コントローラ
- メモリリフレッシュロジック
- EISA/PCI バスマスタカード

バスアービトレーション

バスアービトレーションは、バスコントローラの機能の 1 つです。バスコントローラは、ある時点でどの構成要素がバスを制御するかを決定します。

各バスはそれぞれ独自のバスコントローラを持っています。メモリバスやローカル I/O バスのようなシステムバスのコントローラは、チップセット内にあります。拡張デバイスのコントローラは、システムボード上の別の ASIC (Application-Specific Integrated Circuit) または拡張カード内にあります。

ISA バス



ISA バスは 16MB のアドレス空間が特徴の Intel 80286 プロセッサと共に導入されました。I/O デバイスを含むすべてのシステム構成要素は、プロセッサバスに直接接続され、約 8MHz で動作します。ISA バス向けに策定された規格では、公差は 8MHz 弱～8.33MHz です。

IBM XT や HP ポータブルコンピュータで使われていた初期の 8ビット PC バスは、ISA 仕様の先行バージョンです。ISA は、80286 以降の実装方法を指します。

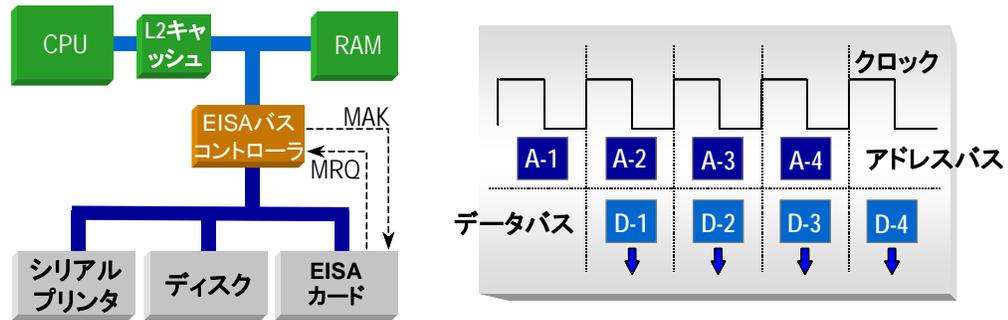
ISA バステクノロジ

現在製造されている ISA バス、EISA バスおよび拡張ボードの大部分は 8.33MHz に設定されています。これらは、ISA バスと EISA バスが接続されている 33.3MHz PCI バスの 4 分の 1 の速度で動作します。ISA および EISA システムでは拡張バスとシステムバスは同じです。ISA ボードは 8ビットまたは 16ビットです。

ISA バスには 24 本のアドレスライン (A0～A23) があり、アドレス範囲は 16MB です。16 本のデータライン (D0～D15) で 16ビット転送が可能なので、一度に 2 バイトを転送できます。

ISA の最大転送レートである 8MB/s は、データ転送毎に 2 クロックを必要とする 80286 プロセッサにより実現されます。1 番目のクロックでアドレスが作成され、2 番目のクロックでデータが転送されます。

EISA バス



EISA バスは 1988 年 HP など 9 社により導入されました。ISA を拡張した EISA は、ISA バスの機能をすべて想定し、ISA 構成要素に対して互換性を備えていました。

EISA バスは新しい業界標準ハイパフォーマンスバスに切り替えられました。この新しいバスの開発においても HP は最先端を担っています。

EISA バステクノロジ

EISA システムは 32 ビットのアドレス範囲をサポートします。プロセッサ、DMA コントローラ、EISA バスマスタは EISA システムを使って、4GB を直接アドレス指定できます。

EISA バス上での転送はすべて 8.33MHz のバスクロックに同期化されます。EISA 構成要素間の標準的な転送は 2 クロックサイクルかかり、最大 16.5MB/s の転送レートが実現可能です。バーストモードでの転送は、転送毎に 1 クロックサイクルが必要なだけで、最大 33.3MB/s の転送レートを達成します。

バスマスタのパフォーマンス

Central Arbitration Controller (CAC) は、プロセッサ、EISA カード、DMA など、バスを必要とする構成要素すべてのアクセスを制御する特別なチップです。

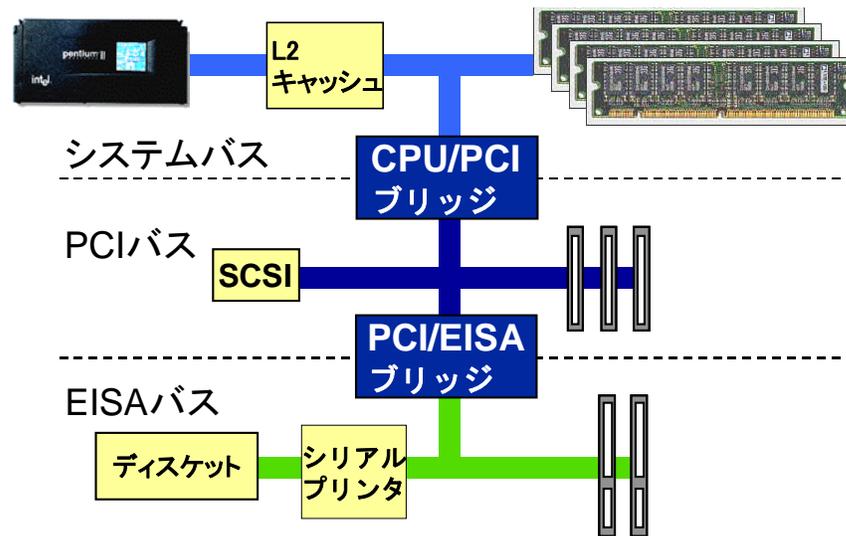
ハードウェアによってバスを調停するので、ISA バスマスタカードでのタイミング上の制約も EISA では発生しません。EISA は公平なローテーション原則でバスを共有します。

バースト転送

バースト転送を開始すると、33.3MB/s の転送レートを達成できます。バースト転送では、データ転送の段階ではアドレスラインは必要ありません。ある転送が実行中に、次の転送用アドレスが発行されますので、次のサイクルも別のデータサイクルとなります。

EISA バスの最大クロックレートは 8.33MHz です。バーストモードでは、システムは全サイクルをデータ転送に使用し、毎秒 833 万転送、33.3MB/s (833 万転送/秒 x 4 バイト/転送 = 33.3MB/s) を達成します。

PCI バス



PCIバス仕様は1992年、業界標準のインディペンデントバスとして導入されました。PCI仕様1.0は、33MHzで動作する32ビットバスを定義しました。1994年のPCI仕様2.1では、66MHz/64ビットPCIが定義されました。

幅	速度	スループット	仕様
32ビットカード	33MHz	133MB/s	1.0
32ビットカード	66MHz	267MB/s	2.1
64ビットカード	33MHz	267MB/s	2.01
64ビットカード	66MHz	533MB/s	2.2
64ビットカード	100MHz	800MB/s	PCI-X
64ビットカード	133MHz	1066MB/s	PCI-X

PCIバスには次の特徴があります。

- 完全な独立性
- プラグ&プレイアーキテクチャ
- マルチプレクシング
- デバイスのロード
- PCIバスでの転送
- カードスロットのサポート
- 32/64ビット33/66MHzカード

完全に独立したバス

PCI は、プロセッサローカルバスから完全に独立しているため、すべてのシステムで使えます。

- Intel ベースのシステム
- Alpha ベースのシステム
- アップルコンピュータ
- その他のシステム

ホスト/PCI 間ブリッジは、I/O バスをプロセッサのフロントサイドバスに接続する PCI 設計の基本パーツです。ブリッジチップはデータバッファリングをサポートしているので、I/O バス アクティビティとプロセッサ/メモリ間のバス アクティビティを並行して実行できます。またこのブリッジにより、アクティビティはプロセッサ信号および ISA や EISA の I/O バス信号から独立したものになるので、すべてのプロセッサを PCI バスにつなぐことができます。接続には、ISA、EISA、および Micro Channel などのシステムの I/O 部分が使用されます。

プラグ&プレイアーキテクチャ

プラグ&プレイアーキテクチャでは、システム BIOS または SCU (System Configuration Utility) が自動構成を行い、システムリソースは自動的に割り当てられます。PCI 拡張カードは、必要なリソースを自分でコンピュータに伝えるので、構成ファイルを必要としません。結果として、リソース割り当て時の競合の可能性を大幅に減らすことができます。

IRQ 共有

PCI では、IRQ ラインを共有できます。PCI は IRQ ラインのかわりに INT ラインを使います。その後 INT ラインは IRQ ラインに経路指定されるか、またはプロセッサの高性能プログラマブル割り込みコントローラ (Advanced Programmable Interrupt Controller: APIC) バスを通して直接経路指定されます (フルテーブルモード)。IRQ 共有は、PCI 2.1 以降の仕様に基づくカード使用時のみ必須です。

マルチプレクシング

PCI 規格は、ピン数を削減し、PCI スロットのサイズを縮小するマルチプレックス アドレス/データバスを使って、コストを削減します。

デバイスのロード

PCI は、バス 1 個につき最大 32 デバイス、システム 1 台につき最大 256 バスをサポートします。ただし、実際にバスが処理できる周辺デバイスの数は、負荷に左右されます。負荷は、インダクタンス、静電容量などの電気的な特性により変化します。また電気的な要求のため、バス毎のスロット数は 10 に減少します。PCI チップセットは 3 スロットを使い、7 スロットは周辺機器用に残されています。システムボード上のコントローラは 1 デバイスを使い、拡張スロットにプラグインするコントローラは 1.5 デバイスを使います。PCI/PCI 間ブリッジを使うと、複数の PCI バスを接続できるので、たくさんのスロットを使用できます。

PCI バスでの転送

PCI バスは、すべてのデータ転送をバースト転送として実行します。通常の転送では、各データ転送の前に新しいアドレスが要求されます。この特徴により、各クロックサイクルでデータ転送が 1 回実行されます。バーストはメモリまたは I/O 空間に対して行われます。

33MHz で動作する 32 ビット PCI バスのピーク転送レートは、133MB/s (転送 3330 万回 x 4 バイト) です。各 PCI デバイスには、他の PCI デバイスがバスを使おうとしたときに、バスを伝える最大時間を定義した待ち時間タイマーがあります。

カードスロットのサポート

PCI カードおよび PCI スロットの多くは、5V または 3.3V のシングルボルテージでのみ動作します。PCI スロットおよび PCI カードは、5V カードが 3.3V スロットに挿入できないように、また 3.3V カードが 5V スロットに挿入できないようになっています。

ユニバーサル PCI カードはどのスロットにも適応しています。ユニバーサル PCI カードは 5V スロットでも 3.3V スロットでも使用でき、電圧も問題ありません。

▲ 注記

66MHz PCI は、3.3V で動作するカードだけをサポートしています。

32/64 ビット 33/66MHz カード

PCI は、標準的なハードウェアコストで最高のパフォーマンスを実現します。物理的に EISA より小さい PCI スロットで、EISA の 4 倍のパフォーマンスを達成し、信号線の数も半分で済みます。

PCI は、33MHz または 66MHz で動作する 32 ビットバスまたは 64 ビットバスとして構築できます。PCI33MHz/32 ビットシステムは最大 90MB/s で、標準 PCI バス上のデータを移動します。残りの時間は非効率的なバスの使用により生じるオーバーヘッドと待ち時間に費やされます。

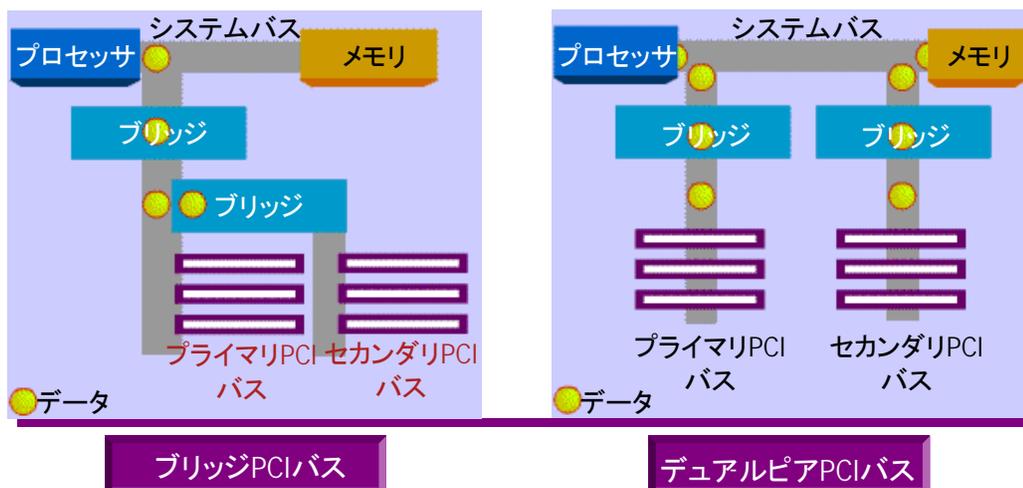
66MHz で動作する 64 ビット PCI

66MHz で動作する 64 ビット PCI バスのピーク転送レートは、533MB/s です。64 ビット PCI バスのほうがバスの使用効率が高く、バンド幅は 32 ビット PCI バスの 2 倍です。64 ビットアドレス空間は巨大 (40 億 x 4GB) です。32 ビット PCI カードはデュアルアドレスサイクルとして知られる連続する 2 つの 32 ビット PCI サイクルを使って、64 ビットアドレスをサポートできます。

32/64 ビット 33/66MHz カードのルール

- 66MHz PCI カードは、33MHz PCI バスで使える
- 66MHz PCI バス内の 33MHz カードは自動的に 33MHz で動作する
- 32 ビット PCI カードは、64 ビット PCI スロットにインストールできる
- 64 ビットカードは 32 ビットスロットにインストールでき、その場合 32 ビットモードで動作する

複数の PCI バスシステム

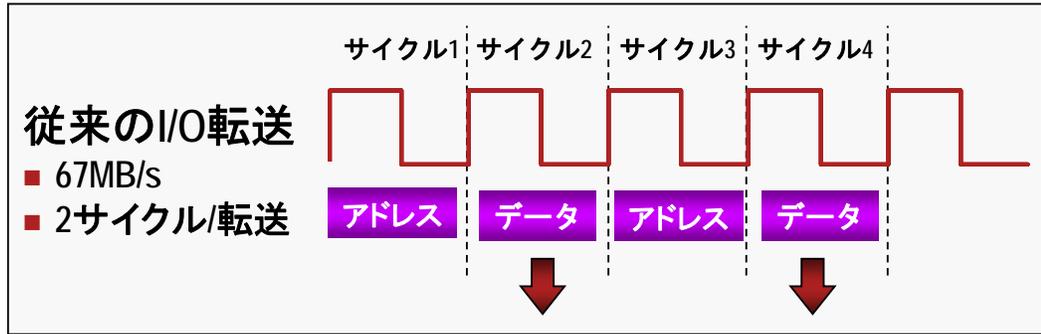


PCIバス設計の多くでは、ブリッジが2台あるPCIバスシステムが採用されています。このシステムでは、バスが飽和状態になるとパフォーマンスが低下するおそれがあります。HPサーバでは、1台のシステムに2、3台の別々のPCIバスがあるピアPCIバス設計が採用されています。

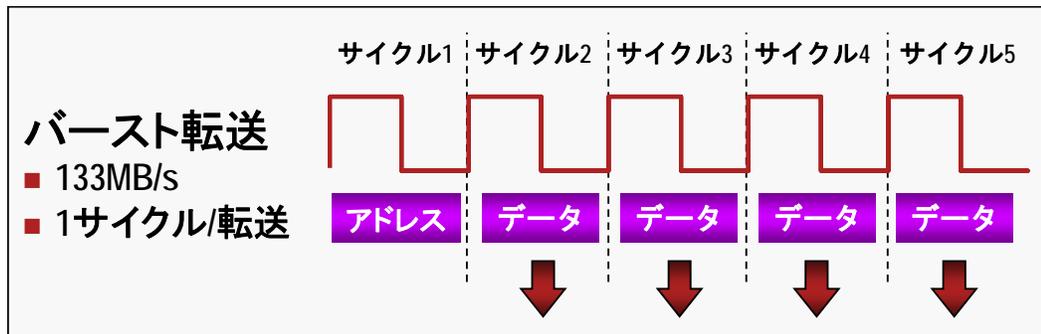
ブリッジ/シリアル配列のパフォーマンスは全体で133MB/sが限界です。ピアPCIバスアーキテクチャのシステムパフォーマンスは、PCIバスが追加される度に向上します。スロットを2台以上のバスシステムで分けると、各バスが4～6スロットしか持たず適切な負荷分散となるので、安定性が高まります。HPピアPCIバスアーキテクチャシステムの特徴は、64ビットPCIスロットが11個ある64ビットピアPCIバスが3台あることです。

33MHzで動作するデュアルピア32ビットPCIバス設計のシステムのパフォーマンスは、266MB/s (2 x 133MB/s) です。ProLiant 8500サーバには3台のピアPCIバスがあります。2台は64ビット/33MHzバスで、1台は64ビット/66MHzバスです。これにより累計PCI帯域幅は1.06GB/s (266 + 266 + 533MB/s) になります。

PCI バスでのデータ転送



従来の転送



バースト転送

初期の PCI バスでは、次の 2 形態のデータ転送のみが可能です。

- 従来の I/O 転送 — 最大 67MB/s (転送 1 回につき 2 サイクル)
- バースト転送 — 最大 133MB/s (転送 1 回につき 1 サイクル)

例

転送 3330 万回 x 4 バイト = 133MB/s

33MHz で動作する 64 ビット PCI バスでは、大幅に速い転送レートが実現できます。

例

転送 3330 万回 x 8 バイト = 267MB/s



注記

対象デバイスが対象アドレスを計算します。

64 ビット/66MHz の PCI 拡張仕様もありますが、この仕様を採用しても大部分のシステムとアダプタでは、PCI 周波数は 33MHz で動作します。

PCI バスのパフォーマンスプラン

PCI バスは使いすぎると、パフォーマンス上のボトルネックになる場合があります。適切な PCI デバイスのパフォーマンス構成プランを立てることをお勧めします。

構成を始める前に、次の事項を考慮してください。

- サーバで使用可能な PCI バスとスロット
- 構成されるデバイスがサポートしている PCI プロトコル
- 各デバイスが受信または送信する予想データ量(スループット)

スループットには PCI バスの割り込みおよび PCI プロトコルのオーバーヘッドは含まれていませんが、十分適切なバスプランを立てることができます。

最適な構成を実現するには、

1. 66MHz スロットを 66MHz デバイスと合わせる
 - a. 66MHz スロットはあるが 66MHz デバイスがない場合、66MHz スロットを 64 ビットスロットとして使う
 - b. 66MHz デバイスはあるが 66MHz スロットがない場合、66MHz デバイスを 64 ビットデバイスとして使う
 - c. 66MHz スロットと 66MHz デバイスがある場合、最大スループットの 66MHz デバイスを 66MHz スロットに配置する。残りのデバイスは、64 ビットデバイスとして使う
2. 32 ビットスロットを 32 ビットデバイスと合わせる
 - a. 32 ビットスロットと 32 ビットデバイスがある場合、最大スループットの 32 ビットデバイスを 32 ビットスロットに配置する。残りのデバイスは、64 ビットスロットに配置する。いくつかの 32 ビットスロットは 64 ビットバス上にあるので適切に使うこと
 - b. 32 ビットデバイスはあるが 32 ビットスロットがない場合、32 ビットデバイスを 64 ビットスロットに配置する

3. 残ったデバイスについては、
 - a. 使用可能な 32 ビットスロットがある場合、最小スループットのデバイスを使って最低限の 64 ビットデバイスを 32 ビットスロットに配置する
 - b. それでもまだ、使用可能な 64 ビットスロットよりデバイスのほうが多い場合、66MHz スロットは 33MHz (64 ビット) で動作する必要がある
4. 残ったデバイスを使用可能な 64 ビットスロットに振り分ける。この手順では、データ転送ビットの半分だけしか使われないので、32 ビットデバイスのスループットを 2 倍にする必要がある。
 - a. 最大スループットの未割り当てデバイスを、スロット数が最小のバスに配置する
 - b. 残りの未割り当てデバイスの中で、最大スループットのデバイスを次にスロット数が少ないバスに配置する。ここでも、32 ビットデバイスのスループットを 2 倍にする
 - c. このように最大スループットのデバイスを次に使用可能なバスに順々に割り当て、ラウンドロビン方式ですべてのバスを使う。

PCI バス周波数の調整

PCI バス周波数はサーバの初期化時に調整され、変更するにはサーバのリブートが必要です。バス周波数は、最も遅い PCI デバイスまたは最大 PCI バス速度のどちらか遅いほうに調整されます。すべての 32 ビットおよび 64 ビットデバイスは、同一の 64 ビット PCI バスで制限なしに動作します。32 ビットデバイスは、データバスの 32 ビットを使いますが、64 ビットデバイスはデータバスの 64 ビットすべてを使います。

PCI バス番号の割り当て

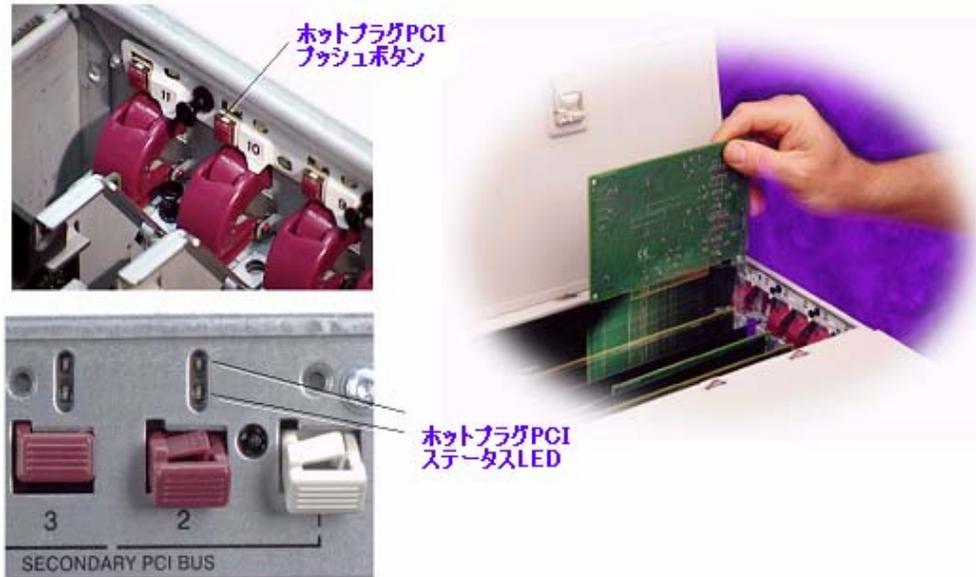
各 PCI バスはバス番号に割り当てられます。(Smart コントローラなどの) PCI ブリッジのある PCI ボードがプライマリバスに追加されると、通常デュアルピア アーキテクチャのセカンダリバスに割り当てられているバス番号は変更されます。これは次のような 450GX ベースの Pentium Pro サーバにとっては重大な問題です。

- ProLiant 5000
- ProLiant 5500
- ProLiant 6000
- ProLiant 6500
- ProLiant 7000

この種のボードが追加された場合、セカンダリバス上の HP NIC を再構成する必要があります。他社のブートコントローラがセカンダリバス上に存在する場合、ブートに関する問題が起こるおそれがあります。

新しい Xeon ベースのシステムは、PCI バス番号を各スロットに事前に割り当てます。事前割り当てにより、PCI ブリッジを持つカードがインストールされても、セカンダリバス、ターシャリバスの PCI バス番号は変更されなくなりました。

PCI ホットプラグのサポート

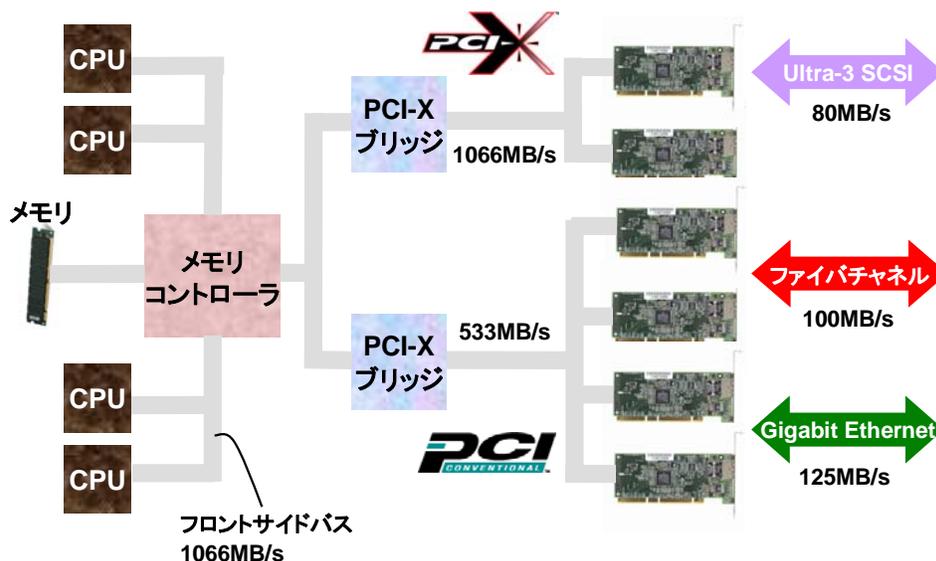


PCI ホットプラグを用いると、システムの電源を切ることなく新しい PCI カードを交換・インストールできます。PCI ホットプラグは Compaq により開発され、オープンな業界標準として受け入れられています。

PCI ホットプラグテクノロジーには様々な長所があります。

- サービスの中断がない — ホットインサージョンまたはホットリムーバルにより、ネットワークまたは他の I/O コントローラボードをシステム稼動中に追加・交換可能です。
- 幅広く一般的な互換性 — 提案された実装方法は、既存の PCI 規格と互換性があります。システムハードウェア、オペレーティングシステム、またはアダプタドライバへのどんな変更も、既存のホットプラグ未対応システムの機能に影響を及ぼしません。
- 以前のバージョンとの下位互換性 — このテクノロジーはホットプラグに対応していない構成要素に対して完全な下位互換性を持ち、システム内での構成要素の共存を可能にします。
- ホットプラグ — 3 つの PCI ホットプラグ機能が使用可能です。
 - ホットリプレース — 既存の PCI アダプタの交換
 - ホットリムーバル — 既存の PCI アダプタの取り外し
 - ホットアド — 新しい PCI アダプタの追加

PCI-X バス



PCI-X は互換性を維持した既存の PCI バスの拡張バージョンで、Compaq、HP、および IBM により共同開発されました。PCI-X は、PCI2.1 プロトコルに対して完全な下位互換性があります。PCI-X により、64 ビットバス幅を使い最大 133MHz の周波数で動作することで、1GB/s 以上のバースト転送レートを実現するシステムの設計が可能になりました。

PCI-X 仕様はオープンな業界標準で、PCI ローカルバス 2.2 仕様の補足資料として PCI Special Interest Group (SIG) から入手できます。

PCI-X は PCI バスと比べて次のように拡張されています。

- 以前のバージョンとの下位互換性 — PCI-X は PCI と下位互換性があるので、既存の PCI カードは PCI-X 環境でも動作します。ただし、PCI-X バスの速度は PCI カードの速度 (一般的に 33 または 66MHz) になります。
- 以後のバージョンとの上位互換性 — PCI-X カードはユニバーサルカードとして設計されており、3.3V I/O スロット (66MHz 以上) または 5V I/O スロット (33MHz) のどちらにもプラグイン可能です。ただし既存の PCI バスでは、PCI-X カードの追加機能と速度を完全には実現できません。
- 共存と混在 — PCI-X と PCI カード両方を同一のバスで使えますが、バスの実行速度は最も遅いカードの速度となります。たとえば、66MHz カードと 133MHz カードを同一の PCI-X バスに配置した場合、バスは 66MHz で動作します。PCI-X バスと PCI バスを同一サーバで使うのも一般的な実装方法と言えます。

- 速度 — PCI-X は 64 ビットバスで、最大 133MHz の速度で動作します。設計が改良されたのでバスの効率は向上し、I/O スループットも大幅に増えました。32 ビット/33MHz の一般的な PCI の実行方法のパフォーマンスと比べると、PCI-X のパフォーマンスは最大 10 倍向上しています。既存の PCI バスも 64 ビット/66MHz で設計できますが、実際の実装例はほとんどありません。
- スロット数 — 66MHz システムで使用可能なスロット数も PCI の制限事項でした。厳しい電気的な要求のため、66MHz の PCI システムがサポートできるのは 1 スロットか 2 スロットだけでした。PCI-X は 66MHz で 4 スロット以上サポートでき、この問題を解決しています。

64 ビット PCI-X アーキテクチャは最大 133MHz の速度で動作し、1GB/s 以上のバースト転送レートを実現します。エンタープライズアプリケーションを実行する次のような業界標準サーバは、この重要な I/O 帯域幅を必要としています。

- Gigabit Ethernet
- ファイバチャネル
- Ultra3 SCSI

パフォーマンスの改善に加えて、PCI-X は、オペレーティングシステムが効率的にアダプタと連携することで適切にエラー状態を管理する助けとなり、PCI バスの障害分離性を高めています。

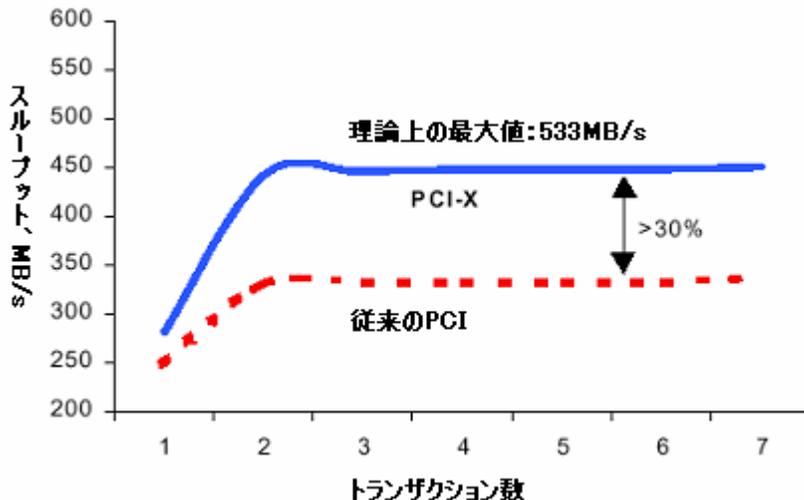
PCI-X はパフォーマンスと柔軟性を高め、PCI スロットとバス速度のトレードオフを最小化します。

例

予備シミュレーションテストによると、66MHz で 4 スロットの PCI-X の実装例では、従来の PCI バスの最大構成 (64 ビット、66MHz) 時に比べ、25% 以上の I/O パフォーマンスの改善が見られました。また、速度が異なる複数の PCI-X バスを使って、最大級のパフォーマンスと最大のスロット容量を同時に実現することもできます。

PCI はオープンな仕様なので、業界で絶え間なく続くテクノロジー革新の土台となってきました。しかしコンピューティングの進歩により、PCI 2.2、PCI-X などマルチドロップ パラレルバスアーキテクチャの性能では、I/O 帯域幅要件を満たせなくなりつつあります。

PCI-X のパフォーマンス



予備シミュレーションテストによると、66MHz の PCI-X の実装例では、従来の PCI バスの最大構成 (64 ビット、66MHz) 時に比べ、30% 以上の I/O パフォーマンスの改善が見られました。

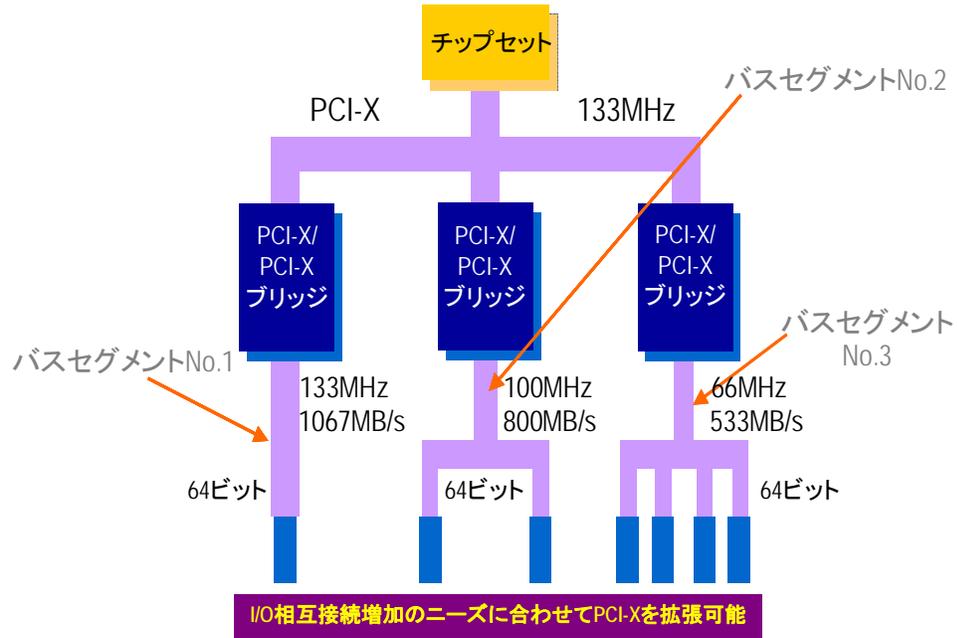
Gigabit Ethernet や Ultra3 SCSI コントローラに一般的な 4KB ブロックサイズを使った場合、PCI-X プロトコルは従来の PCI プロトコルのパフォーマンスを最大 34% 向上させています。なお、この向上率は、32 バイトプロセッサ キャッシュラインの理想的なメモリコントローラと理想的な 64 ビット PCI アダプタが使われていることを前提にしています。

PCI-X がパフォーマンスを改善できたのは、従来の PCI と PCI-X の 2 つの決定的な違いによります。

- 直接的なプロトコルではなくレジスタツーレジスタ プロトコルで実現された高いクロック周波数
- アトリビュートフェーズ、スプリットトランザクション、最適なウェイトステート、および標準ブロックサイズの移動など新たに拡張されたプロトコル

INTERNET PCI-X のパフォーマンスの詳細は、PCI SIG Web サイト (<http://www.pcisig.com/>) をご覧ください。

PCI-X の階層型構造



PCI-X ブリッジにより PCI は最大 256 のバスセグメントにリンク可能

異なる速度で動作する複数の PCI-X バスを用いて、パフォーマンスを向上させるか、スロット数を増やす選択が可能です。

例

133MHz で動作する PCI-X バスは、図に示されているように、PCI-X/PCI-X ブリッジ間で異なる速度の複数のスロットをサポートできます。

PCI-X バスセグメントに PCI-X デバイスのみ存在する場合、バスは PCI-X モードで動作します。PCI-X システムは自動的にバス周波数を調整し、バスセグメントの中で最も遅いアダプタに周波数を合わせます。

例

バスに 66MHz PCI-X アダプタがある場合、バスセグメントの最大クロック周波数は 66MHz です。

PCI-X は最大 256 のバスセグメントをサポートします。各セグメントは別々に初期化されるので、異なる動作周波数を用いることができます。

PCI-X システムの柔軟性

パフォーマンスの向上に加え、PCI-X は柔軟なシステム設計を実現します。最大の接続性が必要なソリューション向けに、PCI-X は 66MHz で動作する 4 つの拡張スロットをサポートできます。また最速のスピードが必要な場合は、PCI-X は 133MHz のスロットを 1 つサポートできます。

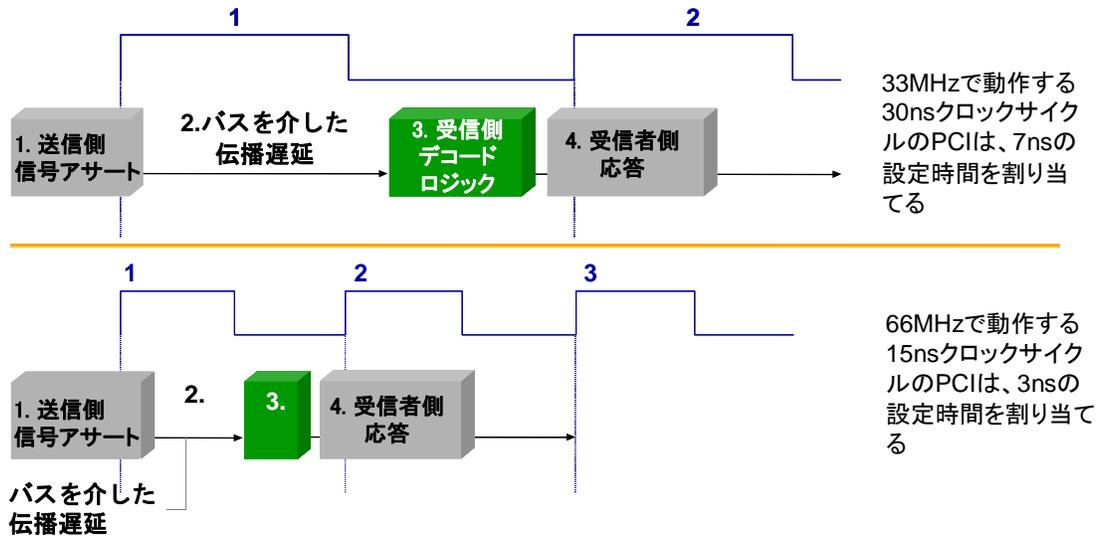
バス幅	バス周波数	バス帯域幅	PCI スロット	PCI-X スロット
32 ビット	33MHz	133MB/s		なし
64 ビット	33MHz	267MB/s		
64 ビット	66MHz	533MB/s	 または 	
64 ビット	100MHz	800MB/s	なし	
64 ビット	133MHz	1067MB/s	なし	

▲ 注記

PCI-X は、従来の PCI システムに対して完全な下位互換性を持っています。PCI-X を導入するにあたって、既存のハードウェアのデバイスドライバやオペレーティングシステムを変更する必要はありません。デバイスドライバを変更しなければならないのは、トランザクションバイトカウントなどの新しい機能を利用する場合だけです。

PCI と PCI-X の比較

PCI プロトコル



従来の PCI では、トランザクションのデコードに使用可能な時間は減り、バス周波数は 33MHz から 66MHz に増えました。

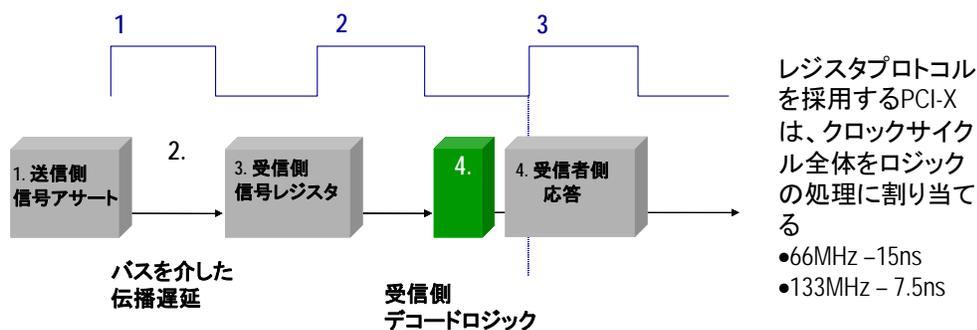
従来の PCI バスでは直接的なプロトコルが使われており、デバイスがコントロール信号を切り替えると次の操作が開始されます。

1. 立ち上がりクロックエッジで、デバイスは PCI バス上で信号をハイまたはローのステートに切り替える
2. 信号がバスを伝わる(伝播遅延)
3. 同一のクロックサイクル内で、受信側デバイスは信号をデコードし、信号が受信側デバイス向けのものか、また出力の 1 つを切り替えて応答しなければならないかを判断する
4. 受信側デバイスは、直ちに、すなわち次のクロックサイクルで応答する

クロック周波数が 33MHz の場合は、クロックサイクル全体が 30ns であるのに対して、デコードロジックに割り当てられる時間は 7ns です。33MHz の場合、この時間は受信側デバイスが次の立ち上がりクロックエッジで応答するのに十分です。しかし、バス周波数が倍の 66MHz (クロックサイクル時間は 15ns) になると、ロジックに割り当てられる時間は 3ns に短縮されます。

従来の PCI 仕様が抱えるこれらの時間的制約によって、従来の PCI バスまたはアダプタでは 66MHz の設計が困難でした。

PCI-X のレジスタツートレジスタ プロトコル



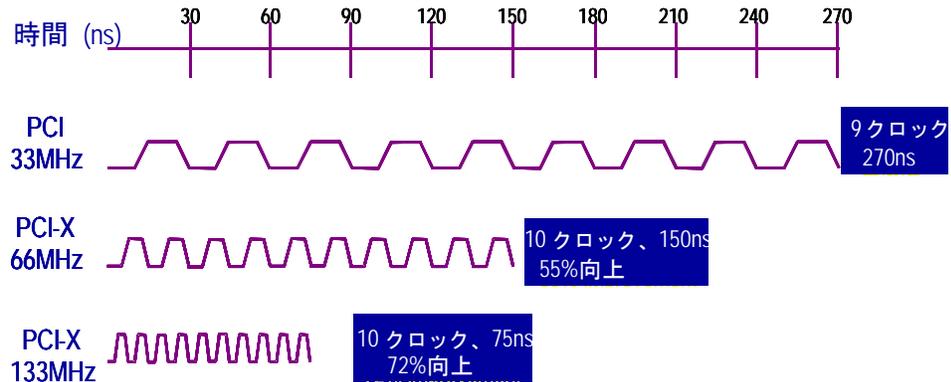
PCI-X プロトコルではクロックサイクル全体でいつデコードロジックが発生してもよい

PCI-X のレジスタツートレジスタ プロトコルでは、次の操作が実行されます。

1. 立ち上がりクロックエッジで、デバイスは PCI-X バス上で信号をハイまたはローのステートに切り替える
2. 信号がバスを伝わる
3. 信号はレジスタ、すなわちフリップフロップに送られ、次のクロックサイクルまでここで信号のステートが保持される
4. 受信側デバイスは、クロックサイクル全体を使って信号をデコードし、適切な応答を指定する
5. 受信側デバイスは、送信側デバイスが信号を切り替えた後、2クロックサイクル全体で応答する

PCI-X ではクロックサイクル全体でいつデコードロジックが発生してもよいので、タイミング制約が緩和されます。重要な相違点は、PCI-X トランザクションは一般に従来の PCI トランザクションよりも 1 クロックサイクル余計に必要なことです。従来の PCI で 9 クロックサイクルの書き込みトランザクションは、PCI-X では 10 クロックサイクルかかります。

PCI-X のタイミングの違い



典型的な書き込みトランザクションでの PCI と PCI-X プロトコルの実際のタイミングの違い

タイミング上の制約が緩和されたことで、66MHz 以上で動作するアダプタやシステム的设计が非常に簡単になりました。システム設計者は、タイミング上の制約の緩和によって、PCI-X バス上で、133MHz の PCI-X スロットを 1 つ使って最高のパフォーマンスを得るか、スロットを追加して最大の接続性を得るかを選択できます。

システム設計者がバス周波数の向上を選択すると、クロックサイクルは 1 つ増えますが、実際にトランザクションにかかる時間は大幅に短縮されます。33MHz で 9 サイクルかかるトランザクションは終了するのに 270ns かかりますが、133MHz で 10 サイクルかかる PCI-X トランザクションは 75ns で終了するので、トランザクション時間は 72% 短縮されたことになります。

またシステム設計者が 66MHz の周波数の保持を選択すると、バスセグメントにスロットを追加できます。あるデバイスから別のデバイスへの信号送信に、より多くの時間が割り当てられるので、信号は複数のスロットにまたがる長い経路を移動することができます。

プロトコルの拡張

PCI-X は、バスおよびデバイスの効率を改善するスプリットトランザクションやトランザクションバイトカウントなどの新しい特徴を提供します。

アトリビュートフェーズ

PCI-X プロトコルには、アトリビュートフェーズという新しいトランザクションフェーズがあります。アトリビュートフェーズは 36 ビットのアトリビュートフィールドを使っており、そのフィールドは従来の PCI 仕様よりも詳細にバストランザクションを表します。アトリビュートフェーズはアドレスフェーズの直後に続き、以下の情報を表す数個のビットが割り当てられています。

- トランザクションのサイズ
- トランザクションの順序
- キャッシュスヌープ要求
- トランザクション イニシエータに関する情報

アトリビュートフェーズには、次の拡張機能があります。

- ゆるやかな順序付け
- ノンキャッシュコヒーレントトランザクション
- トランザクションバイトカウント
- シーケンス番号

スプリットトランザクションのサポート

従来の PCI プロトコルは遅延トランザクションをサポートしていました。遅延トランザクションでは、データを要求するデバイスはターゲットにポーリングを行い、いつ要求が完了してデータが使用可能になるのかを確認する必要があります。

PCI-X でサポートするスプリットトランザクションでは、データを要求するデバイスがターゲットに信号を送信します。ターゲットデバイスは、要求を受理したことを要求側に通知します。ターゲットデバイスが新しいトランザクションを開始して要求側にデータを送信するまで、要求側は他の情報を処理できます。このように、スプリットトランザクションによってバスをより効率的に利用できるようになります。

最適なウェイトステート

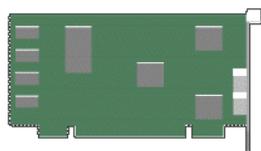
従来の PCI デバイスには、多くの場合、トランザクションが開始されるまでの余分なクロックサイクル、すなわちウェイトステートがありました。ウェイトステートが加わると、PCI デバイスがトランザクション処理に対応できない場合に、バスが「停止」します。この結果、バスのスループットが大幅に低下するおそれがあります。

PCI-X では、イニシャルターゲット待ち時間以外にウェイトステートがありません。PCI-X デバイスに転送データがない場合は、このデバイス自体をバスから削除するので、別のデバイスがそのバス帯域幅を利用できます。これにより、バスとメモリリソースの効率的な使用が実現します。

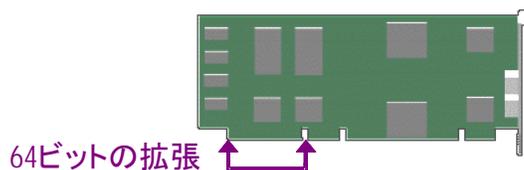
標準ブロックサイズの転送

PCI-X では、アダプタとブリッジ（ホスト/PCI-X 間と PCI-X/PCI-X 間）は本来の整列による 128 バイト境界上でのみトランザクションを切断します。したがって、より長いバーストが促進され、プロセッサバスやメインメモリなどのキャッシュラインベースのリソースをより効率的に利用できるようになります。また、PCI-X デバイス内でパイプラインを多用するアーキテクチャが使いやすくなります。

アダプタカードの選択



コスト重視型32ビットカード



64ビットの拡張 ↑↑
ハイパフォーマンス64ビットカード

PCI-X カードは現在の PCI システムでは、従来の 66MHz カードと同じように動作します。

アダプタカード	従来 の速度と帯域幅	PCI-X の速度と帯域幅	カード電圧
コスト重視型 32ビットカード	<ul style="list-style-type: none"> ■ 33MHz 133MB/s ■ 66MHz 256MB/s(オプション) 	<ul style="list-style-type: none"> ■ 66MHz 256MB/s(オプション) ■ 133MHz 533MB/s(オプション) 	<ul style="list-style-type: none"> ■ 3.3V ■ ユニバーサル
ハイパフォーマンス 64ビットカード	<ul style="list-style-type: none"> ■ 33MHz 256MB/s ■ 66MHz 533MB/s(オプション) 	<ul style="list-style-type: none"> ■ 66MHz 500MB/s(オプション) ■ 133MHz 1066MB/s(オプション) 	<ul style="list-style-type: none"> ■ 3.3V ■ ユニバーサル

PCI-X 133MHz アダプタは、PCI-X 向けに 66MHz、100MHz、および 133MHz の周波数をサポートしています。100MHz PCI-X スロットで最高のパフォーマンスを得るには、133MHz PCI-X アダプタを使用します。

相互運用性マトリクス

		従来のPCIカード		PCI-X カード		
		33MHz (5Vまたはユニ バーサル)	66MHz (3.3Vまたはユ ニバーサル)	66MHz (3.3Vまたはユ ニバーサル)	133MHz (3.3Vまたはユ ニバーサル)	
従来型システム	33MHz	33MHz	33MHz	33MHz	33MHz	
	66MHz		66MHz	66MHz	66MHz	
PCI-Xシステム	66MHz				100MHz	133MHz
	100MHz					
	133MHz					

PCI-X デバイスの動作速度は、33～133MHz です。PCI-X システムは、システム内の最も低速なデバイスの速度で動作します。以前のバージョンとの互換性を実現するため、すべての PCI-X デバイスは、古いシステムおよび低速デバイスや PCI デバイスを持つ PCI-X システムでも動作するために、より遅い速度でも動作する必要があります。

PCI-X バスは、133MHz の PCI-X デバイス 1 台、100MHz のデバイス 2 台と 66MHz のデバイス 4 台以上にのみ対応可能です。

プロトコル	バス幅	バス周波数	バーストスループット
32ビット PCI	32ビット	33MHz	132MB/s
64ビット PCI	64ビット	33MHz	264MB/s
64ビット/66MHz PCI	64ビット	66MHz	528MB/s
50MHz PCI	64ビット	50MHz	400MB/s
66MHz PCI	64ビット	66MHz	528MB/s
100MHz PCI	64ビット	100MHz	800MB/s
133MHz PCI	64ビット	133MHz	1056MB/s

PCI-X での PCI ホットプラグのサポート

PCI-X テクノロジは PCI ホットプラグをサポートしたため、ホットプラグコントローラ的设计の許容範囲はとてつも広がりつました。PCI-X に移行する際はハードウェアとソフトウェアへの影響をいくつか考慮する必要があります。

ハードウェア関連の影響

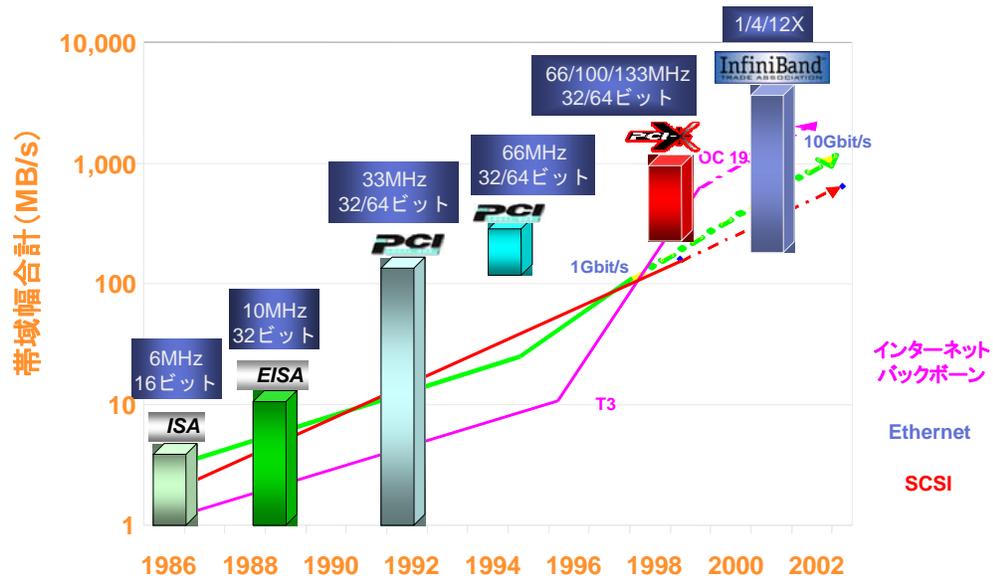
ホットプラグコントローラは、

- PCIXCAP ピンをチェックし、PCI-X アダプタを特定する手段をホットプラグシステムドライバに提供する必要があります。
- そのスロットの RST# の立ち上がりエッジの前に、適切なタイミングでバス上の PCI-X 初期化パターンを駆動する必要があります。
- アービタ (ホットインサクション中のバスの所有権を調整する機能) と連携する必要があります。

ソフトウェア関連の影響

ホットプラグシステムドライバは、挿入されたアダプタがそのバスの周波数と動作モードを確実にサポートするように、挿入されたカードの M66EN と PCIXCAP ピンを読み取る必要があります。

PCI-X 2.0

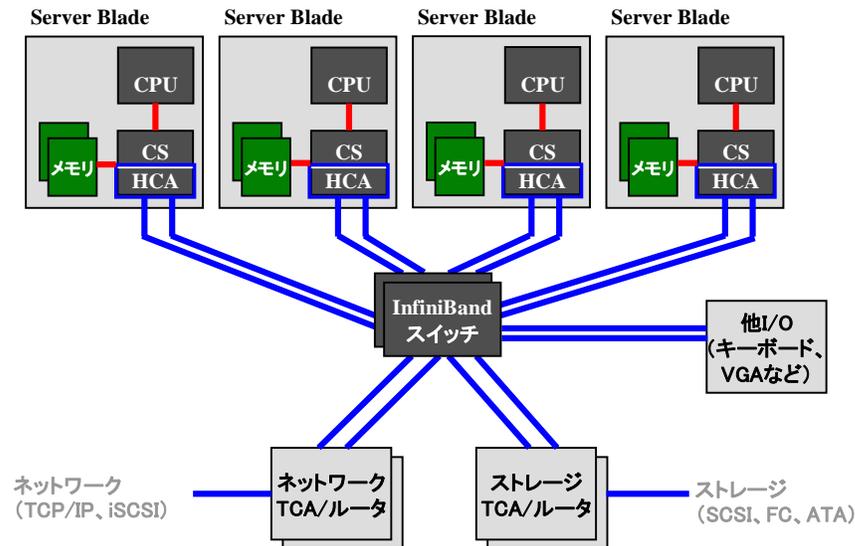


PCI-X 2.0 は PCI ローカルバス仕様のハイパフォーマンス拡張バージョンです。

PCI-X 2.0 の帯域幅は PCI-X の 4 倍ですが、ピン数は同じです。PCI-X 2.0 は、PCI や PCI-X のハードウェアおよびソフトウェアと下位互換性があります。PCI-X 2.0 の主な特徴は次のとおりです。

- 2 倍または 4 倍の帯域幅
- 次の既存要素を使用
 - フォームファクタ
 - ピン配列
 - コネクタ
 - バス幅
 - プロトコル
 - 電気信号
- ECC (error checking and correcting) をサポート

InfiniBand



InfiniBand アーキテクチャの構成要素

1999年、Compaqなどのコンピュータ業界の主要企業は、チャンネルベースのスイッチ付きファブリック I/O 相互接続のための新しい業界標準仕様を開発するために、InfiniBand Trade Association を設立しました。InfiniBand Trade Association の目標は、既存の相互接続テクノロジーの長所を *InfiniBand* アーキテクチャと呼ばれる 1 つの仕様にまとめることです。

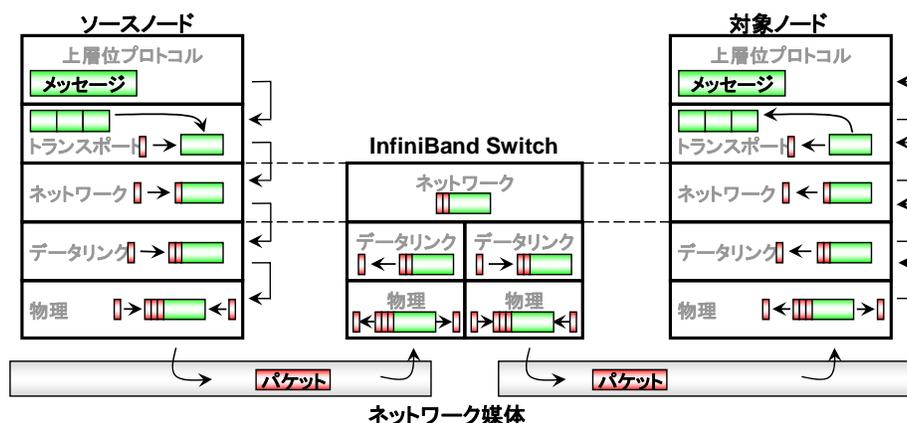
InfiniBand Trade Association の主要メンバーとして、HP はこの仕様に重要なテクノロジーを提供しています。

INTERNET InfiniBand Trade Association の詳細は、Web サイト (<http://www.infinibandta.org/>) を参照してください。

InfiniBand は、次の特徴によりサーバの信頼性を高めます。

- チャンネルベースのアーキテクチャ — ファブリックノード間の接続を含むチャンネルベースの I/O モデルをベースにしたアーキテクチャ
- メッセージパッシング構造 — データ転送
- 自然な冗長性 — 冗長性を得るためにノードはファブリックリンクに接続されており、ファブリックは複数レベルの冗長性を持つよう設計されている

InfiniBand テクノロジ



InfiniBand アーキテクチャの層

InfiniBand テクノロジはサーバとリモートストレージやネットワークデバイスを接続します。InfiniBand テクノロジを用いると、高帯域幅デバイスは、他の PCI バストラザクシオンに影響を与えることなくスイッチ付きファブリックアーキテクチャを使用できます。

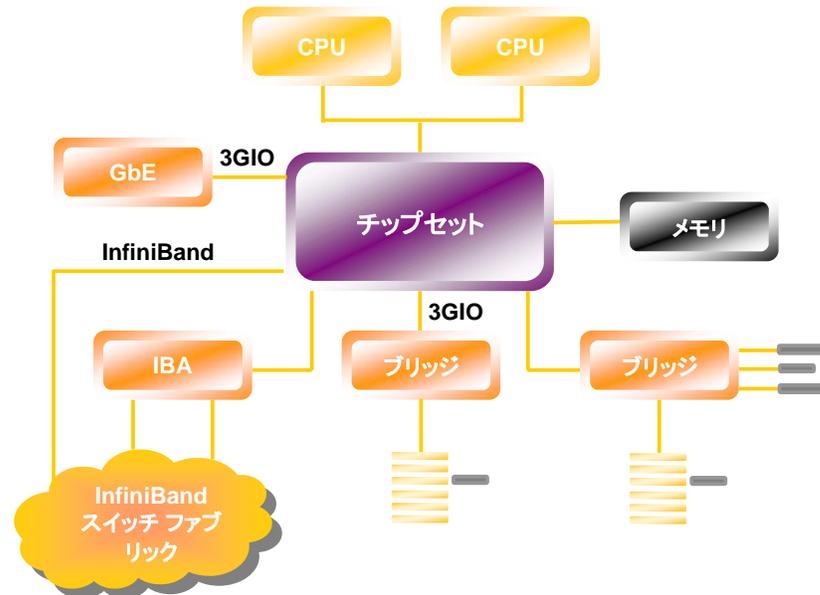
ホストノードには、1 つ以上の双方向リンクを経由してメモリコントローラをファブリックに接続するホストチャンネルアダプタがあります。ホストチャンネルアダプタには、ハードウェアでリンクプロトコルを実現するリンクプロトコルエンジンがあります。ターゲットチャンネルアダプタを用いると、エンドデバイスをファブリックやスイッチに直接接続できます。ターゲットチャンネルアダプタはホストチャンネルアダプタよりシンプルで、InfiniBand アーキテクチャやデバイス固有の機能を最小限サポートするのに必要なことだけを実現します。

InfiniBand ファブリックは、スイッチやルータを経由してホストとデバイスを接続します。スイッチはサブネットワークを作り、ルータはスイッチを相互接続します。各ホストチャンネル、ターゲットチャンネルアダプタ、スイッチ、およびルータは、インターネット対応のため、IPv6 アドレス指定をベースにした Globally Unique Identifier (GUID) を持つこととなります。

InfiniBand では、各ホストとターゲットチャンネルアダプタ間で高速の双方向シリアルリンクが使われています。双方向リンクには、データを同時に送受信する専用チャンネルがあります。

InfiniBand はまず、リンク毎のチャンネル数を増やすことで理論上の帯域幅を増やします。つまり、InfiniBand 仕様では 1、4 および 12 幅のデュプレックス差分ペアリンクが必要となります。各リンクのチャンネルの半数はデータを送信し、残りの半数はデータを受信します。各チャンネルは 1 ビット幅で、理論上の信号レートは 2.5Gb/s です。したがって、2、8、および 24 チャンネル双方向リンクの理論上の帯域幅は、それぞれ 5Gb/s、20Gb/s、および 60Gb/s となります。InfiniBand テクノロジの開発が進むにつれ、信号レートは高速化し、リンク毎の帯域幅も拡張されます。

PCI Express



PCI Express ベースのサーバ/ワークステーションシステム

Third Generation I/O (3GIO) Work Group は、HP、Dell、IBM、Intel、Microsoft、および PCI-SIG が創設メンバーの非営利特別利益団体で、多数の主要企業と連携し、PCI Express と呼ばれる新しいシリアル I/O 相互接続を定義します。

PCI Express は、パケット化されたプロトコルとロード/保存アーキテクチャを定義します。層状のアーキテクチャなので、銅メディア、光学メディア、新しい素材を使った信号メディアへの接続が可能になります。PCI Express は、拡張されたグラフィック帯域幅向けのグラフィック I/O 接続ポイントおよび InfiniBand などの他の相互接続向けの接続ポイントとして、アダプタカードへの接続を実現するチップツーチップやアドインカードで用いられます。

PCI Express は、複数のポイントツーポイント接続を使うレーンと呼ばれるリンクに、フルシリアル相互接続します。PCI Express の初期速度は一方方向 2.5Gb/s なので、一方方向 200MB/s 通信チャンネルを 2 つ実現しています。これは従来の PCI の速度のおよそ 3 倍です。また、レーン数を増やすことで、帯域幅は簡単に拡張できます。

PCI Express は、長い目で見た拡張性を念頭に設計されています。主な特長は次のとおりです。

- ピン毎の高い帯域幅
- 低いオーバーヘッド
- 短い待ち時間
- 内蔵クロックアーキテクチャ

内蔵クロックのタイミングおよび差分信号方式により、大容量メタルテクノロジーを採用した場合は 10~15Gb/s と考えられるメタル信号方式の限界まで PCI Express のパフォーマンスを向上できます。

内蔵クロックアーキテクチャではパラレルアーキテクチャよりピン数が少なくすむので、経路指定が簡素化されます。この特徴は、コストを最小限に抑え、構成要素、マザーボード、アダプタ、システム設計の柔軟性を高めるのに役立ちます。また信号が少ないと、少ないボード空間と小さなコネクタでシステム設計が可能になるので、より小さく革新的なフォームファクタをサポートできます。

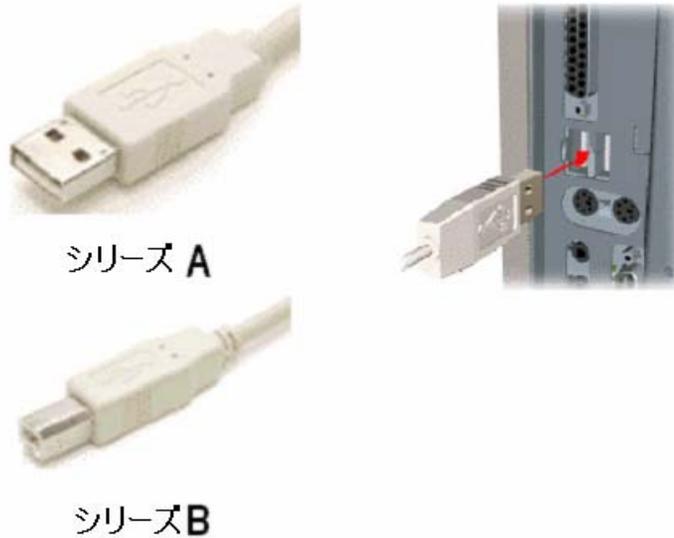
PCI Express は現在の PCI ソフトウェア環境と互換性があります。ゆくゆくは PCI から PCI Express に切り替わることが予想されますが、InfiniBand アーキテクチャと PCI Express は共存できます。InfiniBand アーキテクチャは主に、堅固なメッセージパッシングアーキテクチャを採用するマルチコンピュータ環境の共有 I/O に重点を置いています。一方、PCI Express はロード/保存アーキテクチャを採用するコスト重視型ローカル I/O に重点を置いています。PCI Express は、InfiniBand アーキテクチャの電気層とシリコンテクノロジー上に構築されます。これは、この 2 つの I/O アーキテクチャの将来的な統合を実現するための重要な成功要因です。

次のようなテクノロジーの進歩により、PCI バスの現在の性能と費用を重視する潜在的な拡張性では I/O 要件に対応できなくなることが予想されます。

- 10GHz を超えるプロセッサ速度
- 高速化するメモリ速度
- 高速グラフィック処理
- ギガビット Ethernet

ハイパフォーマンスで、拡張性が高い汎用 I/O アーキテクチャを実現する PCI Express は、開発者にとって貴重です。さらに、PCI Express はデスクトップ、モバイル、サーバ、通信、内蔵アプリケーション、開発中のアプリケーション、将来のアプリケーションの要件を満たす長期的な汎用 I/O 相互接続として設計される予定です。

USB



Universal Serial Bus (USB)は、コンピュータおよびテレコム業界により開発された周辺バスの規格です。この規格を採用すると、システムを再起動または再構成することなく、周辺デバイスをホットプラグできます。USB デバイスは物理的に接続されるとすぐ、自動構成されます。

USB を用いると、多数の周辺機器を同時に接続できます。多くのコンピュータには、USB ポートが 2 つあります。USB ハブと呼ばれる特別な USB 周辺機器には、複数のデバイスをダイジーチェーン方式で接続する追加ポートがあります。

USB はまた、電力を多数の周辺機器に振り分けます。USB を用いるとコンピュータが自動的に必要な電力を感知してデバイスに供給するので、見た目の悪い電源ボックスは不要になります。

USB には次のような特徴があります。

- **業界の幅広いサポート** — コンピュータとテレコム分野の主要企業が策定した規格
- **使いやすさ** — 周辺機器は自動構成され、ホットプラグ可能
- **柔軟性** — 様々なデバイスを同一のコネクタを使って同時にコンピュータに接続可能

USB の仕様には次のものがあります。

- 12Mb/s の低コストデータ転送レート
- 最大 127 台のデバイスをサポート
- 同期データ転送と非同期データ転送両方をサポート
- 最大 5m のケーブルセグメント

USB コネクタ

USB コネクタには次の 2 種類があります。

- シリーズ A
- シリーズ B

シリーズ A およびシリーズ B の機能は同じですが、USB アーキテクチャトポロジに違反する接続を防止するために形状は違います。

シリーズ A

シリーズ A コネクタは、平らな長方形で、USB ホストやハブの電源に近いポート受け口に差し込まれます。

オスコネクタとメス受け口には、直列の接続点が 4 つあります。シリーズ A コネクタはすべての USB デバイスで使用できます。ほとんどの場合、USB ケーブルは周辺機器に付属しています。

シリーズ B

シリーズ B コネクタは角が丸い正方形に近い形で、USB デバイスやハブの電源から遠い受け口に差し込まれます。

オスコネクタとメス受け口には、2 対 2 で重なった接続点が 4 つあります。付属のケーブルが制限される場合に使うために、シリーズ B コネクタは作られました。

Windows 98 および Windows 2000 が提供する USB クラスドライバおよび WDM (Windows Driver Model) は、特別なデバイスクラス仕様に適合するデバイスをサポートします。デバイスは、オペレーティングシステムが提供する一般的なクラスドライバを使う場合もあります。また、ハードウェアに追加された独自の特徴を利用するために、メーカーがドライバや WDM ミニドライバ (デバイスクラスによる) を提供している場合もあります。

USB 2.0

USB 2.0 は、インタフェースの性能を 12Mb/s (200 x 56Kb/s) から 120~240Mb/s に拡張しました。

復習問題

1. バス幅とは何ですか。
 - a. 1秒間に発生するバスサイクル数
 - b. バス内の電気回線の数
 - c. データ転送のプロセス
 - d. 一定時間内にバスを流れるデータ量
2. バスの最大転送レートを計算するには、どの式を使いますか。
 - a. (速度 x 幅)/バスサイクル毎クロックサイクル
 - b. (速度 x バスサイクル毎クロックサイクル)/幅
 - c. (幅 x バスサイクル毎クロックサイクル)/速度
 - d. 速度/(バスサイクル毎クロックサイクル x 幅)
3. システムパフォーマンスを向上させるための4つの基本的な対策を教えてください。

.....

.....

.....

.....
4. バスマスタは、バス上の他のデバイスと直接通信するデバイスです。
 - 正
 - 誤
5. PCIバスは_____ビットまたは_____ビットのバスで、一度に_____または_____バイトのデータを転送します。その時のクロックレートは_____MHzまたは_____MHzで、最大バーストスループットの範囲は、_____~_____MB/sです。
6. ブリッジPCIバスアーキテクチャのシステムパフォーマンスは、PCIバスが追加された分だけ向上します。
 - 正
 - 誤

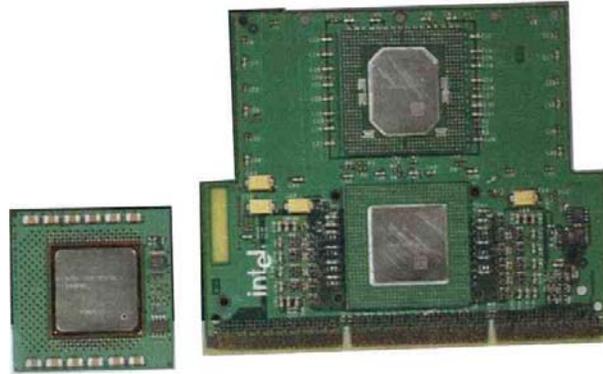
7. PCI-X バスセグメントは 66MHz で動作する 4 つの拡張スロットをサポートできます。
- 正
 - 誤
8. PCI-X はクロックサイクル全体でデコードロジックを発生させ、タイミング上の制約を緩和します。
- 正
 - 誤
9. 各用語とその説明を対応付けてください。
- a. PCI Express 高帯域幅デバイスがスイッチファブリックアーキテクチャを使い、サーバをリモートストレージやネットワークデバイスに接続できる
 - b. InfiniBand パケット化プロトコルとロード/保存アーキテクチャを定義する
 - c. USB システムを再起動または再構成することなく、周辺デバイスのホットプラグを可能にする

目的

ここでは、次の内容について学習します。

- システムアーキテクチャにおけるチップセットの重要性
- サーバアーキテクチャの進化と主要なアーキテクチャテクノロジー
- ProLiant サーバで使用されている以下のチップセットの種類と特徴
 - ServerWorks ファミリのチップセット
 - ハイリーパラレル システムアーキテクチャ
 - ProFusion
 - F8

チップセット



サーバの全体的なアーキテクチャは、チップセットのアーキテクチャに大きく依存します。性能上のボトルネックを解消し、スケーラビリティを確保しながら、性能を向上させるためには、チップセットが最適かつバランスの取れたアーキテクチャを備える必要があります。

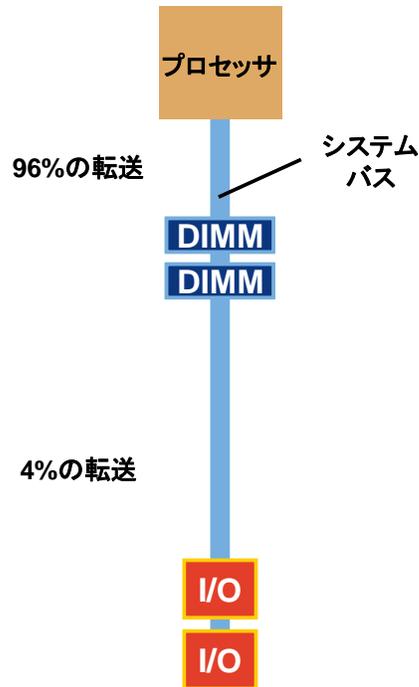
チップセットはプロセッサを動かす、使用可能なメモリ容量や、コンピュータがサポートできるインターフェースを決定付けます。したがって、全体の性能にとって、チップセットはプロセッサ自体と同じくらい重要です。

チップセット開発におけるパートナーシップ

HP は Intel および ServerWorks と協力し、3 社のテクノロジーを結集することによって、最高性能のチップセットテクノロジーを統合した最良のサーバアーキテクチャを提供しています。

サーバアーキテクチャの進化

初期のアーキテクチャ

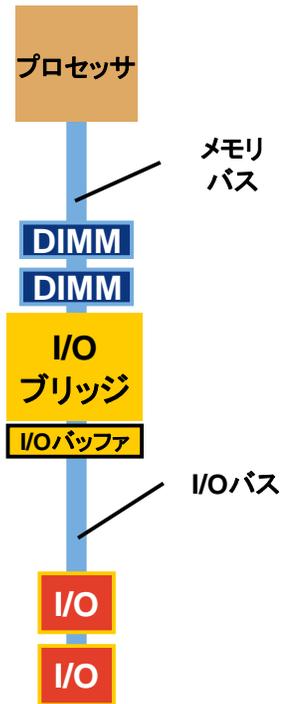


初期の PC システムアーキテクチャではシングルバス方式が採用され、すべての周辺機器に対して 1 つの速度と 1 つのバス幅で対応していました。

この方式には次のボトルネックがあることが、明らかになりました。

- すべてのデバイスに対して、1 つのシステムバスしか使用できない
- 同時に 1 つのデバイスしかシステムバスを使用できない
- バスの転送速度は、それがシステムか I/O かにかかわらず、同じバス速度を上限とする
- データ転送の約 96% は、プロセッサとメモリ間で行われる

並列バス

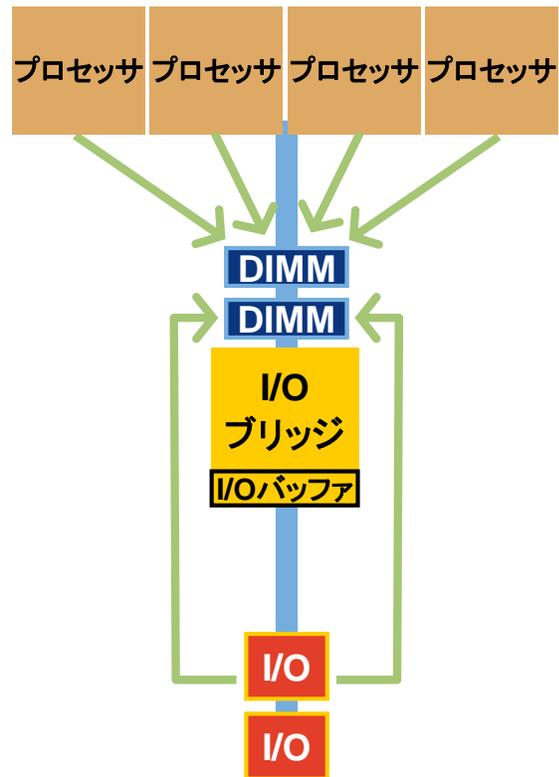


次の段階では、バスが2つの独立したバスに分割されました。

この設計では、プロセッサとメモリ間のメモリバスの速度を、低速のI/O拡張バスよりも高速にすることができました。2つのバス間でのデータ転送の同期には、I/Oブリッジが採用されました。

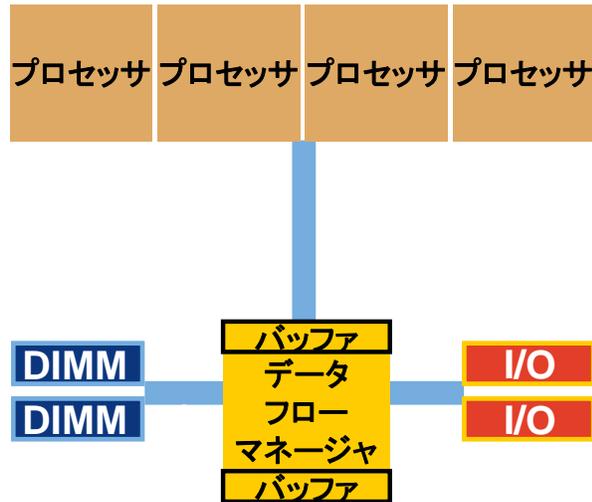
これにより、ボトルネックは、ホストバスをめぐってI/Oが競合するI/Oブリッジに移っていきました。そして、ホストバスをめぐるI/Oデバイスの競合によって生じる遅延を低減し、ブリッジの効率を改良するために、I/Oバッファリングが追加されました。

システム並列性の拡大



サーバの進化における次の段階は、バス マスタリングテクノロジーの登場でした。これにより、複数のバスマスタデバイスが、同時にシステムメモリにアクセスできるようになりました。また複数のプロセッサが、データをメインメモリに直接転送できるようになりました。I/O アダプタも、データを直接メインメモリに転送できるようになりました。この結果、性能が飛躍的に向上したために、メモリアクセス待機状態が増大し、これが新しいボトルネックとなったのです。

システム並列性の拡大—MIOC



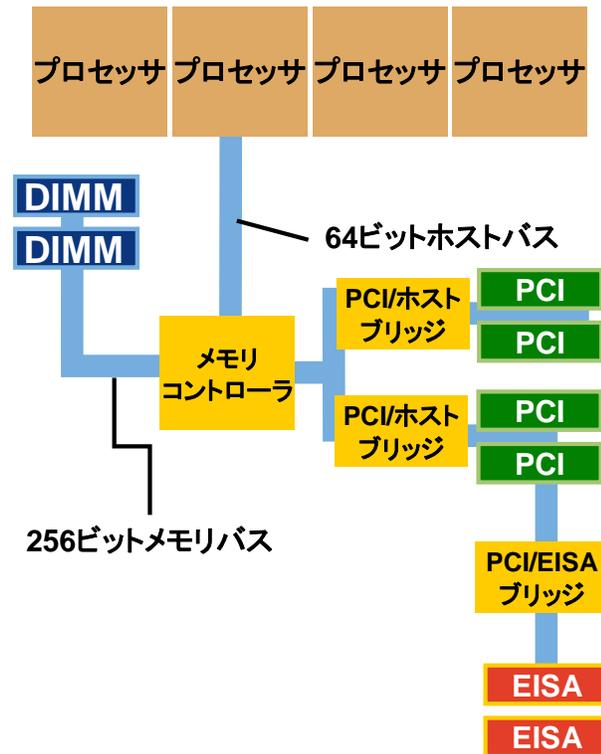
これに対する解決策は、データフローマネージャでした。これによって、バスマスタを原因とするボトルネックが解消しました。データフローマネージャとは、以下の処理を行うメモリ兼 I/O コントローラ (MIOC) です。

- バスアービトレーション
- タイミング
- プロセッサ、メモリ、および I/O の間のバッファリング

さらにメモリバス幅が 128 ビットに拡張されたので、複数の並列 I/O トランザクションが可能になりました。

次にボトルネックとなったのは、I/O アクセスと I/O バス速度でした。サーバに搭載される I/O デバイス数の増加にもかかわらず、バスが 1 つしかなく、バス速度も 1 種類のみだったからです。

並列 I/O バスによる I/O 並列性の拡大



システムアーキテクチャの次の段階では、デュアルピアやトリプルピア I/O バスの追加によって並列性が拡張されました。この設計を可能にしたのは、ブリッジと呼ばれる I/O コントローラの追加でした。この設計では、どのバスにある周辺機器も、独立してプロセッサやメモリにアクセスできるようになりました。

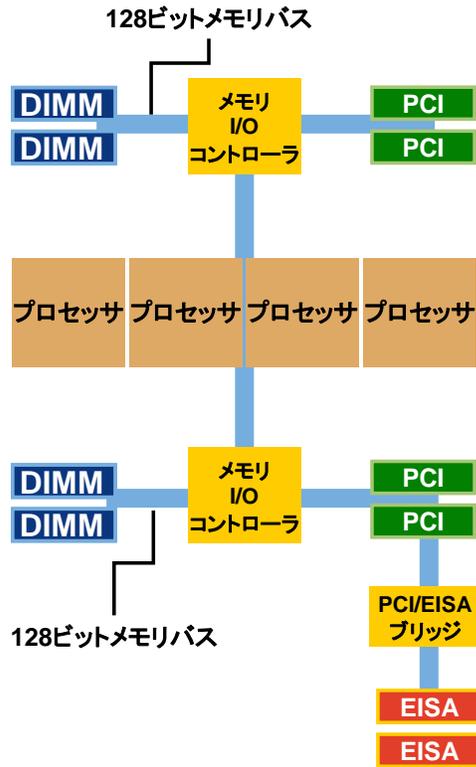
この設計は、I/O デバイスの速度が高速化したときに、複数の I/O バスがそれぞれ異なる速度で動作できるので、低速の I/O と高速の I/O の分離が可能です。またブリッジ内にバッファが用意され、I/O 転送の待ち行列ができるので、遅延時間が減少します。

この設計の主な利点は次のとおりです。

- シングルバス システムの 2 倍の I/O 幅 — 133MB/s (267MB/s) から 267MB/s (533MB/s) に高速化
- シングルバス システムより多くの PCI デバイスをサポート
- 使用頻度の高い周辺機器 (グラフィックコントローラやディスクコントローラなど) を別のバスに配置することによる I/O 負荷と性能のバランス

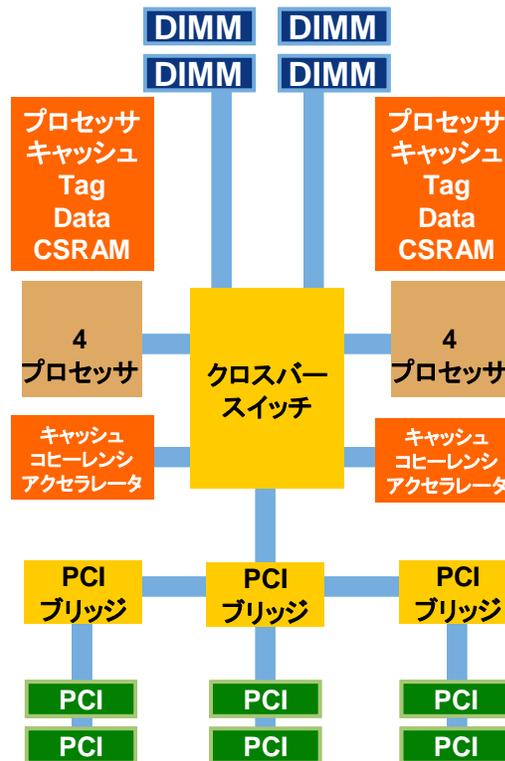
しかしこの設計では、マルチプロセッサ方式で命令を処理するサーバにおいて、重要なメモリや I/O サブシステムへのアクセスをプロセッサ間で競合することになるので、帯域幅の制限が問題になります。

デュアル メモリ コントローラ



ピア I/O バスの登場によって、ボトルネックは I/O からホストバスメモリコントローラへ戻りました。チップセットの進化における次の段階は、デュアルメモリ コントローラの登場でした。この HP システムアーキテクチャ設計は、ハイリーパラレル システムアーキテクチャ (HPSA: Highly Parallel System Architecture) と呼ばれました。このアーキテクチャもデュアルピア PCI バスを備えていました。HP はこの強力なテクノロジーを ServerWorks と共同開発し、世界で初めて市場に投入したのです。

クロスバースイッチ



HP チップセットの進化における次の段階は、メモリ I/O コントローラでのデータスループットのさらなる向上でした。HP は Corollary (現在は Intel に統合) と協力して、スイッチ コントローラを使用する ProFusion チップセットを開発しました。

メモリ コントローラにおける競合ボトルネックを減少させるため、HP はメモリ I/O コントローラの代わりにクロスバースイッチを採用しました。メインフレーム技術を採用したクロスバースイッチを使用することで、プロセッサ、メモリ、および I/O 間で並列に読み書きすることが可能になるので、5 つのメインポートのそれぞれから他の各ポートへの高速データ転送を実現します。

物理システムバスは 2 つ存在しますが、オペレーティングシステムからは、これらのバスが論理的に 1 つのシステム イメージとして認識されます。

改良型クロスバースイッチ

HP は ProFusion チップセットで培った経験を生かして、クロスバースイッチ設計を採用し、さらに改良を加えて HP F8 チップセットを完成させました。

ServerWorks チップセット

ServerWorks にとって HP は提携企業であり、共同開発者であり、最初の顧客です。HP が初めて ServerWorks と提携したのは 1995 年のことで、ServerWorks はそのときすでにハイリー パラレル システムアーキテクチャを設計していました。このアーキテクチャは、ServerSet I の開発に利用されました。1997 年、HP はメインストリームサーバおよびワークステーション製品上で ServerSet I を使用した最初のメーカーとなりました。

HP は ServerWorks と協力して、ServerWorks ファミリの各チップセットを設計しました。その中には HP 特許のサウスブリッジ設計もあり、現在でも ServerWorks チップセット設計で使用されています。HP はこの設計によって、ProLiant サーバファミリ全体に統一的な I/O 設計を実現し、ハードウェアを再検証する必要性を減少させました。

HP と ServerWorks の技術提携関係は、ServerWorks が第 4 世代のチップセットを開発している今日も続いています。HP は第 4 世代の ServerWorks チップセット実装のために、PCI-X テクノロジーとホットプラグ PCI-X テクノロジーを ServerWorks にライセンス供与しています。

また ServerWorks は現在、チップセットと周辺機器間で高性能なデータパスを実現する業界標準の PCI-X ブリッジを開発中ですが、HP は独自の I/O 専門技術によってこれを支援しています。

チップセット	搭載している ProLiant サーバ	サポートされているプロセッサ
ServerSet I	PL 5500 PL 3000	Pentium Pro Pentium II
ServerSet II LE	PL 3000	Pentium II Pentium II Xeon
ServerSet III LE	ML330、ML330 G2、 ML350、ML350 G2、 ML370、ML530、 DL320、DL360、DL380	Pentium III Pentium III Xeon
ServerSet III LE LP	BL10e	ULV Pentium III
ServerSet III HE	ML570、 DL580	Pentium III Pentium III Xeon
ServerSet III HE SL	ML370 G2、DL360 G2、 DL380 G2、BL20p	Pentium III Pentium III Xeon
ServerSet GC-SL	ML310、ML330 G3、 DL320 G2、 BL10e G2	Pentium 4 ULV Pentium M Xeon
ServerSet GC-LE	ML350 G3、ML370 G3、 DL360 G3、DL380 G3、 DL560、 BL20p G2、BL40p	Xeon Xeon MP
ServerSet GC-HE	ML530 G2、ML570 G2、 DL580 G2	Xeon Xeon MP

ServerSet I

ServerSet I チップセットは、最大 4 個の Intel Pentium Pro プロセッサをサポートする初期 ProLiant 5500 サーバ、および最高 2 個の Pentium II プロセッサをサポートする初期 ProLiant 3000 サーバに実装されています。

ServerSet II LE チップセット

ProLiant 3000 サーバの最初のシステムアップグレードでは、デュアル PCI/メモリコントローラを持つ ServerSet II LE チップセットが使用されました。

ServerSet III LE チップセット — シングル ノース ブリッジ

ServerSet III LE チップセットを使用する ProLiant サーバは次のとおりです。

- ML330
- ML330 G2
- ML350
- ML350 G2
- ML370
- DL320
- DL360
- DL380

ServerSet III LE チップセット — デュアルノースブリッジ

ProLiant ML530 サーバは、ServerSet III LE チップセットとデュアルノースブリッジ構成を使用しています。

ServerSet III LE LP チップセット

ProLiant BL10e は、ServerSet III LE チップセットとシングルノースブリッジ構成を使用しています。

ServerSet III HE チップセット

ProLiant ML570 サーバと ProLiant DL580 サーバでは、ServerSet III HE チップセットが使用されています。

一般に、シングルノースブリッジ搭載の対称型マルチプロセッサ (SMP) サーバのシステム性能は、デュアルノースブリッジ搭載の SMP サーバより劣ります。しかしながら ServerSet III HE チップセットでは、メモリと I/O 機能が拡張されているので、ボトルネックが解消され、全体的なシステムの性能が大幅に向上しています。

ServerSet III HE SL チップセット

ServerSet III HE SL チップセットを使用する ProLiant サーバは次のとおりです。

- ML370 G2
- DL360 G2
- DL380 G2
- BL20p

ServerSet GC SL チップセット

ServerSet GC SL チップセットを使用する ProLiant サーバは次のとおりです。

- ML310
- ML330 G2
- DL320 G2
- BL10e G2

ServerSet GC LE チップセット

ServerSet GC LE チップセットを使用する ProLiant サーバは次のとおりです。

- ML350 G3
- ML370 G3
- DL360 G3
- DL380 G3
- DL560
- BL20p G2
- BL40p

ServerSet GC HE チップセット

ServerSet GC HE チップセットを使用する ProLiant サーバは次のとおりです。

- ML530 G2
- ML570 G2
- DL580 G2

ハイリーパラレル システムアーキテクチャ

ハイリーパラレル システムアーキテクチャは、2 個の PCI バスと 2 個のメモリコントローラを並列に使用することによって、マルチプロセッサシステムの性能を飛躍的に向上させます。このアーキテクチャでは、AGTL+バステクノロジーを使用してシステムの性能をさらに高めています。

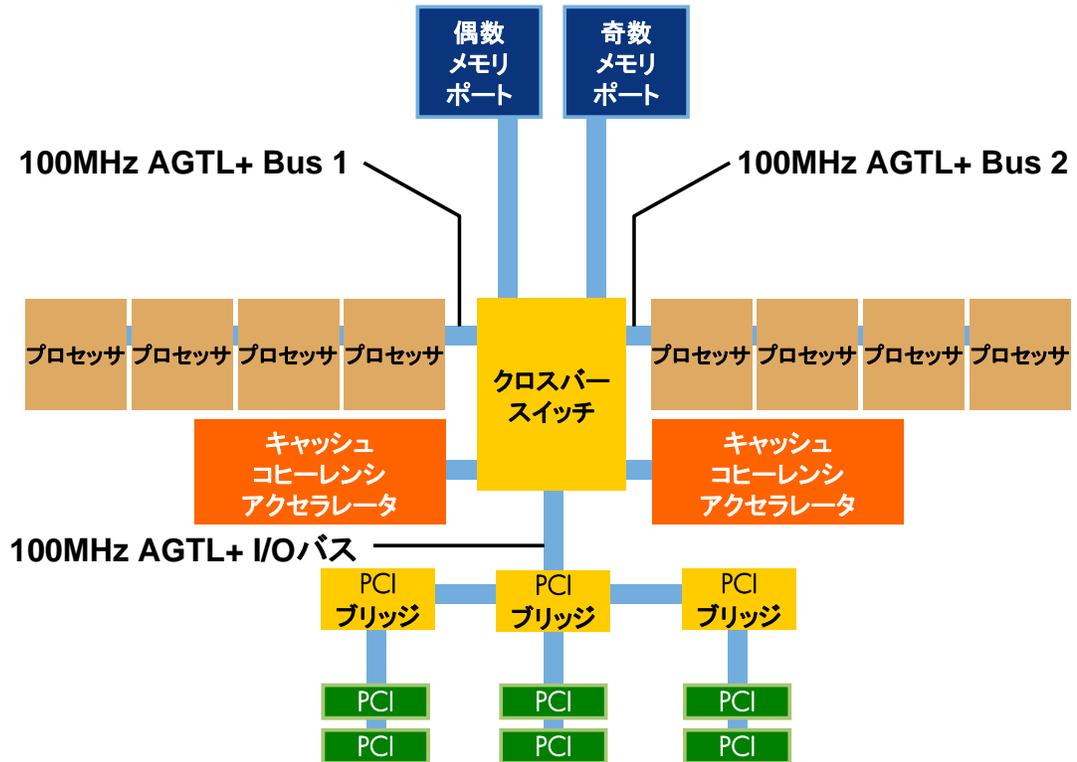
このシステム設計は、メモリ、I/O、およびプロセッサなどの主要サブシステムで、より大きな帯域幅を確保する必要性に対応します。デュアルメモリコントローラとデュアルピア PCI バスを使用して、プロセッサと重要なサブシステム(メモリやディスクドライブなど)の間のデータ転送を行います。

デュアルメモリコントローラを採用しているサーバは、メモリ要求を並列処理できるので、100MHz (133MHz、2.12GB/s) SDRAM において、最高 1.6GB/s のメモリ帯域幅を達成できます。

ハイリーパラレル システムアーキテクチャのコンポーネントは次のとおりです。

- AGTL+バス
- Intel Pentium II および III プロセッサ
- デュアル Wide-Ultra SCSI コントローラ
- デュアルピア PCI バス
- デュアルメモリコントローラ
- インタリーブ メモリ

ProFusion チップセット



8 ウェイ Intel ベース サーバのブロック図

ProFusion チップセットは HP と Intel が共同開発した HP 8 ウェイシステムアーキテクチャです。このチップセットは、1.6GB/s のメモリ帯域幅 (2 x 800MB/s) を持つ独自の 2 ポート (バス) メモリであり、両方のポート上のメモリに同時にアクセスできます。

ProFusion メモリアccess コントローラ (MAC) は、別個のメモリバス上の 2 つのメモリコントローラを管理します。この技術は、キャッシュラインインタリーブを利用した高速 SDRAM (同期 DRAM) を採用しています。片方のコントローラが奇数のキャッシュラインアドレスをすべて管理し、もう片方が偶数のキャッシュラインアドレスをすべて管理します。この結果、2 つの連続するキャッシュラインに、アドレス番号順にアクセスした場合の遅延が減少します。MAC は、2 つのメインシステムバスにある 2 つのプロセッサから、メモリおよび I/O システムへのアクセスを調整します。

ProFusion データインタフェースバッファ (DIB) は、データバス制御、および AGTL+バスとメモリ間のバッファリングを行います。DIB は、プロセッサ、I/O AGTL+バス、およびメモリの間のバッファで構成されます。MAC がメモリサイクルを開始したときに、必要な DIMM がビジー状態である場合、DIB が一時的にアドレスを保存し、次のサイクルでメモリへ転送します。

ProFusion アーキテクチャは、共有メモリ SMP アーキテクチャの単純なプログラミングモデルを利用して、潜在的な性能上のボトルネックを低減しています。

ProFusion クロスバースイッチ

ProFusion チップセットは、5 ポイント、ノンブロッキング クロスバースイッチを使用して以下を結合します。

- 独立した 2 つのプロセッサバス
- キャッシュコヒーレンシフィルタ
- 独立した 1 つの I/O バス
- 2 つのメモリポート

ノンブロッキング クロスバースイッチと I/O コントローラは、特に性能向上の目的で設計されているので、ProFusion チップセットの登場により、市販のアプリケーションやオペレーティングシステム向けの業界標準の 8 ウェイサーバの規模を拡張できます。

標準のベンチマークテストと実績が、ProFusion のアーキテクチャの有効性を証明しています。

例

ProFusion を実装した 900MHz Intel Xeon プロセッサ搭載の ProLiant DL760 は、8 ウェイサーバの 256 プロセッサクラスターで、709,220 トランザクション/分という性能によって TCP-C ベンチマークで第 1 位にランクされました。

ベンチマークテストで、顧客の複雑な環境が完全に再現できることはありません。ProLiant 8000 と 8500 の 2 つのシステムは、CompuBank、Intuit、PeopleSoft などの大企業で、Web コンピューティング、トランザクション処理、およびエンタープライズリソースプランニング用に採用されています。

F8 チップセットのアーキテクチャ

HP によるシステムアーキテクチャ設計の進化における次の段階は、ProFusion チップセットをベースとした F8 アーキテクチャです。HP は、ProFusion の 8 ウェイアーキテクチャで培った経験を利用して、F8 という 8 ウェイマルチプロセッシングアーキテクチャを開発しました。

HP が開発した F8 チップセットは、マルチポート、ノンブロッキングクロスバー スイッチを搭載しており、メモリ、プロセッサ、および I/O サブシステムに同時に効率よくアクセスできます。

F8 チップセットは、複数の PCI-X ブリッジをサポートし、I/O サブシステムにおいて高可用性を実現するために、HP 内蔵 PCI ホットプラグコントローラが組み込まれています。

F8 チップセットの利点

cc-NUMA アーキテクチャと比較した場合、F8 SMP アーキテクチャはプログラミング モデルが単純であり、全体の平均遅延が少なく、標準のオペレーティングシステムやアプリケーションを使用できます。

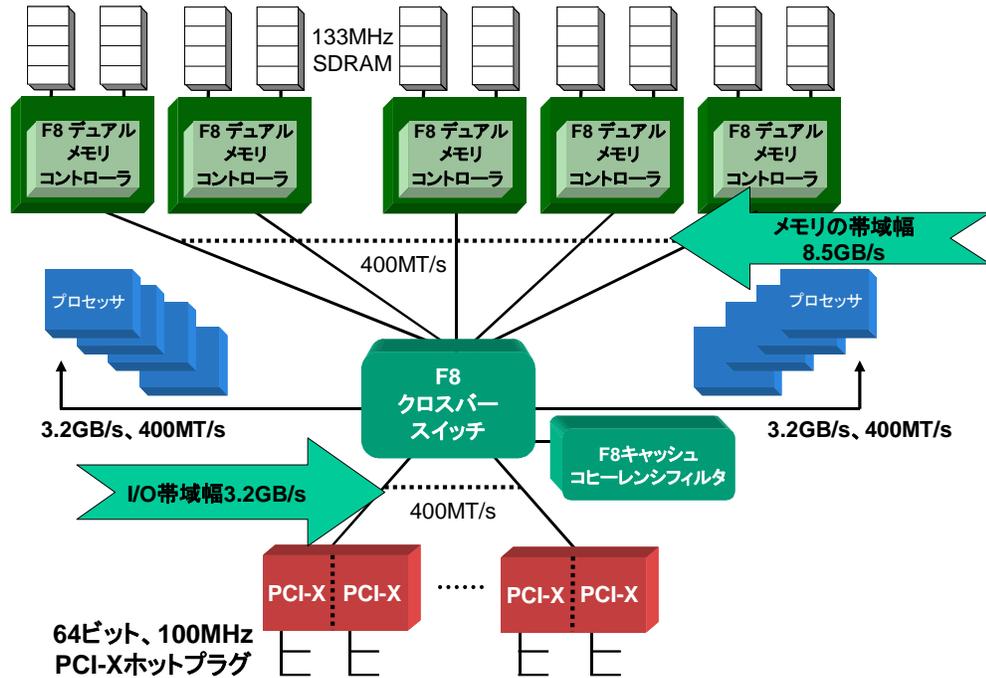
この他、HP 製の F8 チップセットには、以下の利点があります。

- Xeon MP プロセッサの処理能力に適した高い帯域幅を使用することにより、潜在的なボトルネックを除去
- 最適化されたクロスバースイッチ機能を使用して潜在的なボトルネックを除去
- オンライン交換機能の拡張によるホットプラグ RAID メモリの搭載

F8 チップセットは、HP ホットプラグ RAID メモリを持つ 5 つのメモリコントローラとマルチポートクロスバースイッチを搭載しています。F8 チップセットは、以下をサポートします。

- 8.5GB/s の総メモリ 帯域幅
- 3.2GB/s の個別プロセッサ バスの帯域幅
- 64GB のホットプラグ RAID メモリ
- ホットプラグをサポートする最高 4 個の 100MHz の PCI-X ブリッジ
- 8 個の Xeon プロセッサ

F8 の高い帯域幅容量



F8 チップセットは、ProFusion アーキテクチャと同様、すべての潜在的なボトルネックを最大限除去します。F8 チップセットには、ノンブロッキングクロスバースイッチが付いており、プロセッサ、メモリ、および I/O 間の通信パスを同時に使用できます。

バス速度は、MT/s (メガトランザクション/秒) で表します。

例

100MHz で動作し毎クロック 4 パケットを転送するバスは、400MT/s となります。

最適化クロスバースイッチ

HP は ProFusion の 8 ウェイアーキテクチャによって得た経験を生かし、クロスバースイッチコンポーネントを最適化するとともに、Intel Xeon MP プロセッサの処理能力に合わせて帯域幅を高めることによって、さらに高性能の F8 チップセットを設計しました。

このチップセットの設計では、128 のキャッシュラインを持つ大容量バッファ、13 の読み取りポート、および 4 つの書き込みポートによって、スイッチ内の並行トランザクション数を増やし、スイッチ、キャッシュコヒーレンシフィルタ、および Guaranteed Snoop Access アルゴリズムによって、バス間トラフィック量を減らします。これらの機能により、F8 アーキテクチャの効率が向上し、スケーラビリティが高まります。

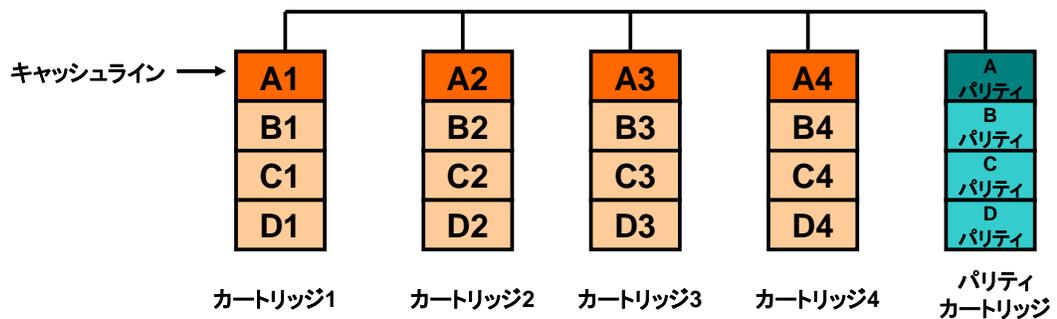
PCI-X ブリッジ

F8 チップセットには、I/O サブシステムに最高 4 個の業界標準デュアル PCI-X ブリッジがあり、それぞれに PCI ホットプラグコントローラが組み込まれています。このチップセットでは、最高 100MHz の速度で動作する PCI-X バスセグメントを各ブリッジが 2 個ずつサポートできるので、ギガビットイーサネットや Ultra320 SCSI などの高速相互接続を使用する周辺機器に容易に対応できます。

スケーラブルなメモリ性能

HP は、総メモリ帯域幅を、プロセッサバスを含めた帯域幅の 33% 増である 8.5GB/s にまで高めることによってスケーラブルなメモリ性能を実現しました。この設計では、各プロセッサバスが Pentium III Xeon プロセッサバスの 4 倍の帯域幅を持つので、コンピューティング上の要求を満たすための十分な余裕をもたらしています。

ホットプラグ RAID メモリ



ホットプラグ RAID メモリ対応のサーバは、RAID DIMM を使用してフォルトトレランスを提供するとともに、サーバ動作時にサーバを停止することなくメモリの交換や追加が行えます。これにより、DIMM の障害時に予定外のダウンタイムが生じるのを防ぎます。

F8 チップセットのメモリコントローラは、メモリにデータを書き込む必要がある場合、データのキャッシュラインを 4 つのブロックに分割します。次に各ブロックが、4 つのメモリモジュールに渡って書き込まれ(ストライプされ)ます。RAID エンジンがパリティ情報を計算し、パリティ専用の 5 番目のカートリッジに格納します。4 つのデータカートリッジと 1 つのパリティカートリッジにより、データサブシステムは冗長構成となり、DIMM のデータが誤っている場合や、カートリッジが 1 つ取り外された場合でも、他の 4 つのカートリッジからデータを再構築できます。

復習問題

1. ハイリーパラレル システムアーキテクチャで、性能を高めるために使用されるコンポーネントは何ですか。

.....
.....

2. バスマスタリングから生じたボトルネックを、HP はどのように解消しましたか。

.....
.....
.....

3. デュアルメモリコントローラ的设计を使用すると、メモリ帯域幅が最高 1.6GB/s になります。

正

誤

4. ProFusion チップセットには、メモリバスがいくつありますか。

.....

5. ProFusion クロスバースイッチには、どのようなコンポーネントが結合されていますか。

.....
.....
.....
.....

6. F8 チップセットを使用するとき、バス速度を表すための尺度は何ですか。

.....

目的

ここでは、次の内容について学習します。

- ATAPI や ATA RAID などの ATA/IDE 技術、従来のサーバでの ATA の実装方法、および新しい ATA 技術
- SCSI と ATA/IDE の相違点、顧客の要請やニーズに応じたこれらの技術の選択
- さまざまな SCSI 標準
- SCSI の以下の特性
 - 転送モード
 - アドレッシング
 - 通信
 - バスフェーズ
 - プロトコル互換性
 - 電氣的インタフェース
 - デバイスタイプ
 - コマンドセット
 - 実装方法
- HP によるストレージソリューションにおける SCSI 技術の実装方法
- iSCSI
- SCSI ベースのコンポーネントのトラブルシューティング

ディスクストレージの概要



サーバシステム業界では、以下の2種類のドライブ技術が一般に提供されています。

- ATA/IDE
- SCSI

ATA/IDE

ATA/IDEは、コントローラを拡張カードではなくハードドライブに統合するディスクドライブ構成です。この構成では、インタフェースのコストが削減されるとともに、ファームウェアの実装が容易になります。

SCSI

SCSIは、多様なデバイスをサーバに接続するための先進技術です。

ATA/IDE ハードドライブ技術

ATA は、最初の IBM AT コンピュータのデバイス接続インタフェースとして開発されました。

Compaq、CDC、Western Digital は、1980 年代に、最初の ATA インタフェースとディスクドライブを開発しました。この ATA インタフェース設計は 1980 年代の後半に、ANSI (American National Standards Institute) 標準に規定されました。

ATA インタフェースは、ドライブとその動作を定義するコントローラインタフェースおよび標準です。IDE は、ATA 標準を実装するために設計されたドライブと 40 ピンのインタフェース、およびドライブコントローラアーキテクチャです。

サーバにおける ATA/IDE

ATA/IDE ハードドライブは、伝統的にデスクトップやラップトップコンピュータ製品に採用され、サーバでは一般的に採用されていませんでした。しかし最近、サーバの使用負荷が比較的軽く、価格を重視して購入を決定する環境や顧客においては、低コストの ATA/IDE ハードドライブを選択する機会が増えてきました。

ATA/IDE ハードドライブ技術には、以下の利点があります。

- ギガバイトあたりの価格が最も安い
- 電力消費が最も少ない
- 発熱量が少ない
- エントリレベルの信頼性

ATA/IDE は、パラレル ATA とも呼ばれています。シリアル ATA 標準が規定されたため、現在では 2 つの標準を区別するために、「パラレル ATA」という用語の方が使われています。

ATA/IDE 標準

ATA の各バージョンは、それより古いバージョンと互換性があります。ATA の新しいバージョンは、わずかな例外を除いて旧バージョンの上に構築されており、旧バージョンの拡張と考えられます。

ATA 標準を以下に示します。

標準	説明
ATA-1	最初の ATA インタフェースとして、ディスクドライブと ISA バスベースのホストシステムとの間の統合バスインタフェースを定義する。
ATA-2	PIO (programmed input/output) のようなパフォーマンス向上機能が含まれている。EIDE は ATA-2 の拡張であり、ATA-2 と ATAPI の両方の標準の上に構築されている。
ATA-3	PIO モード 4 を改良したバージョンであり、単純なパスワードベースのセキュリティ方式、より高度な電力管理機能、および S.M.A.R.T. (Self-Monitoring Analysis and Reporting Technology) を提供する。ATA-3 は、ATA-2、ATAPI、および ATA デバイスと下位互換性がある。
ATA-4 (Ultra-ATA/33)	DMA データ転送を使用し、最高 33MB/s のバーストデータレートを実現する高性能のバスマスタリングを提供する。Ultra-ATA の実装方法は、通常 Ultra-DMA/33 または UDMA/33 と呼ばれる。Ultra-ATA ドライブシステムを実装するには、Ultra-ATA ドライブ、コントローラ、およびサポートする BIOS が必要である。
ATA-5 (Ultra-ATA/66)	DMA データ転送を使用し、最高 66MB/s のバーストデータレートを実現する高性能のバスマスタリングをサポートする。Ultra-ATA の実装方法は、通常 Ultra-DMA/66 または UDMA/66 と呼ばれる。Ultra-ATA ドライブには、コントローラ、ケーブル、およびサポートする BIOS が必要である。 信号ケーブルは、Ultra-ATA/66 または Ultra-ATA/100 用に特に設計された 40 ピン/80 コンダクタのケーブルである。それまでのバージョンの ATA 標準と完全に下位互換性があるが、標準の 40 ピンケーブルでは UDMA の性能は提供されない。オペレーティングシステムが DMA 対応であり、DMA モードをアクティブにする必要がある。
ATA-6 (Ultra-ATA/100)	既存の EIDE/UDMA ハードドライブ、リムーバブルメディアドライブ、および CD-ROM や R/RW ドライブと下位互換性がある。Ultra-ATA/100 ドライブは、Ultra-DMA/100 または UDMA/100 とも呼ばれ、100MB/s のバーストレートを実現する。元々 Intel、Quantum、および Seagate が設計改良によって作成したドライブである。 検出メカニズムにより、ホストは 80 コンダクタケーブルを検出でき、Ultra ATA/66/100 転送レートを有効にするか否かを決定できる。Ultra ATA/100 では、Ultra ATA/66 と同じ 80 コンダクタケーブルが必要であり、Ultra ATA/33 では、40 コンダクタケーブルが必要である。
ATA-7 (Ultra-ATA/133)	ATA のインタフェース速度を 133MB/s まで高速化し、インタフェースレートを PCI バスデータレートまで引き上げた。Maxtor によって導入されたこの標準は、Ultra ATA/33 のダブルエッジクロッキング技術と CRC (巡回冗長検査)、および Ultra ATA/66 インタフェースで導入された 80 コンダクタケーブルの上に構築されている。旧バージョンのすべてのパラレル ATA デバイスと下位互換性があり、ATA/100 に現在使用されているのと同じ 80 コンダクタ、40 ピンケーブルを使う。

ATAPI(ATA Packet Interface)

ATAPI(ATA Packet Interface)は、ATA インタフェースの拡張であり、ハードドライブ以外のデバイスを通常の ATA/IDE ポートに接続するために設計されました。ATAPIは、標準のパケットベースのインタフェースで、SCSI から派生したコマンドセットを持っており、通常の ATA/IDE コネクタに接続する CD-ROM やテープドライブなどのデバイス用のコマンドを提供します。ATAPI ではまた、LS-120 スーパーディスクドライブや内蔵用 Iomega Zip、Jaz ドライブなど、他のリムーバブルストレージも作動します。

ATAPI 仕様は、当初 ATA-2 で定義されましたが、これは現在廃止されています。ATAPI の現行バージョンは、ATA-4 仕様の一部として定義されています。

ATA RAID

ATA RAID 機能によって、SCSI ベースのアレイコントローラを追加購入することなく、ソフトウェア RAID を上回る非常に高レベルなフォールトトレランス、性能、および利便性が提供されます。この技術は、ファームウェアとソフトウェアの統合によって可能になりました。

ATA RAID 0/1 PCI カードは、データをミラー化し、ATA システム上のドライブをブートします。オペレーティングシステムは、あらかじめ構成したミラー (RAID 1 ボリューム) にインストールできます。どのドライブに障害が生じた場合でも、ドライブを交換した後、ユーザやオペレーティングシステムが介入しなくてもミラーを回復できます。RAID をソフトウェアによって実装する方法では、通常、同様の障害から回復するために、ハードウェアやソフトウェアの設定を再構成したり、手動で強制再構築をしたりする必要がありますが、ATA RAID ではこの点が改善されています。

内蔵 ATA RAID は、標準の ATA コントローラ、BIOS、ドライバ、およびプロセッサを統合して、ATA ハードドライブ上で RAID 0、1、または 1+0 機能を実行します。この機能は BIOS によって実行されるので、ATA RAID はブートドライブのミラー化をサポートします。ただし、RAID 機能を実行するためにシステムプロセッサを使うので、ハードウェア RAID に比べてシステムパフォーマンスは低くなります。なお内蔵 ATA RAID は RAID-5 機能をサポートしません。

▲ 注記

内蔵 ATA RAID は、ソフトウェアとハードウェアの両方によって実装されます。サードパーティの ATA RAID は、ハードウェアソリューションとして提供されています。

新しい技術 – シリアル ATA

シリアル ATA 技術によって、将来チップセットを効率的に統合するために不可欠な低電圧、少数のピンによる実装を実現できるようになりました。シリアル ATA は、各デバイスが専用のリンクを使用してホストに直接接続されるポイントツーポイントインタフェースです。パラレル ATA と同様、シリアル ATA は、内蔵用ストレージインタフェースとしてのみ設計されています。

シリアル ATA では、パラレル ATA に比べ、以下の機能が改良、追加されています。

- **ドライブジャンプがない** — プライマリとセカンダリデバイス間の競合がないため、各デバイスはインタフェース帯域幅全部を専有できる
- **低電圧** — シリアル ATA では、それまで 5V であった信号電圧が、およそ 250mV に下がった
- **ピン効率の向上** — 現在、ATA インタフェースでは、26 本の信号ピンがインタフェースチップに接続されている。シリアル ATA では信号ピンが 4 本だけとなり、ピンの効率が向上する
- **ケーブルとコネクタの改良** — シリアル ATA のケーブル長は 1 メートルである。これはパラレル ATA の最大ケーブル長の 18 インチに比べ、大幅に延長された。現行のパラレル ATA では、80 コンダクタリボンケーブルと 40 ピンヘッダコネクタを使っている。シリアル ATA では、はるかに細いケーブルと、わずか 7 ピンの小型ケーブルコネクタを使う
- **ホットプラグ機能** — シリアル ATA には、システムの電源を入れたまま、デバイスのプラグをシステムに直接接続できる機械的、電気的機能が用意されており、デバイスの検出や初期化の処理はプロトコルが保証する
- **互換性** — シリアル ATA は、現行のパラレル ATA とレジスタレベルの互換性を維持。シリアル ATA は、既存の ATA および ATAPI デバイスをすべてサポートする
- **スケーラビリティ** — 現行のパラレル ATA インタフェースは、今後数回の速度倍増をサポートするスケーラビリティがない。シリアル ATA では、1.5Gb/s から 3Gb/s、さらに 6Gb/s に移行するためのロードマップが定義されている
- **信頼性** — シリアル ATA は、8B/10B という特別なエンコード形式を使用して、ケーブルで送信するデータのエンコードとデコードを行う。これにより、回線を流れる信号パルスの電圧の安定が保証されるので、信頼性が向上する

シリアル ATA II

2002 年の第 1 四半期、Serial ATA II Working Group の結成が発表されました。このワーキンググループでは、既存のシリアル ATA 仕様をサーバおよびネットワークストレージ市場の分野に拡張するとともに、第 2 世代の信号速度を実現するシリアル ATA II を開発する予定です。

SCSI とパラレル ATA/IDE

SCSI ハードドライブ技術は、その信頼性と優れたプロトコル機能から、業界の標準となっています。HP の SCSI ハードドライブ技術は、以下の独自の機能を提供します。

- ユニバーサルアーキテクチャ
- 最適化したファームウェア
- 優れた検証および品質標準

より高いデータの一貫性、性能、信頼性、およびそのホットプラグ機能のため、顧客は一般に SCSI ハードドライブ技術を選択します。

サーバ環境における ATA/IDE ハードドライブ技術は、容量あたりの価格が最も安く、ドライブあたりの電力消費が最も少なくなります。また同時に、負荷の軽い環境で必要な程度の信頼性も提供できます。ATA/IDE は、静的な Web ページ配信、エントリレベルのファイル/印刷、共有インターネットアクセスなど、基本的なサーバ機能を必要とする顧客に最適です。

次の表では、パラレル ATA と SCSI について、基本的な相違点を説明します。

特性	パラレル ATA	SCSI
性能	シークタイム : 8ms	シークタイム : 5ms
スピンドル速度	最高 7,200 rpm	最高 15,000 rpm
転送レート	100MB/s	320MB/s (Ultra320)
スケーラビリティ	チャンネルあたり 2 ドライブ	チャンネルあたり 15 ドライブ
最大ケーブル長	19 インチ (約 48cm) のパラレルケーブル	12m (LVD)
信頼性 (平均故障間隔 – MTBF)	30 万 ~ 50 万時間	100 万時間
保守性	ホットプラグ不可	ホットプラグ可能
製品保証	3 年間	3 年間
耐用時間 (5 年間の製品寿命)	3000 時間/年	8760 時間/年
設計上の使用想定温度	35° C	55° C
1 年あたりの世代 (平均)	2	1
キューイング	1 回に 1 トランザクションのみが可能	複数の同時トランザクションが可能
切断/再接続	行わない。トランザクションは、完了するまでバスに接続されたままとなる。	行う。バスは要求の後切断され、データが利用可能になった時点で再接続される。

注記

シリアル ATA は、ATA と SCSI インタフェースのギャップを縮めます。

SCSI の概要

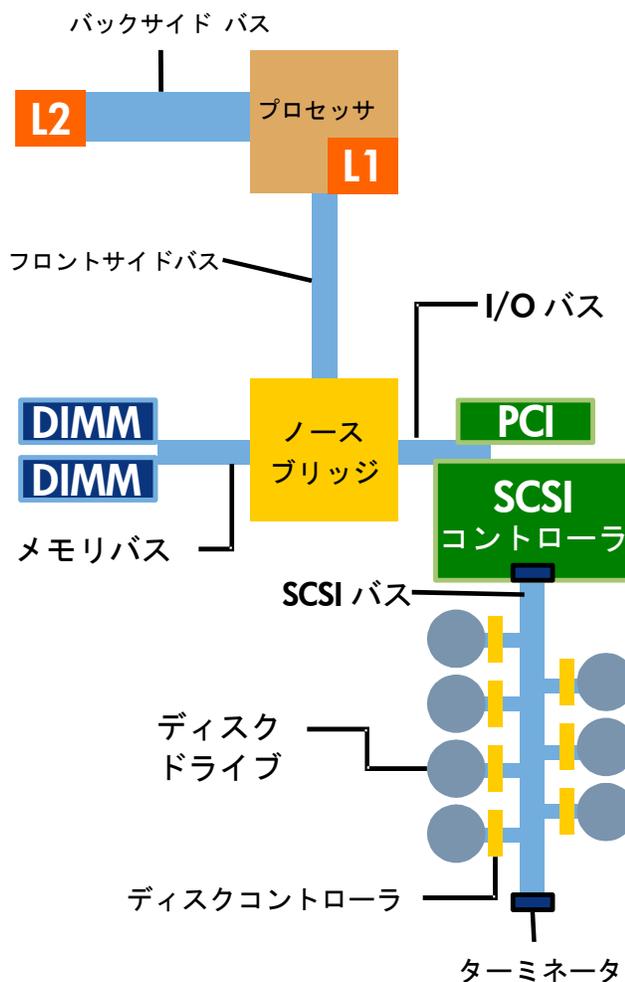


SCSI 技術の標準化作業は、1980 年代前半に始まりました。1986 年、ANSI は、SCSI 周辺機器を接続する際の機械的、電気的、機能的要件を定義する標準として SCSI 規格を正式に承認しました。

しかし、その後のプロセッサや周辺機器の性能向上により、最初の SCSI 規格は限界に達しました。その結果、1994 年に SCSI 標準の改良版が開発され、性能や機能性を向上させる新しい仕様が盛り込まれました。それと同時に、SCSI の各バージョンは、すべての SCSI デバイスとの下位互換性を維持しています。

SCSI 標準の進化に伴って、HP はサーバやストレージオプションでの SCSI 技術のテストや統合を行ってきましたが、これは、エンタープライズコンピューティング環境で SCSI が広く普及する上で重要な役割を果たしています。HP の戦略は、ユーザの要求を満たすスケラブルでマルチベンダの様々な構成を実現できるような、エンタープライズコンピューティングシステムを提供することです。HP のハードディスクドライブの厳しい品質管理とマルチベンダ戦略により、シングルベンダ ソリューションが持つリスクを軽減できます。

SCSI の基礎知識



SCSI のブロック図

SCSI は、オープンシステムの世界で中心的な I/O 技術です。新しいストレージ技術を、市場にある多種多様な製品に統合するのは困難なので、ファイバチャネルのような新しい I/O 技術を短期間で開発するために、多くの場合、いずれかの SCSI をベースにして開発が行われます。

SCSI は、進化し続ける ANSI 標準の電子インタフェースのセットであり、コンピュータがパラレルバスを通じて周辺ハードウェア機器と通信を行います。

最近の変更ではシリアル接続も規定されていますが、SCSI の主流はパラレルバスであることは変わりません。

SCSI コマンドセット

SCSI は、物理的相互接続の標準以外に、ドライブデバイスが従うべき論理コマンドセットの標準も定義しています。共通コマンドセット(Common Command Set)は、ANSI SCSI-1 で開発されました。SCSI-2 および SCSI-3 には、標準の中に改訂版の共通コマンドセットが規定されています。コマンドは使うデバイスの種類によって異なります。

SCSI-1 は、最初、6 種類のデバイスについてコマンドセットを定義しました。しかし、この標準はあまりに寛容で、ベンダ独自のオプションを数多く認めていました。その結果、異なるベンダの製品間の相互運用性に問題が生じました。

そこで SCSI の互換性の問題を解決するために共通コマンドセットが定義されました。その結果、共通コマンドセットが標準のサブセットとなり、そこでは例外が認められませんでした。共通コマンドセットの定義によって、1980 年代に SCSI-1 がサーバディスクサブシステムとテープバックアップ市場に浸透し始めました。

SCSI-2 は SCSI-1 の弱点解決に重点を置き、10 種類のデバイスのサポートを採用しました。また、より効率的なコマンドセットが導入され、切断オプションおよびコマンドのキューイングオプションが規定されることにより、デバイス機能が向上しました。

SCSI 標準

SCSI は 1986 年以来、次の 3 つの主要な標準を経て進化してきました。

- **SCSI-1** — 1986 年に ANSI が承認した最初の SCSI 標準であり、ケーブル長、信号特性、コマンド、転送モードについて最初の SCSI バスを定義
- **SCSI-2** — 1994 年に承認された。SCSI-2 は大幅に改良され、多数の先進機能をサポート
- **SCSI-3** — 異なる物理層上での SCSI プロトコルの実装方法を定義した一連の仕様

SCSI-1

1986 年に ANSI が承認した最初の SCSI 標準であり、次の仕様について最初の SCSI バスを定義しています。

- ケーブル長
- 信号特性
- コマンド
- 転送モード

SCSI-1 のデフォルトの速度は 5MB/s でした。バスは 8 ビットの平行バスで、バスサイクル 1 回について 1 バイトのデータ転送を行いました。機能は以下のとおりです。

- 2 つのデバイス
- 非同期モード
- シングルエンドインタフェース

SCSI-2

SCSI-2 は 1994 年に承認されました。SCSI-2 は大幅に改良され、次のような多数の先進機能をサポートしています。

- **同期モード** — コマンドのキューイングに使われる。256 のコマンドを 1 つのデバイスに送信
- **10MB/s の転送レート** — SCSI-1 の 2 倍のスループット向上を実現
- **Fast SCSI** — 高速転送プロトコル。バス速度は 2 倍の 10MHz
- **Wide SCSI** — SCSI-1 では 8 ビットだったバス幅を 16 ビットに広げたことによって、同じ信号速度でのデータスループットが向上。Fast-Wide SCSI-2 は、最高 20MB/s の転送レートを発揮
- **バスあたりのサポートデバイス数の増加** — Narrow SCSI の 8 個に対し、Wide SCSI バスは 16 個のデバイスをサポート
- **ケーブルとコネクタの品質向上** — 新しい高密度 68 ピン B ケーブルとコネクタを定義
- **アクティブターミネーション** — バス終端の信頼性が向上

この他、SCSI-2 は、すべての SCSI デバイスとの下位互換性を維持しています。

SCSI-3

SCSI-3 は、次のような異なる物理層上での SCSI プロトコルの実装方法を定義した一連の仕様です。

- SCSI-3 パラレルインタフェース
- 高性能シリアルバス
- ファイバチャネル
- シリアルストレージアーキテクチャ

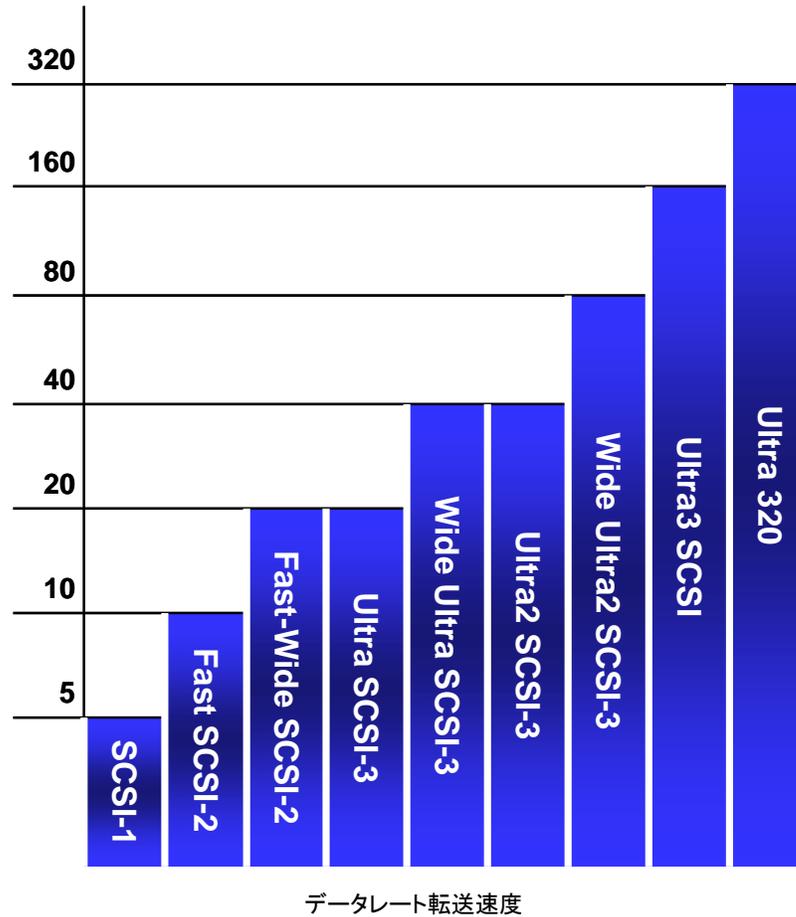
物理層は、それぞれ異なるパフォーマンス特性を持ち、使うハードウェアも異なります。現在、SCSI-3 標準には、SCSI-2 の性能や機能に次のような改良が加えられています。

- **Ultra SCSI-3** — バス速度を 2 倍の 20MHz とし、8 ビットデータパスで 20MB/s の転送レートを実現
- **Wide-Ultra SCSI-3** — 16 ビットデータパスで Ultra SCSI の 2 倍に相当する 40MB/s の転送レートを実現
- **ケーブルの改良** — Wide SCSI の B ケーブルに代わって、新しい 68 ピン、P ケーブルを使用

注記

▲ HP は、Wide-Ultra SCSI-3 技術を HP サーバやストレージオプションと徹底的にテストし統合してきました。なぜならこの技術が、SCSI ホストインタフェースで最高レベルの性能を発揮し、下位互換性によってユーザの投資を保護できるからです。

Wide Ultra2 SCSI-3

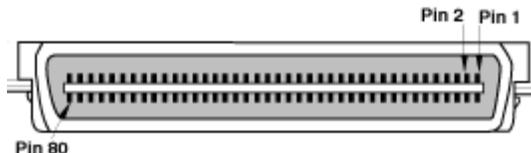


アプリケーション、プロセッサ、ストレージデバイスの処理能力の向上に伴い、ユーザはシステムパフォーマンスを向上させる方法を模索するようになりました。Wide Ultra2 SCSI-3 製品は、ワークステーションやサーバの最適化に必要な速度、柔軟性、互換性を提供します。

Wide Ultra2 SCSI-3 カードは、LVD (low-voltage differential) 技術を採用し、最新の SCSI 規格を実現しています。Wide Ultra2 SCSI-3 プロトコルには、次の特長があります。

- データバーストが Wide-Ultra SCSI の 2 倍の 80MB/s になり、システムスループットが向上
- 最大ケーブル長が Wide-Ultra SCSI-3 バスの 4 倍の 12m になり、外部ストレージの追加や、クラスタ化サーバの構成での柔軟性が向上
- 下位互換性を維持しているため、それまでのすべての SCSI の実装を、同じバス上で使用可能

Wide Ultra2 SCSI-3 コネクタ



SCA-2 80ピンコネクタ

電源ケーブルとデータケーブルが別になった従来の接続プロトコルは、ハイエンドサーバ環境では十分に機能しません。Wide-Ultra や Ultra2 SCSI などのハイエンドドライブは、一般に使われるケーブルシステムであるだけでなく、電源ケーブル、データケーブル、および通常のデバイス ID 割り当て方式に変わる特別な接続システムを使うよう設定できます。

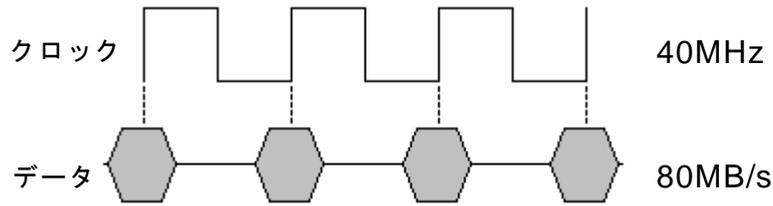
Single Connector Attachment (SCA) がサーバや他のハイエンドシステム用に開発され、従来のケーブルシステムは、シングル 80 ピンコネクタを使うバックプレーンシステムに置き換わっています。

このコネクタは、標準の 68 ピン Wide SCSI ケーブルや標準の 4 ワイヤ D 型電源コネクタの代わりに用いられ、個別のドライブごとにデバイス ID を設定する信号が用意されています。

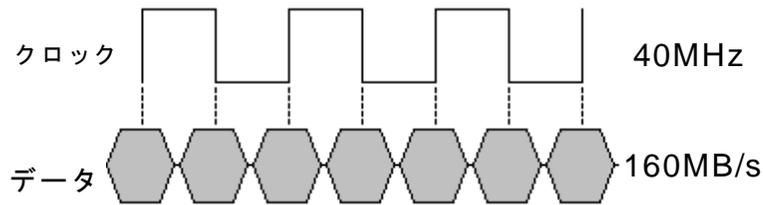
Single Connector Attachment-2 (SCA-2) は、最初の SCA コネクタの改良版です。最大の相違点として、SCA-2 のコネクタは、コネクタの 80 ピンのうちの 2 つのピンが、ドライブを接続するときには最初に接続され、取り外すときには最後に外れる設計になっています。これによって、ドライブが確実にアース(接地)されるようになり、信頼性が向上します。

ハイエンドワークステーションおよびサーバ上で複数のハードドライブが使うことにより、パフォーマンスを向上させ、信頼性を改善できます。これは通常、RAID によって実現されます。また多くの場合、ハードドライブ サブシステムは、ホットスワップ可能、つまり、システムの電源を切断せずに故障したハードドライブを交換できる設計になっています。

Wide Ultra3 SCSI-3



Wide Ultra2 SCSI-3 のデータ転送 (通常のクロッキング)



Ultra3 SCSI-3 のデータ転送 (ダブルランジション クロッキング)

SCSI Trade Association (STA) の認定によると、Ultra3 という名称は次のいずれかの機能を持つ製品を指します。

- **ダブルランジションクロッキング** — Wide Ultra2 と Wide Ultra3 はどちらも、16ビットの wide (ワイド) バスで 40MHz のバス速度を使います。Wide Ultra3 では、同じサイクル時間内でのサンプリング速度が 2 倍になるため、Wide Ultra3 のデータ転送レートは最高 160MB/s です
- **CRC (巡回冗長検査)** — 転送レートの高速化に伴い、データの信頼性に対する要求が高まっています。CRC は、データをロングバイナリの数値として扱い、バイナリ数値を素数で除算します。データの送信側と受信側の両方で計算を行い、求められた剰余を比較します。それまでの SCSI プロトコルでは、パリティ検査でシングルビットエラーしか検出できませんでした。CRC は、シングルビットエラーとダブルビットエラー、奇数個のエラー、および 32 ビットまでのエラーバーストを検出できます
- **ドメインバリデーション** — ホストコントローラは、ターゲット デバイスへの接続に成功するか、サポートする低速接続をすべて試みるまで、データ転送レートのネゴシエーションを行います。ネゴシエーションのプロセスでは、ホストコントローラが、ターゲット デバイスへ、一定のデータを特定の速度で送信します。ターゲット デバイスが同一のデータを返すと、その速度を使います。同一のデータが返されなかった場合、ホストコントローラはそれより遅い転送レートでネゴシエーションを試み、接続に成功するまでこれを繰り返します

以前の SCSI 標準では、ネゴシエートされた速度が使えない場合、そのデバイスがアクセス不能となることがありました

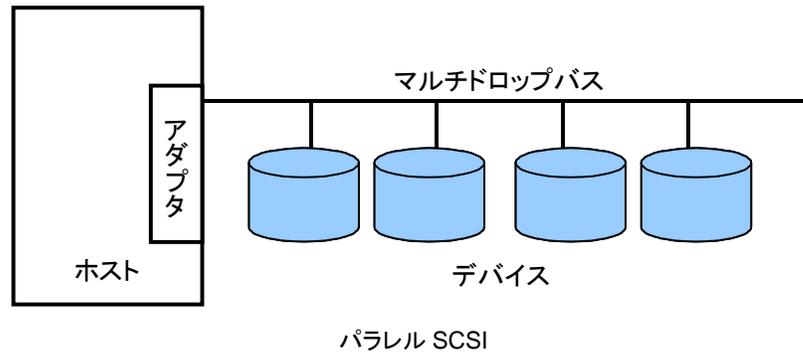
あらゆるストレージシステムに広がるユニバーサルストレージ デバイスの戦略は、業界の先進的技術である Wide Ultra3 SCSI-3 技術をベースにしています。Wide Ultra3 SCSI-3 ドライブとオプションは、旧バージョンの速度の 2 倍である最高 160MB/s のデータ転送レートを提供します。

Ultra320

Ultra320 ではオーバーヘッドを削減してパフォーマンス向上を図る新しいテクノロジーが追加されています。

- **パケット型 SCSI** — これにはパケットプロトコルのサポートが含まれています。パケット型デバイスでは、コマンド、データ、ステータスが、遅い非同期フェーズではなく DT (デュアルトランジション) データフェーズで転送されるためコマンドオーバーヘッドが削減されます。これによりバス使用率が最大になり、コマンドオーバーヘッドが最小になるため性能が向上します。さらにパケットプロトコルでは、1 回の接続で複数のコマンドを転送できます。Wide Ultra3 SCSI-3 ではデータは 160 MB/sec の同期フェーズで転送されますが、コマンドフェーズやステータスフェーズは遅い非同期フェーズで転送され、1 接続あたり 1 転送に制限されます
- **クイックアービトレーションアンドセレクション (QAS)** — クイックアービトレーションアンドセレクション (QAS) により、あるデバイスから他のデバイスへ SCSI バスをコントロールリリースするオーバーヘッドが削減されます
- **リードライトデータストリーミング** — リードライトデータストリーミングにより、ターゲットは 1 データストリーム LUN Q-TAG (LQ) パケットに続けて複数のデータパケットを送信できるため、データ転送のオーバーヘッドが最小になります。非ストリーミング転送では、各データパケットには 1 データ LQ パケットしか入れることができません。ライトデータストリーミングでも、各 LQ とデータパケット間にバスのターンアラウンド遅延 (DT データインから DT データアウトまで) が発生しないので性能が向上します
- **フロー制御** — フロー制御によりイニシエータは、書き込み時にデータをプリフェッチする動作と読み出し時にデータ FIFO をフラッシュする動作を最適化できます。ターゲットはデータストリームの最終パケット転送時期を知らせることができるため、イニシエータは以前のものより早くデータプリフェッチを終わらせたり、データ FIFO のフラッシュ動作を開始したりすることができます

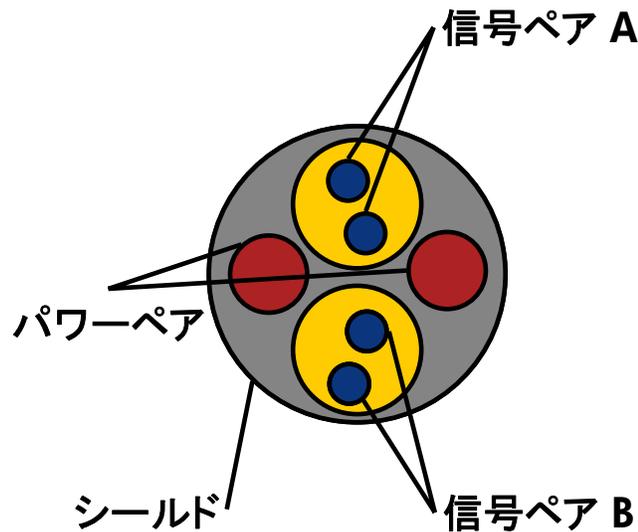
パラレル SCSI



パラレル SCSI は、マルチドロップバスによってデータを転送します。デバイスは、それぞれホストアダプタと直接、データ転送を行います。ホストアダプタとデバイスの間の転送レートは、個別のデバイスごとにネゴシエートされます。

SCSI ホストアダプタは、SCSI バスとコンピュータの内部 I/O バスとの間のゲートウェイとして動作する SCSI デバイスです。これはコマンドの送信と応答を行うとともに、バス上のデバイスとコンピュータ内部との間でデータを転送します。接続された周辺機器がいくら優れていても、すべてのデータやコマンドはホストアダプタを通過するので、ホストアダプタは、システムで SCSI を適切に実装するための鍵となります。

シリアル SCSI



IEEE 1394 ケーブルの断面図

シリアル SCSI は、その規格名である IEEE 1394、または Apple Computer 社のブランド名である FireWire と呼ばれています。これは高性能のシリアルバスであり、多目的で高速、かつ低コストで、さまざまなコンピュータ周辺機器、ストレージシステム、一般ユーザ向け電子機器を相互接続します。シリアル SCSI には、次の特長があります。

- **高速** — 100、200、または 400Mb/s、すなわち 12.5、25、または 50MB/s
- **データ転送の保証** — アイソクロナス転送と非同期転送の両方をサポート
- **一般ユーザにとっての利便性** — 小さなコネクタ、ホットプラグ可能、自動 ID、ターミネーション(終端)が不要
- **拡張性** — 63 個のデバイス(16 個のデジーチェーン接続が可能)、4.5m のフレキシブルケーブル

IEEE 1394 シリアルバスには、次の特長もあります。

- **デジタルインタフェース** — デジタルデータをアナログに変換する必要がないので、データの完全性を保証しやすい
- **コンパクト設計** — 大型で高価なインタフェースの代わりに、細いシリアルケーブル
- **ホットプラグ可能** — バスがアクティブなまま、デバイスの接続や取り外しが可能
- **スケーラブル** — 現在、1 つのバス上で 100、200、400Mb/s のデータレートが混在でき、800Mb/s と数 GB/s のアップグレードバスがある
- **柔軟性** — トポロジはデジーチェーン型と完全なピアツーピア通信の分岐構造をサポート

シリアル SCSI ケーブル

68 ワイヤケーブルの代わりに、シリアル SCSI では 6 ワイヤケーブルを使用します。

IEEE1394 技術を使用した信号管理は単一のデータ線に限定されるので、事前に設定された速度でのデータ転送を保証するアイソクロナスデータ転送をサポートします。シリアル SCSI は、ストリーミングビデオやオーディオなどのマルチメディア I/O 配信にとって理想的です。

パラレル SCSI とシリアル SCSI の比較

パラレル SCSI インタフェースでは、同時に 8 ビットまたは 16 ビットのデータが並列に伝送されます。シリアル SCSI では、一度に 1 ビットが伝送されます。パラレル SCSI とシリアル SCSI の違いは、シリアルポートとパラレルポートの違いに似ています。

SCSI バスの帯域幅は、その幅に直接比例します。8 ビットや 16 ビットの転送幅を使っていたのに、1 ビットずつの転送に変更するのは後退しているように見えるかもしれませんが、シリアル通信では速度が劇的に向上するので、帯域幅での損失以上の効率が得られます。

技術の進歩に伴い、帯域幅に対する需要が高まっています。この高速化の要請に応えるために、SCSI バスは最初のバージョンから Fast、Ultra SCSI へと進化してきました。しかしながら、より高速なバスが作られるたびに、複雑な信号を管理し、データ破壊を防止することがより困難になっていきます。

パラレルバス経由で信号を送信する場合、それらすべてが宛先に同時に到着する保証はありません。非同期データ転送を使用しているので、異なる信号にまぎれてしまうかもしれません。これに対し、シリアル SCSI は、アイソクロナスと非同期の両方のデータ転送をサポートします。

▲ 注記

アイソクロナスデータ転送では、事前設定した速度でのデータフローが保証されるので、アプリケーションは時間どおりにデータを処理できます。

シングルエンド SCSI でデータの完全性を保証しようとする、信号速度が 2 倍になるごとに最大ケーブル長が半分になります。

旧式のシングルエンドバスを使用して得られる最高の伝送速度は、Ultra SCSI の 20MHz です。

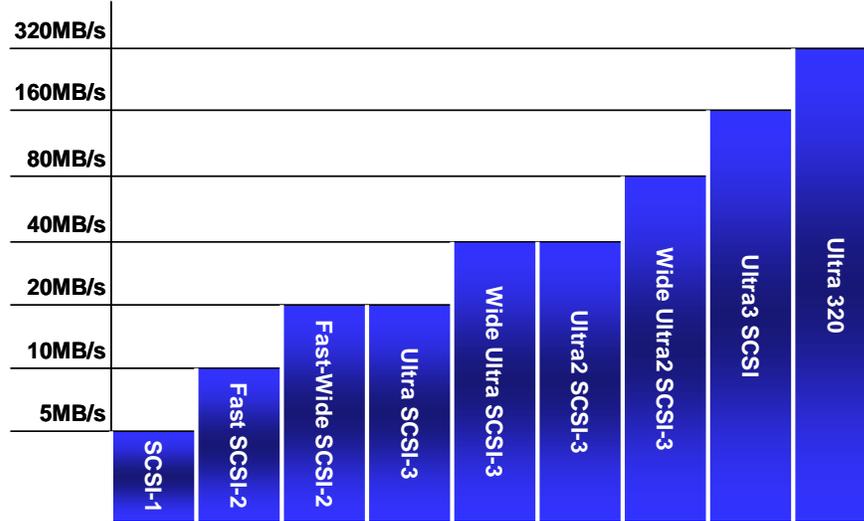
SCSI 標準のまとめ

つぎの表では、SCSI デバイスのそれぞれについて、特性をまとめます。

種類	転送レート	SCSI 標準	バス幅	デバイス数/ バス	コネクタ
SCSI-1	5MB/s	SCSI-1	8	7	25/50 ピン
Fast SCSI-2	10MB/s	SCSI-2	8	7	50 ピン
Fast-Wide SCSI-2	20MB/s	SCSI-2	16	15	68 ピン
Ultra SCSI-3	20MB/s	SCSI-3	8	7	50 ピン
Wide Ultra SCSI-3	40MB/s	SCSI-3	16	15	68 ピン
Ultra2 SCSI-3	40MB/s	SCSI-3	8	7	50 ピン
Wide Ultra2 SCSI-3	80MB/s	SCSI-3	16	15	68 ピン
Ultra3	160MB/s	SCSI-3	16	15	68 ピン
Ultra320	320MB/s	SCSI-3	16	15	68 ピン

8ビットのバス幅を narrow(ナロー)、16ビットのバス幅を wide(ワイド)と呼ぶ

SCSI の転送モード



理論上の最大バースト転送レートは、以下の公式で算出されます。

$$\text{バス速度 (MHz)} \times \text{バス幅 (バイト)} / \text{転送あたりのサイクル} = \text{理論上の最大 MB/s (転送あたりのバーストサイクルが 1 の場合)}$$

次の表では、SCSI 転送モードの比較を示します。

SCSI タイプ	バス幅 (ビット)	デバイス数/バス	周波数 (MHz)	転送レート (MB/s)	コネクタ
narrow (ナロー)					
SCSI-1	8	7	5	5	25/50 ピン
Fast SCSI-2	8	7	10	10	50 ピン
Ultra SCSI-3	8	7	20	20	50 ピン
Ultra2 SCSI-3	8	7	20	40	50 ピン
wide (ワイド)					
Fast-Wide SCSI-2	16	15	10	20	68 ピン
Wide Ultra SCSI-3	16	15	20	40	68 ピン
Wide Ultra2 SCSI-3	16	15	40	80	68 ピン
Ultra3	16	15	80	160	68 ピン
Ultra320	16	15	160	320	68 ピン

SCSI プロトコルの互換性

SCSI プロトコルは下位互換性を提供するように設計されているので、現行のホストアダプタを従来の周辺機器で使うことも、現行の周辺機器を従来のホストアダプタで使うこともできます。しかし、SCSI プロトコルは多様なので、特定のデバイスの組み合わせでの互換性は保証できません。

複数の SCSI デバイスを同じグループで使う場合は、次のガイドラインに留意してください。

- すべての SCSI デバイスはどの SCSI バスでも動作するが、ホストアダプタ、周辺機器、ケーブルなどを含めた全体の動作確認を行ったシステムを同じメーカーから購入しない限り、動作は保証されない
- デバイスの設計時期に開きがあるほど、互換性の問題が発生しやすい
- Narrow デバイスと Wide デバイスを同じバス上に混在させられるが、Narrow のみ、または Wide のみの構成より問題が発生しやすい。使用する SCSI タイプ間の相違が大きければ大きいほど、バスが正しく動作しない可能性が高い。余分の帯域幅を使えるのは、Wide デバイスのみである。ただし、ほとんどの場合、バスの Wide の部分と Narrow の部分を終端する必要があり、終端しなければ Wide デバイスが正しく機能しない
- シングルエンド SCSI と高電圧ディファレンシャル (HVD) SCSI は、アダプタを使って、バス上のすべてのデバイスをどちらかの SCSI に統一しない限り、混在させてはならない

注記

SCSI-2 の作成における重要な設計基準の 1 つは、SCSI-1 との下位互換性でした。ほとんどの場合、SCSI-2 デバイスは、旧バージョンの SCSI-1 デバイスと同じバス上で動作します。しかし、旧バージョンのデバイスが SCSI-2 の拡張機能や高速転送プロトコルをサポートできないので、常にそうであるとは限りません。

SCSI アドレッシング

ID	Bit 3	Bit 2	Bit 1	Bit 0
0	OFF	OFF	OFF	OFF
1	OFF	OFF	OFF	ON
2	OFF	OFF	ON	OFF
3	OFF	OFF	ON	ON
4	OFF	ON	OFF	OFF
5	OFF	ON	OFF	ON
6	OFF	ON	ON	OFF
7	OFF	ON	ON	ON
8	ON	OFF	OFF	OFF
9	ON	OFF	OFF	ON
10	ON	OFF	ON	OFF
11	ON	OFF	ON	ON
12	ON	ON	OFF	OFF
13	ON	ON	OFF	ON
14	ON	ON	ON	OFF
15	ON	ON	ON	ON

ID	Bit 2	Bit 1	Bit 0
0	OFF	OFF	OFF
1	OFF	OFF	ON
2	OFF	ON	OFF
3	OFF	ON	ON
4	ON	OFF	OFF
5	ON	OFF	ON
6	ON	ON	OFF
7	ON	ON	ON

SCSI のデバイス ID 設定

バス上の各 SCSI デバイスには、特定の ID 番号のアドレスがあります。Narrow SCSI デバイスには 0～7 の番号が割り当てられ、Wide SCSI デバイスには、0～15 の番号が割り当てられます。16 個の ID をサポートする Wide デバイスは、4 ビットのジャンパブロックを使います。

旧式のホストアダプタの中には、デバイス ID について厳格で、あるハードドライブのデバイス ID を 0 に設定すると、そのハードドライブからしか起動できないものもあります。新しいハードウェアでは、この制限が修正されています。

ID の構成方法は、具体的なハードウェアによって異なります。多くのデバイスでは、デバイスの筐体の背面にハードウェアスイッチ、ジャンパ、またはロータリーダイヤルが付いており、それによってデバイス ID を設定します。高度なデバイス、特に新しいタイプの SCSI ホストアダプタは、ソフトウェアユーティリティを使用します。

SCSI は、バス上で多数のデバイスをサポートし、デバイス間のデータ転送を同時に実行します。コントローラをはじめとして、すべての SCSI デバイスは SCSI ID を使用してアドレスが割り当てられるため、1 つの SCSI チャンネル上の ID は、すべて違う番号でなければなりません。

ホストアダプタは、ID が必要な SCSI デバイスです。Narrow SCSI は、7 つのデバイスとホストアダプタをサポートします。Wide SCSI は、15 のデバイスとホストアダプタをサポートします。

プラグアンドプレイ SCSI では、バス上で自動的にデバイス ID の割り当てができるので、同じ ID を同時に複数のデバイスが使うことはありません。SCA システムは、SCSI-3 で初めて採用され、特別な 1 つのコネクタによって各ハードドライブに電源とデータを供給します。また、コントローラによって、各ドライブのデバイス ID を設定できます。

SCSI バスがデバイスを区別できるよう、デバイスにはそれぞれ一意のターゲット ID を割り当てる必要があります。

**重要**

選択した 1 つのフェーズに対して 2 つのデバイスが応答した場合、SCSI バスエラーが発生します。**複数のデバイスが同じ ID を使うように構成してはなりません。**

8 ビットバスでは、バス上で 8 つのデータチャネルを使って、最大 8 つのターゲットを設定できます。Wide バスでは、データチャネルが多いので、最大のターゲット数が 16 になります。

**注記**

Narrow SCSI デバイスは、ID が 8 以上の SCSI デバイスとは通信できません。SCSI ホストアダプタのターゲット ID を、**8 以上の数に変更してはなりません。**

SCSI ターゲット ID が大きければ大きいほど、デバイスの優先順位が高くなります。デバイス間のアービトレーションの場面では、一番大きい SCSI ID を持つデバイスが優先されます。

例

ディスクデバイス上の情報は(優先順位が高いので)、バックアップテープからのデータよりもアクセスが容易であるべきです。そのため、Self-Monitoring Analysis and Reporting Technology (S.M.A.R.T.) ディスクコントローラには、テープコントローラよりも大きいターゲット ID が必要です。このことから、ホストアダプタに SCSI ID 7 を使用するのがデファクトスタンダードになっています。

**注記**

Wide SCSI バスでは、0~7 の 8 個の ID の方が、8~15 の 8 個の ID よりも優先順位が高くなります。つまり、Wide SCSI システムでは、ターゲット ID の優先順位が、7、6、5、4、3、2、1、0、15、14、13、12、11、10、9、8 という順になります。

SCSI のデバイス ID に関するガイドライン

SCSI のデバイス ID に関するガイドラインは次のとおりです。

- Wide SCSI と Narrow SCSI のホストアダプタは、ともに 7 個までの SCSI デバイスを確実にサポートするが、Wide SCSI デバイスの中には、7 個までしかターゲット ID をサポートしないものがある
- すべての SCSI デバイスには、一意の ID 番号が必要である。SCSI ホストアダプタは、SCSI デバイスとの間の信号を、デバイスの位置ではなく SCSI ID 番号によって識別する。SCSI ID は通常、ジャンパを使って選択することで ID 番号をドライブに設定する。HP のホットスワップ対応のハードドライブでは、ID が自動的に設定される。HP が SCSI デバイス用に「予約済み」または「利用可能」に設定している SCSI ID 番号は次のとおり
 - 0 — 通常、ブート可能なハードドライブ用に予約済み
 - 1～6 — narrow (ナロー) と wide (ワイド) の両方のデバイスで利用可能
 - 7 — 通常、SCSI ホストアダプタ用に予約済み
 - 8～15 — wide (ワイド) デバイスで利用可能
- すべての SCSI チェーンまたは回路は、両端で終端する必要がある。これには、デバイスの終端機能か、終端ケーブルを使う
- 終端機能のない SCSI デバイスは、ターミネータを使う必要がある
- コンピュータに電源を投入する前に、外部 SCSI デバイスに電源を投入する。これによって、システムボード上の SCSI コントローラが外部 SCSI デバイスを認識して、自動的にリセットできるようになる

論理ユニット番号(LUN)

1つの SCSI ID デバイスが、複数の論理ユニット番号(LUN)を持つことができます。

デバイスの各ディスクは、LUN を使って独立したアドレスを設定できます(たとえば、4 ディスクの CD ジュークボックスに、LUN0~3 を割り当てられます)。

例

1つの CD-ROM ストレージシステムに 6つの CD-ROM ドライブがある場合、論理的に 1つの SCSI デバイスとしてアドレスを割り当てるので、必要な SCSI ID は 1 つです。個々の CD-ROM ドライブには、LUN によってアドレスを割り当てます。

SCSI の通信方法

2つの SCSI デバイスがバス上で通信を行うとき、一方のデバイスがイニシエータ、他方のデバイスがターゲットとして動作します。

- **イニシエータ** — 動作を開始する能力がある SCSI デバイスで、一般には、ホストコンピュータのアダプタまたはコントローラである。イニシエータは動作を開始し、ターゲットの選択やコマンドとデータの提供を担当する
- **ターゲット** — 動作を実行する SCSI デバイス。SCSI ストレージデバイスは固定のロール(通常はターゲット)を持つ。SCSI バス上での情報転送は、一度に2つの SCSI デバイス間のみである。ターゲットは、イニシエータが提供するコマンドを管理し、実行する

SCSI は、アービトレーションまたは要求といった方法を使ってバスを制御します。アービトレーションは、優先順位に基づくシステムで、SCSI バスの使用を要求するデバイスのうち、最も優先順位が高い SCSI デバイスに SCSI バスの制御が与えられます。

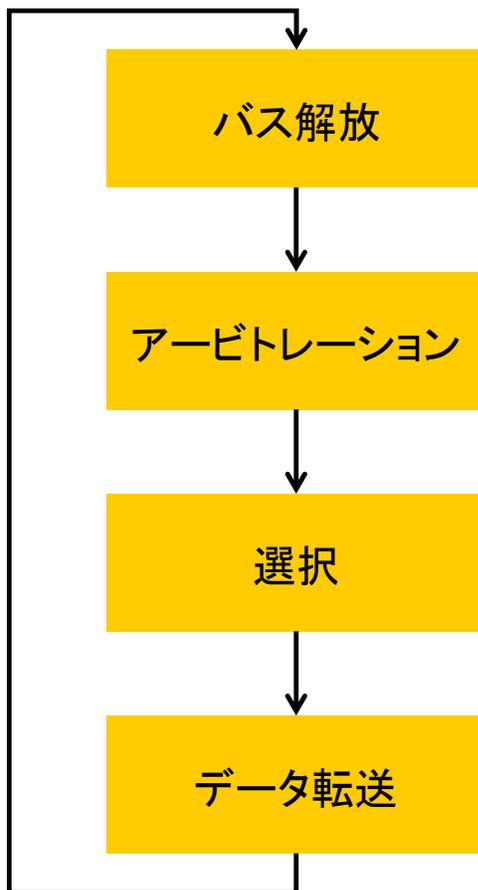
SCSI では、各デバイスがバスの制御についてアービトレーションを行い、バスマスタの優先順位が決定します。イニシエータは、バスが空き状態であることを認識すると、バスビジー信号とイニシエータ自身の ID をアサートします。アービトレーション遅延の後、イニシエータはデータバスをチェックし、自分よりも優先順位が高い ID のデバイスがバス上にある場合は、自分自身をアービトレーションからクリアします。

データバス上の情報転送は非同期で行われ、要求/応答ハンドシェイクに従って動作した後、1バイトのデータが転送されます。

SCSI デバイスとの間の I/O 要求が完了すると、バスは複数のフェーズを移行します。これらのフェーズは主に次の3段階に大別できます。

- **コマンドセットアップ** — セレクションフェーズ、アービトレーションフェーズ、コマンドフェーズからなる。I/O 要求をセットアップする場合、イニシエータのロールにあるホストプロセッサが、ハードドライブを SCSI ID で指定して、SCSI コマンドをターゲットデバイスに発行する
- **データ転送** — データフェーズからなる。ドライブはコマンドを受信すると、データの読み取りの場合はデータをホストプロセッサに転送し、書き込みの場合はホストプロセッサからデータを受け取る
- **コマンド完了** — ステータスフェーズとメッセージングフェーズからなる。要求されたデータが転送されると、ドライブはイニシエータに対して I/O 要求の完了ステータスを伝達し、次の I/O 要求を受け付けられるようにバスをフリー状態に移行させる

SCSI バスのトランザクション



SCSI バスのトランザクションプロセス

一部の SCSI バス機能は、イニシエータまたはターゲットにのみ割り当てられます。一般的な SCSI バスのトランザクションプロセスは次のとおりです。

- イニシエータが SCSI バスの占有権をめぐるアービトレーションを行い、ターゲットを選択する
- ターゲットが、データバス上のコマンド、データ、ステータス、その他の情報を要求し、場合によっては、ターゲットが SCSI バスのアービトレーションを行い、イニシエータを再選択して操作を続ける

ネゴシエーション

ネゴシエーションとは、初期化フェーズで、コントローラが接続先の各デバイスと転送レートについて合意するプロセスです。ネゴシエーション中に、データ転送フェーズで **Fast-Wide** モードで動作することを決定する場合は、コマンドセットアップフェーズ中に、個別のドライブごとに、コントローラとドライブ間で行います。



重要

コントローラとドライブは、処理できる最高のデータ転送レートについて個別にネゴシエートします。これは、8ビットと16ビットの **SCSI** デバイスが同じバス上にある場合に重要です。

ネゴシエーションフェーズで **SCSI** デバイスが応答しない場合を除き、ネゴシエーションを無効にはなりません。Fast SCSI は、Fast SCSI-2 ハードドライブと Fast-Wide SCSI-2 ハードドライブの両方と通信できます。Fast-Wide SCSI-2 コントローラは、いずれのタイプのハードドライブとも使用できます。



重要

SCSI コントローラと各ディスクドライブ間のバス上での最高転送レートは、最も速度が遅いデバイスのネイティブプロトコルによって決まります。

切断と再接続

SCSI デバイスは要求を受信すると、一時的にバスから切断してデータを内部で処理します。要求の処理が終わると、**SCSI** デバイスはバスに再接続します。これによって、1つのバス上で複数のデバイスが同時に動作できます。次に、個々のデバイスからのデータが、バス上をインタリーブモードで転送されます。

接続と切断の時間の長さは、**SCSI** コントローラおよび **SCSI** デバイスの機能によって異なります。**SCSI** ホストとデバイスが、**SCSI** バスから素早く切断する能力があると、パフォーマンスが向上します。

この機能は、タグ付きコマンドキューイング (**TCQ**) プロセスで実行される動作に似ています。ドライブの接続と切断の機能によって、デバイスがバスから素早く切断されるので、他のデバイスが **TCQ** プロセス中に、そのバスにアクセスできます。接続時間と切断時間の短縮によってパフォーマンスが向上するのは、**SCSI** バスが一杯にロードされている場合のみです。切断時間は、技術の進歩に伴って着実に短縮しています。

タグ付きコマンドキューイング (TCQ)

TCQ をサポートする SCSI デバイスは、複数の SCSI コマンドを受け付け、最適な順序でコマンドを実行できます。

例

1 つのハードドライブが 10 個の書き込みコマンドと 5 個の読み取りコマンドを受け付け、ヘッド ポジショニングが最適になるようにそれらをソートします。これによって、スループットが 30% 以上も向上する場合があります。

個々の I/O 要求は、タグによって一意に識別されます。オペレーティングシステムは、タグを使うことで、デバイスドライバキューのどの I/O が、デバイスによって「完了」とレポートされているかを確認できます。イニシエータは、キューのコマンドを追加または削除できます。コマンドをデバイスキューに追加する場合、イニシエータは実行順序を固定したり、実行する次のコマンドを指定したりできます。これらを指定しなければ、すべてのコマンドは、受け取った順序で実行されます。

TCQ を使用すると、ドライブが多くのコマンドを受け取り、SCSI バスを使わずにそれらのコマンドを実行できます。これによって、他のデバイスがそのバスを利用できるようになります。ドライブは、コマンドの処理がすべて完了し、バッファが一杯になり、情報を転送する準備ができたときにそれを通知します。この方法により、一度に 1 つのターゲットしかアクティブにできないバスで、使用効率が向上します。

TCQ を実装することは SCSI-2 ではオプションです。最近の SCSI デバイス、特にハードドライブは、TCQ をサポートしているので、デバイスが同時に複数の未処理 I/O 要求を持つことができます。デバイスのインテリジェント機能により、それ自体の要求キューに応じて、ヘッド ポジショニングなどの操作が最適化されます。

注記

▲ RAID アレイなどの SCSI デバイスでは、TCQ 機能によってデバイスが本来持つ並列性が活用されます。

一部のオペレーティングシステムも、事前にデータをソートする機能をサポートすることによってヘッド ポジショニング時間の短縮を図っていますが、RAID システムではこの種の事前ソーティングが、物理ドライブではなく、論理ドライブへのアクセスを基準に最適化されています。これでは、実際にはヘッドの移動時間が増加することになります。このため、TCQ は、ハードウェアレベルで非常に重要な機能となっています。TCQ を利用するには、デバイスドライバによるサポートが必要です。

TCQ は、オプションの SCSI コマンドセットに属しているため、サポートしないドライブメーカーもあります。サポートされていないデバイスをバスに接続すると、TCQ が完全に無効になり、パフォーマンスが低下します。

注記

▲ HP の SCSI ハードドライブは、すべて TCQ をサポートしています。

SCSI バスフェーズ

SCSI バス上ですべてのアクティビティは、次の 8 つの SCSI バスフェーズのいずれかで行われます。バスフェーズによって、転送方向とデータ線に送信される情報のタイプが決まります。

最初の 4 つのフェーズは、SCSI デバイスが SCSI バスの使用权を得るために使用することから、まとめてネゴシエーションフェーズと呼ばれます。後の 4 つのフェーズは、データや制御情報の転送に使用されることから、まとめて情報転送フェーズと呼ばれます。情報転送フェーズに、決まった順序はありません。

注記

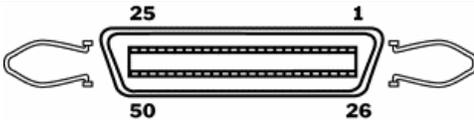
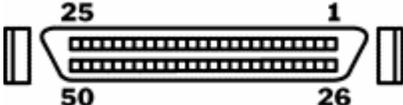
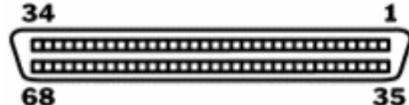
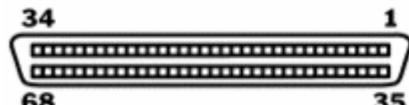
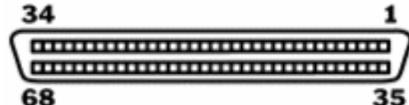
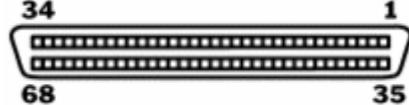
SCSI バスは、同時に複数のフェーズを実行することはできません。

次の表では、SCSI バスフェーズとその機能について説明します。

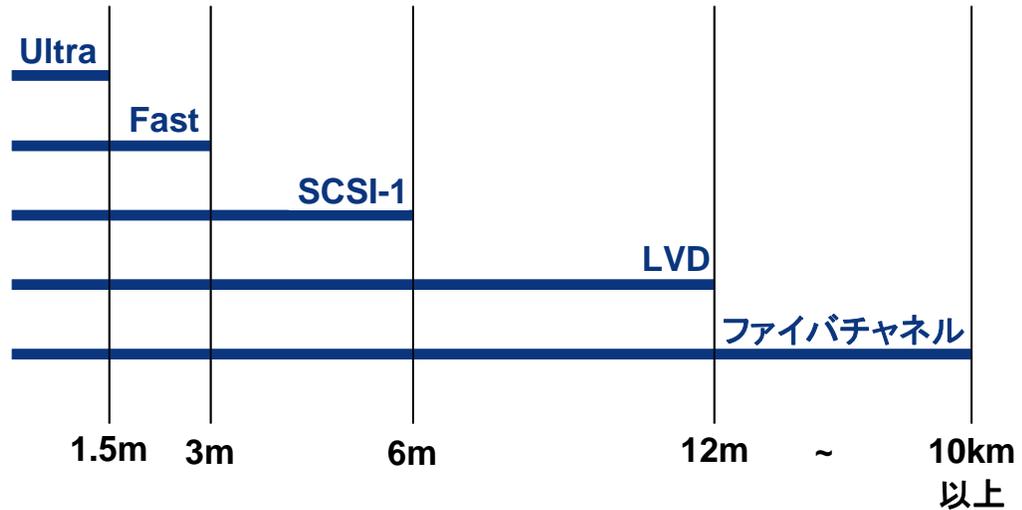
バスフェーズ	説明
バスフリー	このフェーズは、どの SCSI デバイスもバスを使っていないことを示す。SCSI デバイスがアービトレーションフェーズを開始するには、SCSI バスがフリーの状態であることが必要である。
アービトレーション	このバスフェーズでは、SCSI デバイスがバスの制御を獲得し、イニシエータまたはターゲットのロールを実行できる。イニシエータは、バスが空き状態であることを認識すると、BSY 信号とイニシエータ自身の ID をバス上にアサートする。待機時間経過後、イニシエータはデータバスを確認し、自分よりも優先順位が高い ID のデバイスがバス上にある場合は、自分自身をアービトレーションからクリアする。
セレクション	このフェーズでは、イニシエータが、読み取りや書き込みのコマンドなどの機能を実行するターゲットを選択する。この選択は、SEL 行、データバスと対象のターゲット ID、およびイニシエータ ID をアサートして行う。次にイニシエータは、ターゲットがバス上に BSY 信号をアサートできるよう、BSY 信号を解除する。
リセレクション	このフェーズでは、アービトレーションに勝ったターゲットがイニシエータに再接続し、切断の前に開始していたトランザクションを再開する。
コマンド	このフェーズでは、ターゲットが、イニシエータからのコマンド情報の転送を要求できる。コマンド情報とは、Command Descriptor Block というデータパケットになったバイトの集合である。要求にあたって、デバイス間で追加情報の転送が必要なこともある。その場合、追加情報は、Data In または Data Out フェーズで転送される。
データ	Data In フェーズでは、ターゲットがイニシエータへのデータ転送を要求できる。 Data Out フェーズでは、ターゲットがイニシエータからのデータ転送を要求できる。
ステータス	このフェーズでは、ターゲットが、イニシエータへのステータス情報の転送を要求できる。これによってターゲットは、イニシエータに対し、コマンドが完了したかどうかを伝達できる。
メッセージ	Message Out フェーズでは、イニシエータがターゲットデバイスに対し、ターゲットのどの部分がコマンドに対して動作するかを伝達できる。 Message In フェーズでは、ターゲットが、イニシエータへのメッセージ送信を要求できる。このフェーズは、エラーやコマンドの中断を示す場合、またはターゲットデバイスが切断されることをイニシエータに警告する場合に使われる。

SCSI コネクタ

次の表では、SCSI コネクタの例を示します。

SCSI タイプ	バス幅 (ビット)	バス 速度 (MHz)	最大転送 レート (MB/s)	コネクタ
SCSI-1	8	5	5	 <p>セントロニクス 50ピンコネクタ</p>
Fast SCSI-2	8	10	10	 <p>50ピン外部コネクタ</p>  <p>50ピン内部コネクタ</p>
Fast- Wide SCSI-2	16	10	20	 <p>Wide 68ピンコネクタ(内部および外部)</p>
Wide- Ultra SCSI-3	16	20	40	 <p>Wide 68ピンコネクタ(内部および外部)</p>
Wide Ultra2 SCSI-3 (LVD)	16	40	80	 <p>Wide 68ピン内部コネクタ</p>  <p>VHDCI 68ピン外部コネクタ</p>
Wide Ultra3 SCSI-3 (LVD)	16	40	160 (ダブル トランジション クロッキング)	 <p>Wide 68ピン内部コネクタ</p>  <p>VHDCI 68ピン外部コネクタ</p>

ケーブル



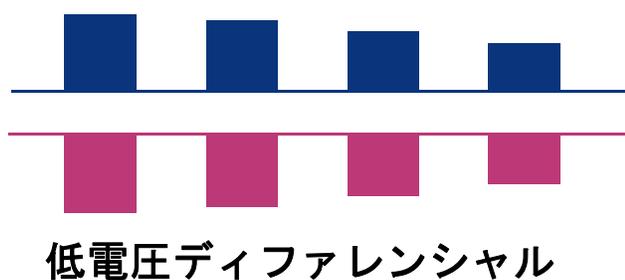
使われる SCSI ケーブルのタイプは、プロトコルと構成によって異なります。SCSI にはさまざまなプロトコルと構成があり、ディファレンシャル SCSI とシングルエンド SCSI のデバイスは同じように見えるので、ケーブルの選択には注意が必要です。

一般にはシングルエンドケーブルが使われますが、シングルエンドケーブルとディファレンシャルケーブルには、次のような共通の特性があります。

- **A ケーブル** — 標準の 50 ワイヤケーブル。そのまま Narrow、8 ビット SCSI に使え、B ケーブルと組み合わせれば、Wide や 16 ビット SCSI に使える。ケーブル両端のコネクタは、デバイスが内部か外部かによって異なる
- **B ケーブル** — B ケーブルは SCSI-2 で採用され、Wide SCSI で使われる。このケーブルは 68 ピンコネクタを持ち、A ケーブルとともに使う。このように 2 本のケーブルを使うのはわずらわしいので、B ケーブルの代わりに開発されたのが P ケーブルである
- **P ケーブル** — このケーブルは、現行の標準 68 ワイヤケーブルであり、すべての Wide SCSI の実装で使われる。P ケーブルは、内部接続と外部接続に同じコネクタを使う

SCSI バス幅でケーブルタイプが決まり、バス速度によってケーブル長が決まります。シングルエンド SCSI を使う場合は、SCSI バスが高速になると長距離での信号破壊の問題が多くなるので、ケーブルを短くしなければならないという制約があります。ディファレンシャル SCSI では、速度に関係なくシングルエンドより長いケーブルが使えます。

電氣的信号方式



SCSI 電氣的信号方式

高速な上に、外部ケーブル接続を行うので、バス上での信号の完全性が常に問題になります。ケーブルが長くなるほど、信号劣化や干渉に伴う問題が起こりやすくなります。またバス速度が大きくなるほど、信号を正確に保つのが難しくなります。

SCSI は、次の 2 つの電氣的信号方式を定義しています。

- シングルエンド SCSI
- ディファレンシャル SCSI
 - 高電圧ディファレンシャル (HVD)
 - 低電圧ディファレンシャル (LVD)

シングルエンド SCSI



シングルエンドインタフェースでは、1本の線とそれに対応するグラウンドワイヤでデータを伝送します。電圧は信号線でのみ伝送されます。シングルエンドデバイスとLVDデバイスは同じバス上で使えますが、その場合、SCSIバス全体がシングルエンドの速度で動作します。



警告

シングルエンド SCSI と HVD SCSI は、電氣的レベルの互換性がありません。シングルエンド SCSI と HVD SCSI のデバイスを同じバスに混在させてはなりません。混在させた場合は、機器に物理的損傷が生じる可能性があります。

シングルエンド SCSI は標準の SCSI であり、他のバス上で使われる従来型の信号方式を使います。各信号は1本の線で伝送されます。シングルエンド SCSI は、すべてのデータや信号の共通のリターン線としてグラウンド電圧を使います。正の電圧は1、グラウンド電圧は0で表現されます。

シングルエンド SCSI は最も一般的なタイプの SCSI で、最も柔軟でコストパフォーマンスの高いソリューションです。しかし、ケーブルの長さは極端に制限されるという欠点があります。シングルエンド SCSI バスは、5V または 0V の信号を使い、それらは共通のグラウンドレベルを基準にします。シングルエンドの 8ビット SCSI バスは、約 25 本のグラウンド線を持ち、それら全部が、すべてのデバイス上にある 1本のレールに接続されています。

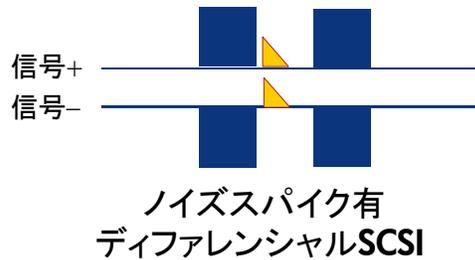
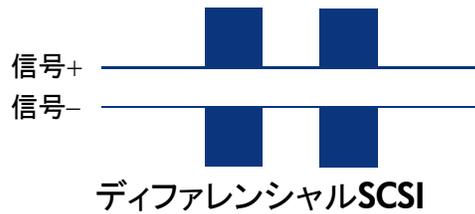
ケーブルの内部が並行線なので、隣り合う線同士で干渉の問題が発生します。グラウンド線は通常、太くて強固ですが、信号線は干渉に対して脆弱です。このため、シングルエンド SCSI は、短いケーブルしか使えません。

標準のシングルエンドバスは、最大長が 6m です。同じバスを Fast SCSI デバイスで使うと、最大長が 3m に短縮されます。Fast-SCSI は、Narrow チャンネルの場合、バス上を 5MB/s ではなく 10MB/s で転送できます。Wide チャンネルの場合は、20MB/s で転送できます。Ultra SCSI と Ultra2 は、それぞれ 20MHz と 40MHz での転送が可能です。したがって、Ultra は 8 ビットバスで 20MB/s、16 ビットバスでは 40MB/s です。Fast-20 では最大バス長が 1.5m、Fast-40 は最大バス長が 12m です。

▲ 注記

バス上に Ultra SCSI または Ultra2 を使って通信するデバイスがある場合、高速なバスの最大長の制約に従う必要があります。

ディファレンシャル SCSI



ディファレンシャル SCSI バスは、SCSI-2 標準で初めて採用され、バス長の制限による潜在的なボトルネックを軽減します。

ディファレンシャル SCSI は、一部の SCSI ドライブで使われている論理信号システムです。このシステムでは、SCSI バス上のノイズの影響を減らすために、お互いのミラーイメージである正と負の信号レベルペアを使います。信号にノイズが混入すると、ノイズは正の状態と負の状態の両方に現れ、相殺されます。ディファレンシャル信号では、バス信号がそれぞれ専用のリターン線を使います。したがって、信号はツイストペアの線で伝送されます。この 2 本の線の電圧差分によって、信号が「真」であるか「偽」であるか判定されます。

「1」は一方の線では正の電圧で表され、他方の線では同じ大きさの負の電圧で表されます。「0」は両方の線で電氣的グラウンド、つまり 0 電圧で表されます。2 つのコンダクタを使って 1 つの信号を表現するので、信号の回復が容易であり、破壊され難くなります。このため、シングルエンド SCSI よりはるかに長いケーブルを使えますが、費用も高くなります。

従来のディファレンシャル SCSI は、現在、HVD と呼ばれています。ディファレンシャル SCSI バスは、最大長が 25m です。シングルエンド Fast SCSI バスは、最大長が 3m です。

ディファレンシャルインタフェースは、すべての SCSI 信号にツイストペアケーブルを使います。データをドライブに渡す前に、信号の完全性がケーブル端で確認されるので、外部接続用に使われます。

ディファレンシャル信号は、ノイズマージンの点で優れています。ディファレンシャルバスは、主にキャビネット間接続に使われます。低コストのシングルエンド SCSI は、主に短い距離のバスで使われています。

例

元の信号が 5V の場合、もう 1 つの信号線では逆の -5V の信号が伝送されます。電圧差分が増幅され、2 つの信号線の差分は $5V - (-5V) = 10V$ となります。差分を使うので、1 番目の線から 2 番目の線を差し引くわけです。

線上のノイズスパイクその他の電氣的ノイズは、同時にほぼ同じ振幅で発生することになります。

例

1V のノイズスパイクが発生した場合、両方の線に 1V のスパイクが同時に伝送されません。差分は $1V - 1V = 0V$ となるので、スパイクは差分増幅器によって相殺されます。この方法は、同じ信号、つまり同位相の信号を除去するので、同相信号除去と呼ばれています。

これらの信号方式システムにも、必要な場合に備えて、転送レートプロトコルが定義されています。だからといって、すべてのディファレンシャルプロトコルが、シングルエンド SCSI と HVD SCSI の両方で簡単に使えるわけではありません。HVD SCSI が使われる頻度ははるかに低く、はるかに高いコストがかかります。

高電圧ディファレンシャル

ディファレンシャルインタフェースは、信号トランスミッタとレシーバを使っているため、シングルエンドバスよりも長距離の信号伝送ができ、ノイズの耐性も大きくなります。信号は正と負のコンポーネントに分解され、ケーブルのペアを使って伝送されます。両方のケーブルで 5V が伝送されます。信号は受信側デバイスで再統合されます。シングルエンドデバイスのケーブル長が 6m 以下であるのに比べ、ディファレンシャル SCSI は、ノイズ耐性が高いため、25m という長い SCSI ケーブルを使うことができます。データ転送レートも速くなります。ディファレンシャル SCSI とシングルエンド SCSI は、SCSI コンバータなどの電子機器を使ってディファレンシャルとシングルエンドの構成を変換しない限り、同じバス上での互換性はありません。

低電圧ディファレンシャル

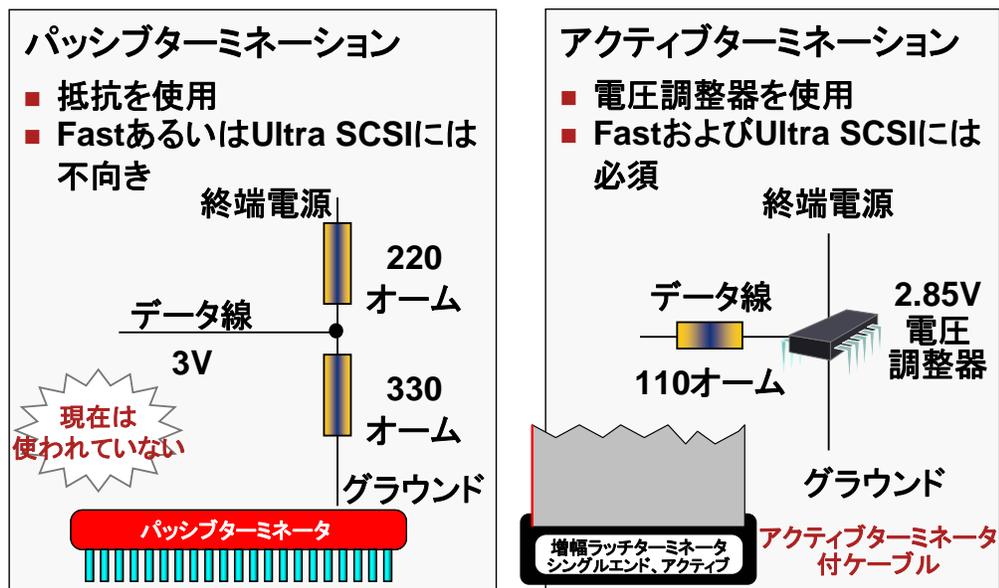
このインタフェースは、3.3V とディファレンシャルモードの動作を使って、12m という長いケーブル長を実現しながら、80MB/s の転送速度で回線ノイズを有効に除去します。つまり、既存のシングルエンド SCSI 構成よりも、高速かつ信頼性が高くなります。LVD は、HVD の高コストという短所を軽減するために開発されたインタフェースです。

ファイバチャネルはコストが合わず、シングルエンド SCSI はケーブル長の制約から使えないといった構成では、LVD SCSI が最適です。低電圧であることは、同時に電子のノイズの発生が少ないことを意味します。

ケーブルについての留意事項

シングルエンド SCSI と HVD SCSI で使うケーブルは、見た目では区別できません。実際に接続する前に、ケーブルタイプを確認してください。シングルエンド SCSI と HVD SCSI 間は、コンバータがあります。シングルエンド SCSI と HVD SCSI は、特別なアダプタを使わない限り、信号方式、デバイスのケーブル接続方式、電気特性に互換性はありません。SCSI ホストアダプタで使うデバイスのタイプに合わせる必要があります。

ターミネータ



SCSI ターミネーションのオプション

物理的な波が糸を通して伝わるように、電気信号はワイヤを通して伝送されます。信号がワイヤの端に到達すると、信号は反射してワイヤを逆方向に戻っていきます。反射が起こると、反射した信号がバス上の他のデータに干渉し、信号の損失やデータの破壊が発生します。これを防ぐためには、SCSI バスの両端を終端(ターミネート)する必要があります。つまり、特別なコンポーネントを使って、バスの端に到達した信号をすべて捕捉し、反射を防ぐわけです。

SCSI バスは、以下の 3 種類の終端を使います。この 3 種類は、バスの終端に使う電気回路が異なります。

- **パッシブターミネーション** — 最も初期の、信頼性が低いタイプの終端方法である。バス終端に単純な抵抗を使う。パッシブターミネーションは、短い低速の SCSI バスには十分であるが、Fast SCSI や他の高度な SCSI 使用には適さない
- **アクティブターミネーション** — 高速の SCSI バスにとって最低限の必要条件。パッシブターミネーションで使われる抵抗に電圧調整器を追加して、バス終端の信頼性と一貫性を高める
- **FPT (Forced perfect termination)** — 高度なアクティブターミネーションの方法で、回路にダイオードクランプを取り付けることによって、確実に正しい電圧で終端する。この終端は、信号の反射や考えられるその他の問題をほぼ完全に防止するので、SCSI バス終端の最適な形といえる

ターミネータ電源

ターミネータが正しく動作するためには、電源を供給する必要があります。SCSI バスでは、1本の線がこの目的専用で使用されます。TERMPWR は、終端ロジックに電源を供給する 5V 信号です。通常、コントローラがこの電圧を供給します。各デバイスは、統合ターミネータソケットを通じて、それぞれのターミネータ電源を供給できます。HP 製の機器を使う場合は、TERMPWR ジャンパを TRMPWR = OFF に設定します。

注記



統合ターミネータ電源を SCSI バス線に供給するデバイスが、意図的にまたは誤って電源供給を停止すると、バスは終端されなくなります。

外部ターミネータと統合ターミネータをデバイス上で同時に使うと、誤ってバスの終端以外の場所にターミネータが構成されたり、1つのバスに2つのターミネータという基準を超えてターミネータが構成されたりといったことが、システムで発生する可能性があります。このように構成されると、システムの故障やエラーが発生します。

イニシエータは、バスの動作を開始し、ターミネータ電源を供給するデバイスです。SCSI デバイスは、ターミネータ電源を供給してもかまいませんが、必ずしも供給する必要はありません。

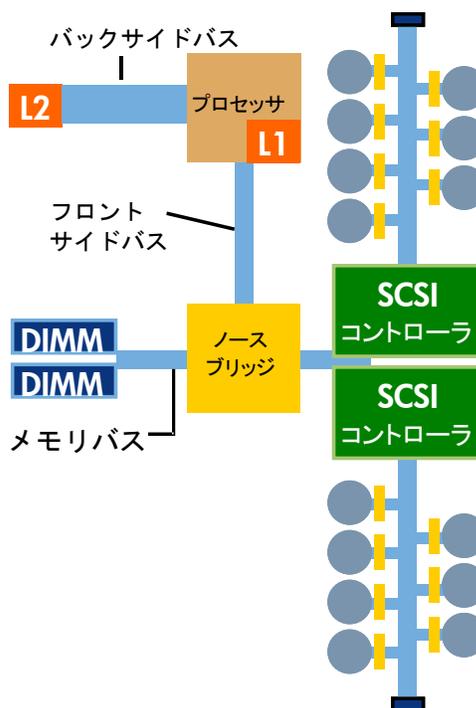
電源が入っていないデバイスがバス上に存在できるように、ターミネータの電源はダイオードを通して供給されなければなりません。これは、電源が入っていないデバイスに電流が逆流するのを防ぐためです。

ターミネータ電源は、普通、ヒューズが入っています。ヒューズが飛んで、バスが機能停止することもあります。複数のデバイスがターミネータ電源を供給していれば、ヒューズが1つ飛んでもバスは機能停止しません。外部ターミネータの中には、ターミネータ電源が供給されていることを LED で示すものもあります。最新の設計では、リセットされ自動復帰するヒューズが使われることもあります。

HP による SCSI の実装方法

HP の技術は、高性能でスケーラビリティを備え、HP プロダクトファミリー内での互換性を維持する SCSI の実装方法を開発してきました。これらの革新的技術は、すべての HP ストレージソリューションと ProLiant サーバに完全に統合され、すべてテスト済みです。

HP による SCSI コントローラチップ



1990 年、Compaq は、サーバシステムボード上でも、スタンドアロンの EISA コントローラとしても利用できる SCSI コントローラを発表しました。現在では RAID や PCI-X に対応しているだけでなく、互換性を保った SCSI チップのファミリーとして進化してきました。

チップ	SCSI プロトコル	チャンネル	スピード	バスタイプ
53C710	Fast	1	10MB/s	EISA
53C810	Fast	1	10MB/s	PCI
53C825	Fast-Wide	1	20MB/s	PCI
53C875	Wide-Ultra	1	40MB/s	PCI
53C876	Wide-Ultra	2	2x40MB/s	PCI
53C895	Wide Ultra2	2	80MB/s	PCI
53C896	Wide Ultra2	2	2x80MB/s	PCI
53C1510	Wide Ultra2	2 x RAID	2x80MB/s	PCI
AIC7892X	Wide Ultra3	1	160MB/s	PCI
AIC7899G	Wide Ultra3	2	2x160MB/s	PCI
AIC7899W	Wide Ultra3	2	2x160MB/s	PCI、PCI-X

HP による SCSI ストレージソリューション



ProLiant サーバは、1 つまたは複数の SCSI コントローラを内蔵しています。

サーバ	SCSI バス	バスタイプ	転送レート	チップセット
ProLiant ML330G2 SCSI, ML350G3, ML370G3, ML530G2, ML570G2	2	Wide Ultra3	2 x 160MB/s	AIC7899W
ProLiant ML370G2	2	Wide Ultra3	2 x 160MB/s	AIC7899G
ProLiant ML310 SCSI, ML330G3, DL320G2	1	Wide Ultra3	160MB/s	AIC7892X
ProLiant DL760, 8500	2 x RAID	Wide Ultra2	2 x 80MB/s	53C1510
ProLiant DL580, ML570	2 x RAID	Wide Ultra2	2 x 80MB/s	53C1510
ProLiant DL380, ML370, DL360	2 x RAID	Wide Ultra2	2 x 80MB/s	53C1510
ProLiant ML530, ML350, ML330	2	Wide Ultra2	2 x 80MB/s	53C896
ProLiant DL320 SCSI	1	Wide Ultra2	80MB/s	53C1510
ProLiant 7000, 6000 Xeon	3	Wide-Ultra	3 x 40MB/s	53C875 + 876
ProLiant 7000, 6000, P-Pro	2	Wide-Ultra	2 x 40MB/s	53C875 + 875
ProLiant 6500 Xeon	2	Wide-Ultra	2 x 40MB/s	53C876
ProLiant 550, 3000	2	Wide-Ultra	2 x 40MB/s	53C876

HP の SCSI アダプタ

HP の SCSI アダプタは、ProLiant サーバおよび管理ソフトウェアの全ラインに完全に統合され、テスト済みです。HP のアダプタは、TCQ を使っているため、複数のコマンドを同時に受け取り、効率的に実行できるようにソートし、応答時間を短縮することによって、パフォーマンスの向上を図ります。

HP の SCSI アダプタとその機能は次のとおりです。

- デュアルチャネル 64 ビット/133MHz PCI-X Ultra320 SCSI アダプタ
 - チャンネルあたり最高 320MB/s のデータ転送レート
 - 64 ビット技術と優れたパフォーマンス
 - 133MHz PCI-X のサポート
 - テープアレイのサポート
 - 最高 28 デバイスのデュアル独立チャネル接続
 - Ultra3 SCSI、Wide Ultra2 SCSI、Wide Ultra SCSI のデバイスとの下位互換性
- デュアルチャネル 64 ビット/66MHz Wide Ultra3 SCSI アダプタ
 - チャンネルあたり最高 160MB/s のデータ転送レート
 - 64 ビット技術と優れたパフォーマンス
 - 66MHz と 33MHz の PCI サポート
 - テープアレイのサポート
 - 最高 28 デバイスのデュアル独立チャネル接続
 - Wide Ultra2 SCSI、Ultra2 SCSI-3、Wide Ultra SCSI-3、Ultra SCSI-3、Fast-Wide SCSI-2、Fast SCSI-2 のデバイスとの下位互換性

HP の SCSI ドライブ

HP のドライブは、製造元が供給するドライブとは異なります。ハードウェアはおおむね同一ですが、ファームウェアが以下の追加機能をサポートするように変更を加えています。

- ドライブジオメトリの再構成
- 自己モニタリング
- 事前予防保証のサポート
- 優れた診断機能
- 優れたエラー検出、修正機能

ユニバーサルディスクドライブ

HP の Wide Ultra2 と Ultra3 ドライブには、強固な機械的設計と優れたドライブ冷却機能を持つ新しいドライブキャリアが採用されています。

ユニバーサルディスクドライブは、Intel ベースおよび Alpha ベースのシステムで使えます。

10K および 15K RPM ドライブ

10K rpm ドライブは高い性能を発揮します。ファイルサーバやプリントサーバ環境では、7200 rpm ドライブに比べて 40% の性能向上が可能です。

15K rpm ドライブはきわめて高性能です。ファイルサーバやプリントサーバ環境では、10K ドライブに比べて 30%、7200 rpm ドライブに比べて 60% の性能向上が可能です。

注記



他社のドライブを使うと、タイムアウトや一般的なデータ破壊などの問題が発生することがあります。HP が供給するドライブのファームウェアは、RAID 環境に合わせて最適化され、障害予測警告メカニズムと高度な診断機能をサポートします。

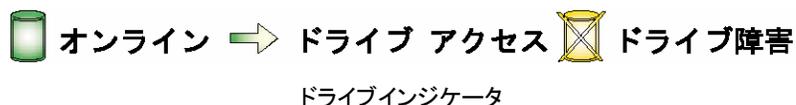
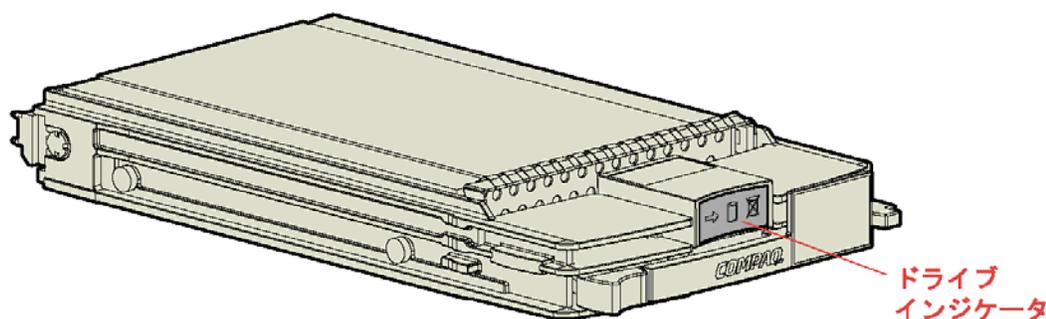
ホットプラグ SCSI ドライブ

特殊なドライブキャリアにマウントされているすべての HP ドライブは、システムの動作中に取り外し、取り付けが可能です。

ホットプラグ可能なドライブの SCSI ID は、自動的に割り当てられます。SCSI ID は、ドライブベイの位置によって自動的に決まります。ID は、バックプレーンボードにハードコードされています。SCSI ID は、ドライブ自体ではなくドライブベイに割り当てられるので、ホットプラグ可能なドライブを取り外すと、その SCSI ID が解除されます。

HP のホットプラグ可能なドライブの LED は、ドライブではなく、SCSI コントローラによって制御されます。他社のコントローラを使う場合は、LED がサポートされません。

ホットプラグ可能な SCSI ドライブの交換



ホットプラグ可能なドライブがサポートされているので、システムの動作中に、フォールトトレランス構成のドライブの取り付け、取り外しができます。

ドライブは、通常の動作中にホットプラグが可能です。ディスクドライブをホットプラグするとき、取り外したドライブを交換して、再構成が完了するまでの間、サブシステムがフォールトトレランスでなくなります (RAID ADG を除く)。



警告

ホットプラグ可能なディスクドライブの交換中に、別のドライブが故障した場合は、アレイ全体が失われます (RAID ADG を除く)。

iSCSI

インターネット SCSI (iSCSI) は、TCP/IP 上で、ストレージネットワークを構築する新しいプロトコルです。iSCSI は、ネットワークおよびインターネット内で、ストレージブロックを伝送します。

iSCSI は、ファイバチャネルに似ていますが、データ伝送の方法が異なります。SCSI ストレージプロトコルは、IP プロトコルの上のレイヤに置かれ、データをネットワークまたはインターネット経由で送信します。

HP は、iSCSI によって統合 SAN を拡張して、リモートサイトとの間の IP トンネリングを実現する予定です。この iSCSI トンネリングは、現在使われているリモートサイトのリプリケーションプロトコルの後継となります。

iSCSI トンネリングでは、次の処理が行われます。

1. ユーザまたはアプリケーションがデータ、ファイルまたはアプリケーションを要求する
2. SCSI コマンドとデータ要求が、オペレーティングシステムによって処理される
3. コマンドがカプセル化または暗号化され、パケットヘッダが追加される。作成されたパケットがイーサネット接続で伝送される
4. パケットが受信され、ディアセンブルまたは復号化される。SCSI コマンドと要求が分離される
5. SCSI コマンドとデータが SCSI コントローラに送信された後、SCSI ストレージデバイスに送信される



注記

要求に対応してデータを返すこともできます。iSCSI は双方向です。

SCSI のトラブルシューティング

以下のリストは、SCSI の一般的な問題を解決するためのガイドラインです。

- コネクタやケーブルがしっかりと接続されているかを確認する
- ターミネータの位置と数を確認する
- 特に外部ターミネータの場合、ターミネータ電源の供給元がバスに最低 1 つ存在することを確認する
- 2つのデバイスが同じ ID を持っていないことを確認する
- システムの起動前に、すべての外部デバイスの電源が投入されることを確認する
- 可能な限りデバイス数を減らし、バス構成を最低限にする
- 可能であれば、バス速度を下げるようホストアダプタを構成する

復習問題

1. 5つの SCSI 標準とは、SCSI-0、SCSI-2、SCSI-5、SCSI-6、SCSI-7 です。
 正
 誤
2. シングルエンド SCSI とディファレンシャル SCSI インタフェースの主な違いは何ですか。
.....
.....
.....
3. パッシブターミネーションとアクティブターミネーションの主な違いは何ですか。
.....
.....
4. Wide Ultra SCSI-3 の最大バースト転送レートは?
.....
5. Wide Ultra2 SCSI-3 の最大バースト転送レートは?
.....
.....
6. 初期化フェーズで、コントローラが接続先の各デバイスと転送速度について合意するプロセスを表すのは、以下のどの用語ですか。
 - a. 切断/再接続
 - b. タグ付きコマンドキューイング
 - c. ネゴシエーション
 - d. 上のどれでもない

7. タグ付きコマンドキューイング (TCQ) の目的は何ですか。

.....

8. ターミネータは、SCSI バスの両端に必要です。

正

誤

9. Wide Ultra2 ハードドライブを Fast-Wide SCSI コントローラに接続できますか。

.....

10. Wide Ultra3 SCSI-3 の 3 つの新機能は何ですか。

.....

目的

この章では、次の内容について学習します。

- RAID テクノロジー
- RAID の実装方法
- RAID のレベル
- ストライプサイズの最適化方法
- Smart アレイコントローラとそれらの機能
- Smart アレイコントローラの管理に使う次のようなテクノロジーとユーティリティ
 - アレイコンフィギュレーションユーティリティ (ACU)
 - アレイコンフィギュレーションユーティリティ XE (ACU-XE)
 - Option ROM Configuration for Arrays (ORCA)
 - Array Diagnostics Utility (ADU)
 - Dynamic sector repair (DSR)
 - ホットプラグドライブのサポート

RAID テクノロジーの概要

RAID は、Redundant Array of Independent Disks (独立ディスクの冗長アレイ) の頭字語です。RAID は、ホストシステムから 1 つの大容量デバイスとして認識される、ディスクドライブのグループです。RAID の目的は、フォールトトレランス機能と高速ディスクアレイを提供することです。

RAID テクノロジーは、複数の小容量のドライブで構成されるグループの方が、大容量の 1 つのドライブより効率が高いという理論に基づいています。フォールトトレランス RAID では、基本的な前提として、インテリジェントなマネージャであれば、アレイ内の個々のドライブに障害が発生しても、データを損失することなく、ドライブアレイを制御できると考えられています。

RAID の利点

RAID ストレージは、非アレイストレージシステムに比べて、次のような主な利点があります。

- **パフォーマンスの向上** — RAID システムは、データと I/O 要求をドライブアレイのすべてのメンバ間に配布します。アレイ内の各ドライブは、I/O 要求の処理を支援します。これによって、同時並行性が高まり、全体のパフォーマンスが向上します。

注記



フォールトトレランス RAID の構成は、フォールトトレランスの構成方法に応じて、パフォーマンスが低下します。

- **データの可用性と信頼性** — RAID システムでは、フォールトトレランス構成を採用して保存データの一部または全部を複製し、ディスクに障害が発生しても情報が損失しないように保護します。RAID ソリューションの多くは、ディスクの障害から自動的に復元でき、ホスティング環境に止まることのない I/O サービスを提供します。
- **ストレージ容量の動的拡張** — RAID システムは、1 つのディスクシステムの容量をはるかに超えるストレージ容量を拡張できます。RAID ソリューションの多くは動的であり、ディスクを追加してアレイを拡張できます。
- **ストレージ管理が容易** — RAID システムは、多くの小容量ディスクドライブを 1 つまたは複数の仮想ディスクに統合するので、容量、データ可用性、パフォーマンスレベルをすべて調整できます。これにより、1 つのアレイに統合された大容量記憶装置の管理が可能になります。

JBOD

Just a Bunch of Disks (JBOD)は、1 つまたは複数のスタンドアロン ディスクドライブをサーバの RAID コントローラやその他のドライブコントローラに接続できるストレージオプションです。このオプションは、容量を拡大し、重要でない業務データに使われます。

JBODドライブは、RAID アレイに組み込まれるのではなく、RAID コントローラによって制御される他のデバイスと同じ相互接続バス上のサーバで利用できるようになります。JBOD ディスクドライブには、データ冗長性やストライプの機能はありません。

ドライブアレイ

ドライブアレイとは、複数のハードディスクドライブまたは物理ドライブのセットを 1 つの大容量ドライブとして構成したものです。アレイは、論理ドライブ (または論理ボリューム) という 1 つまたは複数のサブセットからなります。論理ドライブは、アレイにあるすべての物理ドライブに広がっています。論理ドライブは、数個の物理ドライブで構成されていますが、オペレーティングシステムは、論理ドライブを 1 つの連続ストレージ空間として扱います。アレイのドライブが、素早く正確に構成されれば、ユーザは大量のギガバイト単位のデータにアクセスできます。

ドライブアレイを実装したドライブの利点は、次のとおりです。

- 高速な実効データ転送レート
- 複数の要求の同時処理能力
- ストレージ容量の増大
- データ構成における柔軟性
- 高信頼性

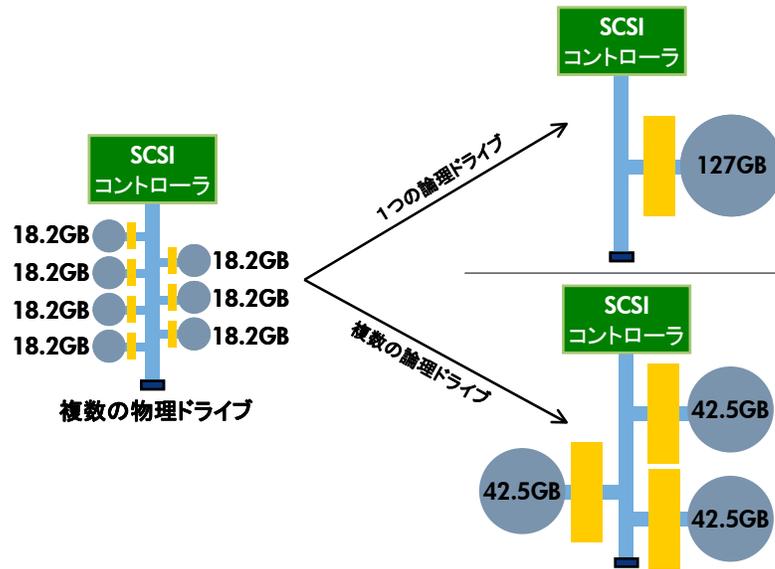
アレイのディスク数

アレイの I/O 帯域幅は、ドライブを追加することで拡大できます。ドライブを追加すると、多数のドライブで同時にデータを読み書きできるようになります。

顧客の多くは、多数の小容量ドライブを購入するより、少数の大容量ドライブを購入しますが、それは価格と管理の面を考慮し、パフォーマンスに与える影響を考慮しないからです。ドライブスピンドルが不足すると、パフォーマンスが低下することがあります。

同じディスク構成でも、RAID のレベルとドライブ数によって、利用できる I/O 帯域幅が異なります。アプリケーションの I/O レートを維持するためには、(ドライブ容量に関係なく) 十分なドライブ数が必要です。

論理ドライブ



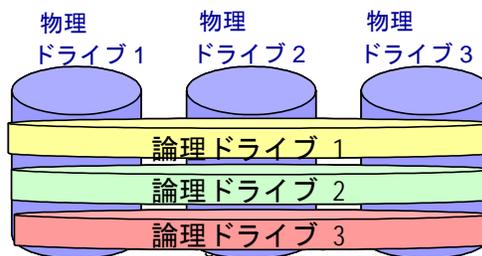
1つの論理ドライブを構成する数個の物理ドライブ

論理ドライブは、オペレーティングシステムから1つのドライブとして認識されるアレイのサブセットです。オペレーティングシステムは、個別のディスクを認識できず、1つの論理ドライブに1つのディスクのみが存在すると想定します。

アレイコントローラに接続すると、複数のSCSIドライブを1つのアレイとして構成し、1つの論理ドライブまたは複数の論理ドライブとして扱うことができます。

例

7つの18.2GBのドライブを、論理的に1つの127GBのドライブとして構成します。



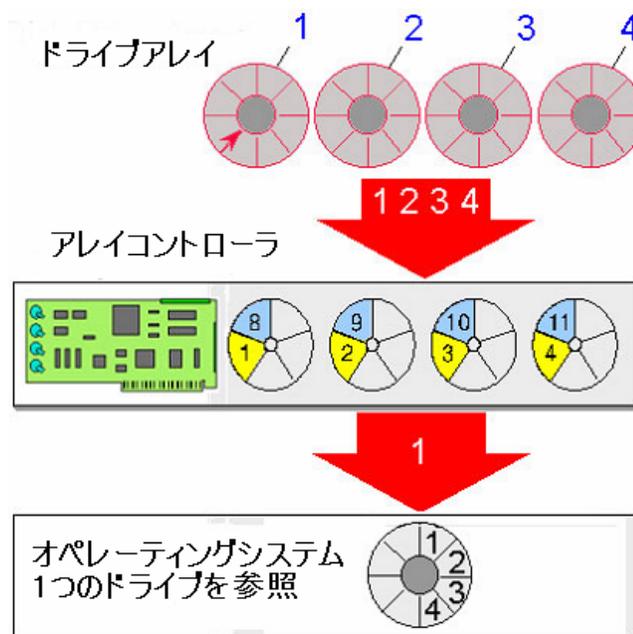
ドライブアレイテクノロジーでは、1つの物理ドライブからよりも複数のドライブからのほうがデータに高速にアクセスでき、アレイ構成のドライブが複数の要求を同時に処理できます。

1つのドライブアレイの中で、ドライブのサイズが異なってもかまいません。すべてのディスクは、最も小さいディスクの容量に合わせてサイズを調整します。すべての大容量ドライブの余った容量の部分にはアクセスできません。

注記

サイズ、速度、SCSIプロトコルが異なるドライブを1つのアレイに混在させることができる場合があります。構成によっては、容量が無駄になるだけでなく、パフォーマンスが低下するため、なるべく避けてください。

ドライブアレイ機能



ドライブアレイには、次の機能があります。

- 複数のドライブにわたるデータストライプ — ファイルがあらかじめ選択した数のセクタに分割された後、一連の複数のドライブにわたって書き込まれます。複数のドライブに分割保存されたファイルを読み書きするプロセスは、1つのドライブで同じプロセスを行うよりも、はるかに高速です。
- 複数チャンネル — ドライブアレイは、複数のチャンネルを同時に使うので(使用するコントローラに依存します)、パフォーマンスが向上します。
- 要求の処理 — 複数のチャンネルにわたって複数のコマンドが発行されるので、すべてのコマンドが同時に処理されます。要求は、タグ付きコマンドキューイング (TCQ) を使って、論理的に最適化された順序で処理されます。

タグ付きコマンドキューイング (TCQ)

TCQ を使用すると、デバイスはホスト I/O コントローラから複数のコマンドを受け付け、ディスクドライブ上で最も効率的な順序に並べ替えることができます。ホスト I/O コントローラとデバイスは、番号付きタグによって I/O 要求のトラッキングを行います。

TCQ では I/O トラフィックがランダムにアクセスされ、パフォーマンスが向上します。データに順次アクセスする場合は、暗黙のソートがすでに適用されているので、パフォーマンスへの影響が最小限になります。

RAID の実装方法

HP は、次の 2 つの異なったレベルの RAID テクノロジーをサポートしています。

- **ソフトウェアベースの RAID** — アレイ管理機能は、オペレーティングシステムなど、ホスト環境で動作するソフトウェアによって実装されます。
- **ハードウェアベースの RAID** — アレイ管理機能は、RAID コントローラ内のハードウェアとファームウェアによって実装されます。RAID コントローラは、1 つのディスクドライブ用コントローラと同じ方法で読み書きアクティビティを管理し、アレイを 1 つのまたは仮想ドライブであるかのように扱います。このレベルは、ホストソフトウェアに対して透過で、ストレージソリューションにすべて内蔵されています。この機能によってホストの主要機能が影響を受けることはないので、全体のパフォーマンスが向上し、RAID 能力の改善の可能性が広がります。

▲ 注記

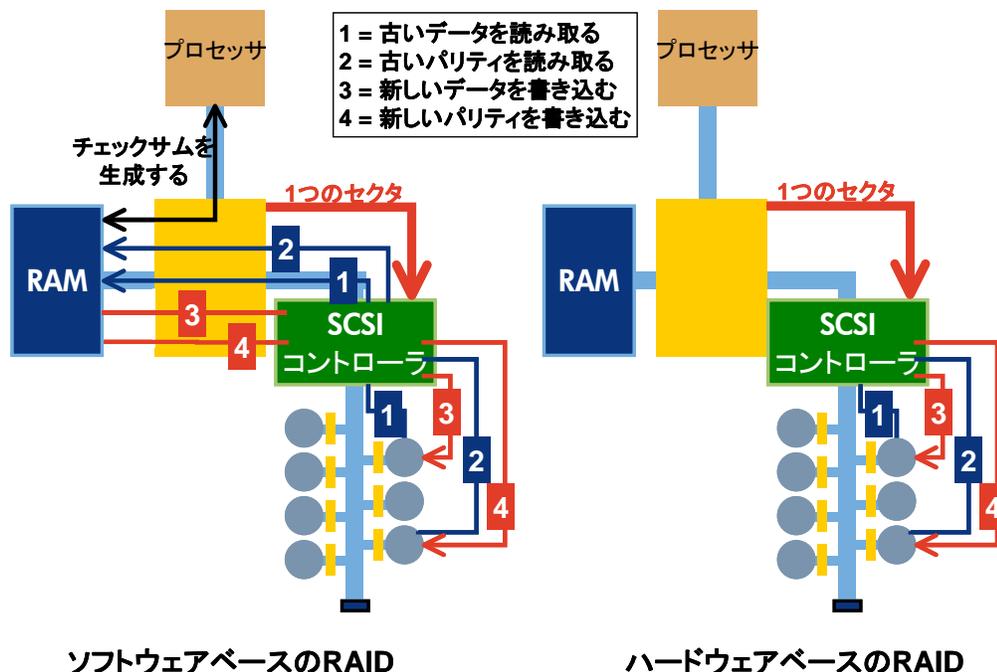
ハードウェアベースの RAID では、オンラインスペア機能が利用できます。オンラインスペアとは、RAID セットのドライブに障害が発生したときに、自動的にオンラインになる予備のドライブです。

ソフトウェアベースの RAID の長所は、次のとおりです。

- 複数のストレージサブシステムや複数のホストにまでわたって機能できる
- 低価格

ソフトウェアベースの RAID の短所は、次のとおりです。

- ハードウェアベースの RAID よりもパフォーマンスが低い
- RAID 5 の保護アレイからのブート機能が使えない (Windows の場合)
- RAID 1+0 と RAID ADG がサポートされない
- NetWare で RAID5 がサポートされない
- RAID 5 の保護アレイに対するホットプラグがサポートされない (Windows の場合)
- オンライン スペアドライブがサポートされない
- プロセッサの負荷が大きい
- 再構築の開始をオペレーティングシステムに指示する必要がある



ソフトウェア RAID ソリューションでは、1 セクタあたり余分の 3 セクタを I/O バス経由で転送する必要があります。メモリバスとプロセッサバスもまた、パリティ生成を余分に実行する必要があります。

ドライブが障害から復旧するとき、ハードウェア保護された RAID 5 論理ドライブは、ソフトウェア保護されたドライブの何倍もの速度で再構築できます。これは、障害が発生したドライブのデータを残りのドライブから再構築する際のオーバーヘッドが異なるからです。

ハードウェア RAID ソリューションでは、データがアレイコントローラに到達した後、I/O バス、メモリバス、フロントサイドバスのいずれでも余分のアクティビティは必要ありません。すべての RAID 関連のアクティビティは、アレイの SCSI バス内で行われます。

HP ハードウェア RAID の長所は、次のとおりです。

- アレイアクセラレータによる安全な書き込みキャッシング
- 異なる RAID レベルでの複数のボリュームなどの優れた管理可能性
- I/O を複数のディスクボリュームに分割できるなどの高い柔軟性
- 自動復旧、ホットプラグ対応、オンラインスペア
- 複数のディスクドライブ間での、自動データ配信と I/O 負荷分散
- ソフトウェア RAID より高性能で、I/O の最適化にも優れている

平均故障間隔

平均故障間隔 (MTBF) は、通常の消耗が原因でハードウェア構成要素に障害が発生するまでの予測時間です。ディスクドライブの MTBF は、ボリューム製造過程で多数のドライブをテストすることによって得られた、理論的な定常故障率から算出されます。

具体的な個別のドライブでは、MTBF はドライブの使用状況と環境条件によって異なります。通常の使用状況を上回る負荷をかけた場合は、そのドライブの予測 MTBF を大きく下回ることがあります。

アレイの MTBF は、個別のドライブの MTBF をアレイのドライブ数で割った値になります。障害が発生する可能性がある物理スピンドルの数が大きくなると、アレイの MTBF は小さくなります。このような場合は、フォールトトレランスをサポートする RAID 構成を使用することをお勧めします。

例

1 つのドライブの MTBF が 200,000 時間で、同様のドライブが 5 つあるアレイの MTBF は、200,000 を 5 で割って求められ、アレイ全体の MTBF は 40,000 となります。

RAID 構成でドライブ数が増加すると、アレイの効率は向上しますが、ドライブメーカーが予測する各ドライブの MTBF は変わらないことに注意してください。

MTBF の最大化

ドライブアレイを使うことでドライブ数が増加すると、ドライブに障害が発生する確率が高くなりますが、これを相殺するために、どのフォールトトレランス RAID レベルにも、オンラインスペアを追加できます。オンラインスペアを追加すると、論理ドライブの障害発生確率が、追加前のレベルの約 1000 分の 1 にまで減少します。

加えて、ProLiant サーバは、広範な障害管理機能を備えています。ProLiant サーバでは、障害処理、障害予防、フォールトトレランス、および迅速な復旧のための独自の高可用性テクノロジーを採用しています。

RAID レベル

業界標準の RAID レベルを次の表で説明します。



注記

RAID レベル 2 と 3 は、現在、業界では使われていません。

レベル	説明
RAID 0	パリティなしのデータストライプ
RAID 1	ディスクミラーリング
RAID 1	デュプレキシング(二重化)
RAID 2	複合エラー修正
RAID 3	パラレル転送、パリティドライブ
RAID 4	同時アクセス、専用パリティドライブ(分散データガーディング)
RAID 5	同時アクセス、分散パリティ(分散データガーディング)
RAID 1+0	パリティなしのディスクミラーリングとデータストライプ
RAID ADG	2 セットのパリティ付き分散データガーディング



注記

HP は、RAID をハードウェアレベルで実装しています。ハードウェア RAID は、一般に、ソフトウェア RAID よりも信頼性が高く、パフォーマンス特性も優れています。一部のオペレーティングシステムは、オペレーティングシステムレベルで、RAID 構成を実装しています。ソフトウェアで RAID を実装すると、プロセッサに余分のオーバーヘッドが必要になるので、ハードウェアによる実装方法に比べて効率が低下します。

ストライピングファクタ

ストライプは、複数の物理ドライブを1つの論理ドライブに結合します。論理ドライブは、データのブロックが、論理アレイ内のすべての物理ドライブに交互に書き込まれるように構成します。ブロックあたりのセクタ数を、ストライピングファクタと呼びます。

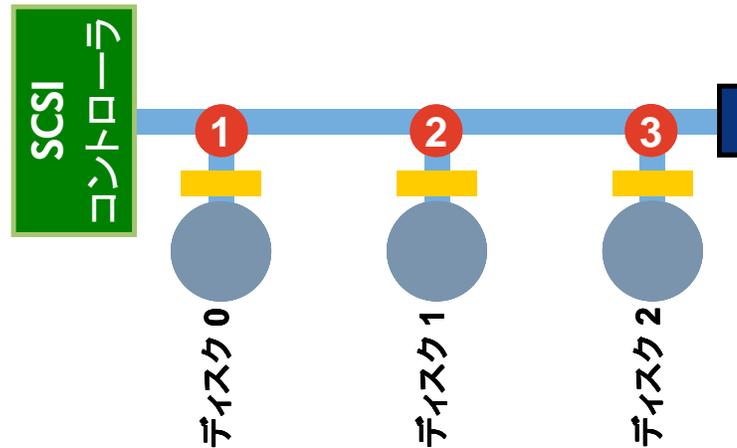
使用しているアレイコントローラによって方法は異なりますが、ストライピングファクタは変更可能です。通常は、メーカー製のシステム構成ユーティリティで変更します。ほとんどの Smart アレイコントローラは、アレイコンフィギュレーションユーティリティ(ACU)を使って変更できます。



警告

Smart アレイ 3100ES 以前に発売された Smart アレイコントローラでは、ストライプサイズを変更すると、データが消去されます。また、論理ボリュームのジオメトリ(ストライピングファクタ、ボリュームサイズ、RAID レベルなど)は、どれを変更しても、データが消去される可能性があります。

RAID 0 — ディスクストライプ



RAID 0はフォールトトレランスではなく、多くはアップタイムよりパフォーマンスと容量の方が重視される、ミッションクリティカルでない状況で使われます。RAID 0は、Smart アレイコントローラがサポートする唯一のフォールトトレランスでない RAID レベルです。

RAID 0では、アレイのすべてのドライブにまたがって、データがストライプ、つまり分散されるので、ストライプセットと呼ばれます。それぞれのドライブに容量一杯まで順次に書き込んでいくのではなく、ブロックまたはチャンクと呼ばれる少量のデータのみを1つのドライブに保存した後、次のドライブに進んで同様に保存します。各ブロックのサイズ(チャンクサイズ)をストライピングファクタと呼び、各物理ドライブの同じ位置にあるブロックの集合をストライプと呼びます。

データはストライプ(またはチャンク)に分割された後、複数のディスクにまたがって書き込まれるので、ディスク待ち時間(ターゲットセクタがディスクヘッドの下に移動するのに要する時間)が大幅に改善されます。また、ディスクスペースを100%データの格納に使えるので、全体のディスクパフォーマンスが向上します。

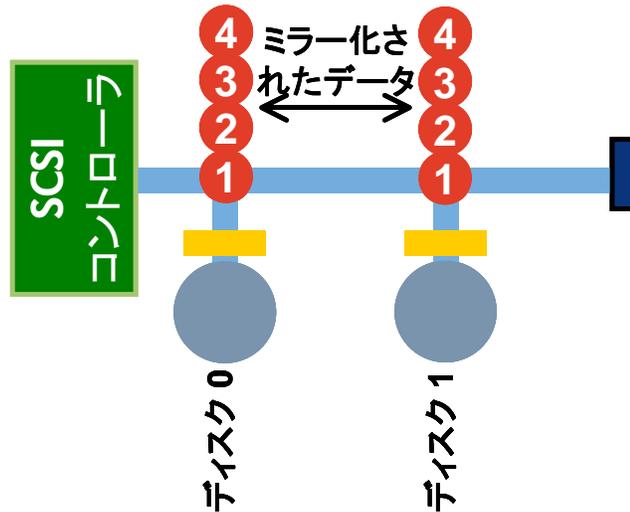
定義上、本当の意味でのストライプセットを構成するためには、RAID 0に2個以上のドライブが必要です。しかし、アレイコントローラでは、1つのドライブでRAID 0の論理ボリュームを作成できます。

RAID 0には、情報の二重化に関するオーバーヘッドがないので、最高のパフォーマンスを実現することができます。読み取り要求と書き込み要求は、すべてのメンバディスクを同時に使います。

RAID 0の制限事項

- データストライプは、1つのディスクへの通常のファイルの書き込みよりも高速であるが、ドライブに障害が発生した場合に、障害保護機能がない
- アレイにドライブを追加した場合、論理ドライブの障害発生率が高くなる

RAID 1 — ディスクミラーリング



RAID 1 は、次のような小さいデータセットに対してパフォーマンスとフォールトトレランスが要求されるミッションクリティカル環境で使用されます。

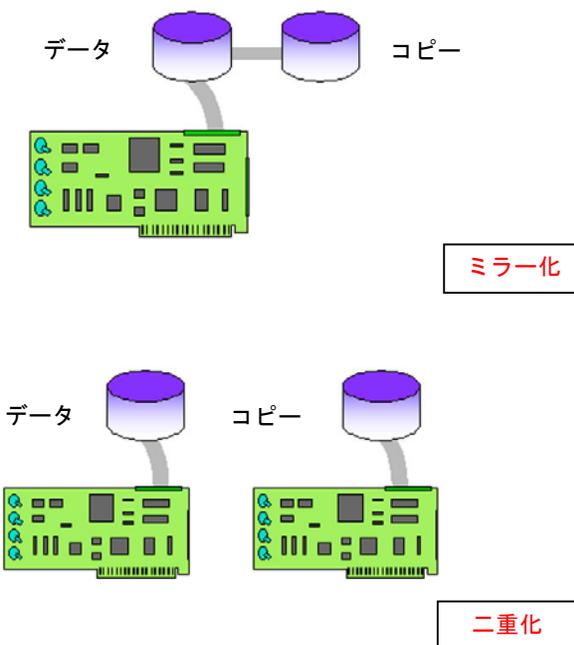
- システムディスク
- ルートマスタファイル
- データベースジャーナル

ディスクミラーリング

ディスクミラーリングには、同じサイズのディスクドライブを 2 つ使います。データが 1 つのドライブに書き込まれ、同一のコピーが 2 番目のディスクに書き込まれます。1 つのディスクに障害が発生した場合、ミラー化ドライブによって、データの損失がなく、読み書き操作が引き続き処理されることが保証されます。

すべての Smart アレイコントローラは、ハードウェアベースのミラー化をサポートします。ソフトウェアベースのミラー化では、同じサイズのドライブが必要ありませんが、オペレーティングシステムのサポートが必要です。データは、均等なサイズのパーティションに書き込まれます。

デュプレキシング(二重化)



デュプレキシング(二重化)は、ミラー化と同じような機能を持ちます。ただし、デュプレキシング(二重化)では、ディスクコントローラを1つではなく2つ使い、これによってフォールトトレランスの向上を図ります。コントローラに障害が発生しても、論理ドライブが利用できます。

▲ 注記

デュプレキシング(二重化)は、オペレーティングシステムが処理する必要があります。

RAID 1 のパフォーマンス

マルチディスク構成では、RAID 1 によって、ディスクのペアが互いにミラー化されます。これらのディスクペアは、ストライプされて、仮想ディスクが作成されます。ミラー化ペアを形成する 2 つのディスクの両方に障害が発生しない限り、RAID 1 は、複数のディスクに障害が発生しても機能し続けられます。

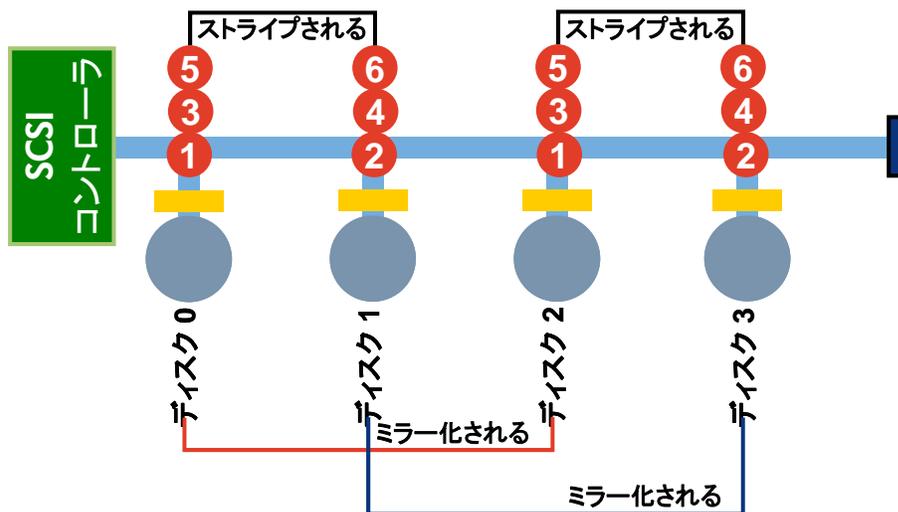
特にアレイコントローラがスプリットシーク(シーク時間が少ない順に、一つのディスクを選んで読み取る)を活用する場合、読み取り環境で高いパフォーマンスを示します。RAID 1 は、書き込み要求を二重化するために、書き込みではわずかにパフォーマンスが低下します。

▲ 注記

HP によるドライブのミラー化は、ハードウェアによって実装されています。ドライブのミラー化は、オペレーティングシステムレベルで、ソフトウェアによって実装することもできます。ソフトウェアによるミラー化は、プロセッサで余分のオーバーヘッドが必要になるので、多くの場合、ハードウェアによるミラー化に比べて効率が低下します。

RAID 1 の制限事項は、データを保存するためにドライブストレージが 2 倍必要になるので、コストの高いソリューションだという点です。データストレージとして利用できるのは、総ディスクスペースの 50%のみです。

RAID 1+0 — ストライプミラーセット



RAID の最も高価な構成は、RAID 1+0 で RAID 1 のミラー化と RAID 0 のストライピングを組み合わせた構成です。ドライブはまずミラー化され、次にメンバディスク間でストライプされます。RAID 1+0 は、高性能で、書き込みの際の機能低下がなく、優れた冗長性を備えています。

このアレイ構成では、同じミラー化ペアに属さない複数のディスクを失ってもかまいません。同じミラー化ペアの 2 つのディスクの両方に障害が発生した場合は、データが失われます。つまり、RAID 1+0 は、2 つのディスクの障害に対する保護は保証できません。

RAID 1+0 プロセスでは、次の処理が実行されます。

1. ディスク 0 がディスク 2 にミラー化される
2. ディスク 1 がディスク 3 にミラー化される
3. ディスク 0 とディスク 1 が同時にストライプされる
4. ディスク 2 とディスク 3 が同時にストライプされる

RAID 1+0 構成では、すべての HP RAID コントローラに次の特徴があります。

- 複数のドライブの障害からデータを保護できる
- ドライブが複数のバスにまたがって等しく分散されている場合は、1つのバス全体の障害からデータを保護できる
- 機能低下した状態でも、I/O 要求は動作を続けているすべてのドライブで処理される
- 各ミラー化ペアのどちらかのメンバが動作している限り、 $n/2$ 個のドライブ (n はアレイにあるドライブ数) の障害からデータを保護できる

▲ 注記

RAID 1+0 は、複数のドライブに障害が発生しても、それらが同じミラー化ペアに属さない限り、それに対応できます。

RAID 1+0 のパフォーマンス

RAID 1+0 のパフォーマンス特性は RAID 1 と同じです。RAID 1+0 は、障害が発生した 2 つのディスクが互いにミラー化されていない限り、複数のディスク障害からデータを保護できます。

読み取りのパフォーマンス

RAID 1+0 の読み取り操作では、ディスクコントローラが使えるのはディスクドライブの半分だけで、後の半分には同じ情報が保存されます。これは RAID 0 に比べて、わずかですがパフォーマンス上、マイナスのインパクトがあります。読み取り要求の処理に利用できるディスクドライブ(スピンドル)数が 50% 低下します。

読み取りパフォーマンスに対するこのマイナスのインパクトは、ドライブからデータを読み取るプロセスで、読み取り/書き込みヘッドに最も近い要求データを読み取ることによって相殺されます。

書き込みのパフォーマンス

書き込み操作については、ディスクコントローラが、同じ書き込み要求を 2 回生成する必要があります。同じデータをミラー化ディスクドライブのセットの両方に書き込む必要があるため、コントローラでは中程度のオーバーヘッドが発生します。アレイコントローラでは、書き込み操作の回数が 2 倍になるため、相当量のオーバーヘッドが発生します。

RAID 2 — ECC によるデータセクタ ストライピング

RAID 2 アレイは、ドライブグループにわたってデータをストライプしますが、いくつかのドライブはエラーのチェックと修正 (ECC) 情報の保存に使われます。現在ではほとんどのハードドライブが、ECC 情報を各セクタに保存しているため、RAID 2 は RAID 3 に比べて特別な利点はありません。

▲ 注記

RAID 2 は今では古いテクノロジーであり、Smart アレイコントローラではサポートしていません。

RAID 3 — パリティチェック

RAID 3 は、基本的な構成は RAID 4 と同じですが、ビット(またはバイト単位)でデータを分割し保存します。各メンバにあるデータを基に算出されたパリティが、パリティディスクに保存されます。

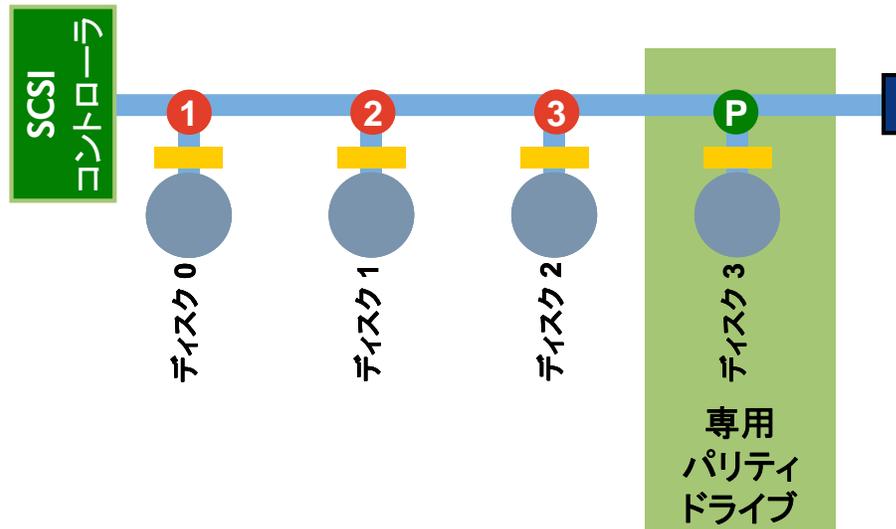
RAID 3 は、大量の連続した I/O を比較的 low コストで処理します。大量の I/O は通常、大量の差分を保存する場合に要求され、RAID 3 のパフォーマンスは連続した I/O のときに最大になります。RAID 3 は、スーパーコンピューティングの場面で、よく使われてきました。

しかし、RAID 3 は、多くのデータベース管理システムのワークロードのようなランダムアクセス I/O には適していません。RAID 3 ストライプの物理的な組織構造では、RAID 3 ボリュームへの書き込みを行うたびにすべてのディスクにアクセスしなければならないため、RAID 3 でのランダムアクセスは効率が大きく低下します。また RAID 3 ではパリティディスクは 1 つしか使わないため、これがボリューム全体のボトルネックになっています。

▲ 注記

Smart アレイコントローラは、RAID 3 をサポートしていません。

RAID 4 — データガーディング



RAID 4 は、データストライプと専用のパリティドライブを組み合わせて使います。RAID 4 のディスクセットには、3 つ以上のディスクが必要です。

1 つのドライブに障害が発生した場合、障害があるディスクのデータは、アレイの他のドライブから再構築されます。チェックサム の保存にはドライブを 1 つしか使わないので、利用できるディスクスペースは、ミラー化よりも大きくなります。すべてのドライブは個別にアクセスできるので、論理ドライブのディスクが増えれば、パフォーマンスが向上します。また複数の操作（読み取りと書き込みが混在した操作も含む）を同時に実行できます。

▲ 注記

RAID 4 ドライブは、同じバスを共有できます。

上図は、複数のディスクにストライプされたファイルと専用のドライブに書き込まれたパリティサムを表しています。RAID 4 は、各データストライプについてのパリティ情報を 1 つのディスクに保存します。RAID 4 がデータを保護できるのは、1 つのディスクで障害が発生した場合のみです。

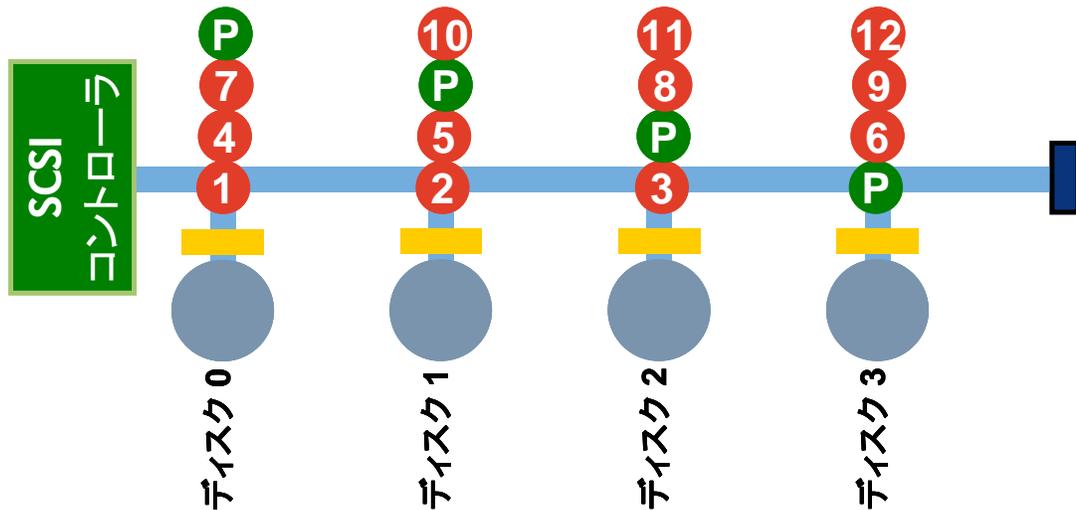
▲ 注記

RAID 4 では、1 つのドライブの障害からデータを失わずに復旧できるので、フォールトトレランスです。RAID メモリは、RAID 4 で実現されます。

制限事項

書き込みが大量に発生する環境では、パリティアップデート操作が、1 つのディスクに対して順次実行されます。すべての書き込み操作でデータをパリティディスクに書き込む必要があるため、このディスクがパフォーマンスのボトルネックとなる可能性があります。このため、RAID 4 は急速に使用されなくなりました。

RAID 5 — 分散データガーディング



RAID 5 の分散データガーディングは、パリティ付きストライプセットとも呼ばれ、データをチャンクに分割し、パリティを計算し、それらのデータチャンクをディスクドライブに「ストライプ状」に書き込み、ドライブごとに 1 つのストライプをパリティデータ用に確保します。冗長性（パリティデータ）のために使うディスクスペースの合計量は、1 つのドライブの容量に等しくなります。

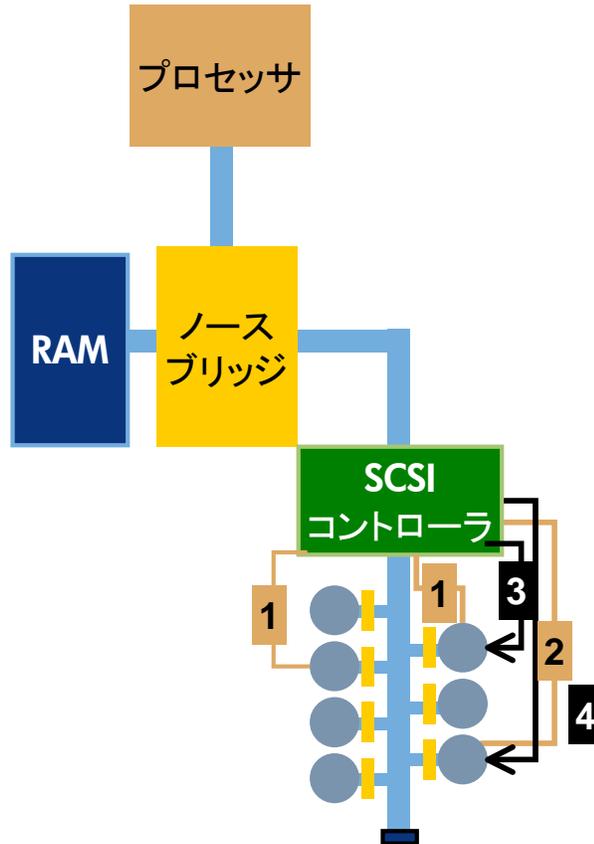
RAID 5 は、パリティ情報をすべてのディスクメンバに分散して保存します。これにより、パリティディスクでのパフォーマンスのボトルネックが解消されるとともに、1 つのドライブに障害が発生した場合には、障害のあるディスクを交換すれば、データをすべて復旧できます。

RAID 5 は、書き込み操作の実行時に新しいチェックサムを生成するという RAID 4 の原則を基にして構築されています。RAID 4 とは異なり、パリティデータを専用のドライブに保存するのではなく、すべてのドライブに均等に分散します。

RAID 5 には、最低 3 つのドライブが必要です。利用可能なディスクスペースは、すべてのディスクのサイズ合計から 1 つのディスクのサイズを引いた値です。たとえば、9GB のディスクが 5 個ある場合、論理ドライブのサイズは、 $(5-1) \times 9\text{GB}$ で 36GB となります。

RAID 5 の短所は、同時に 2 つのドライブに障害が発生するとデータが損失する点です。最初に障害が発生したドライブの交換と復旧が完了する前に 2 つ目のドライブに障害が発生すると、データが失われます。

RAID 5 のパフォーマンス



RAID 5 は、フォールトレランス RAID ソリューションの中で、コストパフォーマンスは最も高くなりますが、RAID 1+0 よりもパフォーマンスが低く、メンバディスクに障害が発生した場合、さらにパフォーマンスが低下します。

RAID 5 は、I/O における読み取り要求の方が多い環境で、性能を発揮します。読み取り操作が並行して実行される上、パリティが分散されているので、すべてのドライブを読み取り操作に利用できます。RAID 5 は、データベースランザクシオン処理やマルチユーザファイルサービス環境でよく使われます。

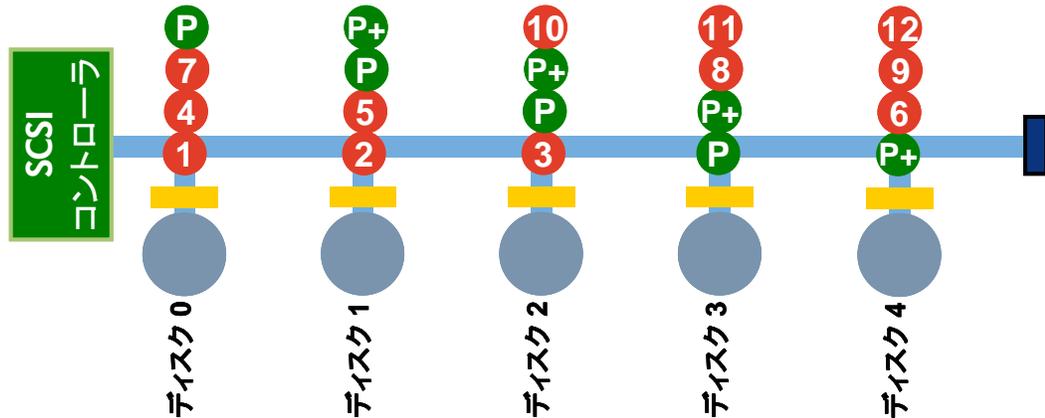
書き込み操作のたびに、アレイコントローラは平均 4 回の I/O 操作を実行する必要があるため、書き込みオーバーヘッドの負荷によって、システム全体のパフォーマンスが低下します。複数の書き込み操作では、インタリーブパリティにより、パリティアップデートが並行して実行できます。

4 回の I/O 操作とは、次の 4 つです。

1. コントローラが古いデータブロック(上書きされるデータブロック)を読み取る
2. コントローラが古いパリティブロックを読み取る(このパリティブロックは、新しいデータを反映してアップデートされる)
3. コントローラが新しいパリティ情報を再計算して書き込む
4. コントローラが新しいデータブロックを書き込む

RAID 1 のミラー化では、データ保護のために容量を 100% 追加する必要がありましたが、RAID 5 では、50% (2 + 1) に過ぎず、一般には 20% (5 + 1) の容量を追加するだけで済みます。RAID 5 のセットは、容量が大きくなるほど管理の負荷が大きくなり、またパフォーマンスが低下しやすくなります。

RAID ADG — アドバンスド データガーディング



RAID アドバンスド データガーディング (ADG) は、RAID 1 に近い高度のフォールトトレランスを備えながら、RAID 5 と同様に高い容量使用率を維持します。2 個のハードドライブに同時に障害が発生した場合も、データの損失やダウンタイムなしに動作を維持できる能力を持ち、複数のドライブの障害からデータを保護します。

RAID ADG では、複数のディスクにわたってストライプされた 2 セットのパリティを使います。この方法により、56 個のドライブからなるアレイを保護しながら、パリティ情報の保存に必要な容量はドライブ 2 個分にとどまります。

RAID ADG のパフォーマンス

ADG は、2 セットの独立したパリティ方式を採用しています。2 セットのパリティは、論理書き込み操作のたびに変更が必要で、次の 6 回の物理的転送が必要になります。

- 3 回の読み取り
- 3 回の書き込み

既存のデータと 2 セットのパリティが物理ドライブから読み込まれた後、既存のデータ、書き込みデータ、古いパリティを操作することによってパリティの各セットが変更されます。次に、新しいデータと変更されたパリティデータの両方のセットが物理ドライブに書き込まれます。

RAID ADG にはフォールトトレランスと大容量の 2 つの利点がありますが、これは他の RAID レベルを下回るパフォーマンスの犠牲の上に成り立っています。

読み取りパフォーマンスは RAID 5 と同じですが、パリティデータが多いので、書き込みパフォーマンスは低下します。

RAID ADG の用途

RAID ADG は、RAID 1+0 に近い高度のフォールトトレランスを備えながら、RAID 5 と同様に、高い容量使用率を維持します。保護レベルが高いため、大容量の論理ボリュームが必要な場合には理想的です。

ADG には次の利点があります。

- 最高で合計 56 のドライブを安全に保護できる。またパリティドライブ数が増加するのでフォールトトレランスが高まる
- 2 つのドライブに同時に障害が発生しても、ダウンタイムやデータ損失なしに動作を維持できる
- RAID 4、RAID 1+0、RAID 5 よりもフォールトトレランスが高い
- RAID 1 や RAID 1+0 よりも、実装コストが小さい
- オンラインスペアドライブをサポートしている
- RAID 1、RAID 1+0、RAID 5 のいずれかからのオンライン RAID レベルの移行をサポートしている
- 多くの物理ドライブからなる大容量の論理ドライブが必要なアプリケーションに理想的

RAID ADG は、次の基準を満たす場合に最適です。

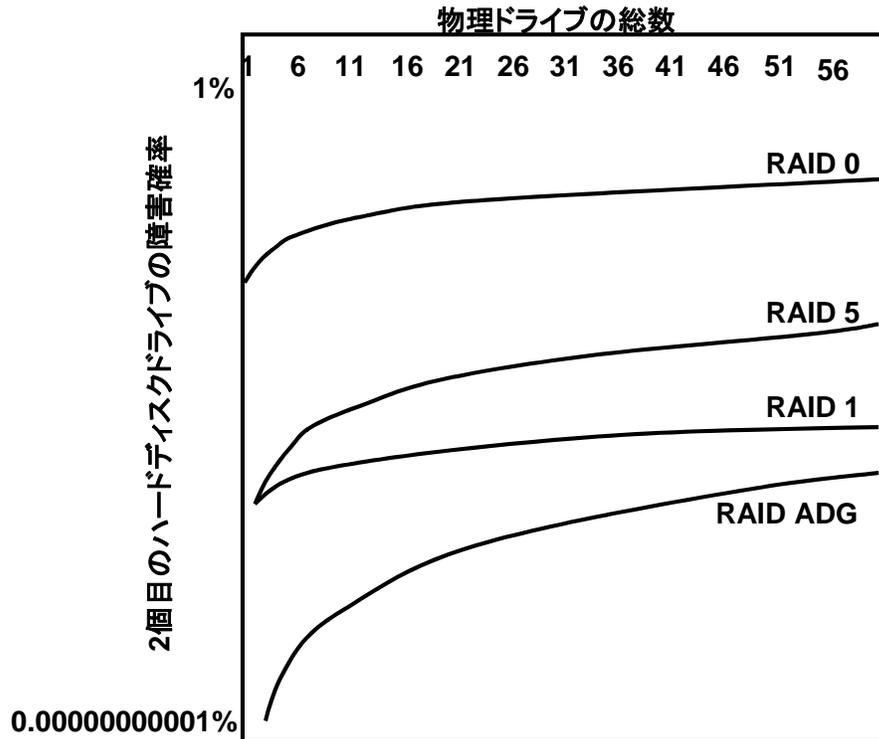
- ミッションクリティカル データ
- 容量使用率を高める必要がある場合
- 大容量ボリュームを必要とするアプリケーション
- RAID 1 や RAID 1+0 よりも高い容量使用率が必要な場合
- RAID 1、RAID 1+0、RAID 5 よりも高いフォールトトレランスが必要な場合

Smart アレイコントローラの ADG サポート

ProLiant サーバで ADG をサポートするには、次の基準を参考にしてください。

- Smart アレイ 5304/128、または 6402 コントローラでの最低 4 つのハードドライブの使用
- Smart アレイ 5302 コントローラでの最低 64MB キャッシュと ADG アップグレードキットの使用、かつ最低 4 つのハードドライブの使用
- ディスクが 9 個以上の場合には、RAID ADG または RAID 1+0 を推奨。ディスクが 8 個以下の場合には、RAID 5 を推奨

RAID ADG のフォールトトレランスの比較



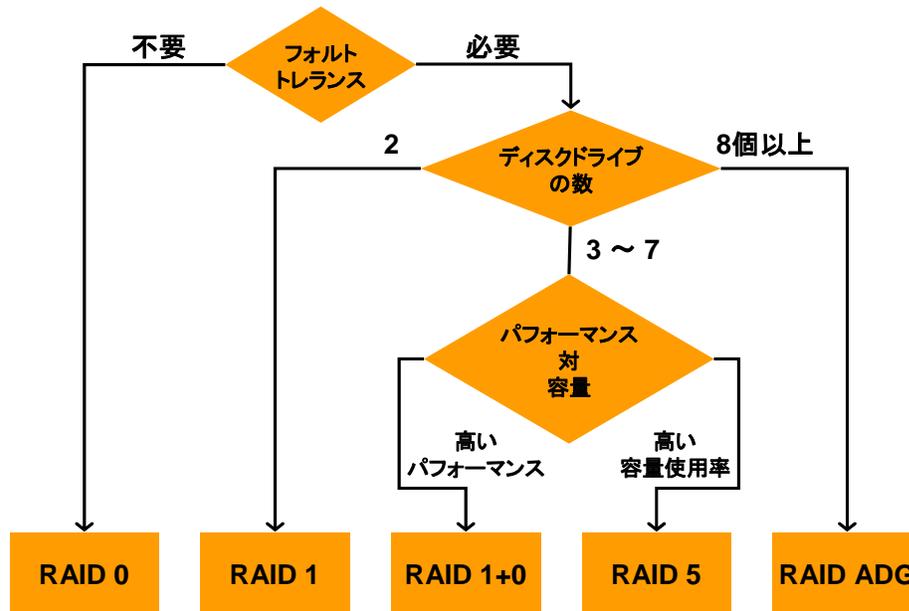
上の図は、オンラインスペア機能がないと想定した場合の、論理ドライブの相対的な障害確率を、RAID レベルと論理ドライブサイズごとに示したものです。

1 つのコントローラ上の 1 つの論理ボリュームに必要な個別ディスクドライブ数が増えても、RAID ADG は RAID 5 や RAID 1 に比べて、データの高信頼性と優れたフォールトトレランスを提供します。

- RAID 0 では、1 つの物理ドライブに障害が発生すると、論理ドライブ全体の障害となる
- RAID 5 では、2 つの物理ドライブに障害が発生すると、論理ドライブ全体の障害となる
- RAID 1 または 1+0 では、 $n/2$ 個までのハードドライブに障害が発生しても論理ドライブの障害につながらない。ただし、RAID 1+0 論理ドライブで、2 つのハードドライブに障害が発生し、それらが互いにミラー化されている場合は、論理ドライブの障害となる
- RAID ADG では、3 つのハードドライブに障害が発生すると、データ損失が生じる

RAID 5 については、1 つの論理ドライブあたり 15 個以上の物理ドライブを使わないことをお勧めします(最適数は 8)。ただし、RAID ADG は論理ドライブ障害の可能性がはるかに低いので、このフォールトトレランス方式が動作するシステムにおいて、Smart アレイコントローラはドライブアレイあたり最高 56 個の物理ドライブの使用をサポートします。

RAID レベルのパフォーマンスの比較



RAIDレベルの決定ツリー

RAID 0

RAID 0は、全体のパフォーマンスが一番高くなります。しかし、アレイにある1つのドライブに障害が生じただけで、すべてのデータが損失します。フォールトトレランス用のストレージ空間は必要ありません。RAID 0は、サーバでの使用は奨められませんが、高速ディスクアクセスが必要なワークステーションには適しています。データは、できるだけ頻繁にバックアップする必要があります。

RAID 1 および RAID 1+0

RAID 1 または RAID 1+0 は、3 個以上のディスクをアレイ構成にした場合、複数ドライブの障害からデータを保護します。すべての RAID レベルの中で、フォールトトレランス用の容量が最も大きくなります。RAID 1 または RAID 1+0 は、RAID 5 より高い書き込みパフォーマンスを発揮します。

RAID 4

RAID 4 は奨められません。RAID 5 では、利用できるスペースとフォールトトレランスが同じでありながら、読み書きが高速化されているからです。RAID 4 は、一部の Smart アレイコントローラでのみサポートされています。

RAID 5

RAID 5 は、1 個のドライブ障害に対処でき、使われるフォールトトレランス用のストレージ容量が最小です。読み取りパフォーマンスは、RAID 0 や RAID 1 とほぼ同等ですが、書き込みパフォーマンスは、通常、RAID 1 の書き込みパフォーマンスを下回ります。

ADG

RAID ADG は、RAID 1 に近い高度のフォールトトレランスを備えながら、RAID 5 と同様に、高い容量使用率を維持します。RAID ADG は、複数のドライブ障害からデータを保護しますが、パリティ情報の保存用にドライブ 2 個分の容量のみを必要とします。保護レベルが高いため、多数の物理ドライブを使った大容量の論理ボリュームが必要な場合には理想的です。

RAID での障害ドライブの復旧時間

ドライブに障害が発生した場合のデータ復旧時間は、以下の要因に左右されます。

- ドライブのタイプとサイズ
- RAIDレベル
- システムのワークロード
- コントローラのタイプ
- アレイアクセラレータの設定
- アレイ復旧作業の優先順位レベル

システムを使用中にドライブを再構築する場合は、復旧時間が、アクティビティのレベルに大きく依存することがあります。ほとんどのシステム、特に、RAID 1は、負荷のない中程度のアクティビティとほぼ同じ時間で復旧します。RAID 5は、障害のあるシステムの I/O 要求がかなりの量にのぼるので、復旧の際のシステム負荷に大きく影響されます。



重要

ディスク障害によるダウンタイムが一切許されないミッションクリティカルなアプリケーションを実行している場合は、Smart アレイコントローラによってサポートされている**オンラインスペアドライブ**の使用を検討してください。オンラインスペアは、ドライブ障害が発生したときにコントローラが使用するドライブです。ドライブに障害が発生した場合、障害ドライブ上のデータは、コントローラがオンラインスペア上に再構築します。コントローラはまた、通常であれば障害が発生したドライブに保存されるはずのデータを、直接オンラインスペアに送信します。

ストライプサイズの最適化

適切なストライプ (チャンク) サイズの選択は、アレイ内のパフォーマンスを最適にするために重要です。ストライプサイズは、アレイコントローラがデータ要求を処理する際に、アレイ内の各ディスクで読み書きされるデータ量です。

注記

チャンク、ブロック、セグメントの 3 つの用語は、どれも同じ意味です。「チャンク」は、ストレージの説明でよく使われます。

次の表では、利用可能なストライプサイズとその特性について説明します。

フォールトトレランス方式	利用可能なストライプサイズ (KB)	デフォルトサイズ (KB)
RAID 0	128, 256	128
RAID 1 または 1+0	8, 16, 32, 64, 128, 256	128
RAID 5 または RAID ADG	8, 16, 32, 64	16

デフォルトのストライプサイズは、ほとんどの環境で良好なパフォーマンスを発揮します。高性能が重要視される場合は、ストライプサイズを変更する必要があるでしょう。

ストライプサイズによって、次のような影響が出ることがあります。

- 大きすぎると、ドライブ間での負荷分散機能が低下する
- 小さすぎると、ストライプ間転送 (スプリット I/O) が多くなり、パフォーマンスが低下する

スプリット I/O は 2 つのディスクで行われ、両方のディスクがシーク、回転、データを転送します。応答時間は、遅い方のディスクに依存します。スプリット I/O では、少ないドライブ数で着信要求を処理するので、要求レートが低下します。

サーバアプリケーションのタイプ	推奨されるストライプサイズの変更
読み書きが混在	デフォルト値をそのまま使用
主に読み取り (データベースやインターネットアプリケーションなど)	ストライプサイズを拡大
主に書き込み (画像処理アプリケーションなど)	RAID 5、RAID ADG の場合は、ストライプサイズを縮小 RAID 0、RAID 1、RAID 1+0 の場合は、ストライプサイズを拡大

Smart アレイコントローラ

SMART-2ファミリ



SMART-2SL
Wide Ultra



SMART-2DH
Wide Ultra

エントリーレベル



**Smartアレイ
221**
Wide Ultra2



**シングルチャネル
RAID
(RAID LC2)**
Ultra3



**Smartアレイ
431**
Ultra3



**Smartアレイ
5i**
Ultra3



**Smartアレイ
641**
Ultra320



**Smartアレイ
642**
Ultra320

ハイエンド



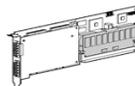
**Smartアレイ
3200**
Wide Ultra2



**Smartアレイ
4200**
Wide Ultra2



**Smartアレイ
5302**
Ultra3



**Smartアレイ
5304**
Ultra3



**Smartアレイ
5312**
Ultra3

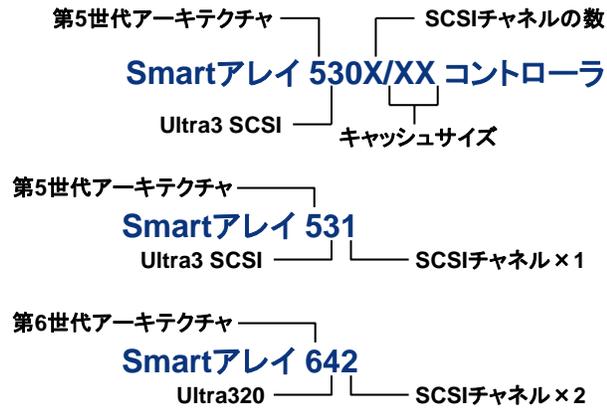


**Smartアレイ
6402**
Ultra320

HP は 1989 年以降、エンタープライズカスタマのストレージに関するニーズに対応するため、優れたドライブアレイコントローラを提供してきました。

Smart アレイコントローラは、業界標準の RAID レベルにはない独自の高可用性機能を備えています。

Smart アレイコントローラの命名方法



ハイエンドの Smart アレイコントローラには 4 桁の番号が付き、エントリーレベルのコントローラには 3 桁の番号が付きます。最初の数字は、プロダクトファミリーまたは Smart アレイコントローラファミリーにおける世代を表します。2 番目の数字は、プロトコルを表します (1=Ultra SCSI、2=Ultra2、3=Ultra3)。3 番目の数字は、他の製品との区別に使います。

例

Smart アレイ 4250ES で、3 番目の数字は、Smart アレイ 4200 以外のコントローラであることを示します。最後の数字はプレースホルダです。「ES」は、ProLiant 6000、7000、8000 サーバのケーブルフリー環境に合わせて特別設計された特別なコントローラであることを示します。

注記

SMART は SCSI Managed Array Technology の頭字語です。初期のアレイコントローラ (特に SMART-2 プロダクトライン) を指すときは、SMART を大文字で表記します。

アレイコントローラの比較

SMART-2 ファミリ

アレイ コントローラ	サポートして いる SCSI	キャッシュ	RAID レベル	最大データ 転送レート	チャンネル 数
SMART-2/E	Fast-Wide	4MB 読み取り/書き込み	0、1、4、5	20MB/s	2
SMART-2/P	Wide Ultra	4MB 読み取り/書き込み	0、1、4、5	40MB/s	2
SMART-2SL	Wide Ultra	6MB 読み取り専用	0、1、5	40MB/s	1
SMART-2DH	Wide Ultra	16MB 読み取り/書き込み	0、1、4、5	40MB/s	2

エントリレベル

アレイ コントローラ	サポートして いる SCSI	キャッシュ	RAID レベル	最大データ 転送レート	チャンネル 数
Smart アレイ 221	Wide Ultra2 Wide Ultra	6MB 読み取り専用	0、1、1+0、5	80MB/s	1
内蔵 Smart アレイ	Wide Ultra2 Wide Ultra	8MB 読み取り専用	0、1、1+0、5	80MB/s	2
シングルチャンネル (RAID LC2)	Ultra3 Wide Ultra2 Wide Ultra	8MB 読み取り専用	0、1、1+0、5	80MB/s	1
Smart アレイ 431	Ultra3 Wide Ultra2 Wide Ultra	16MB 読み取り専用	0、1、1+0、5	160MB/s	1
Smart アレイ 532	Ultra3 Wide Ultra2	32MB 読み取り専用	0、1、1+0、5	160MB/s	2
Smart アレイ 5i	Ultra3 Wide Ultra2 Wide Ultra	32MB 読み取り専用	0、1、1+0、5	160MB/s	2
Smart アレイ 5i Plus	Ultra3 Wide Ultra2 Wide Ultra	64MB 読み取り/書き込み	0、1、1+0、5	160MB/s	2
Smart アレイ 641	Ultra320	64MB 読み取り専用	0、1、1+0、5	320MB/s	1
Smart アレイ 642	Ultra320	64MB 読み取り専用	0、1、1+0、5	320MB/s	2

ハイエンド

アレイ コントローラ	サポートして いる SCSI	キャッシュ	RAID レベル	最大データ 転送レート	チャンネル 数
Smart アレイ 3100ES	Wide Ultra	56MB 読み取り/書き込み	0、1、1+0、4、5	40MB/s	3
Smart アレイ 3200	Wide Ultra2 Wide Ultra	64MB 読み取り/書き込み	0、1、1+0、4、5	80MB/s	2
Smart アレイ 4200	Wide Ultra2 Wide Ultra	64MB 読み取り/書き込み	0、1、1+0、4、5	80MB/s	4
Smart アレイ 4250ES	Wide Ultra2 Wide Ultra	56MB 読み取り/書き込み	0、1、1+0、4、5	80MB/s	3
Smart アレイ 5302	Ultra3 Wide Ultra2 Wide Ultra	32、64、128MB、または 256MB(アップグレード) 読み取り/書き込み	0、1、1+0、5、 ADG(アップグ レード)	160MB/s	2、 4(アップ グレード)
Smart アレイ 5304	Ultra3 Wide Ultra2 Wide Ultra	128、または 256MB 読み取り/書き込み	0、1、1+0、5、 ADG	160MB/s	4
Smart アレイ 5312	Ultra3 Wide Ultra2	128MB、 256MB(アップグレード) 読み取り/書き込み	0、1、1+0、5	160MB/s	2
Smart アレイ 6402	Ultra320 Ultra3 Wide Ultra2	128MB、 256MB(アップグレード) 読み取り/書き込み	0、1、1+0、5、 ADG	320MB/s	2、 4(アップ グレード)

エントリレベルの Smart アレイコントローラ

エントリレベルの Smart アレイコントローラは、ハイエンドファミリより SCSI チャンネル数が少なくなっています。また Smart アレイ 5i Plus を除くこれらのコントローラでは、バッテリーバックアップ キャッシュが省略され、書き込みキャッシュもサポートされません。

Smart アレイ 532 コントローラ

Smart アレイ 532 コントローラは、2 チャンネル、64 ビットアーキテクチャ、Ultra3 SCSI アレイコントローラであり、エントリレベルのハードウェアベースのフォールトトレランスに対応し、最高 28 個の Ultra3 または Ultra2 ハードドライブをサポートします。ProLiant サーバ内部のドライブに接続する内部 SCSI ポートが 1 つあります。2 番目の外部 SCSI ポートは、StorageWorks 4300 エンクロージャに接続して、さらに 14 個の外部ドライブを追加できます。

内蔵 Smart アレイコントローラ

内蔵 Smart アレイコントローラは、インテリジェントアレイコントローラであり、エントリレベルのハードウェアベースのフォールトトレランスに対応し、内部 Ultra2 SCSI 内部ハードドライブをサポートします。これは、ソフトウェア RAID に代わる市場で最もコストパフォーマンスの高い選択肢です。

内蔵 Smart アレイコントローラは、ProLiant DL360、DL380、DL580、8500 の密度最適化サーバ用に設計され、それらに統合されています。ProLiant ML370 および ML570 サーバも、内蔵 Smart アレイコントローラをサポートしています。内蔵 Smart アレイコントローラを使うと、全 Smart アレイコントローラファミリのどのコントローラにでもアップグレードできます。

シングルチャネル RAID (RAID LC2) コントローラ

シングルチャネル RAID (RAID LC2) コントローラは、低コストのシングルチャネル 32 ビット PCI RAID コントローラで、ハードウェア RAID が必要なエントリレベル サーバとワークグループサーバを対象にしています。シングルチャネル RAID (RAID LC2) コントローラは、ProLiant ML330 (G2)、ML350 (G2) および DL320 サーバでサポートされており、エントリレベルの RAID 機能を提供します。また Wide Ultra2 SCSI と最高 5 つの内部ディスクドライブをサポートしており、すべての Smart アレイコントローラとのデータ互換性とアップグレードパスが用意されています。

高性能な Smart アレイコントローラ

Smart アレイ 5300 コントローラには、高性能、大容量の必要性に対応するためのモデルがいくつか用意されています。

Smart アレイ 5300 コントローラ

Smart アレイ 5300 シリーズの高性能 Ultra3 アレイコントローラには、次の機能があります。

- 高度な柔軟性とデータ保護機能
- 革新的なモジュール設計
- ADG テクノロジ

モジュール方式の設計はアップグレードが容易で、パフォーマンスが最適化できるだけでなく、容量を必要に応じて 2 チャンネルから 4 チャンネルへ拡張でき、また 32、64、128、256MB のいずれかのバッテリバックアップキャッシュを搭載できます。バッテリバックアップキャッシュは、停電やサーバ障害、コントローラ障害の場合に、キャッシュデータを保護します。冗長バッテリーが、さらに保護を強化します。最高 4 チャンネルは、PCI スロットあたりの業界最高のストレージ容量です。

Smart アレイ 5300 シリーズは、高いパフォーマンスレベルに加えて、新しいメモリアーキテクチャや RAID エンジンなど、いくつかの拡張機能を備えています。

Smart アレイ 5300 コントローラには、次のモデルがあります。

- **Smart アレイ 5312 コントローラ** — 133MHz PCI-X バスパフォーマンスと 128MB のバッテリバックアップキャッシュを持ち、256MB までアップグレードが可能です。RAID ADG は、このコントローラではサポートされていません。
- **Smart アレイ 5302/32 コントローラ** — Smart アレイ 3200 コントローラの後継機種です。Smart アレイ 5302/32 には、32MB のバッテリバックアップキャッシュと、内部または外部にアクセスできる 2 つの Ultra3 SCSI チャンネルがあります。RAID ADG は、64MB 未満のキャッシュではサポートされていません。
- **Smart アレイ 5302/64 コントローラ** — キャッシュが 64MB に拡張されています。RAID ADG をサポートするようアップグレードできます。
- **Smart アレイ 5302/128 コントローラ** — キャッシュが 128MB に拡張されています。RAID ADG をサポートするようアップグレードできます。
- **Smart アレイ 5304/128 コントローラ** — Smart アレイ 4200 コントローラの後継機種です。Smart アレイ 5304/128 では、キャッシュが 128MB に拡張されており、さらに 256MB までアップグレードできます。外部アクセス可能な Ultra3 SCSI チャンネルが 2 個追加され (合計 4 個)、RAID ADG をサポートします。
- **Smart アレイ 5304/256 コントローラ** — キャッシュが 256MB に拡張されています。RAID ADG をサポートします。
- **Smart アレイ 5302/FC コントローラ** — Smart アレイ 5302/32 に SAN アクセスモジュールを追加したものです。SCSI チャンネルの追加はできません。

Smart アレイ 5300 シリーズのすべてのモデルには、次の共通機能があります。

- 4GB 超のメモリを持つサーバをサポートする 64 ビットのメモリアドレッシング
- SA5302 と SA5304 は、バーストスループットが 533MB/s の 64 ビット、66MHz PCI
- SA5312 は、バーストスループットが 1GB/s の 64 ビット、133MHz PCI-X
- ROM の障害または破損から保護するためのリカバリ ROM
- RAID 5 書き込みパフォーマンスの向上。旧バージョンの Smart アレイコントローラの 6 倍の速度
- ADG の採用により、旧バージョンの Smart アレイコントローラでの RAID 5 から、フォールトトレランスとパフォーマンスが向上
- Ultra3 と Ultra2 ハードドライブの共存と混在をパフォーマンスの低下なしにサポート。Ultra3 SCSI テクノロジーによる高性能とチャネル帯域幅あたり最高 160MB/s のデータ帯域幅
- モジュール設計による SCSI チャネル数の増加とキャッシュサイズのアップグレード

Smart アレイ 4250ES コントローラ

Smart アレイ 4250ES コントローラは、ProLiant 8000 と ML750 サーバでのストレージ可用性とパフォーマンスの最適化を目指して設計された、64 ビット Wide Ultra2 SCSI アレイコントローラであり、フェイルオーバー冗長性を備えています。Smart アレイ 4250ES コントローラは、ケーブルフリーの拡張 SCSI コントローラ設計であり、ホットプラグ可能な冗長アレイコントローラを 2 枚搭載できます。

Smart アレイコントローラについての留意事項

Smart アレイについて、次の留意事項があります。

- Smart アレイコントローラは、ACU を使って構成します。ACU は、SmartStart を使ってオフラインで、または Windows 2000/2003、Windows NT、NetWare のいずれかを使ってオンラインで実行します。
- ドライブアレイの構成情報がドライブ上に保存されるので、コントローラを交換しても再構成が不要です。SMART-2 以降のアレイコントローラは、ドライブ構成を不揮発性 RAM (NVRAM) には保存しません。
- アイドリング状態の間、アレイコントローラに接続されたドライブは、Dynamic Sector Repair (DSR) を実行します。これは通常の動作です。
- コントローラを追加した場合、またはコントローラやサーバでハードウェアを変更した場合は、必ずコントローラオーダを検証する必要があります。ブート可能なドライブに接続されているコントローラは、コントローラオーダの最初に配置する必要があります。
- すべての Smart アレイコントローラは、バスマスタ デバイスです。アレイコントローラを追加すると、システムリソースが解放されて他のアクティビティに利用できるだけでなく、ディスクの読み取り/書き込みパフォーマンスが向上します。
- 旧バージョンの Smart アレイコントローラでは、ホットプラグ可能なドライブの位置を変更すると、SCSI ID の割り当てが混乱して、ドライブアレイ構成に問題が生じる可能性があります。このために、その影響を受けるすべての論理ドライブですべてのデータが消去されることがあります。新しいコントローラではドライブを移動できますが、推奨はできません。



重要

容量の拡大、拡張、RAID の移行 (オフラインまたはオンライン) などの操作を行う前に、重要データをバックアップすることをお勧めします。

-
- SMART-2 アレイコントローラの多くは、ドライブに問題が生じ、データを消去せずに復元できる場合に、エラーを表示します。
 - SCSI デバイスがアレイコントローラに接続されているときは、電源投入時の自己診断テスト (POST) で SCSI ID が表示されません。
 - Smart アレイ 532、Smart アレイ 5312 および Smart アレイ 6402 コントローラは、3.3V、PCI/PCI-X スロットのみをサポートします。コントローラは、特別なスロットコーディングを持っているので、旧バージョンの ProLiant システムまたは 5V の PCI スロットには適合しません。
 - Smart アレイ 5i、532、5312、641、642、6402 コントローラは、シングルエンド デバイスをサポートしません。
 - Smart アレイコントローラは、Boot-ROM 機能を備えており、ProLiant サーバでブート可能です。ブートプロセス時に、ドライブは、INT13 呼び出しを使ってアクセスされます。

注記

▲ サーバにディスクコントローラがいくつかある場合、プライマリに設定されたコントローラがブートコントローラになります。

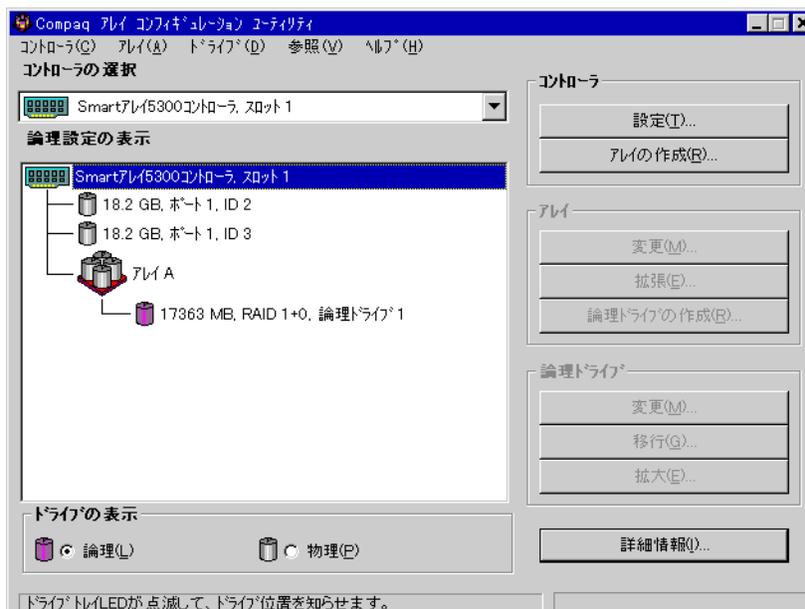
アレイコントローラ テクノロジとユーティリティ

すべての Smart アレイコントローラは、標準のテクノロジセットと管理ソフトウェアやユーティリティソフトウェアを使って、ストレージ管理を容易にします。

この章では、次のアレイコントローラの機能について復習します。

- アレイコンフィギュレーションユーティリティ (ACU)
- アレイコンフィギュレーションユーティリティ XE (ACU-XE)
- Option ROM Configuration for Arrays (ORCA)
- Array Diagnostics Utility (ADU)
- オンラインスペアドライブ
- オンラインでのドライブアレイ拡張
- 論理ドライブ容量の拡張
- オンラインでの RAID レベルの移行
- オンラインでのストライプサイズの移行
- ハードドライブ障害予測テクノロジ
- Dynamic sector repair (DSR)
- ホットプラグ ドライブのサポート
- Automatic Data Recovery (ADR)
- アレイアクセラレータ (読み取り/書き込みキャッシュ)
- データ保護

アレイコンフィギュレーションユーティリティ(ACU)



ACUは Smart アレイコントローラのインテリジェント機能へのインタフェースを表示して、アレイの構成を簡単にします。ACUの起動は、オペレーティングシステム、SmartStart CD、ブート可能ディスクのいずれかから行います。

ACUには次の機能があります。

- 最適なアレイ設定のための構成ウィザード
- 高速とカスタムの2つの初期構成オプション
- 容量拡張、論理ドライブ拡張、RAIDレベルの移行、ストライプサイズの移行などのツールを使った簡単な構成変更
- RAIDレベル0、1、1+0、5、RAID ADGのフォールトトレランスのサポート
- ユーザによるストライプサイズを選択が可能
- コントローラのパフォーマンスを調整するために、キャッシュの読み取り/書き込み比率とストライプサイズの変更が可能
- ドライブ再構築と容量の拡張についての優先順位の設定
- オンラインスペア(ホットスペア)の構成
- 論理ドライブ別に分割したフォールトトレランス構成
- ストレージを素早く識別するためのドライブトレイLEDの点滅

ACU を使ったドライブアレイ容量の拡張

ドライブアレイ容量の拡張とは、物理ディスクを既存のドライブアレイに追加することをいいます。アレイコントローラは、論理ボリュームのデータを再配布して、新しいドライブに保存します。追加した容量は、それを使って新しく論理ボリュームを作成することも、既存の論理ボリュームの容量を拡張することもできます。

ドライブアレイの拡張は、論理ドライブレベルではなく、アレイコントローラレベルで実行されます。ほとんどの場合、あるコントローラに接続されたすべてのディスクドライブは、1つのアレイとしてグループ化する必要があります。これにより、RAID のフォールトトレランスを最も効率的に利用できます。

注記

Windows 2000/2003、Windows NT、NetWare では、このユーティリティがオンラインで起動できます。ディスクの構成時にサーバの電源を切断する必要はありません。

アレイコンフィギュレーション データの位置

アレイコンフィギュレーションを ACU から保存すると、情報は各ハードドライブの RAID 情報セクタに保存されます。これによって、データを失わずにコンピュータ構成要素を交換できます。データを失わずに、ドライブのセットをマシン間で移動することもできます。

サポートされているコントローラ

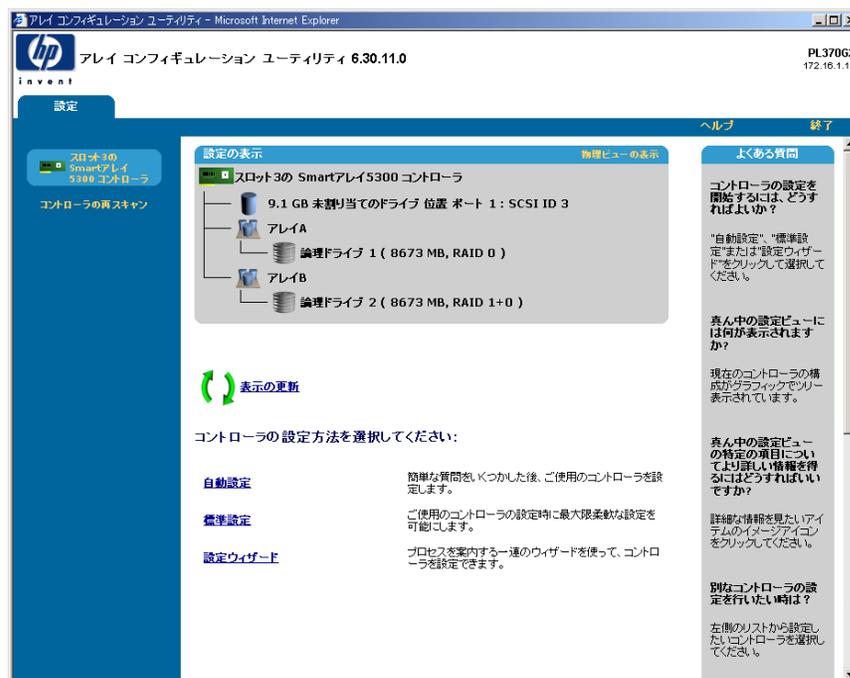
ACU は、次の Smart アレイコントローラとストレージシステムをサポートします。

- | | |
|---------------------|----------------------------|
| ■ Smart アレイ 5302 | ■ 内蔵 Smart アレイ |
| ■ Smart アレイ 5304 | ■ Smart アレイ 3200 |
| ■ Smart アレイ 5312 | ■ Smart アレイ 3100ES |
| ■ Smart アレイ 5i | ■ Smart アレイ 221 |
| ■ Smart アレイ 5i Plus | ■ SMART-2/E |
| ■ Smart アレイ 532 | ■ SMART-2/P |
| ■ Smart アレイ 4250ES | ■ SMART-2SL |
| ■ Smart アレイ 4200 | ■ SMART-2DH |
| ■ Smart アレイ 431 | ■ StorageWorks RA4000/4100 |

注記

ACU は、シングルチャネル RAID (RAID LC2) コントローラをサポートしていません。シングルチャネル RAID (RAID LC2) をサポートしているのは、ORCA のみです。

アレイコンフィギュレーションユーティリティ XE (ACU-XE)



アレイコンフィギュレーションユーティリティ XE (ACU-XE)は、ACUと同じ機能が利用できますが、ブラウザを使って、アレイをローカル、リモートの両方で構成できます。

ACU-XEは、オンラインのリモート管理と構成の機能を提供します。またすべての Smart アレイコントローラを一元的な場所から管理して、次の作業を実行します。

- オンラインでの論理ドライブ容量の拡張
- オンラインでの RAID レベルの移行
- オンラインでのストライプサイズの移行
- 再構築と拡張の優先順位の変更
- マルチシェルフ構成でのストレージの素早い識別
- アレイアクセラレータの構成

Option ROM Configuration for Arrays (ORCA)

ORCA は、シングルチャネル RAID (RAID LC2) コントローラなどのアレイコントローラ上にある Option ROM から実行します。これは、構成を最小に抑える必要があるユーザのために設計されています。POST 実行時に、ORCA をサポートする Smart アレイコントローラは、初期化プロセスの 1 つとして、コンピュータシステムコンソールにプロンプトを表示します。

アレイコントローラに構成済みの論理ドライブが存在しない場合、このプロンプトは 10 秒間表示された後、ORCA をバイパスして POST を続行します。論理ドライブが構成されている場合、プロンプトは 5 秒間表示されます。ORCA をサポートする Smart アレイコントローラがシステムにインストールされている場合は、コントローラベースのオプション ROM によって、POST の実行中に 10 秒間プロンプトが表示されます。

ORCA を起動する場合は、[F8]キーを押します。

ORCA の主な特徴は以下のとおりです。

- 実行のためのディスクや CD が不要
- サーバの電源投入時に起動できる
- 論理ドライブの作成、削除が可能
- コントローラオーダが構成できる
- 作成した論理ドライブに対して、オンラインスペアが割り当てられる
- RAID レベルが指定できる
- ストライプサイズは設定できない
- 英語のみをサポート

さまざまな構成方法

ORCA をサポートするのは、最新の Smart アレイコントローラのみです。他のすべての Smart アレイコントローラは、ACU または ACU-XE のみをサポートします。シングルチャネル (RAID LC2) コントローラは、ORCA によってのみ構成できます。



注記

ORCA は、ドライブ拡張、RAID レベルの移行、ストライプサイズの設定はサポートしません。

Array Diagnostics Utility

Array Diagnostics Utility (ADU)を使用すると、不適切なエラーレート、アレイアクセラレータボード上でのバッテリー障害などの問題が素早く識別できます。

SmartStart and Support Software リリース 4.10 以降は、Drive Array Advanced Diagnostics (DAAD) Utility に代わって ADU が提供されています。DAAD Utility は SmartStart CD に収録されておらず、最新のアレイコントローラをサポートするためのアップデートも行われていません。

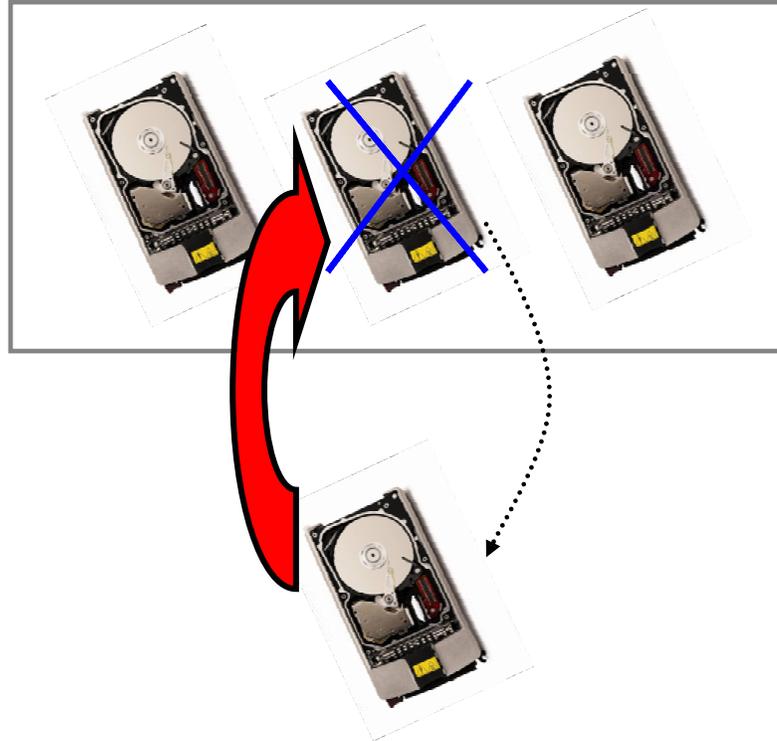
ADU を正しく実行するには、SmartStart CD からシステムを起動します。

ADU は、以下の Smart アレイコントローラをサポートします。

- SMART-2/E、SMART-2/P、SMART-2SL、SMART-2DH
- Smart アレイ 221
- 内蔵 Smart アレイ
- シングルチャネル RAID (RAID LC2)
- Smart アレイ 431
- Smart アレイ 532
- Smart アレイ 5i、5i Plus
- Smart アレイ 3100ES、Smart アレイ 3200
- Smart アレイ 4200、Smart アレイ 4250ES
- Smart アレイ 5300 シリーズ
- Smart アレイ 5312
- Smart アレイ 641、642
- Smart アレイ 6400 シリーズ

オンラインスペアドライブ

RAID 5が論理ドライブを保護



オンラインスペアドライブ

オンラインスペアドライブは、障害が発生したドライブに対する一時的な代替ドライブとして動作します。すべてのフォールトトレランス論理ドライブには、オンラインスペアドライブを追加できます (RAID 0 はサポートされません)。ドライブ容量の効率的な使用が重視される環境では、1つのオンラインスペアを複数のアレイに割り当てられます。オンラインスペアの容量は、アレイの他のドライブ以上でなければなりません。

データドライブに障害が発生すると、オンラインスペアドライブが、障害デバイス上のデータを自動的に再構築し始めます。オンラインスペアドライブが再構築を完了した後は、別のドライブに障害が発生しても、データを損失せずに対処できます。オンラインドライブが再構築を完了する前に別のドライブに障害が起こる可能性は非常に低く、2つのドライブに同時に障害が起こっても常に対処できるのは ADG のみです。

障害が発生したドライブを交換すると、データが新しいドライブ上に自動的に再構築されます。新しいドライブ上でデータが完全に再構築されると、オンラインスペアの役割はオンラインスペアドライブに戻ります。このため、オンラインスペアドライブが代替として使われたままになることはありません。

注記

データがオンラインスペアに再構築される速度は、1GBあたり10～20分です。これは、再構築に割り当てられた優先順位と、アレイにあるドライブ数の合計によって異なります。

オンラインスペアドライブは、パーティショニングやフォーマットを行う必要はありません。オンラインドライブは、使用されていないときでも、常にアクティブで稼動状態にあります。

Insight マネージャ7では、他のアクティブドライブとまったく同様に、オンラインスペアドライブが監視できます。

オンラインスペアドライブは、RAID 1、RAID 1+0、RAID 4、RAID 5、RAID ADG で利用可能です。

重要

再構築の優先順位を高く設定すると、再構築の処理中、サーバのパフォーマンスが低下します。再構築の優先順位を低く設定すると、サーバがアイドル状態のときのみ再構築が実行されるので、通常のサーバのパフォーマンスが維持されますが、システムのアクティビティによっては、再構築にかなり長い時間がかかることがあります。

オンラインでのドライブアレイ拡張

ほとんどの Smart アレイコントローラは、データの損失がないオンラインでのアレイ拡張をサポートします。オンラインでのアレイ拡張を実行するには、新しいドライブをホットプラグ可能なドライブベイにインストールし、ACU を使って新しいドライブを既存のアレイに追加します。

すべてのデータは、拡張プロセスが開始された後に再配置されます。すべてのドライブにわたってデータを再配布すると、各ドライブに空きスペースができます。すべてのドライブ上にあるこれらのゾーンを使って、新しい論理ドライブを作成したり、既存の論理ドライブの容量を拡張したりすることができます。

最後に、新しい論理ドライブがオペレーティングシステムに認識されます。この結果、オペレーティングシステムは、拡張プロセスの後、ドライブの容量が増えたとは認識しません。認識されるのは、以前の論理ドライブと新しい論理ドライブです。

注記



物理ドライブを拡張しても、論理ドライブの容量が大きくなるのではなく、新しい論理ドライブが作成されます。この新しい論理ドライブは、拡張プロセスが完了した後に表示できるようになります。

データの再配置は、バックグラウンドプロセスとして実行されます。データの再配置には、それに必要なパフォーマンスに応じて、「高」、「中」、「低」のいずれかの優先順位を割り当てられます。データの再配置に必要な時間は、論理ドライブのサイズによって異なります。再配置の実行中、RAID の保護が維持されます。

注記



次のアレイコントローラは、物理ドライブの拡張をサポートしません。

SMART、SMART-2SL、Smart アレイ 221、シングルチャネル (RAID LC2)

Smart アレイコントローラは、最高 32 個の論理ドライブが定義できます。アレイ内のすべてのドライブは、同じサイズでなければなりません。1 つのアレイに、他より容量の大きいディスクをインストールすると、差分の容量が利用できなくなります。オペレーティングシステムによっては、サポートする論理ドライブが 32 個より少ないこともあります。

論理ドライブ容量の拡張

論理ドライブを物理的に拡張すると、新しい論理ドライブが作成され、容量が追加されたことが示されます。新しい空き容量は、すぐにオペレーティングシステムで使えるわけではありません。

ディスクがパーティション分割された後、フォーマットのプロセスによって、既存のパーティションのサイズに適したファイルシステムが作成されます。オペレーティングシステムの多くは、ドライブをフォーマットしなければ、新しいファイルシステムを作成できませんが、ドライブをフォーマットすると、パーティション内のすべてのデータが消去されます。ボリューム拡張をサポートするオペレーティングシステムだけが、データを失わずに追加した容量を利用できます。

Windows 2000/2003、Windows NT、NetWare、および他の高度なオペレーティングシステムは、既存のパーティションと空きスペースを容量の大きなパーティションにまとめる、ボリューム拡張をサポートしています。

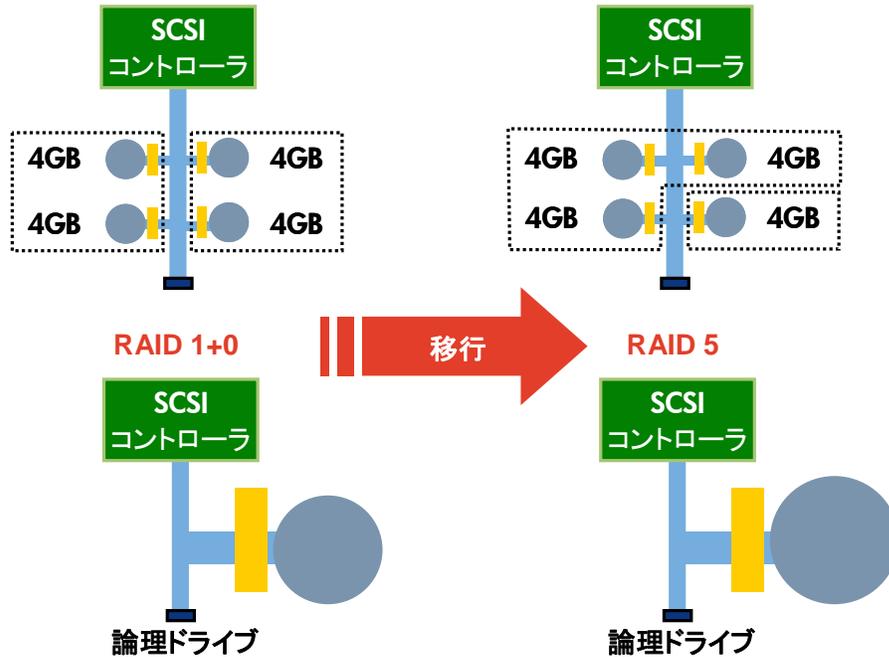
また他のソフトウェアメーカーから、たとえば Server Magic のような、データを失わずにディスクのパーティションを設定しなおせるユーティリティが発売されています。これらのユーティリティは、ほとんどがオフラインで機能します。

▲ 注記

次のアレイコントローラは、論理ドライブの拡張をサポートしません。

SMART、SMART-2/E、SMART-2/P、SMART-2/SL、SMART-2/DH、Smart アレイ 221、シングルチャネル RAID (RAID LC2)

オンラインでの RAID レベルの移行



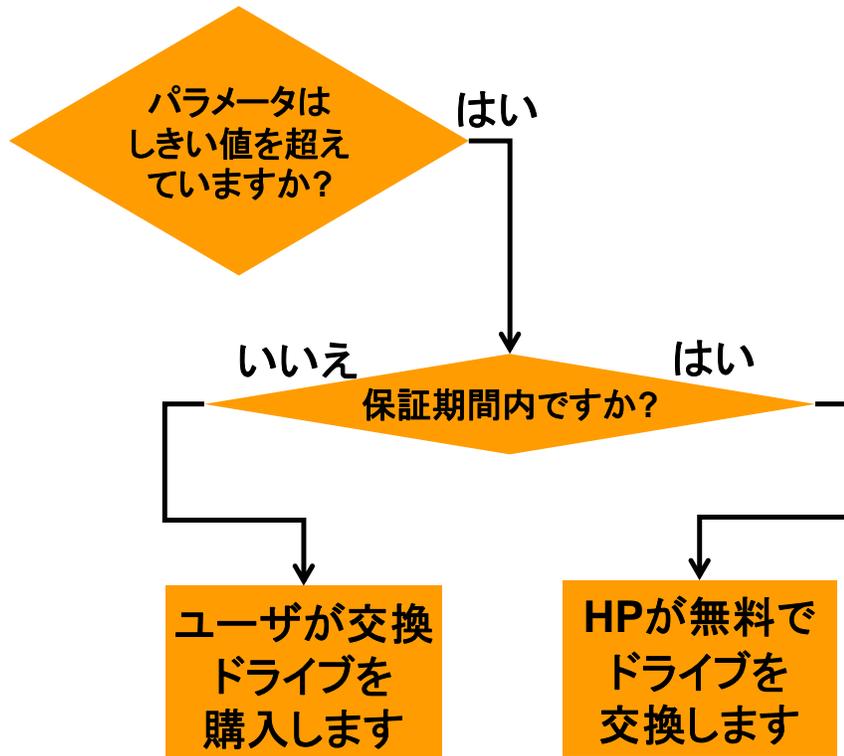
現行のアレイコントローラは、すべて RAID レベルの移行をサポートします。論理ドライブは、新しい RAID レベルに簡単に移行できます。Windows と NetWare に関しては、システムの運用を中断したり、データを損失したりすることなく、オンラインでこの操作を実行できます。オフラインでの移行は、どのオペレーティングシステムでも実行できます。

オンラインでのストライプサイズの移行

最新のアレイコントローラは、すべてストライプサイズの移行をサポートします。ユーザは、既存の論理ドライブのストライプサイズを簡単に変更できます。WindowsとNetWareでは、システムの運用を中断したり、データを損失したりせずにオンラインでこの操作を実行できます。次の表は、いくつかのコントローラのデフォルトストライプサイズを表しています。

コントローラ	RAID 0	RAID 1	RAID 5	移行のサポート
Smart アレイ 6402	128KB	128KB	16KB	○
Smart アレイ 641、642	128KB	128KB	16KB	○
Smart アレイ 5302、5304、5312	128KB	128KB	16KB	○
Smart アレイ 5i、5i Plus	128KB	128KB	16KB	○
Smart アレイ 3200、4200	128KB	128KB	16KB	○
Smart アレイ 3100ES、4250ES	128KB	128KB	16KB	○
Smart アレイ 431、532	128KB	128KB	16KB	○
内蔵 Smart アレイ	128KB	128KB	16KB	○
シングルチャネル RAID (RAID LC2)	128KB	128KB	16KB	×
Smart アレイ 221	128KB	128KB	16KB	×
SMART-2SL、2DH	128KB	128KB	16KB	×
SMART-2/E、2/P	128KB	128KB	16KB	×
SMART	16KB	16KB	16KB	×

ハードドライブ障害予測テクノロジー



HP は Smart アレイコントローラがモニタテストを実行する形式の、ハードディスクドライブの障害予測テクノロジーを他に先駆けて開発しました。これらのテストは **Monitoring and Performance (M&P)** または **Drive Parameter Tracking** と呼ばれ、ハードドライブの属性であるシーク時間、スピンドル時間、メディア障害 (20 以上のパラメータ) などを外部から監視することによって、障害の可能性を示す変化を検知します。

HP はハードドライブ業界と協力して、**Self-Monitoring Analysis and Reporting Technology (S.M.A.R.T.)** として知られる診断/障害予測機能の開発にあたりました。S.M.A.R.T. が完成するまでの数年の間、HP はハードドライブ障害予測テクノロジーのサポートに M&P と S.M.A.R.T. の両方を採用して、事前予防保証によるハードドライブ交換を実現してきました。

S.M.A.R.T. は今では高い完成度を示し、これを受けて HP では現在、事前予防保証をサポートするハードドライブ障害予測テクノロジーとして、もっぱらこのテクノロジーを採用しています。

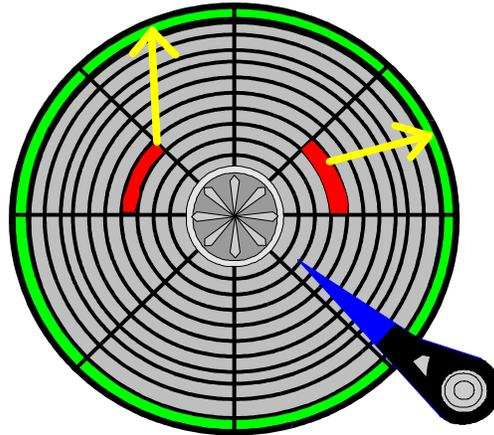
▲ 注記

2001 年以降、HP の販売する SCSI ハードドライブは、Smart アレイコントローラ上の M&P テストを無効化する設定で出荷されています。これによって誤った障害予測が排除されるとともに、M&P コントローラによる 1 時間ごとのテスト実行がなくなるので、パフォーマンスが向上します。

S.M.A.R.T.は、モニタ機能をハードディスクドライブ内に配置することによって、障害予測テクノロジーの向上を図っています。S.M.A.R.T.の監視ルーチンは、特定のドライブタイプに合わせて設計されており、内部パフォーマンス、較正、エラー測定に直接アクセスできるので、以前のM&Pテストよりはるかに正確です。SMARTは、内部パフォーマンスインジケータとリアルタイムモニタ分析機能を使って、元々M&Pテストが持っていたデータ保護と障害予測の能力をさらに強化しています。また、Smartアレイコントローラは、アイドル時間にハードドライブメディアを予防的にスキャンし、メディアの障害が検知された場合はそれに対処します。

S.M.A.R.T.は、多くの場合、障害の発生前に問題を予測できます。Smartアレイコントローラは、S.M.A.R.T.エラーコードを認識し、ハードドライブ障害が間近に予測される場合はそれをシステムに通知します。Insight マネージャ7には、潜在的な問題が発生したときにはすぐに通知されます。事前予防保証では、Insight マネージャ7で劣化と判定されたHPドライブは交換の対象になります。

DSR (Dynamic Sector Repair)



通常の運用では、当初障害がなかったメディアにも障害が発生することがあります。これは一般的な現象です。ディスクのビット密度と回転速度が年々大きくなるにつれ、問題が発生する可能性も高くなっています。通常ドライブは、外部からの支援がなくても、各セクタの最後に保存された CRC (巡回冗長検査) チェックサムを使って、障害セクタを内部的にマップしなおすことができます。

すべての Smart アレイコントローラは、他のディスクアクティビティがないときには、バックグラウンドジョブとして表面解析を実行します。完全に読み取り不可能なセクタも、コントローラの RAID 機能を使って再構築や再マップすることが可能です。

DSR には次の特徴があります。

- ハードウェアベースのフォールトトレランスによって自動的に実現されます。ハードウェアのフォールトトレランスが使われない場合、DSR は利用できません。
- ドライブサブシステムのフォールトトレランスを使って、障害セクタをスペアセクタと交換します。正しいデータは、同じドライブのスペアセクタに書き込まれます。
- トリガは自動的に行われます。Smart アレイコントローラは、アイドル時間が 30 秒経過すると DSR をトリガします。
- 同じトラック内で 2 つのセクタに障害がある場合は、隣接するトラックのセクタにデータを配置します。
- DSR はバックグラウンドタスクなので、ディスクサブシステムのパフォーマンスには影響を与えません。オペレーティングシステムが要求を行うと、DSR は中断されます。

注記

▲ DSR の実行中は、ディスクドライブアクティビティ LED が点灯します。

ホットプラグ ドライブのサポート

Smart アレイコントローラが提供するいくつかの機能には、ホットプラグ可能な SCSI ドライブが必要です。ドライブがホットプラグ可能な場合は、ドライブをオンラインにしたまま、以下の操作が可能です。

- フォールトトレランス アレイでの障害ドライブの交換
- ドライブとアレイの追加
- アレイの拡張

アレイコントローラの主要な利点の 1 つに、サーバをオフラインにせずに、ドライブ障害から完全に復元できる機能があります。この機能を利用するには、ホットプラグ可能なドライブをアレイコントローラとともに使う必要があります。

ホットプラグ可能なドライブの LED

Smart アレイコントローラのファームウェアでは、接続されたホットプラグ可能なハードドライブが劣化ステータスになった場合に、ハードドライブにあるアンバー（琥珀色）の LED が点滅するようになっています。これにより、特に Insight マネージャ 7 などのシステム管理ユーティリティによって通知されたときに、劣化した物理ハードドライブを簡単に探して交換できます。劣化したホットプラグ可能なハードドライブは、オンラインのまま、次の表に示す組み合わせで LED が点灯します。

ステータス	LED の状態
動作状態	点灯、消灯、点滅のいずれか
オンライン	点灯または消灯
故障	アンバー（琥珀色）の LED が点滅

注記

この機能は、RAID 0 の非フォールトトレランス構成ではサポートされません。コントローラは、RAID 1、RAID 1+0、RAID 5、RAID ADG のいずれかのフォールトトレランス構成である必要があります。

Automatic Data Recovery

ドライブを交換すると、Smart アレイコントローラはそれを自動的に検知します。RAID レベルが 1、1+0、4、5、ADG のいずれかに設定されている場合は、データは新しいドライブに自動的に再構築されます。必要な作業は、ドライブの交換だけです。ホットプラグ可能なドライブをサポートしているシステムでは、システムの稼動中にドライブを交換できます。再構築作業の優先順位の設定や変更は、ACU を使っていつでも行えます。

アレイアクセラレータ(読み取り/書き込みキャッシュ)

Smart アレイコントローラのアレイアクセラレータは、I/O パフォーマンスを劇的に改善します。サイズは、コントローラに応じて、4、16、32、64、128、256MB のいずれかです。

デフォルトでは、アレイアクセラレータのキャッシュ容量は、読み取り用と書き込み用に二等分されます。サーバアプリケーションで、読み取りが書き込みよりはるかに多い場合(またはその逆の場合)は、この設定を変更してパフォーマンスを改善する必要があります。この変更は、システムをリブートすることなく、オンラインで実行できます。設定する最適な割合は、アプリケーションに依存します。

アレイアクセラレータは、インテリジェントな先読みアルゴリズムを使って、必要なデータを予測し、待ち時間を短縮します。1 つまたは複数の I/O スレッドで読み取りアクティビティが順に検知され、次に続く要求が予測されます。データは収集されて、高速キャッシュに保存されます。オペレーティングシステムがデータを要求するとすぐに、そのデータは、ディスクによるデータ転送の 100 倍の速度で転送されます。

ランダムアクセスパターンが検知されると、先読み機能は常に無効化されます。ランダム I/O で先読みを行うと、システムパフォーマンスが低下するからです。

ディスクがビジー状態の場合、新しい書き込み内容がキャッシュに保存され、後でアクティビティが減少したときにディスクに書き込まれます(ライトバック)。通常、いくつかの小さなブロックは、大きなブロックにまとめることができますが、それは少数の大きなブロックをディスクに書き込むことになるので、パフォーマンスが向上します。

▲ 注記

Smart アレイ 5300 と 6400 コントローラファミリは、アップグレード可能なキャッシュモジュールを持つ数少ないアレイコントローラファミリです。

アレイのパフォーマンス調整

アレイのパフォーマンスを最適化するには、以下のいくつかの方法があります。

- 読み取りと書き込みに割り当てるアクセラレータキャッシュの割合を変更する
- システムへの一般的なデータ転送のタイプに応じて、それに適したストライプサイズに変更する
- オーバーヘッド要求が少ないフォールトトレランスモードに変更する
- 4つのコントローラチャンネル(コントローラによって異なる)すべてにわたって、論理ドライブを構成する

データ保護

書き込みキャッシュのデータには、特別な保護が必要です。

高度なエラーチェック

高度な ECC 機能は、メモリチップのビットエラーとメモリチップの障害を処理できます。

バッテリーのバックアップ

アレイアクセラレータには、バックアップ電源として、再充電可能なバッテリーが付いています。このバッテリーによって、キャッシュのデータは 4 日間まで保持できます。Smart アレイ 5300、641、642、6402 アレイコントローラは、冗長バッテリーをサポートします。

リカバリ ROM

Smart アレイには、次のようなリカバリ ROM 機能があり、ファームウェアの障害からデータを保護します。

- コントローラが、ファームウェアのコピーを 2 部、ROM に保存する
- 新しいファームウェアがコントローラにフラッシュされると、それまでに動作していたファームウェアが保存される
- 障害が発生すると、コントローラはスタンバイファームウェアにロールオーバーする
- リカバリ ROM は、新しいファームウェアをコントローラにフラッシュする際のリスクを軽減する

復習問題

1. 非アレイストレージシステムと比べた場合、RAID ストレージの 4 つの主な利点は何ですか。
.....
.....
.....
.....
2. RAID 実装の 2 つの主なタイプは何ですか。
.....
.....
3. RAID レベルの中で、読み取り、書き込み操作が最も高速なのはどれですか。
.....
4. POST の実行中、Smart アレイコントローラに接続されたすべてのハードドライブについて SCSI ID が表示されます。
 正
 誤
5. ホットプラグ可能なハードドライブでサポートされるのはどの機能ですか。
 - a. フォールトトレランス アレイでの障害ドライブの交換
 - b. ドライブとアレイの追加
 - c. アレイの拡張
 - d. サーバがオンラインの状態でのアレイコントローラの交換
6. Smart アレイ 532 の特徴に当てはまるのはどれですか。
 - a. 32 ビットアレイコントローラ
 - b. Wide Ultra2 コントローラ
 - c. オンラインでの RAID レベルの移行のサポート
 - d. Ultra3、Ultra2、Wide-Ultra SCSI-4 のデバイスのサポート

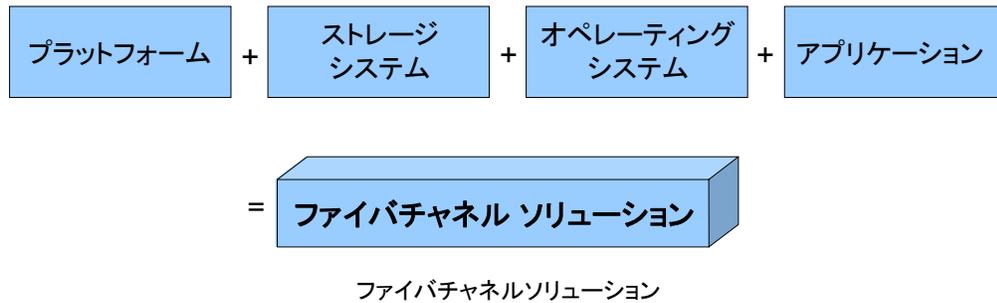
-
7. 次のアレイコントローラのうち、SCSI チャンネルが 3 つ以上あるのはどれですか。
- a. SMART-2DH
 - b. Smart アレイ 5302
 - c. Smart アレイ 5304
 - d. Smart アレイ 5i
8. RAID 1 は、完全なフォールトトレランスを提供する安価なソリューションだと考えられています。
- 正
 - 誤
9. オンボードバッテリーバックアップキャッシュを搭載した Smart アレイ 5300 コントローラでは、サーバ障害の場合にデータが保持されます。このバッテリーによってデータが保持される時間はどれくらいですか。
-
10. RAID ADG は、同時に何個のドライブ障害までデータを失わずにサポートできますか。
-

目的

この章では、次の内容について学習します。

- ファイバチャネルの特徴と利点、一般的なファイバチャネルに関する用語
- ファイバチャネル ソリューションで使われるハードウェア構成要素、各構成要素の機能、HP ファイバチャネル プロダクト
- スイッチとハブの違い、4 種類のファイバチャネル スイッチのゾーニング、ハブとスイッチのカスケード接続の利点
- 光ファイバケーブルの特徴と光ファイバケーブルに関する安全上の注意
- 減衰、マクロバンド、マイクロバンド
- 3 つの一般的なファイバチャネルトポロジとそれらの技術的特徴
- 以下のような新しいファイバチャネルテクノロジー
 - ATM (非同期転送モード)
 - IP and ARP over Fibre Channel

ファイバチャネルの概要



ファイバチャネルは、American National Standards Institute (ANSI) が開発したいくつかの標準を統合したセットの一般名です。ファイバチャネルテクノロジーは、業界標準の相互接続シリアルデータ転送アーキテクチャであり、サーバ業界において、高度の信頼性、スループット、距離における柔軟性を実現します。

ファイバチャネルは、デバイス間の相互接続構造として光ファイバを使います。光ファイバは、FDDI (fiber distributed data interface) や SCSI のような高性能のプロトコルをサポートしており、大量のデータを高速伝送できます。

図に示すように、完全なファイバチャネルソリューションには、ハードウェアプラットフォーム、ストレージシステム、オペレーティングシステム、アプリケーションが含まれます。

ネットワークは分散ノード (たとえば、ワークステーション、ファイルサーバ、周辺機器) の集合で、独自のプロトコルがあり、ノード間の相互作用をサポートします。オペレーティングシステムが、いくつかのネットワークレイヤでデータを変換するので、ネットワークはファイバチャネルシステムより広範なタスクを処理します。ネットワークはソフトウェアを集中的に使うので、オーバーヘッドが比較的大きくなります。

一方、ファイバチャネルソリューションは、ハードウェアを集中的に使い、事前にアドレスを定義した少数のデバイスとの間でのみ機能します。ファイバチャネルシステムでのデータ通信は、通信デバイス間の直接接続または交換ポイントツーポイント接続によって行われます。その結果、ファイバチャネルシステムは、少ないオーバーヘッドでデータを高速伝送できます。

ファイバチャネルの利点

ファイバチャネルは、パフォーマンス、可用性、スケーラビリティ、柔軟性、ノード間の距離の点で優れています。

パフォーマンス

現在のファイバチャネルのスループットの標準は、正味ユーザペイロードで 100MB/s です。ANSI は、200MB/s と 400MB/s 用のファイバチャネルのスループット標準も規定しています。

注記



上記の速度は、半二重伝送の場合です。全二重の場合は、スループットがこの 2 倍になります。

可用性

ファイバチャネルテクノロジーは、Storage Area Network (SAN)、Enterprise Network Storage Architecture (ENSA)、および多くの高可用性やクラスタリングを実装するための基本的なビルディングブロックを提供しています。

ファイバチャネルはまた、可用性に関して次の利点があります。

- ホットスワップ対応のサポート
- 組み込みの強固な機能 (8 ビット/10 ビットのエンコード/デコードによるエラー検知と修正) による実装コストの削減とプロトコルオーバーヘッドの縮小
- 業界における広範なサポート

スケーラビリティ

ファイバチャネルテクノロジーは、次のように高いスケーラビリティを持ちます。

- 拡張スロットごとのストレージ容量の増加
- Fibre Channel Arbitrated Loop (FC-AL) における 126 ノードのサポート
- Fibre Channel Switched Fabric (FC-SW) における 1,600 万ノードのサポート

柔軟性

ファイバチャネルテクノロジーは、次の利点を備えることによって、柔軟性が大きく向上しています。

- プロトコル独立
- 柔軟な配線と小型のコネクタ
- バスの終端が不要
- 複数のトポロジに対応 (ポイントツーポイント、アービトレーテッドループ、スイッチファブリック)
- ネットワーキングと I/O チャネルデータ通信の統合
- 柔軟な伝送サービス
- 専用の帯域幅
- 複数の送信元または送信先ポートとの応答を伴う多重化伝送
- 応答を伴わないベストエフォートの多重化データグラム伝送

距離

ファイバチャネルには、距離に関して次の特徴があります。

- 銅線ケーブルの場合、ノード間の距離は 30m
- マルチモードファイバと短波レーザの場合、ノード間の距離は 500m
- シングルモードファイバと長波レーザの場合、ノード間の距離は 10km

INTERNET ファイバチャネル標準の詳細は、Fibre Channel Industry Association の Web サイトを参照してください。 <http://www.fibrechannel.org/>

ファイバチャネルで使われる用語

ファイバチャネルテクノロジーの説明で使われる一般的な用語を理解することは、このテクノロジーの各方式を把握するために重要です。

ノード

ファイバチャネル デバイスは、ノードと呼ばれます。ノードは、コンピュータ(ホスト)、アダプタ、1つまたは複数のファイバチャネル ポートを持つデバイスのいずれかです。

ノード ID

ノード ID とは、ノードの一意の識別子です。

ポート

ノードをファイバチャネルトポロジに接続する場合は、他のデバイスにアクセスするためのポートが、各ノードにつき最低 1 つ必要です。ノード上のポートは、「N_Port」と呼ばれます。ポートの命名規則では、*N* がノードを、*L* がループを、*F* がファブリックを示します。

以下の表に、ファイバチャネルのポート名、タイプ、トポロジ、それぞれの説明を示します。

ポート	タイプ	トポロジ	説明
N_Port	ノード	ポイントツーポイントまたはファブリック	特定のノード上のポート
NL_Port	ノード	アービトレーテッドループ	アービトレーテッドループに接続されているノード
F_Port	ファブリック	スイッチファブリック	ファブリックポート
FL_Port	ファブリック	スイッチファブリック	アービトレーテッドループに接続されているファブリック
L_Port	ループ	アービトレーテッドループ	アービトレーテッドループ上のハブポート

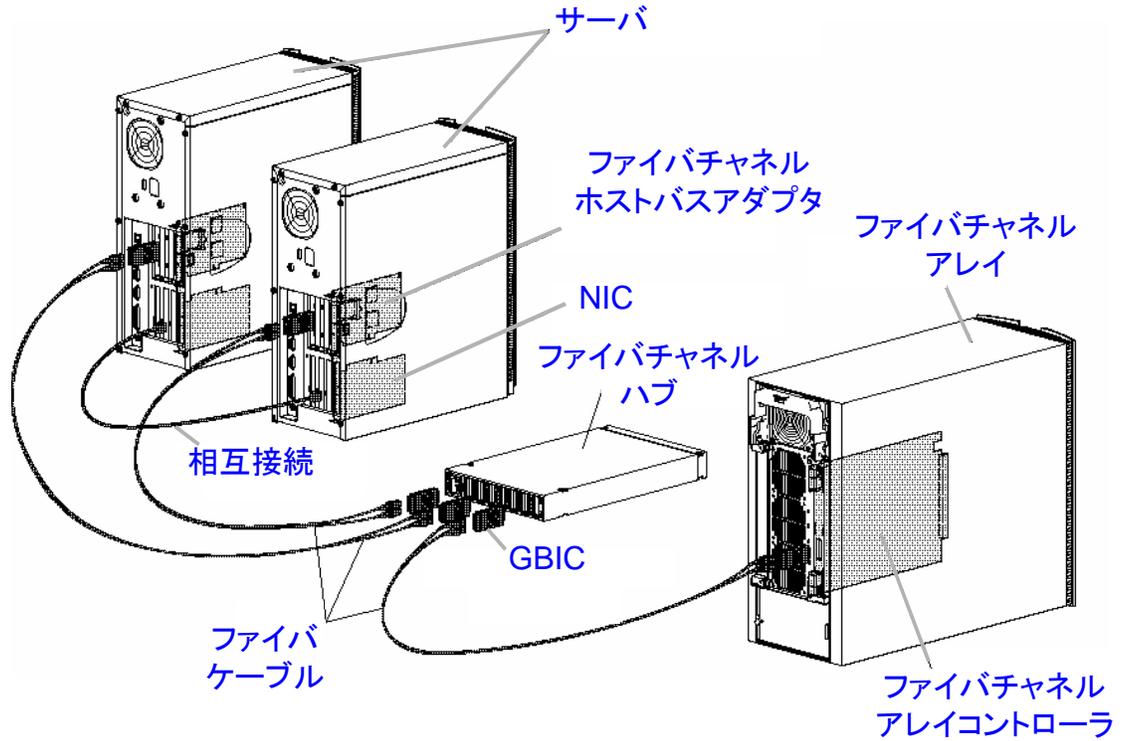
リンク

ファイバチャネル ポートでは、1 対の光ファイバを使います。片方の光ファイバは、受信側ポートへの情報を伝送し、もう片方のファイバは、送信側ポートからの情報を伝送します。この光ファイバのペアをリンクと呼びます。

トポロジ

ファイバチャネルトポロジとは、複数の N_Port を接続する相互接続構造です。ファイバチャネルトポロジには、ポイントツーポイント、FC-AL、FC-SW の 3 種類があります。

ファイバチャネルの構成要素

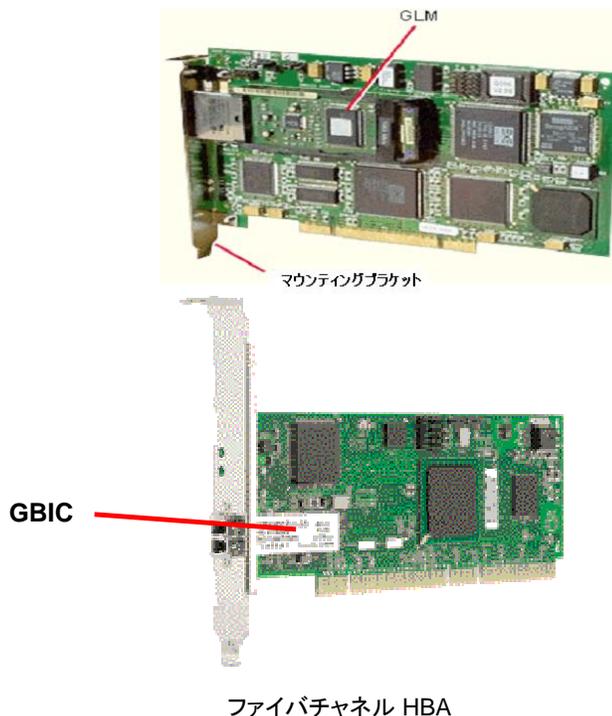


RA4X00 の構成

ファイバチャネルの実装には、次のハードウェア構成要素が必要です。

- ホストバスアダプタ (HBA)
- ギガビット インタフェースコンバータ (GBIC)
- ギガビット リンクモジュール (GLM)
- ストレージアレイとドライブアレイ
- ファイバチャネルアレイコントローラ
- ファイバチャネルハブまたはスイッチ

ホストバスアダプタ



ファイバチャネル HBA は、SCSI コマンドをシリアルデータに変換し、それを GBIC が光信号に変換します。GBIC は、シリアル電気信号とシリアル光信号を相互に変換するトランシーバです。ネットワークの中では、GBIC はファイバチャネル 媒体間でデータを伝送するために使われます。

ほとんどのファイバチャネル HBA は、GBIC か GLM を使用します。GLM は、GBIC と同様、電気信号と光信号を相互に変換しますが、それに加えてシリアル信号とパラレル信号間の変換も行います。

一般的な HBA は PCI バスを使い、この PCI バスが高度に統合されたアプリケーション固有の統合回路 (ASIC) を使って、ファイバチャネル プロトコルを処理し、ホストとの I/O を管理します。

▲ 注記

オペレーティングシステムは、HBA をファイバチャネル NIC としてではなく、SCSI コントローラとして検知します。したがって、HBA は他の SCSI コントローラカードを構成するのと同様、オペレーティングシステムで構成する必要があります。

GBIC (ギガビット インタフェースコンバータ)



光ファイバケーブルに接続される GBIC

GBIC (Gigabit Interface Converters) は、電気インパルスと光ファイバ媒体で使われる光信号の間のトランザクションを実行します。GBIC には、光信号を発するデバイスがあり、それが光ファイバケーブルで伝送されます。

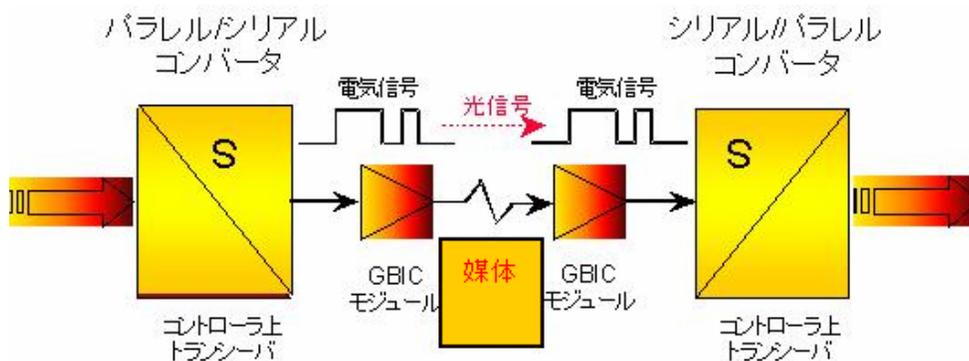
ファイバチャネルの各リンクには、図に示すように、光ファイバケーブルの両端に 1 つずつ、合計 2 個の GBIC が必要です。

GBIC モジュールは、次のコンポーネントの特別なコンセントに接続します。

- ファイバチャネル HBA
- ファイバチャネルコントローラ
- ファイバチャネルストレージハブ
- ファイバチャネルスイッチ

GBIC には、ケーブル (リンク) 内の光ファイバごとに 1 つずつ、合計 2 つのチャネルがあります。したがって、GBIC モジュールは、すべて全二重伝送デバイスです。

GBIC のしくみ



GBIC は、ファイバチャネル 媒体間でデータが伝送できるよう、シリアル電気信号とシリアル光信号を相互に変換します。

ファイバチャネルを通信媒体として使う場合、信号は、次の 2 種類の変換を経由する必要があります。

- **パラレルからシリアル** — ホストまたはターゲットのストレージデバイスが生成したパラレル信号を、シリアル信号に変換する必要があります。
- **電気信号から光信号** — ホストまたはターゲットのストレージデバイスが生成した電気信号を、光ファイバチャネル 媒体が使う光信号に変換する必要があります。

GBIC-SW

GBIC-SW は、マルチモードファイバをサポートする GBIC モジュールの短波バージョンです。以下の表に、GBIC-SW の仕様を示します。

機能	詳細
準拠仕様	ファイバチャネル FC-PH-2 物理層オプション 100-M5-SN-I
ポーレート	1062.5MB/s
ファイバ短波	直径 50 μ m (推奨) または 62.5 μ m マルチモード ファイバ
レーザ波長	780nm (非 OFC)
光コネクタのインタフェース	デュアル SC
距離	<ul style="list-style-type: none"> ■ 50μm: リンクあたり 2~300m (国際規格) ■ 62.5μm: リンクあたり 2~500m (米国規格)

GBIC-LW

GBIC-LW は、GBIC モジュールの長波バージョンであり、シングルモードファイバをサポートします。GBIC-LW の仕様は次のとおりです。

機能	詳細
準拠仕様	ファイバチャネル FC-PH-2 物理層オプション 100-M5-SN-I
ポーレート	1062.5MB/s
ファイバ長波	9 μ m シングルモードファイバ
レーザ波長	1250nm
光コネクタのインタフェース	デュアル SC
距離	リンクあたり 10km

GLM(ギガビット リンクモジュール)



GLM

GLM は、高度に統合された光ファイバトランシーバであり、高速な双方向連続スループットを提供します。GLM は、電気信号と光信号を相互に変換する点で、GBIC と類似しています。しかし GLM は、それに加えてシリアル信号とパラレル信号間の変換も行います。

GLM の両端には、ポイントツーポイント二重構成のプラグが付いているので、それらを HBA やファイバチャネルハブまたはスイッチなどのホストカードに挿入します。64 ビット/33MHz HBA は、GBIC ではなく GLM を使います。

GLM のポートには、シングルモードまたはマルチモードのファイバチャネル SC 二重コネクタが挿入できます。

GBIC と同様、GLM も LW (長波) 構成と SW (短波) 構成の両方が可能です。

ストレージアレイとドライブアレイ



Enterprise Virtual Array 5000

ストレージアレイは、外部ドライブエンクロージャで、コントローラ、電源、ファンアセンブリ、ディスクドライブハウジングが入っています。アレイ構成要素の中には、ホットスワップ可能で、インストールや交換がダウンタイムなしで行えるものもあります。

ドライブアレイとは、論理的に1つの大容量ドライブとして扱われる複数の固定ディスクドライブです。1つのアレイが複数のドライブを格納するので、コンパクトなユニットで大容量ストレージを供給します。アレイは、タワーマウントユニットとラックマウントユニットの両方で利用できます。

ドライブアレイの利点は、次のとおりです。

- 高速データ転送
- 同時に複数の要求を処理
- ストレージ容量の拡大
- 柔軟なデータ構成
- 高信頼性

ファイバチャネルアレイコントローラ



HSG80



MSA1000



RA4100

ファイバチャネルアレイコントローラは、アレイのインタフェースポイントとなります。ファイバチャネルは、サーバに高速接続するために、アレイコントローラ/ホスト間インタフェースを提供します。このタイプのインタフェースは、100MB/sのバーストレートでサーバにデータを転送できます。

ファイバチャネルハブ



ファイバチャネルハブ

ハブは、個々の要素をリンクして、帯域幅を共有する1つのループを形成します。ハブは、次のような共通の特徴を持っています。

- 複数のアレイを1つのホストアダプタに接続できるので、サーバが利用できるストレージが大幅に拡大する
- ループデバイスを物理的スター型トポロジに相互接続する。これにより、すべてのワークステーションが1台のコンピュータを中心にネットワーク化されるので、配線とケーブルの管理が容易になる
- 統合が簡単で、比較的低コスト
- FC-ALトポロジを実装
- 共有 100MB/s FC-AL への接続(サーバ/ディスク間における帯域幅共有の半二重接続)が可能
- ループの初期化時のエラー修復が複雑
- 他のノード間のすべてのトラフィックを検知

ハブには、次のような機能があります。

- シングルモードまたはマルチモードの光ファイバケーブルのサポート
- 保守または修理の際、相互接続ケーブルのホットスワップが可能
- ケーブルまたはノードに障害が発生した場合の障害を隔離する
- 障害ステータスとトラフィックモニタ機能
- SANの実装が可能

ファイバチャネルスイッチ



ファイバチャネル SAN スイッチ

スイッチは、複数のノードを相互接続します。ファイバチャネル環境でのスイッチのネットワークはファブリックと呼ばれ、最高 1,600 万までのノードを設定できます。ノードは、このファブリックに接続することによって、他のノードにアクセスします。

スイッチは、通常、パケットスイッチングを使います。スイッチは、メッセージをパケットと呼ばれる小さなセクションに分割し、個々のパケットに送信元と送信先のネットワークアドレスを追加します。パケットは、送信先までどの経路を通ってもよく、再アセンブリしてから配布されます。

スイッチの主要機能は、送信元ノードからフレームを受信して、送信先ノードにルーティングすることです。各ノードには一意のファイバチャネル アドレスがあり、スイッチはこのアドレスを使って、フレームをルーティングします。スイッチは、それが担当する各ポートのステーション管理の負担を軽減します。各ノードが管理するのは、それ自体とスイッチの間の単純なポイントツーポイント接続のみです。ノードは、サーバ、ストレージデバイス、ネットワーク経由で通信を行う相手方デバイスのいずれかです。

ファイバチャネル スイッチには、次の特徴があります。

- FC-SW を実装
- 500m の長さの光ケーブルをサポート(GBIC-LW では、10km)
- 保守または修理の際、相互接続ケーブルのホットスワップが可能
- ケーブルまたはノードに障害が発生した場合に障害を隔離する
- ファイバチャネルテクノロジーを使ったストレージネットワークの実装が可能
- カスケード接続によって、トポロジを拡張できる
- SNMP (Simple Network Management Protocol) を使って管理や制御ができる
- ファイバチャネル ファブリックで全二重 100MB/s のポイントツーポイント接続を実現するので、サーバがループを使用できる(結果として、ループのスケーリングが改善され、パフォーマンスが最適化される)
- ファブリック内のノードを再構成したり、エラー修復したりする場合に、そのノードを他のノードから隔離する
- インテリジェント機能を内蔵する
- 拡張が可能
- 複数の方法でスイッチ管理が可能

ファイバチャネル SAN スイッチ



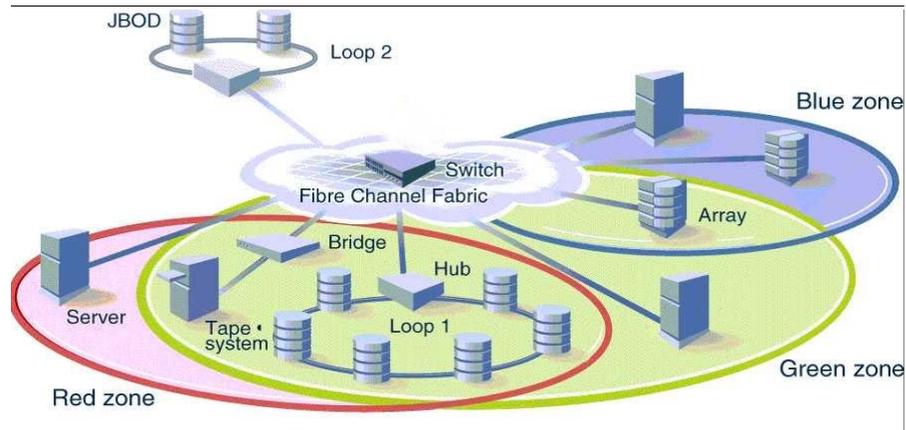
SAN スイッチ 16EL

ファイバチャネル SAN スイッチは、ネットワークサーバを、MA8000 や Enterprise Virtual Array (EVA) などのストレージデバイスに接続します。

以下の表に、2 種類のファイバチャネル スイッチの機能と使用に適した環境について説明します。

相互接続デバイス	機能
FC-AL スイッチ	使いやすくて手頃な価格の高性能 x86 ベースの SAN で推奨される(異種オペレーティングシステム)。ワークグループや部門の環境に最適である。将来的には、ファブリックスイッチに接続できる。ポートあたりのコストが低い。
SAN スイッチ	大規模、高性能異種(x86 および RISC)ファブリックベースの SAN において、業務の継続的遂行の付加レベル(ミラー化や災害耐性)が求められる場合に推奨される FC-AL スイッチより、ポートあたりのコストは高いが、スケーラビリティや機能性の水準も高い。

ファイバチャネル スイッチのゾーニング



ゾーニングとは、サーバとそれに結合したストレージサブシステムをグループとして扱い、そのグループ間のストレージ I/O トラフィックを分離する手法です。

ゾーニングのタイプは次のとおりです。

- **ソフトゾーニング** — ファイバチャネル スイッチは、ゾーンに属するポートを、ゾーンに属さないポートからマスクするためのフィルタリングを行います。
- **ハードゾーニング** — ハードゾーニングは、ハードウェアによって実装され、外部ノードがそのゾーンにアクセスできないよう物理的にブロックします。
- **ポートゾーニング** — ポートゾーニングとハードゾーニングは同じだと思われていることがよくあります。しかし、ポートゾーニングは、ソフトウェアによる実装方法であり、未許可の外部アクセスを可能にします。
- **ブロードキャストゾーニング** — ブロードキャストゾーニングは、ブロードキャストメッセージをフィルタリングして、それらが指定されたポートに送信されないようにします。

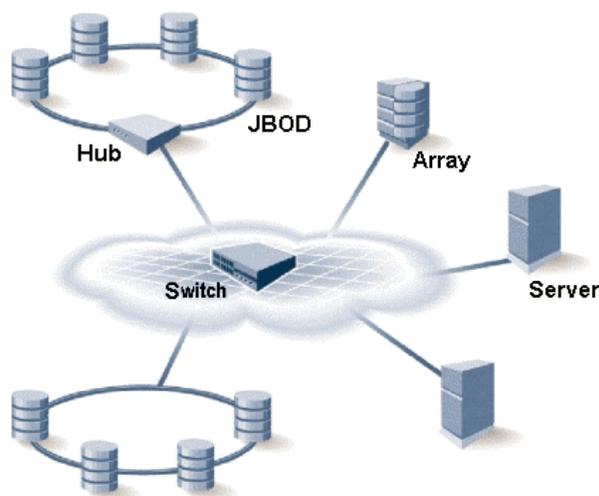
各ゾーンには、1 つまたは複数のゾーンメンバからなるメンバリストがあります。ゾーンメンバは、WWN (World Wide Name)、ポート WWN、またはスイッチのドメイン ID とポート番号によって指定できます。

1 つのデバイスを、複数のゾーンメンバに指定できます。しかし、ゾーニングを有効にした時点で、ファブリックのすべてのデバイスが 1 つまたは複数のゾーンのメンバでなければなりません。

▲ 注記

SAN でゾーニングを実装している場合は、サーバやストレージサブシステムを追加または削除するたびに、ゾーニング情報を変更する必要があります。また HBA を交換した場合もゾーニング情報のアップデートが必要です。

ハブとスイッチの比較



ハブは比較的低コストで統合が容易なので、最近までは、デバイスを接続するためのエントリレベルのソリューションでした。

スイッチは、ハブとまったく同じくらいインストールが簡単ですが、接続機能はハブより優れています。ハブに 127 ノードという上限があるのに対し、ファイバチャネル スイッチは、ほとんどあらゆるサイズにシステムを拡張できます。

以下の表では、ファイバチャネルネットワークでのハブとスイッチを比較します。

ハブ	スイッチ
FC-AL トポロジを実装。すべてのデバイス間で帯域幅を共有する。	FC-SW トポロジを実装。ファブリックデバイスかアービトラード ループデバイスのポートを使う。
共有 100MB/s FC-AL に接続し、一度に 1 つの 100MB 転送を行う。	デバイス間を 100MB/s のポイントツーポイントで接続し、一度に複数の 100MB 転送を行う。複数のデバイスアービトレーション時間はない。
ノードが増えると、アービトレーション時間が増加するので、パフォーマンスが低下する。	ノードを追加しても、パフォーマンスは低下しない。
ファイバチャネルループ上のノードは、他のノード間のトラフィックもすべて検知する。	ファブリックチャネル上のノードは、それ自体に与えられたデータのみを検知する。
エラー修復のためにループ初期化が必要。	SAN 内のノードを再構成したりエラー修復したりする場合に、そのノードを他のノードから隔離する。
再構成時にループ初期化が必要なため、トラフィック中断やエラーが生じやすい。	制御機能によってループの初期化イベントが管理されるので、再構成時のトラフィック中断やエラーが生じにくい。
HP の 12 ポートハブに、オプションの管理機能がある。	ファイバチャネル インフラストラクチャの管理制御機能を備える。
7 ステップのループ初期化プロセス (LIP)。デバイスでエラーが発生すると、LIP が開始される。エラー修復中は、ループをデータ転送に使えない。	エラー修復中の安定性が高い。

ハブは次の環境で使います。

- 1～4 台のサーバを使うアプリケーション
- 同種のプラットフォームとオペレーティングシステム環境で動作する、コスト重視の Windows 2000 アプリケーション
- アービトレーテッドループ構成の共有帯域幅が制限要因にならないアプリケーション

スイッチは次の環境で使います。

- 5 台以上のサーバを使うアプリケーション
- 再構成や修復のためのトラフィック中断の影響が大きいマルチノードクラスタ構成
- 異種プラットフォームと異種オペレーティングシステムのアプリケーション
- アービトレーテッドループトポロジが適さない高い帯域幅のアプリケーション

ファイバチャネルケーブル



マルチモードファイバチャネルキット、ケーブルおよび GBIC 付き

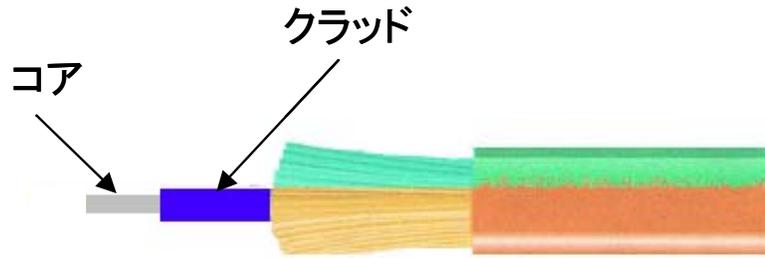
ファイバチャネル仕様には、銅ファイバと光ファイバの 2 種類のケーブルが規定されています。銅ファイバは、普通、メートル単位 (キロメートル単位ではなく) の短い距離に使用します。銅ファイバは、光パルスではなく電気パルスを使い、光ファイバより安価です。



重要

HP は、銅ケーブルをサポートしていません。ファイバチャネルの実装では、光ファイバケーブルを使っています。

光ファイバ



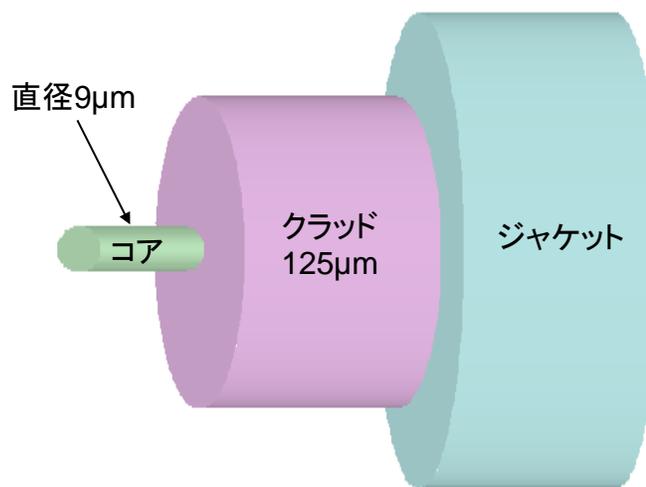
光ファイバケーブルは、コアとクラッドという2つのアクティブな要素で構成されています。

コアは、中心部にある光を伝達する部分で、光信号を伝達します。クラッドは中間の層で、光の屈折率がコアより大きいので、光はコアに閉じ込められます。

光ファイバの一番外側の層は、ジャケットまたは被覆と呼ばれています。これは、コアとクラッドを損傷から守るための緩衝材です。

光ファイバは、シングルモードかマルチモードのどちらかです。シングルモードファイバとマルチモードファイバの主な違いは、コアの物理的な直径です。

シングルモードファイバ



シングルモードファイバ

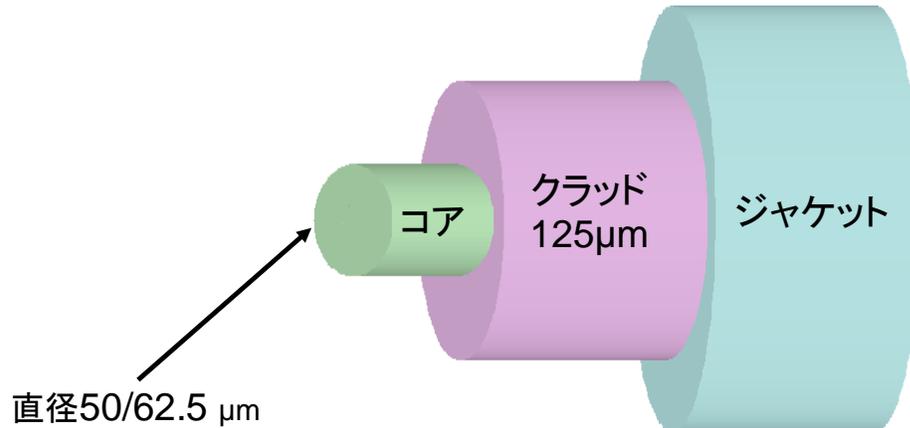
シングルモードファイバは、帯域幅が最大で、パフォーマンスの損失が最小です。コアは非常に小さいので、通過できるのはシングルモードの光だけです。したがって、色分散やモード分散が非常に少ないです。

シングルモードファイバの情報伝達能力は無限です。シングルモードファイバは、数十 GB/s の速度をサポートし、多くのギガビットチャネルを同時に伝達できます。それぞれのチャネルは、異なる波長の光を伝達し、干渉は起こりません。

シングルモードファイバは、長距離通信に適した媒体です。また、ビル間の配線や高速バックボーンのネットワークとしても利用できます。

ネットワークやデータ通信で使われるシングルモードファイバは、シングルモード ステップインデックスファイバです。

マルチモードファイバ



マルチモードファイバ

マルチモードファイバは直径が大きく、複数の光の流れが送信側から受信側までの異なる経路を通ります。短波レーザはマルチモードファイバを使って、中距離を伝送します。

ネットワーク (FDDI とイーサネット) で最もよく使われているマルチモードファイバには、次の特徴があります。

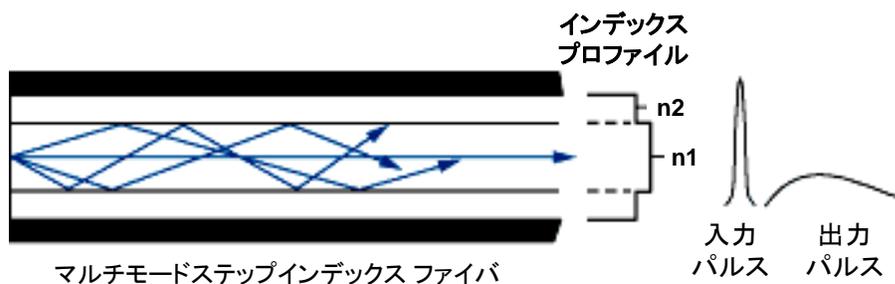
- 50µm/125µm (主に欧州で使われ、世界で最も一般的)
- 62.5µm/125µm (主に米国で使われている)

マルチモードファイバは、100MB/s (1062.5Mb/s) の速度で、最高 500m まで情報を伝送します。

マルチモードファイバには、次の 2 つのタイプがあります。

- ステップインデックス ファイバ
- グレーデッドインデックス ファイバ

マルチモードステップインデックス ファイバ

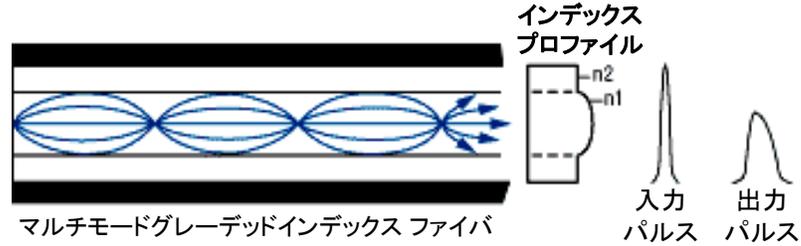


ステップインデックスファイバは、モード分散が大きくなります。コア材料 (n_1) が、連続的に屈折率を変化させる(グレード)構造になっていないため、光ビームの焦点がコアの軸に合いません。同時に送信元から発せられた光線が、送信先に到着するときにはかなりの時間差を生じるので、ファイバの帯域幅が小さく、距離が短くなります。

ステップインデックスファイバには、次の特徴があります。

- 安価
- 帯域幅が小さく、距離が短い
- グレーデッドインデックスファイバよりモード分散が大きい
- ネットワークやデータ通信には、ほとんど使用されない

マルチモードグレーデッドインデックス ファイバ



光ファイバのコアが連続的に屈折率を変化させる(グレード)構造になっているので、光ビームの焦点がコアの軸に近づきます。伝達される光線が軸に近づくので、軌跡が短くなり、到着速度の差が小さくなります。

グレードは、コア材料(n_1)の化学組成を変えることで実現されます。

グレーデッドインデックス ファイバには、次の特徴があります。

- ステップインデックス ファイバより高価。製造技術の進歩と光ファイバケーブルの普及によって、製造コストが大幅に削減された
- 帯域幅が大きく、距離が長い
- 信号の伝送精度が高いため、ステップインデックス ファイバよりモード分散が少ない
- ネットワークやデータ通信でよく使用される (ネットワークで使用されるマルチモードファイバのほとんど全部がグレーデッドインデックス)

INTERNET グレーデッドインデックス ファイバの詳細は、次の Web サイトを参照してください。
<http://www.fibrechannel.org/>

安全上の注意

レーザから出力される光によって目を損傷する可能性があるため、使用できる光学出力が規則で定められています。たとえば、クラス 1 レーザ安全基準は、許容できる安全レベルを規定しています。

ファイバチャネル業界は、オープンファイバ制御 (Open Fiber Control) テクノロジーを採用して、これらの安全要求事項を満たしています。オープンファイバ制御 (Open Fiber Control) とは、ラインドライバ構成要素が光回路の故障を検出して、レーザ出力を抑制する機能です。

▲ 注記

オープンファイバ制御 (Open Fiber Control) ケーブルは高出力レーザを使用し、非オープンファイバ制御 (Non - Open Fiber Control) ケーブルは低出力レーザを使用するため、信号方式が異なり互換性はありません。

減衰

減衰とは、信号が伝送される間にパワーを損失することです。減衰は、kmあたりのデシベル (dB/km)で測定します。

一般に市場で入手できるファイバの減衰は、シングルモードファイバの約 0.5dB/km から、大容量コアプラスチックファイバの 1000dB/km の範囲です。

高品質で高価なシングルモードファイバでは減衰が小さく、低品質で安価なマルチモードファイバでは減衰が大きくなります。

パワーの損失は、次の原因によって生じます。

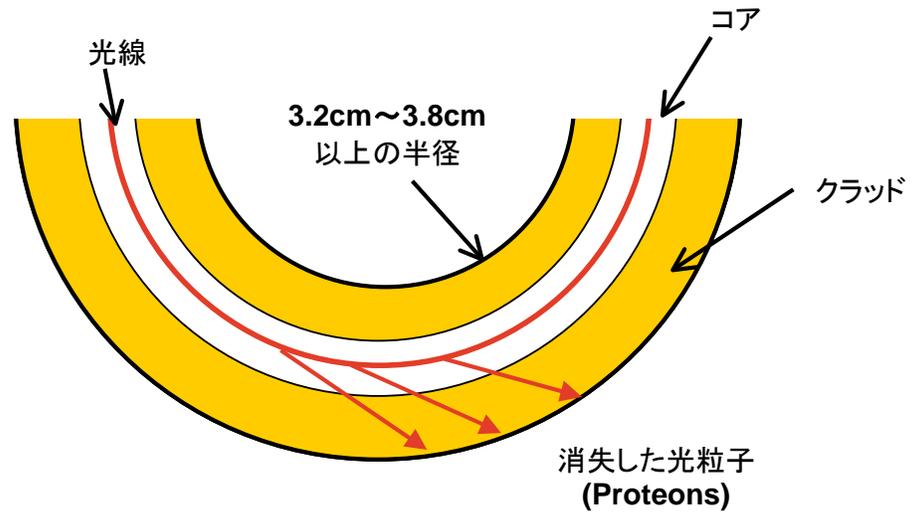
- 材料に含まれる不純物によって生じる光の吸収
- 材料に含まれる不純物や、コアとクラッドの接合部分の不具合、媒体(シリカ)の分子の散乱によって生じる光の散乱
- マクロベンド(規定の半径を超えたケーブルの湾曲)とマイクロベンド(ケーブルのねじりや締め付け)
- ケーブル接合部分での散乱と反射

減衰量は光の波長によって異なります。低損失ウインドウには、次の3つがあります。単位はナノメートルです。

- **780~850nm** — 780~850nm のデバイスは、価格が手頃なので、最もよく使われています(短波)。
- **1200~1300nm** — 損失は減少しますが、LED のためのコストが少しかかります(長波)。
- **1550nm** — 主に長距離電気通信アプリケーション用です(VLD – Visible Laser Diode)。

光ファイバの減衰は、主に、コア材料内の散乱と吸収が原因です。接続部分の端面が荒い場合や、接続状態が悪い場合も、光強度が減少します。したがって、長距離伝送や接続点では損失が起こります。またコネクタでは、その品質に関わらず、挿入損失が発生します。

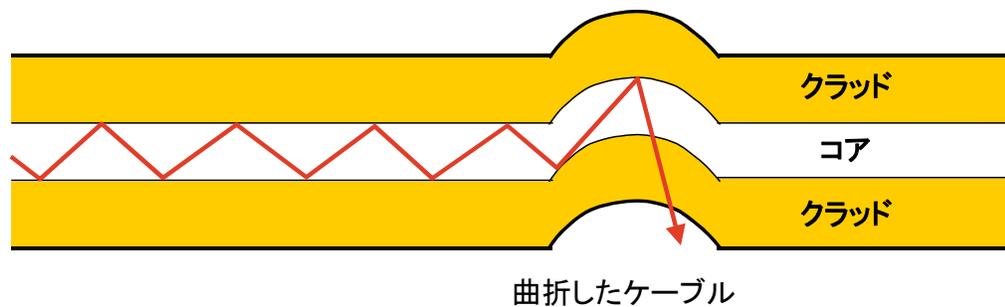
マクロベンド



マクロベンドとは、規定の半径(3.2~3.8cm)未満にファイバケーブルを物理的に屈曲させることです。

ファイバが規定の半径未満に曲げられると、光粒子の一部が消失し、減衰量が増加します。

マイクロベンド



マイクロベンド損失は、ケーブルが紐で結んであったり、クラッドが締め付けられたりして、ビームが完全にリニアパスを通過しないときに起こります。

光ファイバの軸でマイクロベンドが起こると、それを通過する光伝送量が大幅に減少します。

ファイバチャネルトポロジ

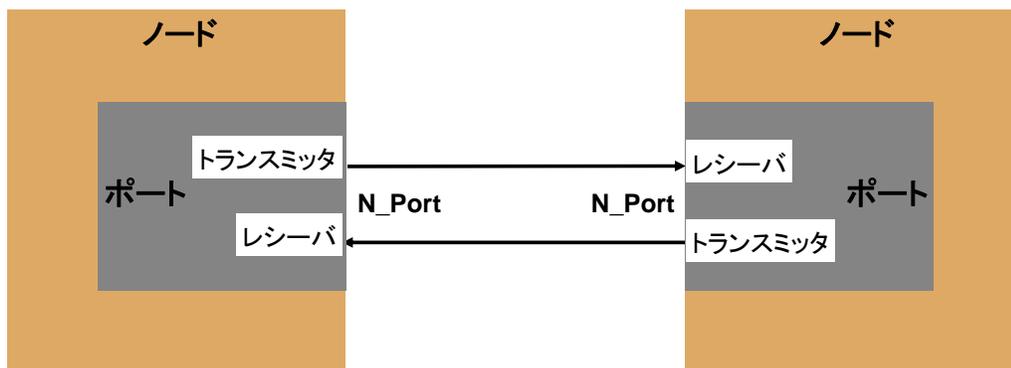
ファイバチャネルは、次の 3 つのトポロジをサポートします。

- ポイントツーポイント
- FC-AL
- FC-SW

ここでの説明を理解するには、以下の表に示すさまざまなファイバチャネル ポートのタイプを把握する必要があります。3 つのトポロジについて学習する際、必要に応じてこの表を再確認してください。

ポート	タイプ	対応するトポロジ
N_port	ノード	ポイントツーポイント、またはファブリック
NL_port	ノード	アービトレーテッドループに接続されているノード
F_port	ファブリック	ファブリックポート
FL_port	ファブリック	アービトレーテッドループに接続されているファブリック
L_port	ループ	アービトレーテッドループ上のハブポート
L_C_F	ループ	Link Control Facility (L_C_F) は、リンクの両端に接続されたハードウェア装置で、データの送受信を管理する。これは各ポート内に取り付けられ、トランスミッタとレシーバを持ち、ノードへの論理インタフェースとなる。
Rx		レシーバ (Rx) は、link control facility の一部であり、エンコード化されたビットストリームを媒体から受信し、それを伝送キャラクタに変換する。FC-PH によって指定されたルールを使ってこれらのキャラクタをデコードする。
Tx		トランスミッタ (Tx) は link control facility の一部であり、伝送コードルールに従って有効なデータバイトと特殊コードの伝送キャラクタへの変換処理をする。Tx はこれらのキャラクタをビットストリームに変換し、そのビットストリームを光または電気の伝送媒体へ送信する。
G_port		E_port モードまたは F_port モードで動作する汎用スイッチポート。
E_port		このスイッチ間拡張ポートは、もう 1 つのスイッチの E_port に接続してファブリックを拡張する場合に使用する。

ポイントツーポイントトポロジ

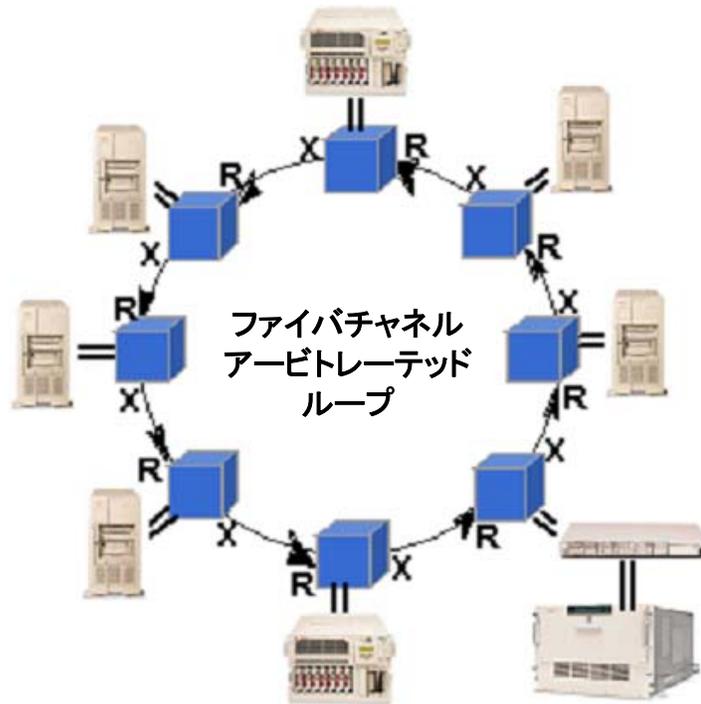


最も単純なファイバチャネルトポロジは、1つのリンクが2つのポートのみを接続するポイントツーポイントです。このトポロジは、ハブが不要なので低コストです。

大規模なポイントツーポイント構成を構築するために、各ノードに複数の N_port を設置することもできます。

ポイントツーポイントの各接続では、N_port によってサポートされる全帯域幅が利用できます。2つのノード間の最大距離は、リンクのタイプ(マルチモードファイバまたはシングルモードファイバ)によって、500m(マルチモードファイバ)か 10km(シングルモードファイバ)のいずれかです。

ファイバチャネル アービトレテッドループトポロジ



ノードをファイバチャネルハブに接続することによって、FC-AL を構築できます。

FC-AL では、容量は拡張されますが(1つのループ上で最高 126 ノードをサポート)、パフォーマンスはそのままです。帯域幅は、ループ上のアクティブなノードすべてが共有します。

このトポロジで、同時に接続できるのは、1 対のポート間のみです(2つのデバイス間が通信する)。セッションが開始されると、ループに接続されている他のデバイスは、その接続が終了するまで接続できません。

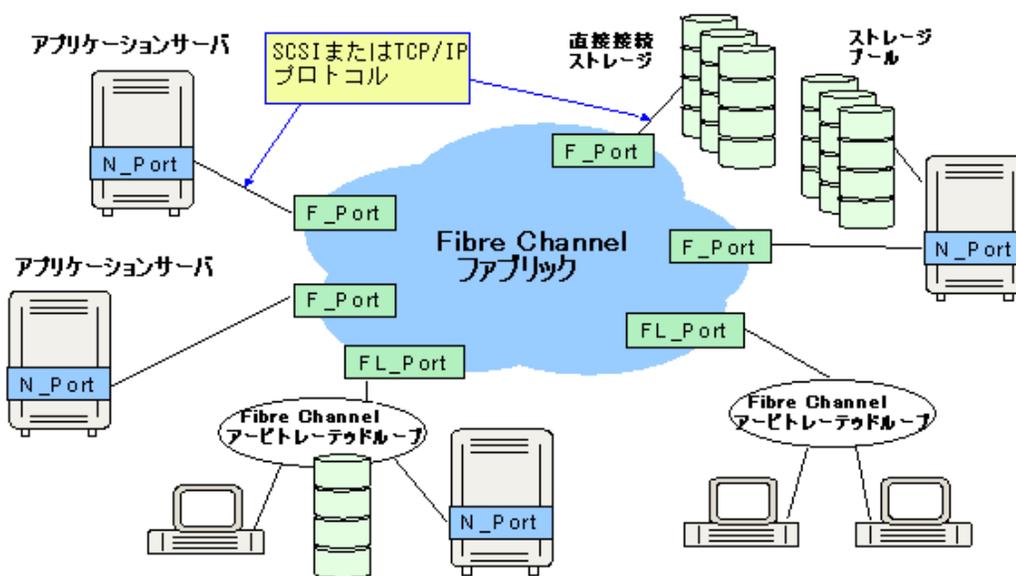
アービトレテッドループに属するデバイスは、アクセスを共有しますが、アクティブなリンクが全帯域幅を使用します。そのループ上で同時に通信できるのは、1 対のノードのみです。

アービトレテッドループトポロジでは、複数のデバイスが、それらをファイバ接続する 1 つのループの帯域幅を共有できます。FC-AL 標準は、N_port を NL_port ポートに変更することによって実装されます。各 NL_port は 1 本のリンクに接続されています。情報は、アービトレテッドループを一方方向にのみ流れます。

アービトレテッドループは、論理的にはループであり、物理的にはスター型です。上の図では、X が 1 つのノードの出力を、R がもう 1 つのノードの受信側ポートを表します。1 つのノードの送信側ポートが、別のノードの受信側ポートに接続されています。情報がループ上で一方方向に流れていることに注意してください。

アービトレテッドループトポロジでは、各ループポートがそれぞれルーティング機能を持っています。これは、送信ファイバと受信ファイバの分離によって、実現されています。

スイッチファブリックトポロジ



ファイバチャネルスイッチファブリック

ファイバチャネル環境におけるスイッチネットワークをファブリックと呼びます。ノードは、このファブリックに接続することによって、他のノードにアクセスします。非常にオープンなアーキテクチャでは、インテリジェントスイッチを使って多くのポートに接続します。

ファイバチャネルファブリックは、ノードと物理層との間の汎用インタフェースとして設計されています。このインタフェースに従うことによって、ファイバチャネルノードは、そのノードについて情報がなくても、ファブリックを経由して他のノードと通信を行えます。

ファブリックは、しばしば「スイッチトポロジ」と呼ばれます。フレームが各ファブリック要素に到達するたびに、フレーム内の送信先アドレス識別子がファブリック要素によって解釈され、それをもとにフレームがさまざまなスイッチをルーティングされます。

ノード上のポートは、同じファブリックに接続されている他のノード上のポートと通信を行えます。ファブリックトポロジでは、同時に多くの接続をアクティブ化できます。

ファブリックは、any-to-any (どこからでも、どこへでも)の接続サービスとピアツーピアの通信サービスを提供しており、これがファイバチャネルアーキテクチャの基礎となっています。ファイバチャネルは、チャネルプロトコルとネットワークプロトコルを同時にサポートできます。

サービス使用のクラスによって、フレームのルーティングが左右されます。サービス要求のクラスは、各フレームのフレームデリミタの初めに生成されます。

ファブリックの特徴

スイッチファブリックポロジには、次の特徴があります。

- 最高 1,600 万ノード
- スイッチがルータのように動作 — 通信を行っているポートのみがトラフィックを検知
- 集約性の高い帯域幅 — 複数のパスを同時に設定できる
- 電氣的隔離と論理的隔離
- FC-AL より高コスト

スイッチファブリックは、次のように動作します。

- ホストにあるファブリックへの直接接続は、NL_port プロトコルをサポートするインタフェースを介して行われます。
- スイッチ自体には F_port があり、これに他のデバイスからのケーブルが接続されます。FC-AL グループ全体を、FL_port 経由でスイッチに接続できます。
- ほとんどの FC-AL グループは、それらのハブを介してスイッチに接続します。また一部のスイッチでは、ハブを使わず、FL_port によって直接ループを接続できます。

ファブリック要素

ファブリックポロジでは、ポート間に少なくとも 1 つのアクティブな要素 (スイッチ、リング、ハブなど) が必要です。この要素は、ファブリックログイン要求への応答、ファブリックのサービスクラスの管理、アドレスの割り当てなどの機能を実行します。

ファブリック要素には、次の機能があります。

- N_port の接続
- アービトレテッドループの接続
- 分散ファブリック要素のルート要素として機能

ファブリック要素は、完全なファブリックポロジとして機能できる最小のエンティティです。ファブリック要素には、ルーティングを決定するために、最低 3 つの F_port または FL_port が必要です。

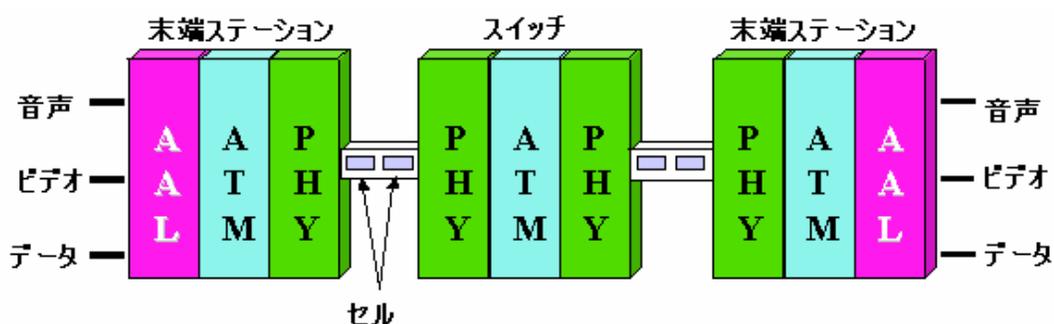
複数のファブリック要素を相互接続すると、それらは協同ユニットを形成し、それがまた 1 つのファブリックとなります。ヘテロジニアス (異種) ファブリックという用語は、複数のファブリック要素で利用できる多様な機能を意味します。

新しいテクノロジー

革新的技術は、世界的な標準の欠如から生まれることが少なくありません。現在、エンドシステムや情報の種類にかかわらず、情報の相互運用が可能な通信プロトコルが特に求められています。

ATM (Asynchronous Transfer Mode) の登場とともに、新しいテクノロジーは 1 つのメーカーの視点や戦略によってではなく、国際的な合意の中から生まれるようになっていきます。IP over Fibre Channel が開発されたのは、既存のファイバチャネル標準では、ファイバチャネル経由で IP パケットを伝送する方法や、IP アドレスをファイバチャネルアドレスに解決する方法が十分に規定されていないからです。

ATM (Asynchronous Transfer Mode)



ATM (Asynchronous Transfer Mode) は、非常に高速のデジタル通信方式で、音声、ビデオ、データ、および画像の伝送に使われます。ATM は、LAN テクノロジーと WAN テクノロジーの両方の基礎として使えます。

音声、データ、ビデオ情報は、それぞれ別のネットワークで伝送することが多いのですが、それは、これらのトラフィックタイプの特徴が異なるから、というのが大きな理由です。ATM を使えば、ネットワークを分離する必要はありません。ATM は、データ、音声、ビデオの同時伝送に対応することを最初から想定して設計された、唯一の標準ベースのテクノロジーです。

LAN 通信を電気通信と統合しようとする試みが広がるにつれ、ATM が支持されるようになってきています。ATM は公衆データ網での広域バックボーンテクノロジーとしても普及しており、これらの両方に対応するハードウェアと通信プロトコルが、ATM をベースとして開発される可能性があります。

標準の ATM 通信は小さなパケットサイズ (53 バイト) の上に構築されており、データファイルの転送に使う場合には、これが非常に大きな技術的課題となっていました。また、ATM パケットのドロップは、音声やビデオデータの転送には重大な影響がなくても、データ通信では許容できないものでした。

これらの問題を解決するために、ATM の改良版 — AAL-5 (Asynchronous Adaptation Layer 5) — が規格化され、コンピュータデータ用のスケーラブルなシステム間通信を提供しています。

ゲートウェイ インタフェースによって、ファイバチャネルスイッチが ATM ネットワークに接続され、ローカルサイトとリモートサイト間ファブリックの距離の延長が可能です。ATM では、メガビットからギガビットまでのさまざまな速度が可能です。

IP and ARP over Fibre Channel

IP and ARP over Fibre Channel は、Internet Engineering Task Force が開発した IP ベースのストレージネットワークングテクノロジーで、ファイバチャネルトンネリングとも呼ばれます。IP and ARP over Fibre Channel では、SAN ファシリティ間のデータを IP ネットワークでトンネリングすることでファイバチャネル情報の伝送が可能になります。この機能により、地理的に分散された拠点を持つ企業内でのデータ共有が容易になります。

ノード間のファイバチャネル経由の IP 通信は、単方向交換を作成することで行われます。受信側ノードが応答する (双方向の相互作用) には、交換をもう 1 つ作成する必要があります。

IP アドレスが 4 バイトで、ファイバチャネルアドレス識別子が 3 バイトなので、これら 2 つのアドレスを相互にマッピングするために、Address Resolution Protocol (ARP) サーバが必要です。通常、ARP サーバはファブリックレベルで実装され、それ自体のアドレス (x 'FF FFFC') を持ちます。

IP パケットセットは、ファイバチャネルレベルで順次に処理されます。IP パケットの最大サイズは 65,280 バイトなので、64KB のバッファの中に、パケットと 255 バイトのオーバーヘッドが入ります。

フレームエラーが発生すると、デフォルト交換エラーポリシー (中断、1 つのシーケンスを廃棄) は、そのパケットを廃棄し、再送信を行いません。後続のシーケンスには影響はありません。再送信が必要な場合は、IP および伝送制御プロトコル (TCP) レベルで処理されます。この処理はファイバチャネルには認識されません。

復習問題

1. ファイバチャネルの 5 つの利点を挙げてください。

2. 各用語とその説明を対応付けてください。

a.	ノード	ファイバチャネルポートの 1 対の光ファイバ
b.	ノード ID	ホストやアダプタなどのデバイス
c.	リンク	2 つ以上のポートを接続する相互接続構造
d.	トポロジ	デバイスに割り当てられている一意の識別子
3. ファイバチャネルは、業界標準の相互接続と高性能シリアル I/O のプロトコルであり、サーバ業界において、これまでになく水準の信頼性、スループット、距離についての柔軟性を実現します。
 - 正
 - 誤
4. 次のうち、ストレージシステムソリューションで使われる 3 つのファイバチャネルトポロジはどれですか。
 - a. パラレルツーシリアル、SCSI ID、スイッチファブリック
 - b. ポイントツーポイント、アービトレーテッドループ、スイッチファブリック
 - c. ポイントツーポイント、アービトレーテッドループ、RAID 0
 - d. ポイントツーポイント、ファブリックアービトレーテッドループ、ストレージファブリック

5. GBIC (Gigabit Interface Converter)とは何ですか。
.....
.....
6. ファイバチャネルハブを使うのは、どのファイバチャネルトポロジですか。
.....
7. ファイバチャネルスイッチの E_port にはどのような機能がありますか。
.....
.....
8. 2種類のファイバチャネルケーブルは何ですか。
.....
.....
9. 次のうち、減衰についての最も適切な記述はどれですか。
 - a. 特定の半径を超えたケーブルの屈曲
 - b. 信号を伝送する間に生じるパワーの損失
 - c. 真空内での速度比率
10. ファイバチャネル環境で、スイッチのネットワークは何という名前と呼ばれますか。
.....

目的

この章では、次の内容について学習します。

- 以下のデータストレージテクノロジーの利点と特性、およびそれらの違い
 - Direct Attached Storage (DAS)
 - Network Attached Storage (NAS)
 - Storage Area Network (SAN)
- HP の標準 SAN トポロジで使用されている設計
 - カスケード
 - メッシュ
 - リング
 - バックボーン SAN ファブリック
- Enterprise Network Storage Architecture (ENSA) の目的
- VersaStor テクノロジーの機能、利点、および特性
- ENSAextended VersaStor テクノロジーの機能、利点、および特性
- テープバックアップテクノロジーの目的と利点
- テープバックアップ方法とバックアップテープのローテーション方式
- テープストレージテクノロジー
- エンタープライズバックアップソリューション (EBS) の利点
- 災害復旧計画を準備する際に考慮する要因と実行する手順

データストレージテクノロジー

ストレージテクノロジーは、以下の 3 つのグループに分割できます。

- **DAS (Direct Attached Storage)** — DAS ソリューションとしては、SCSI プロトコルを使用する専用の内蔵または外付けストレージサブシステムで顧客独自のアプリケーションを実行する、オープンシステムサーバがあります。このタイプのテクノロジーでは、複数のクライアント用の専用ストレージがあり、サーバとストレージの関係は 1 対 1 になっています。
- **NAS (Network Attached Storage)** — NAS サーバは、既存の LAN に直接接続される独立したインテリジェントデバイスです。ファイルシステムは、NAS デバイス上に配置されて管理され、データは業界標準のネットワークプロトコル (TCP/IP または IPX) を介し、業界標準のファイル共有プロトコルを使用して、クライアントに転送されます。NAS デバイスのこの機能により、異機種ネットワーククライアント間の完全なデータ共有が可能になります。
- **SAN (Storage Area Network)** — SAN ソリューションは、オープンなオペレーティングシステム上で顧客のアプリケーションを実行する複数のオープンシステムサーバです。ソリューションには、共有の外付けストレージリソース、ネットワークインフラストラクチャ用のコンポーネント (ファイバチャネルハブとスイッチ)、高度なストレージ管理とデータ管理のための付加価値のあるソフトウェアが含まれます。このテクノロジーでは、スケーラビリティとフォールトトレランス性の高い整理統合され仮想化されたストレージを実現できます。

Direct Attached Storage (DAS)



DAS/SAN SDLT

ほとんどの顧客は、独立したストレージシステムとストレージ管理ソフトウェア製品を所有していて、それらは個別のサーバに接続されています。この方法では、分散化された IT 環境全体のあちこちにデータの塊が発生します。このようなデータをつなぎ合わせるのは困難で、管理には莫大な労力が必要となります。

この方法の利点は以下のとおりです。

- 低価格
- 高速

この方法の欠点は以下のとおりです。

- **リソースの使用効率が低い**— 分離されたプールに記憶スペースが存在する場合、あるサーバには空きスペースがなくなっても別のサーバには 200GB の空きスペースがある、ということがあります。
- **計画外の冗長性** — 同じファイルの重複したコピーが、別のサーバに存在することがあります。
- **サーバベースの管理モデル** — すべてのストレージ管理作業は、サーバごとに実行する必要があります。

Network Attached Storage (NAS)

独立したストレージシステムの管理、運用の難しさから、別のストレージ方法として NAS が開発されました。NAS デバイスにより、異機種ネットワーククライアント間の完全なデータ共有が可能になります。これは主に、小規模ビジネス、支社、ワークグループ、または部門のネットワークにおけるストレージソリューションに使用されます。

NAS 機器はネットワークサーバに似ていて、クライアントはそこからファイルを取得することができます。これは、要求に応じてファイルを送信し、IP プロトコルを使用してクライアントへファイルを供給します。

小規模なプラグアンドプレイ NAS デバイスまたは NAS 機器は、ファイルの供給に特化した独自のオペレーティングシステムを持っています。大規模な NAS 構成の容量は、約 1TB です。NAS では、新しいサーバアーキテクチャが採用されているため、通常、ローカルディスクアクセスと同等かそれ以上の性能を持っています。

ネットワークにではなく、サーバに直接接続された RAID ストレージアレイが存在する環境では、ネットワーク内のすべてのデバイスは、RAID システムにアクセスするためにはサーバにアクセスする必要があります。NAS がネットワークに接続されている場合、すべてのネットワークユーザは、格納されているデータに平等にアクセスでき、サーバにアクセスする必要はありません。

NAS の利点

NAS の利点は以下のとおりです。

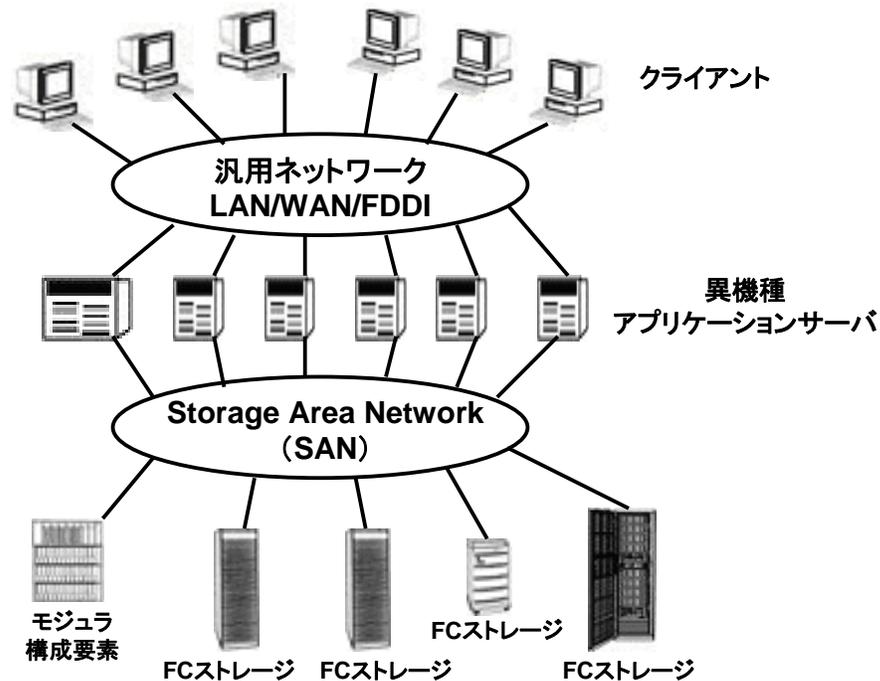
- インストールが簡単
- 比較的低価格
- SCSI などの成熟した技術を使用
- NAS デバイスを追加することで容易に拡張可能
- サーバからのファイル処理作業の負荷軽減による、ネットワークパフォーマンスの向上
- リモート管理

NAS の欠点

NAS の欠点は以下のとおりです。

- 限定的なパフォーマンスと記憶容量
- デバイスの増加による管理コストの増大
- コントローラと記憶装置間の距離制限
- 中央のサーバがシャットダウンされると、データにアクセスできなくなる
- ネットワークの問題が増加し、迅速なデータアクセスを阻害するボトルネックが発生する

Storage Area Network(SAN)



SANは、一元管理された専用の安全な情報インフラストラクチャで、これにより共通の記憶装置またはストレージプールへの直接的な物理アクセスが可能になります。

Storage Network Industry Association では SAN を「コンピュータシステムとストレージ要素間、およびストレージ要素間でのデータ転送を主目的としたネットワーク。SAN は物理接続を提供する通信インフラストラクチャと、データ転送が安全かつ確実に行われるように接続、ストレージ要素、およびコンピュータシステムを編成する管理層で構成される。」と定義しています。

SAN がサポートしている機能は、以下のとおりです。

- ディスクミラーリング
- バックアップと復元
- データのアーカイブとアーカイブデータの取得
- 記憶装置間でのデータの移行
- ネットワーク内の異なるサーバ間での重要なデータの共有

SAN は、NAS システムのあるサブネットワークを含むことができます。

Storage Network Industry Association では、ネットワークがファイバチャネルテクノロジーをベースにしている場合、ファイバチャネル SAN という用語を使用することを推奨しています。一般的に SAN はファイバチャネルをベースにしていますが、Enterprise System Connection (ESCON) などの他のテクノロジーをベースにしていることもあります。

SAN の利点

SAN の利点は、以下のとおりです。

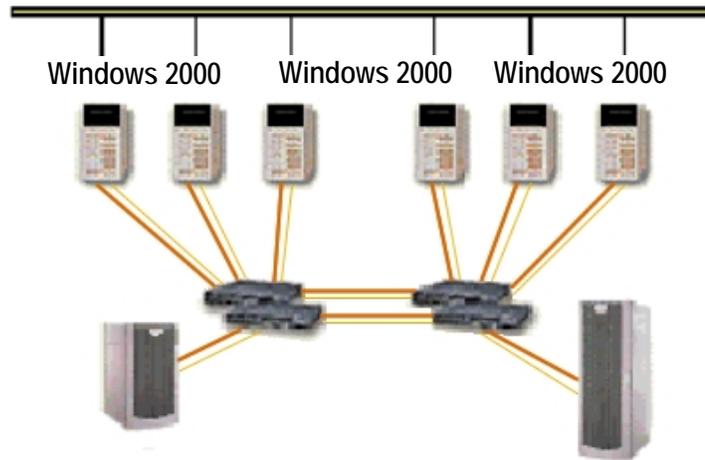
- ネットワーク転送経路でのストレージ I/O 方式の安全な実装として機能
- サーバと記憶装置の高速、スケーラブル、堅牢なネットワークを提供
- サーバクラスタ化アプリケーション用の標準プラットフォームとして機能
- 以下のようなストレージサービスを提供
 - 企業全体に散らばるサーバと接続
 - ある特定の場所にデータを集中
 - データを使用したり複製できるようにデータを自動的に移動
 - 高速なデータアクセスを提供
 - 高度なストレージ管理をサポート
 - 可用性の高い構成に対応
- 以下のストレージ管理機能を含んでいる
 - データの複製
 - 拡張とストレージの再割り当てが容易
 - 災害時復旧
 - 分散化されたストレージの一元管理

NAS と SAN の比較

NAS と SAN の違いを以下の表に示します。

機能	NAS	SAN
ストレージのタイプ	共有ファイル	仮想化されプールされたブロック
プロトコル、転送方式	TCP/IP、イーサネット	ファイバチャネル
対象	クライアントまたはサーバ	サーバ
容量	10TB 以上	数百 TB
管理	容易	容易

同機種 SAN



同機種 SAN ストレージポロジは、以下で構成されます。

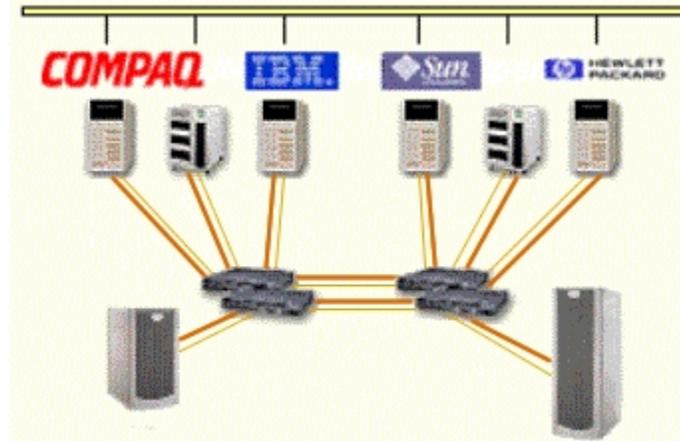
- 同機種オペレーティングシステム (SCO UNIX または Windows 2000 など) で動作する単一ベンダの複数のサーバに接続されている、プールされた共有ストレージ

▲ 注記

同機種 SAN は、SAN 内のすべてのサーバ上で同じオペレーティングシステムが実行されるネットワークです。

- 共通のシリアルスイッチ付きファブリックとネットワークスイッチを使用した、サーバおよび I/O (ディスクとテープ) 間の any-to-any (いつでも、どこでも) 接続
- サーバおよび I/O リソースの静的割り当て

エンタープライズ(異機種混在)SAN



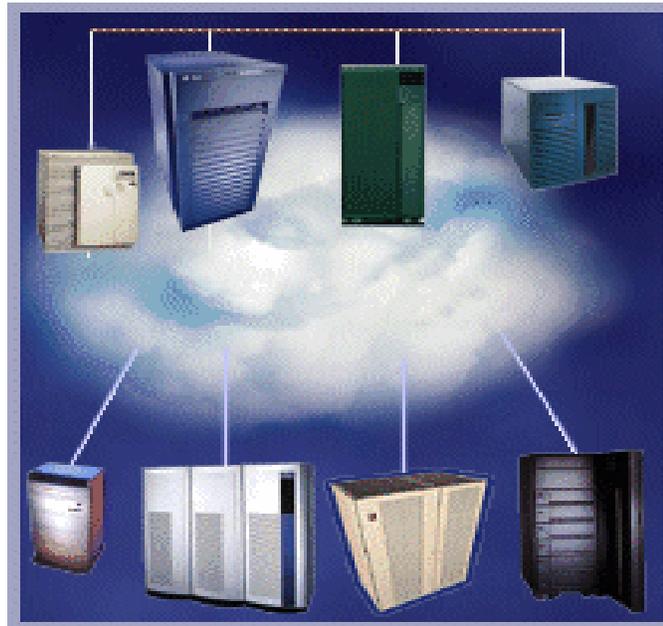
エンタープライズ(異機種混在)SAN ストレージトポロジは、以下で構成されます。

- UNIX (複数バージョン)、Windows NT、2000 または Windows 2003 オペレーティングシステムなどで動作する複数のサーバに接続されている、プールされた共有ストレージ
- 共通のシリアルスイッチ付きファブリックとネットワークスイッチを使用した、サーバおよび I/O (ディスクとテープ) 間の any-to-any (いつでも、どこでも) 接続
- サーバおよび I/O リソースの静的割り当てと、向上された管理機能

▲ 注記

異機種混在 (ヘテロジニアス) SAN では、SAN 内の複数のプラットフォーム上で複数のオペレーティングシステムが実行されます。

Open SAN



Open SAN ストレージテクノロジーは、以下で構成されます。

- 複数のベンダのストレージデバイスを使用し複数のオペレーティングシステムで動作するサーバに接続された、プールされた共有ストレージ
- サービスレベルと可用性を最適化するための、サーバと I/O リソースの動的割り当て制御機能
- 統一 SAN オペレーティングシステムまたはグローバルオペレーティングシステムのサポート

Open SAN は業界標準を採用し、以下の種類に依存しません。

- アプリケーション
- ファイルシステム
- オペレーティングシステム
- サーバプラットフォーム
- ストレージシステム
- テープライブラリ
- SAN 相互接続デバイス

SAN ベースの高可用性 - ビジネスの継続性

今日のビジネスコンピューティングニーズを満たすためには、継続的な可用性を実現できるように設計されたコンピューティング環境を実装することが重要になってきました。

SAN は、シングルポイント障害 (単一機器の障害がシステム全体の障害となること) を排除して、ビジネスアプリケーションの可用性を大幅に向上させることができます。ほとんどの高可用性設計には、クラスタリング、デュアルまたはマルチパス化ソフトウェア、および高可用性トポロジ設計の組み合わせが含まれ、HP の SAN はこれらすべてを完全にサポートしています。

SAN アーキテクチャのカテゴリ

SAN アーキテクチャでは、ファブリックの可用性は以下の 4 つのカテゴリに分類されています。

レベル 1: メッシュ化されていない単一のファブリック/単一のサーバストレージパス

これらの設計では、ファイバチャネルスイッチを単一リンクで接続して実装します。スイッチは、サーバとストレージが 1 つのパスを使用してファブリックに接続するように配置されます。これらの設計は、ファブリックまたはファブリックパスの冗長性は提供しませんが、ポート数と比較して最高の接続性レベルを提供します。

レベル 2: メッシュ化またはカスケード化された単一のファブリック/単一のサーバストレージパス

これらの設計では、スイッチ間に複数の ISL (inter-switch link) があり、ファブリック内のすべてのスイッチへの複数のパスまたはルートがあります。サーバおよびストレージは、1 つのパスを使用してファブリックに接続し、ファブリックの回復機能を提供します。1 つのスイッチポートまたは 2 つのスイッチ間のリンクで障害が発生した場合、ファブリックは自動的にデータを別のファブリックリンクまたはルートに再ルーティングします。サーバの I/O フローが中断することはありません。

レベル 3: メッシュ化またはカスケード化された単一のファブリック/複数のサーバストレージパス

これらの設計では、レベル 2 と同じ設計に加え、1 つのファブリックに接続するサーバとストレージ間に複数のデータパスがあります。レベル 3 には、ファブリックの回復機能と、複数のサーバストレージパスが存在するという利点があります。スイッチ、ホストバスアダプタ、またはパスで障害が発生した場合、データは自動的に代替パスに再ルーティングされます。サーバの I/O フローが中断することはありません。レベル 3 では、ファブリックゾーニングを使用して、1 つのファブリックに 2 つの別々のパスを定義する必要があります。

レベル 4: 複数のファブリック/複数のサーバストレージパス

レベル 4 は、レベル 3 と同様に、サーバとストレージ間の複数のデータパスを提供しますが、これらのパスは別々のファブリックに接続されています。この設計は、最高レベルの可用性を提供し、シングルポイント障害を防止するための最高の機能を提供します。ファブリックのパフォーマンスまたは有用性に影響を与えるあらゆるアクティビティは、代替ファブリックへデータをルーティングすることで改善されます。サーバの I/O フローが中断することはありません。

レベル 4 では、人的エラーなど、ファブリック障害を引き起こす原因が排除されます。これらの設計では、通常動作時にすべてのファブリックに同時にアクセスして使用できるため、最高レベルのパフォーマンスが実現され、使用可能なポート数が多くなります。

このレベルの保護は、選択した設計を複数の別々のファブリックに複製することにより、すべての HP 標準 SAN トポロジで使用できます。これにより実装の全体コストは増加しますが、データ可用性の向上と使用可能なポート数の増加という利点があります。

例

1 つのファブリック、複数のサーバストレージパスのある 4 つのスイッチメッシュで設計を実装すると、サーバとストレージ接続用のポートは最大 52 個になります。2 つのファブリックを使用して同じトポロジを実装すると、サーバとストレージ接続用のポートは最大 104 個になります。また、2 つのファブリックを使用した場合は、あるトポロジタイプから別のトポロジタイプへ安全に移行することもできます。すべての操作を 1 つのファブリックにフェイルオーバーし、サービスを停止することなく、別のファブリックを新しいトポロジに移行できます。

HP 標準 SAN トポロジ

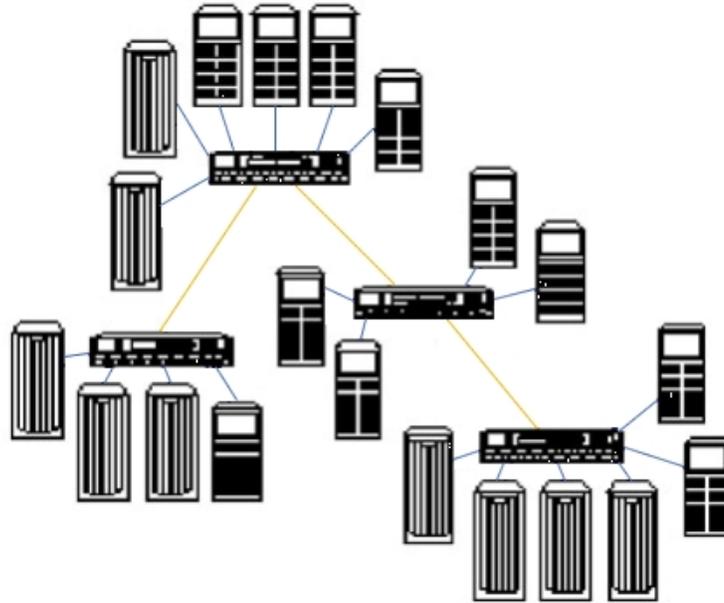
HP 標準トポロジは、SAN ファブリック内でのファイバチャネルスイッチの配置を指定し、特定のデータ位置へのニーズと一般的な作業負荷に対して最適化がなされています。これらのトポロジは、様々なレベルの接続性を提供するため、設計者は理想的な SAN 実装を選択することができます。

HP 標準 SAN トポロジには、以下の 4 つのマルチスイッチ SAN ファブリック設計があります。

- カスケード
- メッシュ
- リング
- バックボーン SAN

最初の 3 つのファブリック設計は、ファブリック内のすべてのスイッチをサーバとストレージの接続に使用する SAN 実装です。通常これらのタイプのファブリック構成では、スイッチ間接続に使用されるスイッチポートの合計数は少なくなります。サーバとストレージに使用されるポートの合計数を減らすことで、ISL の数を増やすことができます。

カスケードファブリック

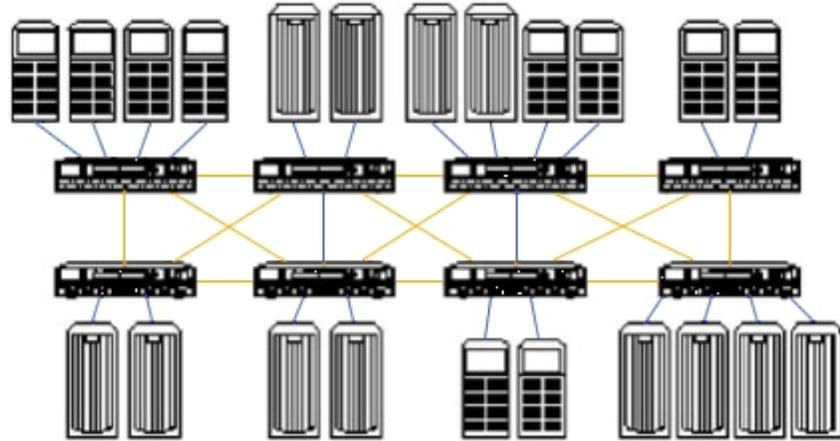


カスケードファブリック SAN は、1 つまたは複数の ISL によって接続された一列のスイッチまたは複数レベルのスイッチ群で構成されます。

カスケードファブリック設計は、サーバ(またはサーバグループ)に対するアクセス要求が基本的に同じストレージシステムまたはストレージセットになされるような、ローカルデータアクセスを使用するアプリケーションに適しています。アクセスされるサーバグループおよびストレージは、同じスイッチに接続することで最高レベルのパフォーマンスを実現できます。

カスケード接続により、SAN を拡張してサーバとストレージの接続性を高め、一元管理/バックアップを実行できるようになります。

メッシュファブリック

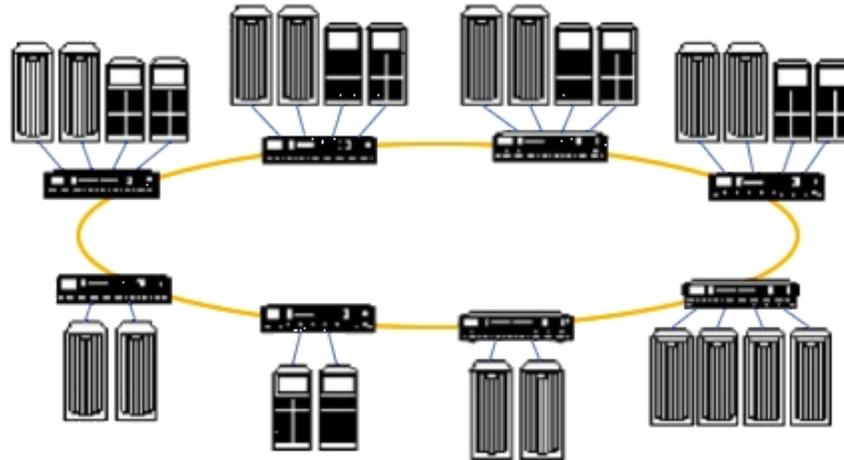


メッシュファブリック設計では、ファブリック内のすべてのスイッチは、上図のように2つのパスで相互接続されています。このタイプの接続は、ある程度のファブリック回復機能を提供します。

1つのISLまたはISLスイッチポートインタフェースで障害が発生した場合、ファブリックは代替パスを通じて自動的にデータを再ルーティングします。代替パスがファブリック内の別のスイッチを通過する場合があります。

メッシュファブリックは、ローカルアクセスと分散アクセスのデータアクセスが混在するアプリケーションに適しています。完全な接続性により、多対多アクセスが保証され、個別のスイッチ、サーバ、およびストレージへのローカルアクセスが可能になります。

リングファブリック



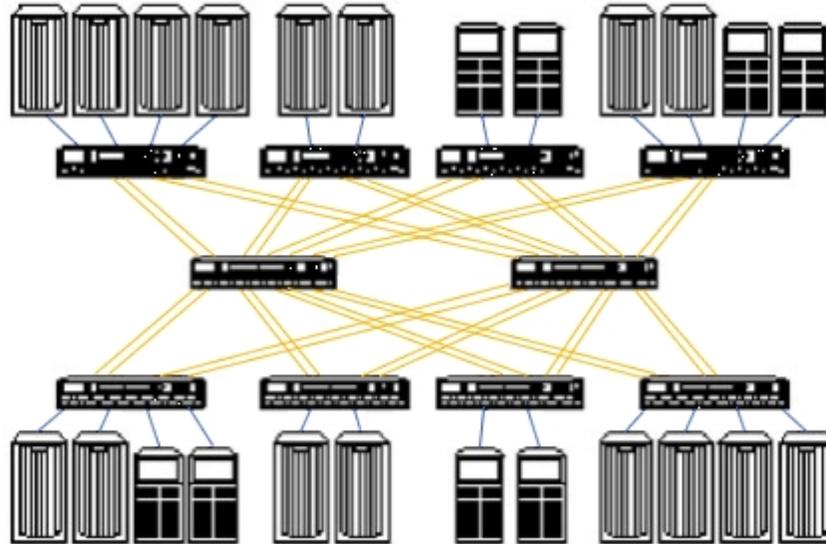
リングファブリックでは、連続したスイッチのリングが 1 つのファブリックとして接続されています。このスイッチ構成は、メッシュ設計とほぼ同等のファブリック回復機能を提供し、ファブリックの完全な接続性と少なくとも 2 つの内部ファブリックパスまたはルートを提供します。

リングファブリック設計は、データアクセスが常に局所化されるアプリケーションに適しています。アクセスされるサーバとストレージは同じスイッチ上にあり、データトラフィックの大半は同じスイッチ内を移動します。この実装では、スイッチ、サーバグループ、およびストレージをセルとして追加することにより、ファブリックをモジュール形式で拡張することができます。つまり、SAN のサイズを段階的に増加させるために、ビルディングブロック手法を使用します。

リングファブリックは、サーバの要件が未定でも、事前に構成してインストールすることができます。この機能により、追加する各ストレージシステムやサーバのインストールが簡単になります。

リング内のスイッチを相互接続することで、SAN を一元管理したり、ある特定の場所からバックアップを実行できるようになります。この設計では、必要に応じてリングを通じたデータアクセスが可能です。ただし、設計内で十分な数の ISL が指定されている必要があります。リングファブリックは、多対多の接続を必要とするアプリケーションにはおすすめしません。

バックボーン SAN ファブリック

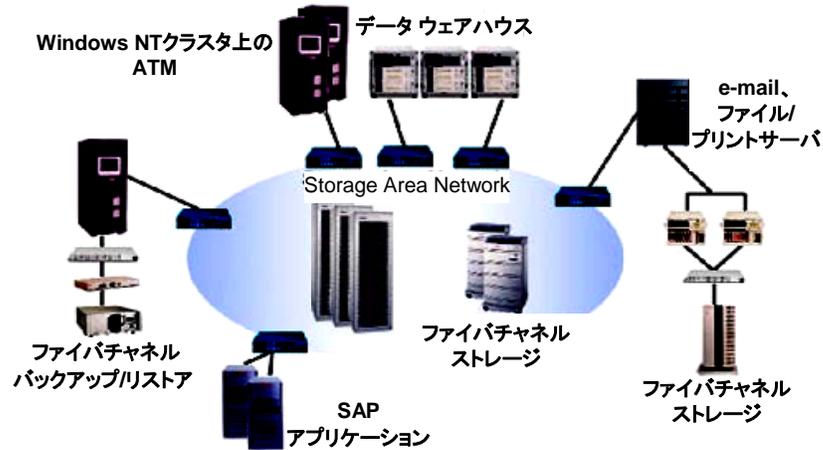


バックボーン SAN ファブリック設計では、ファブリック内の他のスイッチに接続するためのファブリックバックボーンとして機能する 1 つまたは複数のファイバチャネルスイッチがあります。バックボーンスイッチは、全帯域幅と他のすべてのスイッチへの冗長接続機能を持ったルータとして機能します。この実装では、多対多接続が可能で、帯域幅はファブリック全体で均等に分散されます。

バックボーン設計は、データトラフィックパターンがランダムまたは変化するが、多対多の完全なネットワーク接続性と高いパフォーマンスが必要な場合に適しています。このトポロジは、SAN 全体のストレージプーリングと共有化を使用する実装にも適しています。

バックボーン SAN トポロジには、Fat Tree と Skinny Tree の 2 つのタイプがあります。Fat Tree と Skinny Tree の主な違いは、エッジスイッチをバックボーンスイッチに接続するのに使用する ISL の数です。ISL の数は、エンドポートの数から差し引かれ、特定の構成で必要なスイッチの合計数に影響を与えます。Fat Tree は、半分の数のエッジスイッチポートを ISL 接続として使用し、Skinny Tree は半分未満の数を使用します。

Enterprise Network Storage Architecture (ENSA)



ENSA は、ストレージをエンタープライズユーティリティサービスに変換するパスを定義します。エンタープライズユーティリティサービスは、ユーザによって確実に透過的にアクセスされ、組み込みツールとテクノロジーによって完全に管理されます。

アプリケーションサーバは、環境全体に分散配置できるゲートウェイを通じて、透過的なプールにある記憶容量にアクセスします。各アプリケーションから見た可用性と信頼性は、地理的な場所、サーバ、やアプリケーションに依存しません。これらの属性や他の属性は、中央で設定可能なポリシーによって決定され、ストレージユーティリティによって保守されます。

Open SAN をサポートするソリューション

Open SAN をサポートするソリューションは、複数のベンダが提供しているネットワーク化されたストレージ製品の相互運用性をサポートする標準ベースのソリューションです。

ソリューションのフレームワークは、ベンダーデータゾーンの概念をベースにしています。HP データゾーンは、Open SAN 内の領域で、これには StorageWorks と OpenView 製品、および StorageWorks 製品にアクセスする異機種サーバプラットフォームが含まれます。このデータゾーンにより、HP ストレージ製品は、同一の SAN 内で他のベンダのストレージ製品として共存できるようになります。

ストレージバーチャライゼーション

ENSA は、ストレージバーチャライゼーションをより広範に適用することで、柔軟性の高い共有ストレージリソースを作成します。あらゆる場所から、大量の物理容量を企業全体にわたってプールして、アプリケーションサーバに提示することができます。

ENSA バーチャライゼーションテクノロジーにより、一連の物理ディスクから様々な容量とデータ保護レベルをもつディスクを作成できます。バーチャライゼーションはストレージプール内で発生しますが、そこには HP や他のベンダ製のあらゆるストレージ資産と、強力なバーチャライゼーションエンジンがあります。

プールは、時間の経過とともに増加します。最初は、あまり多くない数のアプリケーションサーバのニーズを満たすように拡張されます。その後、ENSA は、長距離延長技術を使用して、プールの範囲を広域に広げます。ストレージプールは、1 つの管理可能なエンティティを構成し、ここから信頼性の高い記憶容量とパフォーマンスをアプリケーションに提供します。

最新の相互接続テクノロジーにより、高速スループットと消費サーバへのストレージパフォーマンスが提供されます。ストレージプールは、DAS および他のストレージ形式と調和した状態で機能します。

動的スケーラビリティ

一般的なディスクとは異なり、ENSA ディスクはストレージプール内のデバイスの物理的なディスク容量によっては制限されません。その代わりに、アプリケーションに適切な容量が指定されています。

ENSA は、実際に使用されている仮想ディスク容量に基づいて、指定されている論理容量に達するまで物理記憶容量を個別のアプリケーションに自動的に割り当てます。ENSA は、サーバに必要な容量を絶えずモニタし、ストレージプールから容量を割り当てます。つまり、プールに提供したロー (raw) 容量により、ENSA ディスクの物理容量が動的かつ透過的に増えることがあります。

簡素化された管理

ENSA の自動化された管理機能により、ストレージ管理の観点は、分散された個別ボリューム、物理デバイス、バックアップ/復元、および他のポリシーを管理する履歴的観点から、ストレージプールのより上位の観点へと引き上げられます。ENSA では、システム内のあらゆる場所にあるディスクを自動的に制御するための中央ポリシーを作成する機能を提供しています。

この機能には、以下が含まれます。

- アプリケーションに完全に透過的な、ディスク間でのストレージの動的割り当て
- 論理ボリューム間での物理容量の再配備
- 複製、通常のバックアップ、またはその両方の組み合わせによる、柔軟性および信頼性の高いデータ保護
- ストレージプール内でのパフォーマンスの調整
- 容量計画、アカウンティング、チャージバック、その他のアクティビティのためのストレージの使用率のモニタ
- ハードウェア、パフォーマンス、容量使用状況など、主なシステム属性のステータスのモニタと通知

VersaStor バーチャライゼーション

HP は、VersaStor テクノロジを組み込むことで、ストレージバーチャライゼーション機能を StorageWorks および OpenView 製品群全体にわたって実装します。

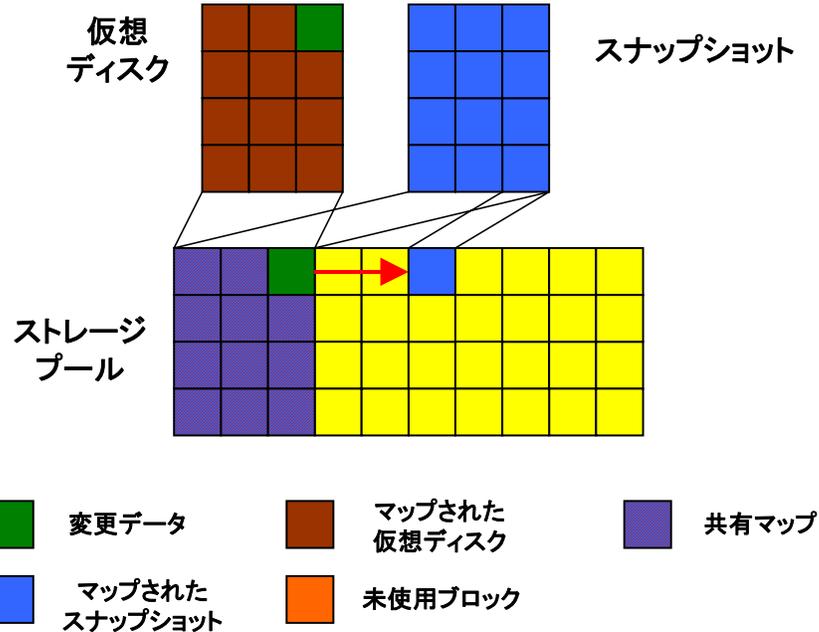
VersaStor テクノロジは、以下の 3 つのタイプのストレージバーチャライゼーションをサポートしています。

- **OpenView Storage Virtual Replicator** — (サーバベースの仮想サーバストレージ) 直接接続または SAN に接続されている仮想化されていないストレージと StorageWorks NAS ソリューションを使用する Windows サーバに VersaStor 機能を提供します。
- **Enterprise Virtual Array (EVA)** — (コントローラベースの仮想アレイ) ディスクドライブから、プールされているストレージ容量へのディスクドライブの構成を自動化します。データ保護と可用性を保証し、統合された負荷分散と自動データストライピングによりパフォーマンスを最適化します。
- **CASA (OpenView continuous access storage appliance)** — (SAN ベース) SAN ベースのストレージシステムの集まりを、相乗的能力のある 1 つの統合されたストレージ容量のプールへ昇格させます。

これらは、ストレージインフラストラクチャ全体にわたるストレージ管理を単純化および自動化することが可能な相補的なソリューションです。

これらの 3 つの実装間の主な違いは、ストレージプールの範囲および境界と、バーチャライゼーションソフトウェアを実行するデバイスです。

Virtual Replicator



Virtual Replicator は、バックアップおよび復元の時間を数時間から数分に短縮します。データおよびアプリケーションの可用性が向上し、ストレージ管理のコストと複雑さを低減します。

Virtual Replicator は、運用データのスナップショットを使用して、安全なバックアップとユーザー起動の復元をサポートします。

Virtual Replicator は、従来の SAN ベースのアレイと DAS を仮想化されたストレージプールに組み込みます。サーバベースのソフトウェアは、サーバがアクセスするすべてのストレージを仮想化します。これは、StorageWorks NAS およびアプリケーション固有のソリューション用、または部門用の Windows サーバ用の真のソフトウェアバーチャライゼーション手法です。

Virtual Replicator は、2 ノードクラスタと 4 ノードクラスタを含む Windows 環境用の Virtual Server Storage を作成するためのソフトウェア製品です。

Virtual Replicator は、VersaStor の以下のバーチャライゼーション機能を提供しています。

- ストレージプール容量の動的な増加
- ディスクパーティション
- 仮想ディスク

OpenView Storage Virtual Replicator はホストベースの製品です。Array Controller Software (ACS) 8.5S 以降は、コントローラベースのクローンとスナップショットをサポートしています。

Virtual Replicator は、他の仮想ディスクと同様に使用可能な、容量制限のない **Vsnap** を数秒で作成します。これらは、運用アプリケーションに影響を与えることなくアクセスすることが可能な、仮想ディスクの独立したコピーです。**Vsnap** は、通常、運用仮想ディスクをすばやく復元するのに使用可能なオンラインバックアップコピーとして作成されます。または、アーカイブ目的でテープライブラリに書き込むことができます。**Vsnap** を、LAN 経由で共有し、その後それを SAN にテスト用にコピーすることで、DAS から SAN ストレージへのサーバデータの移行を単純化することもできます。

Virtual Replicator は、サーバボリュームのストレージを仮想化されたストレージプールと組み合わせます。

例

サーバは、DAS のボリュームと、SAN ベースのストレージアレイから提供されたボリュームを持つことができます。**Virtual Replicator** は、2 つのボリュームから容量のプールを作成し、プールから複数の仮想ディスクを作成できます。追加ボリュームは、ストレージアレイからサーバに提供され、仮想化されてストレージプールの容量を増やします。新しい容量を使用して、仮想ディスクを動的に拡張したり、新しい仮想ディスクを作成したり、**Vsnap** を作成できます。

Enterprise Virtual Array(EVA)



EVA の機能は以下のとおりです。

- サーバとストレージコントローラ間だけにファイバチャネル接続速度をもつのではなく、真のファイバチャネルドライブを備えたネイティブファイバチャネルラックマウント型ストレージおよびストレージコントローラ
- それぞれのデータ転送速度が 2Gb/s のデュアル Fibre Channel Arbitrated Loop (FC-AL) バスのサポートにより、SCSI ベースのファイバチャネルよりも優れたパフォーマンスを実現
- システムあたり最高 240 個のドライブをサポートする HSV110 コントローラ
- 以下のことが可能な仮想化されたストレージプール
 - 物理ディスク間での動的負荷分散
 - 仮想ディスクの動的拡張
 - 柔軟性の高い RAID レベル Selective Storage Presentation (SSP)
- 管理作業専用の Management Appliance

EVA のパフォーマンス

EVA は、ランダムアクセスデータベースやリッチメディアなどのストリーミングアプリケーション用にパフォーマンスが最適化されたストレージを提供します。VersaStor テクノロジは、EVA の業界トップのパフォーマンスを実現するのに重要な技術です。RAID セット制約に依存することなく、複数のディスクドライブに分散されているデータで仮想ディスクを作成します。モジュール形式のアーキテクチャにより、容量に応じてパフォーマンスと接続性が変化します。

EVA は、キャッシュされたトランザクション処理の作業負荷で 1 秒あたり 168K I/O (RAID 5) という業界一のパフォーマンスを提供します。

完全にランダムなトランザクション処理の作業負荷は、1 秒あたり 55K I/O です (RAID 5)。ストリーミングメディアなどのシーケンシャル作業負荷は、読み取り帯域幅 400MB/s、書き込み帯域幅 134MB/s です (RAID 5)。非ミラー化キャッシュの RAID 0 を使用したシーケンシャル作業負荷では、読み取り 400MB/s、書き込み 390MB/s のパフォーマンスを得ることができます。これらのスループットは 1Gb SAN の場合の値であり、2Gb SAN の場合は、400MB/s が 675MB/s になります。

EVA を共通コンポーネントとして使用することで、数百ギガバイトから数百テラバイトに至るストレージ容量を配備できます。これらはすべて、HSV Element Manager Management の機能で管理されます。EVA は、中断されることがないデータアクセスと広範なデータ保護および復旧ソリューションを提供します。

CASA (OpenView continuous access storage appliance)

CASA は、複数のアレイシステムのストレージを組み合わせ、スケーラビリティに関して実質上制限がない大規模な容量のプールを提供します。運用アプリケーションに影響を与えない動的なストレージ再構成を許可することで、ストレージの利用率を向上させます。

顧客は、中央にある拡張容量のプールを保守し、仮想ディスクのサイズを最小化します。その後、アプリケーションのデータセットが大きくなったらそれらを動的に増加します。また、1つのストレージアレイの容量に制限されずに、プールの容量を使用して容量無制限の Vsnap を作成できます。

CASA は、複数のアレイシステム間の境界を消去します。使用可能なあらゆる容量から仮想ディスクを作成して、個別のアレイ上の取り残された未使用領域 (複数のアレイシステムにわたる非連続の容量を含む) を排除します。複数のアレイシステムにわたってデータをストライプする (スピンドル数を増やす) ことでパフォーマンスを最適化します。サーバは論理ブロックアドレスにアクセスするため、アレイシステムの境界は透過的です。

例

ストレージ容量を増やしたい場合は、新しいアレイシステムをストレージプールに登録します。これにより、動的に増加する既存の仮想ディスク、新しい仮想ディスクの作成、Vsnap の作成、または古いストレージからのデータの移行のために容量を使用できるようになります。

古いストレージシステムを除外したい場合は、除外するシステムを指定します。Virtual SAN Executor は、古いシステムから新しいシステムに (または、指定した属性をもつ他の使用可能な容量に) データを自動的に移行します。

仮想ディスクを RAID 0 から RAID 5 に変更したい場合は、冗長性属性を変更します。Virtual SAN Executor は、データを RAID 5 ストレージに移行します。運用アプリケーションは、データ移行中も中断されることはありません。

ENSAextended VersaStor テクノロジ

ENSAextended VersaStor テクノロジにより、ストレージを属性によって容易に管理可能な容量のプールとして編成する、自動化されたストレージソリューションを配備できます。自動化により、1つのストレージ管理構造下において、ディスクドライブや RAID セットなどの詳細が管理者に透過的になります。

管理者は、サーバの仮想ディスクの容量、パフォーマンスレベル、RAID 保護、およびデータ可用性などの属性を指定します。VersaStor テクノロジは、容量のプールから、指定した属性で自動的に仮想ディスクを作成し、指定したサーバにそれを提供します。

単純なツールにより、動的な容量増加や実質的に容量無制限のスナップショットコピーまたは Vsnap などの高度な機能を使用できます。これらの機能により、復元、テープへのアーカイブ、テスト、データ検索などの操作用に仮想ディスクをオンラインにしたままにすることができます。VersaStor バーチャライゼーションテクノロジーでは、以下の機能を提供しています。

- 管理インタフェース
- デバイス間でのデータの自動的な移動
- 基礎となるストレージアーキテクチャ

VersaStor Vector

サーバ接続ハードウェアである VersaStor Vector は、論理ブロックアドレスを、ターゲットアレイシステムを指定する物理アドレスに自動的に変換します。各 VersaStor Vector は、独自の変換ハードウェアを通じて物理ディスクブロックに直接アクセスします。これにより、集中化されたアドレス変換ゲートウェイを使用するバーチャライゼーションアーキテクチャに存在するパフォーマンスおよびスケーラビリティに関する制限が排除されます。

Virtual SAN Executor

Virtual SAN Executor は、EVA によって作成される Vsnap と類似した特性を持っている Vsnap を Virtual SAN 上に作成します。Executor は、管理対象のあらゆる仮想ディスク用の Vsnap を作成できます。ストレージアレイ自体は、Vsnap 機能を必要としません。Executor は、容量のプール全体から Vsnap を作成するため、Vsnap は 1つのアレイシステムの境界によって制限されることはありません。

Virtual SAN Executor は、管理者がストレージプールを管理するための、完全な冗長構成の集中的ユーザインタフェースを提供します。これは、指定した属性に一致する物理ブロックを選択することで仮想ディスクを作成し、仮想ディスクをホストサーバに提供します。仮想ブロックから物理ブロックへのマッピングテーブルの冗長コピーを保守し、VersaStor Vector にテーブルを通知します。

Vraid ソフトウェア

EVA は、Vraid という、次世代の分散仮想 RAID ソフトウェアをベースにしています。この仮想 RAID ソフトウェアは、ストレージプールを作成および管理します。Vraid は、複数のドライブ障害に対する耐障害性を向上させ、自動的にストレージプール全体にデータをストライプすることで、パフォーマンスを最大にし、取り残された容量を最小化します。

EVA は、Vraid によって容易に作成および管理可能な複数のストレージプールに分割できます。各ストレージプールには、RAID 0、RAID 1、および RAID 5 アルゴリズムに従った冗長構成で仮想ディスクを混在させることができます。ストレージ管理者は、仮想ディスクの属性を指定し、Vraid は、適切な特性の仮想ディスクを自動的に作成します。

例

管理者は、RAID 5 レベルのデータ保護機能をもつ 1TB の仮想ディスクを 10 個 (合計 10TB) 必要としています。管理者は、12TB のストレージプールを指定します。これにより、仮想ディスクの動的容量増加用に 2TB 残ります。Vraid は、12TB の使用可能な容量を提供できるように、適切なディスクドライブ数でストレージプールを自動的に作成します。Vraid は、自動的に 10 個の仮想ディスクをストレージプール内のすべてのドライブに分散します。各仮想ディスクへのアクセスは、ストレージプール内のすべてのスピンドルに分散され、パフォーマンスが大幅に向上されます。

管理者が 2 つの仮想ディスクの動的増加量を 2TB にする必要がある場合、ユーザは仮想ディスク容量属性を変更する必要があります。これにより、Vraid は自動的に変更を行います。動的ボリューム増加をサポートしているサーバは、すぐに 2TB の仮想ディスクにアクセスできます。

ディスクドライブを追加して、増加するストレージプールを指定することで、ストレージプールに新しい容量を動的に追加できます。Vraid は、新しい容量を自動的に追加し、ストレージプール内のデータブロックを新しいディスクドライブ全体にわたって再配布します。

また、Vraid は、デュアル仮想スペア用の容量もストレージプールから提供します。仮想ディスクと同様、スペア容量はストレージプール内のすべてのスピンドルに分散されます。これにより、再構成時間が減少し、冗長性を復元する時間が短縮されます。

ドライブで障害が発生した場合、Vraid は、自動的に障害が発生したドライブをプールから除外し、スペア容量上の冗長データを復元して、ストレージプールを完全なフォールトトレランス状態に戻します。データブロックもストレージプール全体で再配布され、パフォーマンスが最適化されます。

Vraid は、利用可能な容量以上の仮想スペア容量が必要になった場合、自動的に未割り当ての容量を使用して冗長性を復元します。バックグラウンドタスクによって、ディスクドライブを絶えずテストして障害発生前の状況を確認し、予防保守を行います。データブロックの広域分散、スペアリングの深さと柔軟性、高速復元、連続的なドライブ検査により、複数のドライブで障害が発生した場合の耐障害性は、Vraid は非仮想化ストレージよりも優れています。

EVA は、Vsnap、スナップショット、スナップクローンを作成します。これらは、ミラーコピーのように読み取りおよび書き込みでき、短時間で作成できます。

Vsnap

Vraid と Vsnap により、ストレージの再構成とバックアップが完了するまでの計画的ダウンタイムが不要になります。Vsnap と SAN ベースのデータリプリケーションは、ローカルまたはリモートディスクから直接復旧できるようにすることで、復旧操作に必要な計画外のダウンタイムを短縮します。

実質的に容量無制限の Vsnap (運用仮想ディスクに割り当てられている容量の一部しか使用しません) またはスナップショット (運用仮想ディスクと同じ容量を割り当てる以外は Vsnap と同じ) を選択できます。Vraid は、多数のディスクにわたってデータをストライブするため、Vsnap とスナップショットは従来のスナップショットよりもアクセスの競合が少なくなります。

スナップクローン

スナップクローンは、スナップショットとして開始しますが、運用仮想ディスクのある時点での物理コピーになります。アレイコントローラは、スナップショットを作成してスナップクローンを作成します。これは、読み取りおよび書き込み用にサーバにすぐ提示することができます。コントローラは、バックグラウンドタスクを実行して、運用仮想ディスクのデータをスナップクローン仮想ディスクにコピーします。結果として、スナップクローンの作成が開始された時点での運用仮想ディスクと同じコピーが作成されます。スナップクローンでは、スナップショットとミラーの最も優れた機能が組み合わされています。

HSV Element Manager

EVA 用の HSV Element Manager は、サーバとは独立してストレージ環境をある特定の場所から一元管理します。

ユーザは、サイズや冗長性などの仮想ディスク属性で容量を規定します。ソフトウェアは、物理的な詳細の管理を自動化し、ユーザが容量をプールとして管理できるようにします。また、ソフトウェアは、データ分散と負荷分散を自動的に最適化して、管理者の数多くのパフォーマンスチューニングタスクを軽減します。

テープバックアップテクノロジー

エンタープライズストレージは、あらゆるビジネスが対応しなければならない最も重要な IT 要素の一つです。今日のアプリケーションは、今まで以上にデータ集約的になってきています。ディスクストレージの消費量は激増していて、以前とは比べものにならないほどデータはビジネスの日常的な運営に必要不可欠なものとなっています。

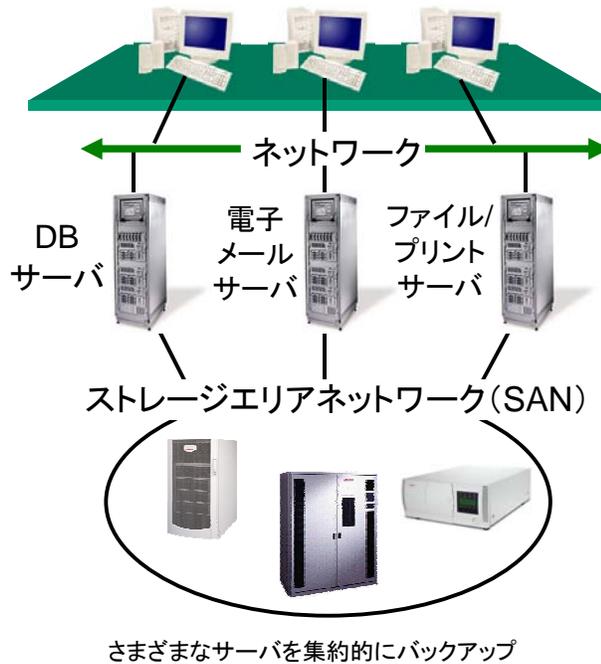
ディスクで障害が発生した場合に備えて RAID または他のフォールトトレランスシステムを使用している場合でも、バックアップシステムは必要です。削除されたファイル、自然災害、または盗難に対する保護を提供できるのはバックアップシステムだけです。

適切に管理されたバックアップポリシーとそれに付随するハードウェアとソフトウェアにより、人的エラー、ソフトウェアエラー、その他の損失からデータを保護できます。バックアップの主な目的は、長期間にわたってデータを維持し、最悪の障害から復旧できるようにセカンダリ媒体上にデータをアーカイブすることです。

復旧処理を実行する必要がある場合は、以下のとおりです。

- ハードウェア障害によるデータベースの損傷
- 間違いまたは悪意によるデータの削除または変更
- 盗難
- ウィルス攻撃または感染
- 火災、洪水、地震、その他の自然災害
- 暴動、政治的動乱、または人によって引き起こされるその他の災害

セカンダリストレージとバックアップの重要性

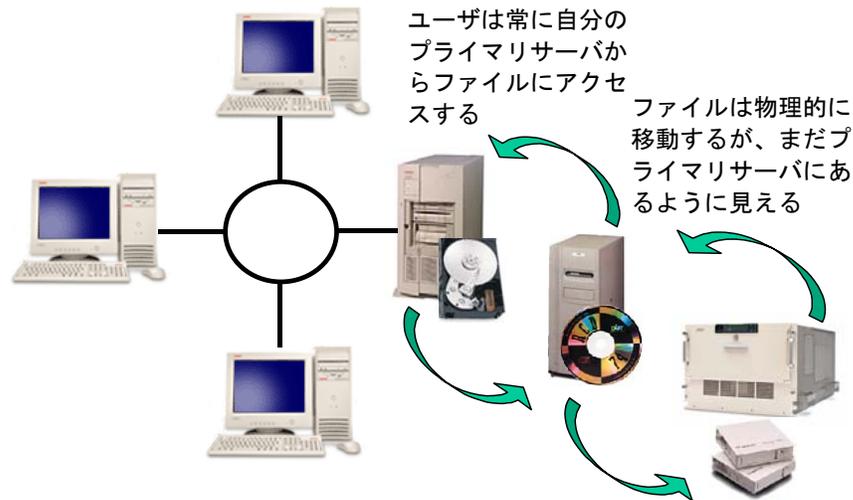


バックアッププロセスでは、重要な情報を磁気テープまたは他の恒久的な記憶媒体にコピーします。これにより、状況に応じて、1つのファイルからシステム全体に至る範囲を復旧できるようになります。バックアップを実行する媒体とデバイスは、即時アクセスして使用できるようにデータが格納されているディスクであるプライマリストレージと区別できるように、通常「セカンダリストレージ」と呼ばれます。

バックアップにより、以下の原因によるデータ損失を復旧することができます。

- 電源サージおよび停電
- 静電気
- 落雷
- その他の自然災害
- テロリストによる爆破
- ユーザによる単純なエラー
- 単純な事故(コーヒーをこぼすなど)
- 妨害行為
- 機器の誤動作
- ウィルス

Hierarchical Storage Management (HSM)



Hierarchical Storage Management (HSM) は、階層型ストレージ管理と呼ばれるユーザの要求を最高の状態で処理できるように、プライマリストレージを最適化するための戦略です。この機能は、Windows 2000 Server で使用できます。HSM は、データ取得速度とディスク上の空き領域のバランスをとります。アクティブなデータをディスクなどの高速な記憶媒体に格納することで高速になり、非アクティブなデータをテープなどのよりコストの低い記憶媒体に格納することで空き領域が増えます。

HSM の基本原則は以下のとおりです。

- **頻繁にアクセスされないデータをよりコストの低い媒体に移動する** — システムに入ってくる新しいデータは、すばやくアクセスできるように高速媒体に格納されます。このデータへのアクセス頻度が少なく、高速媒体に存在している必要がないことが判明すると、データは媒体階層の下位レベルにある媒体に移行されます。
- **頻繁にアクセスされるデータを高速媒体に移動する** — 移行したデータが参照または更新目的で再度アクセスされると、媒体階層の下位レベルから上位レベルにキャッシュされます。これにより、アクティブなデータのアクセス速度要求を満たすことができます。
- **データ移動を透過的に管理する** — 利用可能なストレージオプション間でのアクティブデータと非アクティブデータの移動が透過的に管理され、ユーザにはすべてのデータがローカルで使用可能な状態になります。アクティブであるかどうかにかかわらず、すべてのデータが基本的に同じ状態でユーザとユーザのアプリケーションソフトウェアに提供されます。

バックアップ方法

以下のバックアップソリューションを使用できます。

- 1つのテープデバイスへのバックアップ
- テープデバイスのストライピング (Redundant Array of Independent Tape [RAIT])
- ローカルディスクデバイスへのバックアップ
- ネットワーク経由でのバックアップ

計画とサイズ

複雑なアプリケーションおよびネットワークの構成、カスタマイズしたセットアップ、およびパスワードと ID は、再作成するのが困難でコストが高く、再作成できないこともあります。会社のさまざまな記録や情報は、通常、会社で最も重要な資産です。ミッションクリティカルサーバが突然ダウンした場合には、経営に支障が出る可能性があります。通常、復旧に必要なダウンタイムだけでも経営上の損害は莫大なものになります。

優れた設計のバックアップシステムは、重要な情報を安全に保護し、最も効果的でコストパフォーマンスの高い方法で災害時のデータ、時間、および費用の損失に備えます。

このため、会社全体のバックアップストラテジを開発するためには、ネットワークアーキテクチャとユーザのシステムに対する要求を理解する必要があります。これらの情報を入手したら、システムエンジニアまたは管理者は、以下のことを実施する必要があります。

- ネットワークまたはエンタープライズバックアップのニーズ解析を行う
- 会社全体の効率的なバックアップソリューションを作成および実装する

速度

$$\frac{\text{データ (GB)}}{\text{バックアップウィンドウ (時間)}} = \text{必要な転送速度 (GB/時)}$$

必要な転送速度の式

パフォーマンスのニーズは、バックアップする必要がある情報の量をバックアップウィンドウのサイズ(時間単位)で除算することで判断できます。この簡単な計算により、必要なパフォーマンスを全体転送速度(GB/時)として算出できます。

必要な転送速度の式を使用する前に考慮すべき HP テープドライブの仕様は以下のとおりです。

- 各ドライブが 1 つのカセットに格納できる未圧縮データの最大容量
- 各ドライブの未圧縮データの読み取りおよび書き込み転送速度
- システム内の各ドライブを使用してローカルバックアップを実行するときの一般的なシステム転送速度
- そのシステムが各ドライブを使用して 10GB のバックアップを実行するのに必要な標準時間
- バックアップ操作中にロードおよびアンロードする必要があるテープカートリッジの数

必要なパフォーマンスの計算結果を一般的なバックアップパフォーマンス速度と比較することで、管理者は必要なバックアップパフォーマンスを実現できるかどうかを判断できます。いずれかのドライブタイプのパフォーマンスが十分である場合は、ローカルのオフラインバックアップを選択することをお勧めします。

バックアップテープのローテーション方式

企業のデータは重要であり、代わりとなるものが存在しないため、バックアップはスケジュールに従って規則正しく実行する必要があります。設計する際に十分考慮すべき事柄は以下のとおりです。

- 冗長性
- 記憶スペースの制限
- 災害発生後の復旧速度
- 媒体の効率的な使用方法

定期的なバックアップを設計する際に役立つネットワークに関する情報は、以下のとおりです。

- バックアップのスケジュール
- 部分バックアップまたはフルバックアップのどちらを実行するか
- 部分バックアップの場合は、増分バックアップまたは差分バックアップのどちらを実行するか

「フル」、「部分」、「増分」、および「差分」という用語は、コピーおよびバックアップされる情報の量を表します。フルバックアップまたは部分バックアップのどちらであるかを問わず、すべてのバックアップをオンラインまたはオフラインで実行できます。

フルバックアップ

フルバックアップは、サーバまたはクライアント全体の完全なバックアップです。サーバの場合は、すべてのボリューム、フォルダ、およびファイルが含まれます。クライアントの場合は、すべてのドライブ、フォルダ、およびファイルが含まれます。

フルバックアップには、標準バックアップとコピーバックアップの2つのタイプがあります。

- **標準バックアップ** — データベースファイルをバックアップしてからトランザクションログファイルをバックアップします。その後、ディレクトリからトランザクションログファイルを削除します。バックアップソフトウェアがログファイルを削除するため、循環ロギングを無効にすることができます。このため、定期的なバックアップを実行しているときに、ドライブがログファイルでいっぱいになることはありません。標準バックアップを復旧するには、最後の通常バックアップセットを復旧してサービスを開始します。
- **コピーバックアップ** — 標準バックアップに似ていますが、ドライブ上のログファイルを消去せず、データベースファイル内のバックアップコンテキストを更新しません。コピーバックアップは、標準バックアップのスケジュールを中断することなくデータをバックアップする場合に使用します。

部分バックアップ

以下のいずれかの部分バックアップを実行できます。

- **増分** — 前回のバックアップがどの種類であったか問わず、前回のバックアップ後に変更されたすべてのファイルをコピーします。
- **差分** — 前回の完全バックアップ後に変更されたすべてのファイルをコピーします。

増分

増分バックアップは、ファイルの各リビジョンを維持する必要がある場合に使用します。連続する増分バックアップで同じテープを使用した場合、バックアップファイルの新しいバージョンで以前のバージョンを上書きすることはできません。通常、新しいファイルはバックアップ媒体に追加されます。

通常、バックアッププログラムは、各増分バックアップ後にアーカイブビットをリセットします。増分バックアップは、ログファイルに対してだけ動作し、循環ロギングが無効になっている場合だけ動作します。標準バックアップと同様、増分バックアップはログファイルをバックアップした後にそれらを消去し、復旧能力を損なうことなくログファイルをドライブから削除する別の方法を提供します。

増分バックアップを復旧するには、データベースファイルを含んでいる前回の標準バックアップセットに戻す必要があります。これらのデータベースファイルを復元し、標準バックアップ以降に作成したすべての増分バックアップセットを復元してサービスを起動します。すべてのバックアップセットを復元するまでサービスは起動しないでください。この状態で起動すると、バックアップセット以降に復元されたすべてのログが再生されなくなります。

差分

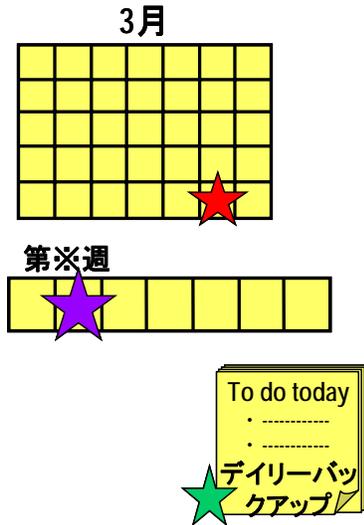
差分バックアップは、各ファイルの最後のバージョンが重要な場合に使用します。連続する差分バックアップで同じテープを使用した場合、バックアップしたファイルの新しいバージョンは、テープ上の同じファイルの古いバージョンを上書きします。通常、バックアッププログラムは、差分バックアップ後にアーカイブビットをリセットしません。アーカイブビットは、次に完全バックアップを実行するまで有効になったままになります。

差分バックアップは、ログファイルに対しても動作するため、差分バックアップを実行する前に循環ロギングを無効にする必要があります。増分バックアップとは異なり、差分バックアップではログファイルは削除されません。

差分バックアップセットを復元するには、前回の標準バックアップに戻し、前回の標準バックアップ後に生成したすべてのログファイルを含んでいる差分バックアップセットを復元します。すべてのバックアップセットを復元するまでサービスは起動しないでください。

Grandfather-Father-Son (GFS) テープローテーション

- マンスリーgrandfatherバックアップ
 - フルバックアップ、オフサイトで保管
 - 通常、月末に実行
- ウィークリーfatherバックアップ
 - フルバックアップ
 - 通常、週の初めまたは終わりに実行
- デイリーsonバックアップ
 - 通常、増分または差分バックアップ
 - 場合によっては、フルバックアップを使用することもある



Grandfather-Father-Son (GFS) テープローテーション方式は、最も一般的な方式で、少なくともサーバ記憶容量の2倍のウィークリーバックアップ容量を必要とします。さまざまなレベルのデータを保持できるように、この方式では3つのバックアップレベルを使用して冗長およびセキュリティ機能を提供します。

システム管理者は、一時バックアップを格納するテープの世代と、アーカイブを格納するテープの世代を選択できます。

GFS バックアップでは、以下のバックアップを実行する必要があります。

- マンスリー(月間) grandfather
- ウィークリー(週間) father
- デイリー(毎日) son

例

システム管理者は、通常フルバックアップを月曜日 (father) ごとに実行し、増分バックアップを火曜日、水曜日、木曜日 (son) に実行します。また、別のフルバックアップを週末 (father) に実行し、さらに別のフルバックアップを月末 (grandfather) に実行します。

注記

GFS テープローテーション方式は、災害が発生した日に失われたデータを復元することを目的としています。災害発生後にオンライン状態に戻すためのサポートを提供している会社もあり、これを利用することですばやく復旧することができます。

ウィークリーおよびマンスリーバックアップを含んでいるテープは、通常、サーバが設置されているサイトとは異なる場所に保管します。媒体のコストを削減できるように、ほとんどの会社が古いウィークリーバックアップテープを再使用しています。

**警告**

最近使用したバックアップテープは、絶対に上書きしないでください。バックアップ中にハードディスクがクラッシュした場合、ディスク上のすべてのデータが失われるだけでなく、テープが使用できなくなり、完全に復元できなくなります。

顧客のニーズによっては、デイリー (son) バックアップで増分バックアップよりも差分バックアップを実行したほうが良いことがあります。また稀に、デイリーのフルバックアップを実行しなければならないこともあります。部分バックアップまたは完全バックアップのそれぞれのタイプで異なる点は以下のとおりです。

- 毎日バックアップするデータの量
- バックアップセッションの所要時間
- 災害発生後、サーバへデータを復旧するのに必要なテープ数

テープ媒体の平均寿命

ほとんどのメーカーが、テープ媒体の平均寿命を、テープのバックアップ回数ではなく、予期されるパス数で指定します。「パス」とは、読み取り/書き込みヘッド上を通過するテープ媒体の順行、逆行の動きを意味します。

フルバックアップの場合は、許容パス回数を媒体上のトラック数で除算することでテープ媒体の平均寿命を見積もることができます。

例

4/8GB SLRドライブは、テープ媒体で46トラック使用します。そのドライブ用の推奨媒体の平均寿命は、最適な環境条件下でフルバックアップを実行した場合に5,000パスです。5,000パスを46トラックで除算すると約108になり、約108回のフルバックアップが可能であることがわかります。

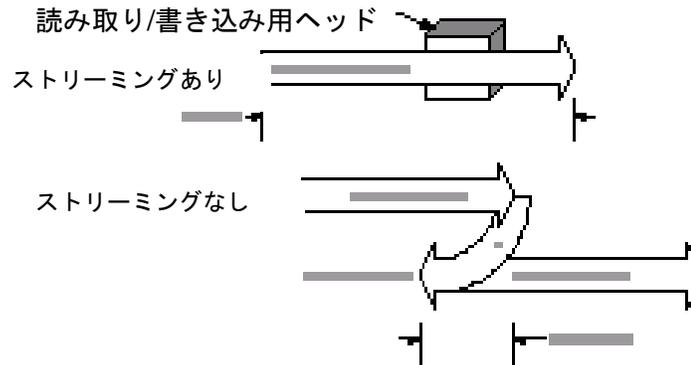
この計算では、ドライブがストリーミングしていて、テープ媒体を端から端まですべて使用すると仮定しています。このような仮定に一致することはほとんどないにもかかわらず、HPが実施した限定的なテストでは、ほとんどのパス見積もり値は控えめな値であることがわかりました。

テープ媒体の実際の平均寿命を判断する要因を以下に示します。

- ヘッドとテープパスの設計
- ドライブが1つの完全バックアップで必要とするパス数またはトラック数
- 湿度と周囲温度
- テープ先頭のヘッダーを読み取りおよび更新する際の追加のパス数
- テープストリーミング

テープストリーミングは、優れたバックアップソリューション設計と媒体の平均寿命の関係を示すため、詳細な検査に匹敵するものです。

テープストリーミング



ストリーミングしないと、追加のテープパスが発生する

テープストリーミングは媒体の平均寿命に影響します。「ストリーミング」とは、「テープ媒体がバックアップ時に読み取り/書き込み用ヘッドを継続的に順行すること」を意味します。テープドライブとバックアップソフトウェアは、ストリーミングするように設計されています。

非ストリーミングとは、「テープが読み取り/書き込み用ヘッドを断続的に順行、逆行すること」を意味します。ストリーミングしないと、媒体の寿命が大幅に短くなり、またヘッドの摩耗が激しくなって、バックアップのパフォーマンスが低下します。

テープドライブは、以下の条件が満たされた場合にテープカートリッジをストリーミングします。

- ディスクドライブまたはディスクドライブアレイが、テープドライブがデータを受け取る以上の速度でデータを供給できる
- オペレーティングシステムのバックアップソフトウェアの構成が、ストリーミングをサポートするよう正しく最適化されている
- プロセッサは、バックアップ作業のトラッキングなどのオーバーヘッドを処理するための十分な処理能力を持っている。また、プロセッサは、テープドライブのストリーミングを処理でき、バックアップ処理中にディスクドライブやディスクドライブアレイにアクセスできる必要がある

ストリーミング中、ドライブは絶えずテープに情報を書き込みます。ホストは、ドライブがテープを停止してデータを受信されるのを待つことがないように、すばやく安定した状態でデータを送信します。テープを停止しなければならなくなった場合、ドライブはテープを逆行し、最後に書き込んだブロックの隣にヘッドを再配置します。データを受信すると、テープの順行が再開されます。テープヘッドの再配置により、テープのその部分のパス数が追加されることになります。

テープ媒体が停止、再配置、再開するのではなく、一方向に継続してストリームする場合、バックアップパフォーマンスは最高になり、読み取り/書き込み用ヘッドの摩耗が最小限に抑えられます。

テープドライブのクリーニング間隔

データバックアップのスケジュールを設定している場合は、バックアップの全体計画にテープドライブの定期的なクリーニング予定を組み込むことが重要です。クリーニングにより、ヘッドの汚れによる問題が発生するのを防止できます。

クリーニングの推奨頻度は、ドライブメーカーの仕様によって異なります。周囲温度、湿度、テープ媒体のメーカーにより、実際のクリーニング間隔は異なります。

バックアップ媒体のオフサイトでの保管

バックアップデータは定期的にオフサイトに移動してください。移動することで、サイトで災害が発生した場合も、安全なオフサイト保管場所にあるバックアップ媒体を使用することができます。

データの重要性に応じて、複数のオフサイト保管場所を設定してください。オフサイト保管サービスを提供している会社は、ローテーションが必要なときにテープの配送/引き取りを行ってくれます。



警告

バックアップ媒体の保管場所を決定する場合は、注意が必要です。磁場により媒体上のデータが破壊されることがあります。媒体は、ビデオモニターや電話などの電気製品から離れた場所に保管してください。

テープストレージテクノロジー

HP が提供しているさまざまなセカンダリストレージ製品は、重要なテクノロジーによってサポートされています。

HP には、以下の 5 つのテープテクノロジー製品があります。

- DAT (Digital Audio Tape)
- AIT (Advanced Intelligent Tape)
- DLT (Digital Linear Tape)
- SDLT (Super Digital Linear Tape)
- Ultrium

能力

無人バックアップにより、これらの管理コストを大幅に削減できます。十分なバックアップ能力と自動化を導入して無人バックアップを可能にすれば、長期間にわたってコストを大幅に削減できます。

無人バックアップの場合、テープドライブと媒体は、ストレージ容量およびバックアップパフォーマンスにおいてバックアップ要件を満たしている必要があります。

パフォーマンスと容量は、選択したドライブテクノロジーによって異なります。容量が最も多く、転送速度が最も速いのは新しい **ultirum** テクノロジーです。

- **DATドライブ** — 容量が 2GB～20GB のサーバに適しています。HP では、DAT テープのオートローダも用意しています。オートローダを使用すれば、ロボット機構によりテープがロード/アンロードされるため、管理コストを削減できます。
- **AITドライブ** — DLTドライブの大容量と高い信頼性で 100GB 以下の容量のサーバを圧縮せずにバックアップすることができます。35/70GB DLT よりもデータ転送速度が速いため、ハイエンド DATドライブからアップグレードするのに最適です。また、DLT の代わりとなる低価格なソリューションです。AIT では、テープアレイおよびライブラリオプションも提供しています。
- **DLTドライブ** — 同時マルチチャネルとマルチヘッド読み取り/書き込みテクノロジーを使用して、最大 40GB の容量を圧縮なしでアーカイブできます。DLTドライブは、さまざまなテープアレイ、ミニライブラリ、および大容量ライブラリオプションに対応していることから、特に 40GB 以上の大容量サーバのハイエンドバックアップソリューションに適しています。
- **SDLT (Super DLT)ドライブ** — DLT テープと互換性があり、記憶容量と転送速度は大きさに従って増加します。ストレージ容量は、圧縮なしで最高 160GB です。SDLTドライブは、その容量と転送速度、および DLT よりも高い信頼性により、エンタープライズクラスのサーバに適しています。
- **ultirumドライブ** — DLT や SDLT テープとの互換性はありません。ストレージ容量は、圧縮なしで 200GB です。短時間で大量のデータをバックアップできるためデータセンタに適しています。

ニアラインテクノロジー

ニアラインストレージとは、オンラインストレージよりもコストパフォーマンスに優れた方法で格納されているデータのことを表します。ただし、取り出しには時間がかかります。データウェアハウスは、ニアラインテクノロジーを使用して大容量のデータを格納します。データ(通常、アーカイブ)は、頻繁には取り出されず、通常その速度も低速です。

ニアラインテクノロジーの例として以下ものがあります。

- CD/DVD
- 光磁気
- RAID

RAIDと同様、RAITは大容量、高性能バックアップソリューションを提供します。RAITは、優れたパフォーマンス、エラー復旧、およびフォールトトレランスを提供する2つ以上のドライブを使用し、インタリーブを利用したストレージ技術と重要なデータのミラーリング機能を備えています。データは、RAIDと同様のストライピング冗長性またはミラーリングで記録できます。

ディスクドライブのRAIDと同様、テープアレイにも以下の特徴があります。

- 帯域幅の増大
- 信頼性の向上



重要

RAITグループ内のテープを分離してはいけません。

クローン化とスナップショット

クローン化

クローンとは、運用ボリュームの物理的なコピーです。クローンは、ボリュームの3番目のミラーとして作成されます。クローンが作成されると、破壊されるまでの間ボリュームのミラーのように機能します。通常、アプリケーションは、ポイントインタイムのデータ整合性を保証できるようにクローンが破壊される前に停止します。アプリケーションが停止しないと、クローンにはクラッシュ時のデータが含まれるようになります。

スナップショット

スナップショットは仮想ディスクの簡易レプリカです。スナップショットを使用すれば、物理的にデータをコピーすることなく運用データの仮想レプリカが作成されます。スナップショットは、読み取りおよび書き込み機能を備えた通常の物理ディスクと同じです。スナップショットディスクには、まったく同じデータが含まれていて、以下の目的で使用されます。

- 実行中のアプリケーションを停止することなくデータのクイックコピーを作成する
- オンラインバックアップとリストアを実行する
- 新しいアプリケーションをテストする
- データウェアハウスまたは Web サーバを追加する

クローンとスナップショットは同じ方法で使用できます。両方ともオフラインバックアップを実行できます。ただし、以下の違いを考慮してください。

- クローン化は、物理的なコピーを作成し、最高レベルの可用性を提供する
- スナップショットは瞬時に完了するが、クローン化には数分かかる(コピーするボリュームのサイズによって所要時間は異なる)
- クローン化は、RAID 3/5 をサポートしていない。このため、環境で RAID レベルが必要な場合は、スナップショットを作成する必要がある

バックアップソリューション



セカンダリストレージソリューション

信頼性とコストパフォーマンスの高いバックアップソリューションを実現するには、すべてのハードウェアおよびソフトウェアコンポーネントのパフォーマンスと信頼性を評価する必要があります。バックアップと復元ソリューションは、チェーン(鎖)にたとえることができます。チェーンの強度は最も弱いリンクで決まるのと同様、バックアップソリューションのパフォーマンスと信頼性は、それらが最も低いコンポーネントに準じ、それ以上にはなりません。

HP では、DAT から **ultirum** テープストレージライブラリに至るさまざまなテープバックアップソリューションを提供しています。

テープバックアップソリューションでは、以下の機能を提供しています。

- 大容量
- 高いパフォーマンス
- 高い信頼性

エンタープライズバックアップソリューション(EBS)

エンタープライズ バックアップ ソリューション (EBS) は、アプリケーションソフトウェアと業界標準のハードウェアを統合して、さまざまな環境のデータのバックアップとリストア専用の完全なソリューションを提供します。

EBS により、同じ Fibre Channel Arbitrated Loop (FC-AL) または Fibre Channel Switched Fabric (FC-SW) に接続されている複数の Windows NT、Windows 2000、Tru64 UNIX サーバ間でテープデバイスを共有することができます。

EBS の利点

EBS には以下の利点があります。

- 専用の FC-AL および FC-SW サーバデータ
 - 100MB/s マルチモードおよびシングルモードファイバでの転送
 - 最適なシステムおよび LAN パフォーマンスで両方の環境を最高の状態にする
- ボトルネックなしで、複数のバックアップセッションとリストアセッションを起動可能
- ファイバチャネルでの最大許容距離の増大。災害復旧や重要なレコードのために、保護されるデータを 10km 離れた場所に配置可能
- 管理インタフェース。エンタープライズアクセス用の Web ベースのハブ、スイッチ、ファイバチャネルネットワークストレージルータ (NSR) を提供する
- すばやい投資回収による TCO (トータル運用管理コスト) の削減
- インテリジェントな媒体管理機能を備えた、集中化された自動共有テープライブラリ

災害復旧計画の準備



最悪の障害が発生する前に災害復旧計画を立てておくことで、間違いを防止し、緊張を和らげることができます。災害復旧計画は、破滅的な障害によりデータが破壊された場合に損害を最小限に抑えるための戦略および手順です。サージプロテクタやパスワード保護 RAID アレイなどの予防策を講じることで被害を最小限に抑えることはできますが、データを 100% 保護することはできません。

災害復旧計画のすべての手順を文書化し、最新版を複数の場所に置いておき、すぐに参照できるようにしてください。定期的に手順を更新してください。システムが変更された場合には、それらを復旧するための手順も変更する必要があります。サーバが起動しない場合は、災害復旧計画を参照して、それに記載されている手順に従ってください。

災害復旧計画には、以下の要素を入れてください。

- バックアップテープローテーション方式
- ハードウェア保護計画
- 災害期間中にビジネス運用を維持するための計画
- データをオフサイトで保管するための計画

災害耐性と復旧

災害耐性

災害耐性とは、システムやコンポーネントで障害が発生した場合に、プロセスやアプリケーションを続行できることです。災害耐性を実装するかどうかは、ダウンタイムのコストと潜在的な損失の危険性によって決まります。このコストが、災害耐性ソリューションとそれをサポートするインフラストラクチャのコストよりも大きい場合は、災害耐性ソリューションを実装する必要があります。

災害耐性ソリューションは、会社のプライマリデータセンターから追加の SAN ロケーションにデータをコピーまたはミラーリングして、継続的なデータアクセスを実現します。

災害保護の範囲は、ミラーサイト間の距離をどの程度にするかに影響を与えます。

- 火災、洪水、通信障害、テロ攻撃などの地域的な災害の場合、300km 未満の短い距離で対応可能
- 台風や地震などの広域災害の場合は、WAN またはそれ以上の距離が必要

最大 10km の距離をサポートしているファイバチャネルテクノロジーは、SCSI バステクノロジーよりも有利です。ただし、ファイバチャネルは、災害耐性の接続要件を十分に満たしていません。

災害復旧

災害復旧とは、障害が発生したシステムを復旧する処理のことです。破局的な障害が発生し、復旧計画を実行する必要がある場合に使用するテクノロジー、サービス、および手順の組み合わせです。

災害復旧と災害耐性を必要とするビジネスの数は、増えつづけています。情報技術への依存度が急激に増加しているため、完全に自動化されたコンピュータシステムを使用して提供されるデータ、情報、およびサービスの価値は今後も増大し続けます。

データレプリケーションマネージャ(DRM)

DRM は、ミッションクリティカルボリュームのリアルタイムリモートミラーリングを提供するアプリケーションです。DRM は、エンタープライズの機能です。SAN ベースのレプリケーション用の仮想ディスクをリモートサイトに指定できます。エンタープライズは、ファイバチャネルリンクまたは高密度波長分割多重 (Dense Wave Division Multiplexing) を使用して、自動的に仮想ディスクを最大 100km 離れた場所に複製します。Fibre Channel-to-ATM (FC-ATM) ゲートウェイまたは Fibre Channel over IP (FCIP) ゲートウェイを使用すれば、この距離を無制限にすることができます。

エンタープライズは、回線が復旧すると、書き込み履歴ログからリモートミラーを更新して、通信回線障害から自動的に復旧します。DRM は、データセンタがシャットダウンされた場合に備えて、データの迅速な復旧を保証します。

復習問題

1. テープストレージバックアップテクノロジーの主な目的を挙げてください。

2. データ損失の原因となるのは、次のどれですか。
 - a. 静電気
 - b. コーヒーをこぼす
 - c. 落雷
 - d. 上記のすべて
3. HP ストレージ製品が使用する 5 つのテープテクノロジーを挙げてください。

4. エンタープライズ バックアップ ソリューション (EBS) は、アプリケーションソフトウェアと業界標準のハードウェアを統合して、さまざまな環境のデータのバックアップとリストア専用の完璧なソリューションを提供します。
 - 正
 - 誤
5. 各用語とその説明を対応付けてください。

a.	DAS オープンなオペレーティングシステム上で顧客のアプリケーションを実行しているオープンシステムサーバ
b.	NAS 複数のクライアント用の専用ストレージがあり、サーバとストレージの関係は 1 対 1 である
c.	SAN 既存の LAN に直接接続される独立したインテリジェントデバイス

6. Enterprise Network Storage Architecture (ENSA) の目的を挙げてください。
.....
.....
7. コピーおよびバックアップされる情報の量を表す 4 つの用語を挙げてください。
.....
.....
.....
8. テープ媒体の実際の平均寿命を判断する要因を挙げてください。
.....
.....
.....
.....
9. 災害復旧計画の 4 つの要素を挙げてください。
.....
.....
.....
.....
10. スナップショットとクローン化は、運用データをオフラインバックアップ用にコピーするために使用します。
 正
 誤

11. HP の標準 SAN トポロジで使用されている 4 つの設計を挙げてください。

.....
.....
.....
.....

12. VersaStor テクノロジがサポートしている 3 つのタイプのストレージバーチャライゼーションを挙げてください。

.....
.....
.....

13. 管理者が設定した属性に基づいて容量のプールから仮想ディスクを自動的に作成するテクノロジーを挙げてください。

.....

14. オフサイトに保管されているデイリー増分バックアップ、ウィークリーフルバックアップ、およびマンスリーフルバックアップを使用するテープローテーション方式の名前を挙げてください。

.....

目的

この章では、次の内容について学習します。

- 可用性のレベルと分類
- 高可用性システムの設計に関する要因
 - システムダウンタイムの主な原因
 - ProLiant サーバで高可用性を実現する方法
- クラスタ化の基本概念、クラスタ化の利点、クラスタの 4 つの基本要素
- クラスタ化に含まれるテクノロジーと 3 つの基本的なクラスタモデル
- HP クラスタソリューション

はじめに

可用性とは、サーバシステムまたはコンポーネントが正常に機能している時間や、ハードウェアやソフトウェアの計画的または予定外の停止にもかかわらずサービスを維持できる能力を表す指標のことを指します。可用性は、1年間に占める割合(パーセント)で表されます。最高値は100%で、この場合システムは常に使用可能であることを表しています。

可用性の要件の比較

インターネットがビジネスを運営するための媒体として急速な成長を遂げたことにより、高可用性ソリューション業界も急激に進化しました。たとえば、オンラインショッピングでは、通常、注文および在庫管理用のデータベースが存在し、これは24時間365日使用可能になっていなければなりません。ダウンタイムが短時間であっても、重大な損失を招き、顧客の満足度が低下します。

例

1日10万ドル(1100万円)、つまり1時間あたり4000ドル(44万円)の収益があるEコマースサイトを運営している企業があるとします。ダウンタイム中に試行された取引の80%が復旧後に完了し、20%が競合他社に流れたとします。この場合、ダウンタイム1時間あたりの損失は800ドル(8万8千円)以上です。

Eコマースでの1日あたりの収益が300万ドル(3億3千万円)の企業の場合、ダウンタイム1時間あたりの損失は25000ドル(275万円)以上です。

損失を防止するためには、システムの可用性を高める必要があります。高可用性を実現するにはコストがかかりますが、システムが使用できないときに発生する損失と比べれば、そのコストは容易に正当化できます。

クリティカルアプリケーションのクラス

クリティカルアプリケーションは、ミッションクリティカルまたはビジネスクリティカルのいずれかに分類できます。

ミッションクリティカル

これらのアプリケーションでは 100% のアップタイムが要求されます。これらのシステムのいずれかが使用不能になると、危機的状況に陥ります。ミッションクリティカルなアプリケーションを以下に示します。

- 航空交通管制
- 110 番、119 番などの緊急コールセンタシステム
- 証券取引所の立会場
- 宇宙飛行管制

ビジネスクリティカル

これらのアプリケーションでは、短時間の停止は許されます。ビジネスクリティカルアプリケーションは企業にとっては非常に重要ですが、多少のダウンタイムが発生しても危機的状況には陥りません。

- 銀行の振込みシステム
- 企業の給与支払システム
- 人事システム
- ワークグループアプリケーション
- 予約システム(座席予約、チケット管理など)
- 現金自動支払システム(ATM)
- 運輸物流システム
- メッセージング(電子メール)システム
- Eコマースサイト

例

ワークグループアプリケーションは、営業時間内であれば必ず稼動している必要がありますが、休日に数分または数時間ダウンしても大きな問題にはなりません。

可用性レベルによるサーバの分類

業界では、4つの可用性レベル(AL:Availability Level)によりサーバの分類が定義されています。

- AL1
 - 作業が停止する
 - 制御不能な状態でシャットダウンされる
 - データの完全性は保証される
- AL2
 - ユーザは切断されるが、再度ログオンできる
 - ユーザは、ジャーナルファイルからいくつかのトランザクションを再実行しなければならないことがある
 - パフォーマンスが低下することがある
- AL3
 - ユーザはオンラインのままである
 - 現在のトランザクションを再起動しなければならないことがある
 - パフォーマンスが低下することがある
- AL4
 - 処理はユーザに透過的である
 - 作業が中断されることはない
 - トランザクションが失われることはない
 - パフォーマンスが低下することはない

スタンドアロンの ProLiant サーバは AL1 に分類されます。サーバがダウンすると、そのサーバで実行されているすべての作業が停止します。ProLiant クラスタでは、AL2 または AL3 を実現できます。この場合、多少の中断またはパフォーマンスの低下が発生するだけで済みます。HP では、NonStop Server などの AL4 のシステムも提供しています。

可用性レベル

業界には、可用性のニーズと機能を比較するための3つのレベルの可用性モデルがあります。

- **標準可用性** — スタンドアロンのサーバで実現される約95～99%のアップタイム。例として、ミラー化された Smart アレイがある
- **高可用性** — 計画的または予定外の短時間の停止は許されるアプリケーション。業界標準の ProLiant クラスタでは、推定値 99.9～99.99%のアップタイムを実現可能
- **連続可用性** — 100%のアップタイムを必要とするアプリケーション。NonStop Server では、99.999～100%のアップタイムを実現可能。

可用性の割合

システムの可用性を評価するマトリックスでは、アップタイムを9の数で表します。つまり、9が1つの場合は90%、2つの場合は99%、3つの場合は99.9%という具合になります。ほとんどのシステムベンダでは、システムが稼動している時間の割合でシステムの可用性を見積もっていて、これを9の数で表しています。一部のベンダでは、99.9%～99.9999%のアップタイムを保証しています。

24時間365日稼動で可用性が99%の場合、年間ダウンタイムは87時間以上になります。可用性が99%のシステムでは、年間3日間にわたって継続する障害が1回、8時間にわたって継続する障害が10回、または1時間にわたって継続する障害が87回発生します。1時間の障害が87回発生することは許されるかもしれませんが、3日間連続してビジネスが行えなくなることは、致命的ではないにしても、その損害は非常に大きなものとなります。

アップタイム%	ダウンタイム%	年間ダウンタイム	週間ダウンタイム
98%	2%	7.3 日	3 時間 22 分
99%	1%	3.65 日	1 時間 41 分
99.8%	0.2%	17 時間 30 分	20 分 10 秒
99.9%	0.1%	8 時間 45 分	10 分 5 秒
99.99%	0.01%	52.5 分	1 分
99.999%	0.001%	5.25 分	6 秒
99.9999%	0.0001%	31.5 秒	0.6 秒

注記

▲ スタンドアロンサーバの可用性には、サーバ自体の可用性だけが含まれ、オペレーティングシステム、アプリケーション、またはネットワーク接続の可用性は含まれません。

クラスタの可用性には、サーバハードウェアの可用性と、オペレーティングシステム、クライアント/サーバアプリケーション、クラスタやクライアントマシン間のネットワーク接続の可用性が含まれます。可用性の高いシステムソリューションを実現するためには、1台のサーバ構成では不十分です。

平均故障間隔

可用性とは、サーバが正常に稼動している時間です。以下の簡単な式と測定値を使用してコンポーネントまたはシステムの可用性を計算できます。

$$\text{Availability} = \text{MTBF} / (\text{MTBF} + \text{MTTR})$$

Availability は、可用性を示す値で、パーセントで表されます。MTBF (Mean time between failures: 平均故障間隔) はアップタイムに等しく、MTTR (mean time to recover: 平均回復時間) は特定の問題を解決するのに必要な時間です。

- MTTR が 0 に近づくにつれ、可用性が 100% に近づく
- MTBF の値が大きくなるにつれ、MTTR の可用性に対する影響が少なくなる
- 式 $\text{Availability} = \text{MTBF} / (\text{MTBF} + \text{MTTR})$ は、合計時間を表す

MTTD (measurement mean time to detect: 平均検出時間) は、故障を検出するための平均時間です。

平均とは、傾向や標準を示す値です。システム管理者がある問題から復旧するのに 20 分かかった場合でも、毎回 20 分かかることにはなりません。把握しておくべき概算値として、許容可能な最大ダウン時間と問題解決に必要な最大時間があります。

例

あるシステムの MTBF が 100,000 時間で MTTR が 1 時間の場合、可用性レベルは $100,000/100,001$ (99.999%) になります。MTTR を 6 分 (1/10 時間) に削減すれば、可用性は 99.9999% になります。

高可用性の設計

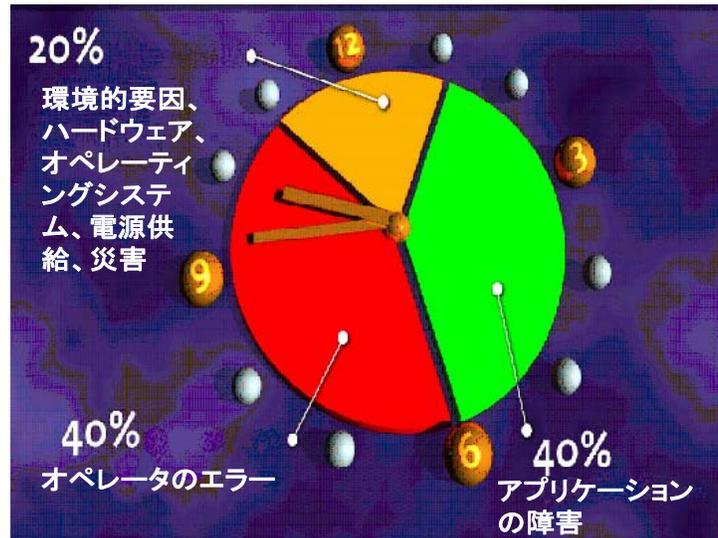
高可用性システムを設計する際には、さまざまな要因について検討、確認、および計画を行う必要があります。障害の原因となるすべての要素を設計に組み込む必要があります。業務の運営方法が可用性に悪い影響を与えないように、手順を設計してそれに従う必要があります。これには、制御の変更手順や災害復旧手順が含まれます。

可用性ドメイン

6つの可用性ドメインは、障害の原因となるすべての要素をカバーするのに必要なサービスの範囲を示します。

環境	物理的なセキュリティ、熱、通気、空調、電源
ハードウェア	物理的なコンピューティング機器
ソフトウェア	オペレーティングシステム、レイヤード製品
ネットワーク	通信インフラストラクチャ
アプリケーション	コア機能を実行している特定のソフトウェア
管理	IT 管理/操作処理

ダウンタイムの原因



ダウンタイム、つまりシステムが使用不能な時間は、いくつかの主な原因によって発生します。

- **ハードウェア障害** — 故障または損傷している機器によるエラー
- **ソフトウェア障害** — オペレーティングシステムとサードパーティ製アプリケーションやドライバとの通信障害
- **計画的なサービスイベント** — ハードウェアのアップグレード(メモリの追加など)後またはソフトウェアのアップグレード(パッチのインストールなど)後のシャットダウンや再起動
- **オペレータのエラー** — 構成を正しく行わなかったり、正しくないサーバをシャットダウンしたり、手順を省略するなどの人為的なエラー
- **環境的要因** — 自然災害、空調、電源供給
- **セキュリティの侵害** — 不正なユーザがシステムに侵入し、ファイルを削除、構成を変更、またはシステムをシャットダウンして使用不能にする

ProLiant サーバでの高可用性の実現

全体的な高可用性を実現するために重要なことの 1 つに、各システムの高可用性を最大化することが挙げられます。HP では、全体的な障害復旧とシステムの可用性に有効な包括的な機能セットを提供しています。

フルスペクトラム障害管理には、以下の 3 つの主なコンポーネントが含まれています。

- **障害予防** — 障害の発生を予防して、ダウンタイムを防止する。これには、潜在的な問題を検出して、それがシステム障害に発展するのを防止する事前予防テクノロジーが含まれる。
- **フォールトトレランス** — 障害発生時にダウンタイムが発生しないようにする。障害は発生するが、フォールトトレランスにより、ユーザや操作にダウンタイムの影響が出ないようになる。障害予防が不可能な場合は、フォールトトレランス機能が重要となる。コンポーネントで障害が発生した場合は、リダンダントコンポーネントが自動的にそれを引き継ぐ。

例

サーバディスクアレイはフォールトトレランスの一つの例で、これによりサーバのユーザや操作にディスク障害の影響が出ないようになります。

- **ラピッドリカバリ** — 問題または障害からの復旧を管理可能。障害管理機能は、潜在的な問題に対処し、システムで障害が発生した場合は、可能な限り迅速にデータを復旧してシステムをオンラインに戻す。

障害予防

HP の障害予防機能の説明を以下の表に示します。

障害予防機能	説明
高度なメモリ保護	エラー検出と訂正 (ECC:Error Checking and Correcting)、オンラインスペア、RAID、ミラーメモリを含む
DSR (Dynamic sector repair)	ハードウェア診断機能を提供し、ドライブアレイの不良セクタを自動的に再マッピングする
Insight マネージャ 7 予測的障害分析	Web ベースのエンタープライズ管理を提供する
アベイラビリティエージェント	一般的なイベントに対する応答を自動化する
UPS	停電時に電源を維持する
温度検出とシャットダウン	システムの温度が危険値を超えたときにそれを検出してシャットダウンを実行する
事前予防保証	管理ソフトウェアを使用して、重要なコンポーネントが機能しなくなる可能性があることを顧客に通知する
マネジメント	リモート Insight ボード Lights-Out Edition (RILOE) と内蔵 Lights-Out (iLO) テクノロジ

フォールトトレランス

HP のフォールトトレランス機能の説明を以下の表に示します。

フォールトトレランス機能	説明
リダンダントコンポーネント - 電源、ファン、NIC、ディスクコントローラ、プロセッサ電源モジュールなど	インストールされている複数のコンポーネントを組み合わせ、サーバコンポーネントの障害に備える
RAID 1、1+0、5、および Advanced Data Guarding (ADG)	ドライブで障害が発生したときのデータ損失を防ぐ
高度なメモリ保護 - オンラインスペアメモリ、ミラーメモリ、RAID メモリ、ECC メモリ	高可用性アプリケーションでのデータ破損とサーバクラッシュを引き起こすメモリエラーを防ぐ
ファンの障害検出とシャットダウン	オペレーティングシステムがファンの障害を検出して自動シャットダウンを実行できるようになる
UPS	外部電源が瞬間的に停電した場合に電源を供給する。停電が長時間にわたる場合は、猶予付きでシャットダウンする
温度モニタ	Intelligent Interface Control を使用して、温度情報を提供する
電圧/電流モニタ	供給電源の電圧と電流の変動をトラッキングする
バッテリーバックアップライトキャッシュ (BBWC)	停電、サーバ障害、またはコントローラ障害時に、キャッシュされているデータを保護する

ラピッドリカバリ

HP のラピッドリカバリ機能の説明を以下の表に示します。

障害管理機能	説明
クラスタ化 - フェイルオーバー構成	サーバの相互自動バックアップを提供することで、さまざまなソフトウェア/ハードウェア障害およびオペレータのエラーを防ぐ
自動サーバ復旧-2(ASR-2)	ASRと同じ機能と、環境復旧機能、高温シャットダウン、UPS シャットダウンを提供する
オンラインリカバリサーバ	2 台のサーバがそれぞれ個別のワークロードを処理しているときに、それらがリダンダントペアとして機能するようになる
Rapid Deployment パック	サーバイメージを短時間で復元する
リダンダント NIC/NIC チーミング	2 つの NIC が同じデバイスドライバコードを共有するため、プライマリ NIC が故障した場合にフェイルオーバーできる
リダンダント電源モジュール	プライマリ電源モジュールが故障した場合に、パワーセーフモジュールをホットスペアとして機能させる
リダンダントパワーサプライ	パワーサプライで障害が発生した場合に、フェイルオーバーリダンダントパワーサプライを提供する
スタンバイリカバリサーバ	2 台のサーバがリダンダントペアとして機能し、1 台がアクティブサーバのホットスペアとして機能するようになる
オンラインスペアドライブ	ドライブ障害が発生した場合に自動再構築を実行して、データ損失のリスクを低減する
テープバックアップ	データの破損や障害が発生したときにデータを取り出せるようにデータを保護する
オフラインプロセッサ復旧	プロセッサで障害が発生した場合に再起動機能を提供する
RILOE/iLO テクノロジ	ブラウザ上の仮想グラフィカルコントロールでサーバを管理する
データリプリケーションマネージャ (DRM)	データをオンラインでミラー化する
ホットプラグ対応コンポーネント - ディスク、電源、ファン、PCI カード、メモリ、アレイコントローラ	サーバまたはワークステーションをシャットダウンすることなく交換可能なコンポーネント

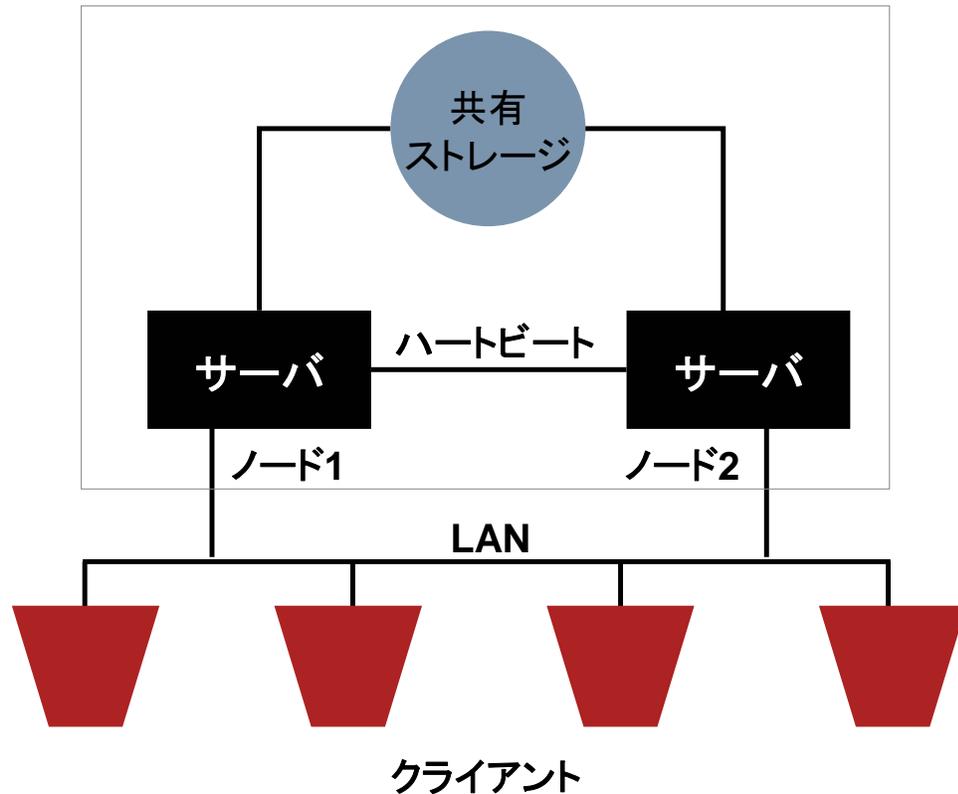
ラピッドリカバリ

高可用性システムでは、障害発生時にダウンタイムを最小限に抑えられるようにラピッドリカバリ方式が採用されています。高可用性システムは、複数のフォールトトレランス機能を備えているため、100%の可用性を実現することも可能ですが、ある程度のダウンタイムは発生します。

ラピッドリカバリ方式の機能は、以下のとおりです。

- **自動ホットリカバリ/再起動** — トランザクションを失わずに障害を検出した時点まで復旧する。または、作業やプロセスを失って障害発生時点まで復旧するウォームリカバリを実行する。
- **クラスタ化** — 複数のプロセッシングシステムが共通のストレージリソースを共有したり、共通のアプリケーションを実行するように構成可能。1つのシステムで障害が発生した場合、ワークロードは自動的に復旧され、別のシステムに移行される。このため、ダウンタイムを最小限に抑えることが可能。
- **アプリケーションモニタリング** — アプリケーションが障害を認識し、復旧を実行できるようになる。

クラスタ化



基本的なクラスタ構成

クラスタ化は、システム障害を迅速に復旧することが可能な高可用性ソリューションです。また、クラスタ化により、プロセッサとI/Oのパフォーマンスを向上させ、ストレージ容量を増やすことができます。

クラスタは、1つの統合されたコンピュータリソースとして使用する複数の疎結合のサーバの集合です。クラスタにより、個別のコンポーネントで障害が発生した場合でも、サーバベースのアプリケーションの可用性を高い状態に維持できます。また、操作の中断を最小限に抑え、処理能力とI/O帯域幅を向上させることができます。システムのハードウェアまたはソフトウェアのいずれのコンポーネントで障害が発生した場合でも、ユーザのパフォーマンスが低下することはありますが、サービスへのアクセスが失われることはありません。

業界標準のProLiantサーバを使用するHPクラスタでは、クラスタ化用の特別なハードウェアは基本的に必要ありません。

クラスタ化の利点

クラスタ化の利点は以下のとおりです。

- リソースの高可用性
- ビジネスの成長に対応したスケーラビリティ
- 集中管理
- 負荷分散
- クラスタ対応アプリケーション

リソースの高可用性

クラスタ化により、リソースの高可用性を実現できます。クラスタ内の 1 つのノードで障害が発生したり、管理者が 1 つのノードをオフラインにした場合、リソースは稼動しているノードにフェイルオーバーできます。フェイルオーバー中、残りのサーバは、オフラインになったノードまたは障害が発生したノードのタスクを自動的に再分散します。

障害が発生したノードまたはオフラインになったノードがオンラインに戻る準備が整うと、リソースは元のノードにフェイルバックできるようになります。この処理は、手動で実行することも、自動的に実行されるように構成することもできます。

ビジネスの成長に対応したスケーラビリティ

クラスタでは、高可用性に加え、高いスケーラビリティも実現できます。クラスタ化を通じて、以下のリソースを徐々に効率良く拡張できます。

- プロセッサ
- I/O
- ストレージ
- アプリケーション

この機能は、システムリソースやデータへの信頼性の高いアクセスを提供するとともに、ハードウェアおよびソフトウェアリソースの投資も保護します。既存のハードウェアと新しいコンピュータをクラスタ化することで、ハードウェアとソフトウェアの両方の投資を保護できます。スタンドアロンコンピュータを置き換える代わりに、同じ能力の別のコンピュータを追加できます。

例

サーバ上のアプリケーションを使用するクライアントの数が増加して、パフォーマンスが低下したとします。そのアプリケーションが複数のノードに処理を分散できるように作られている場合、このサーバを別のサーバとクラスタ化することで、アプリケーションのパフォーマンスと可用性が向上します。

集中管理

一般的なサーバ環境では、さまざまな管理ツールがネットワーク上のサーバを検出し、それらのコンテンツとアクティビティをモニタします。クラスタ環境では、アプリケーションとサービスを集中管理できるため、リソース、サービス、およびアプリケーションをコントロールしやすくなります。

負荷分散

クラスタは、高可用性、スケーラビリティ、および集中化された管理機能を提供するだけでなく、負荷分散機能を提供します。

例

あるサーバ上で大量のアプリケーションが実行されていて、別のサーバはほとんど使用されていないことがわかった場合、通常、システム管理者は、両方のサーバをシャットダウンして再構成する必要があります。それらのサーバが、アプリケーションのフェイルオーバー機能を備えるクラスタグループの一部である場合、システム管理者はアプリケーションを別のサーバに手動でフェイルオーバーして、サーバをシャットダウンすることなくワークロードを分散させることができます。

ほとんどのクラスタ化ソリューションでは、通常の動作中に手動または自動でリソースの負荷分散を行えます。

Web サーバなど、リソースを集約しない環境は、サーバファーム、すなわち複数の独立したサーバで構成されるグループに最適な環境です。各サーバは同時接続を処理し、負荷分散機能を使用して着信リクエストを分散できます。負荷分散機能には、ハードウェア負荷分散機能とソフトウェア負荷分散機能があります。

ハードウェア負荷分散機能の特徴は以下のとおりです。

- ソフトウェア負荷分散機能よりも優れたスケーラビリティ
- 管理が簡単
- アプリケーションサーバのプロセッサに負荷を与えない

ソフトウェア負荷分散機能

- 最大 32 台のシステムを負荷分散クラスタに組み込むことが可能
- すべてのサーバが同じネットワークセグメント上になければならない
- アプリケーションと同じサーバ上で負荷分散が実行される
- 追加のハードウェアが不要なため、ハードウェアの負荷分散機能よりも低価格

クラスタ対応アプリケーション

クラスタ未対応アプリケーションは、手動でフェイルオーバー構成にする必要がありますが、クラスタ対応アプリケーションは、クラスタにインストールされることを認識してクラスタ化に必要なリソースを作成します。これにより、クラスタ未対応アプリケーションの場合よりも簡単にアプリケーションをクラスタ化できます。また、クラスタ対応アプリケーションのほうが管理が簡単で、より多様な障害から復旧することができます。

クラスタ要素

クラスタには、以下の 4 つの基本要素が含まれています。

- **クラスタノード** — クラスタを構成する個別のサーバで構成される
- **クラスタ相互接続** — インターコネクトまたはプライベートネットワークと呼ばれるノード間通信パスで構成される。このパスは、専用または共有のクラスタ内通信インフラストラクチャを形成する
- **共有ストレージ** — 以下の 4 つの主なテクノロジーのいずれかで構成される
 - **ミラードライブ** — データが失われた場合のデータ復旧速度を高速化する
 - **スイッチ付き SCSI** — クラスタノードと共有ストレージシステムとの間に 2 つの SCSI バスを使用する。ストレージシステムにあるスイッチが、2 つのクラスタノードのどちらかがストレージ所有権を持つかを制御する
 - **共有 SCSI** — 同じ SCSI バスを使用して、クラスタノードを共有ストレージシステムに接続する。両方のノードが共有ストレージへのアクティブな接続を同時に維持するが、それぞれが異なるディスクの所有権を持つ
 - **ファイバチャネル** — ファイバチャネルベースの SAN (Storage Area Network) を使用する
- **クラスタ化ソフトウェア** — 共有ストレージ要素を同期して、仮想サーバを作成し、リソースおよびハートビートを管理し、自動フェイルオーバーとフェイルバックを実行する。主要なクラスタ化ソフトウェアには、以下のものがある
 - Microsoft Cluster Service
 - LifeKeeper for Linux
 - Novell Cluster Services

クラスタテクノロジー

クラスタ化にはさまざまな形態がありますが、クラスタは、一般的にはイーサネットで相互接続された一連の標準サーバです。より複雑なタスクを実行したり、より多くのクライアントリクエストを処理できるように、必要に応じてシステムをクラスタに追加できます。クラスタ内の 1 つのシステムで障害が発生すると、そのワークロードは自動的に残りのシステムに分散されます。この転送は、通常、クライアントに対して透過的に実行されます。

用語

業界で使用されているクラスタ化に関するいくつかの概念および用語を以下に示します。

- **ノード** — ノードはクラスタのメンバ。「ノード」と「メンバ」は同じ意味で使用される
- **クラスタサービス** — クラスタ特有のすべてのアクティビティを管理する、各ノード上のソフトウェアの集まり
- **リソース** — クラスタサービスによって管理される認識済みのアイテム。クラスタサービスは、すべてのリソースを同一のオブジェクトとして認識する。リソースには、ディスクドライブやネットワークカードなどの物理ハードウェアデバイス、または論理ディスクボリューム、IP アドレス、アプリケーション全体、データベースなどの論理アイテムがある。リソースは特定のノード上でサービスを提供している場合に、ノード上でオンラインであると呼ばれます。
- **グループ** — 1 つのユニットとして管理するリソースの集まり。通常、グループには、特定のアプリケーションを実行して、提供されるサービスにクライアントシステムを接続するために必要なすべての要素が含まれている。グループにより、管理者はリソースを組み合わせより大きな論理ユニットを作成し、それらをユニットとして管理することができる。グループに対して実行した操作は、そのグループに含まれているすべてのリソースに影響を与える
- **ハートビート** — クラスタ内で使用する通信プロトコルで、ノードが動作しているかどうかを判断する。ハートビートが検出されないと、フェイルオーバが開始される
- **相互接続** — クラスタのノード間の通信リンク。相互接続は、ハートビートまたはクラスタのノード間でのデータ交換に使用可能
- **フェイルオーバ** — 自動または手動で、プライマリサーバノードからセカンダリサーバノードにアプリケーション、サービス、およびデータを移動する処理。自動フェイルオーバは、通常、クラスタノード内でハードウェア障害またはソフトウェア障害が発生した場合に実行される
- **フェイルバック** — 障害が発生したサーバノード、アプリケーション、サービス、またはデータの操作を優先サーバに戻す処理。サービスが実行され、プライマリサーバノードがクラスタに再統合されると、アプリケーションとデータは手動または自動で優先サーバノードにフェイルバックされる

クラスタモデル

クラスタモデルには、以下の 3 つの基本的なモデルがあります。

- シェアードナッシング (Shared nothing)
- シェアードディスク (Shared disk)
- シェアードエブリシング (Shared everything)

シェアードナッシング (Shared nothing)

「シェアードナッシング」モデルでは、サーバノードは共有物理ストレージシステムにアクセスできますが、同じ論理ドライブに同時にアクセスすることはできません。

「シェアードナッシング」ソフトウェアモデルでは、クラスタ内の各システムは、クラスタリソースのサブセットを所有しています。一度に 1 つのシステムだけが特定のリソースを所有およびアクセスできます。ただし、障害が発生した場合には、動的に決定された別のシステムがリソースの所有権を引き継ぐことができます。クライアントからのリクエストは、リソースを所有しているシステムに自動的にルーティングされます。

例

ホストシステムが、データベースからの複数のデータレコードの取得をリクエストしたとします。これにより、多重ディスク読み取りなどの大量のシステムアクティビティが発生します。ただし、最終データが見つかるまで、関連するトラフィックはクラスタ相互接続上に現れません。データベースは複数のクラスタ化されたシステム上で分散されるため、1 台のコンピュータハードウェアによってシステムの全体パフォーマンスが制限されることはありません。

シェアードディスク(Shared disk)

「シェアードディスク」モデルでは、クラスタ内のシステム上で実行されているソフトウェアは、クラスタ内のどのシステムに接続されているリソースにもアクセスできます。2つのシステムが同じデータにアクセスする必要がある場合、ディスクからデータを2回読み取るか、システム間でコピーする必要があります。アプリケーションは、共有データへのアクセスを同期またはシリアル化する必要があります。

通常、分散ロックマネージャ(DLM)を使用して、複数のノードが同じ論理ディスクに同時にアクセスできるようにしてこの同期を行います。DLMは、アプリケーションに提供するサービスで、クラスタ全体のリソースへの参照をトラッキングします。複数のシステムが1つのリソースを参照しようとした場合、DLMは潜在的な競合を認識し、それを解決します。ただし、DLMの調整により、追加のシステムへの関連付けられたシリアル化アクセスが発生するため、追加のメッセージトラフィックが発生し、パフォーマンスが低下することがあります。

シェアードエブリッシング(Shared everything)

「シェアードエブリッシング」クラスタは、シングルシステムイメージクラスタ(SSI)とも呼ばれます。管理者、クライアント、およびオペレーティングシステムからは、システムは1つのユニットとして見えます。ノードは、メモリ、プロセッサ、ディスクを共有します。

HP クラスタソリューション

HP では、1995 年以來、すべてのプラットフォームに対して高可用性ソリューションを提供してきました。初期のソリューションは、複数の HP サーバのデータ可用性を向上させ、サーバをハードウェアエラーから復旧できるようにするものでした。最新の製品は、トップクラスのクラスタ化ソリューションに対する顧客のニーズを満たすように開発されています。

ほとんどの ProLiant サーバが、主要なすべてのオペレーティングシステムでクラスタ化可能であることが認証されています。ProLiant クラスタは、以下のプラットフォームをサポートしています。

- Windows 2000 Cluster Service
- LifeKeeper for Linux
- Oracle Parallel Server for Oracle 8i
- Oracle 9i Real Application Clusters

ProLiant クラスタ

ProLiant クラスタでは、トータル運用管理コストを低く抑えることができ、ハードウェアおよびデータベースや他のアプリケーションで障害が発生した際のダウンタイムを削減できます。

DL380 パッケージクラスタ



ProLiant DL380 パッケージクラスタは、パッケージ化された 2 ノードのクラスタで、ビジネスクリティカルアプリケーション用のクラスタ構成を容易に実現できるように設計されています。2 つの ProLiant サーバノードと共有ストレージで構成されていて、これらはコストパフォーマンスの高い省スペースキャビネットに取り付けられています。

ProLiant DL380 パッケージクラスタの機能は以下のとおりです。

- クラスタ全体にわたる冗長機能で高可用性を実現
 - Smart アレイコントローラの統合
 - 最大 160MB/s の Ultra3 パフォーマンス
 - バッテリバックアップ式キャッシュを標準装備した 14 の共有ドライブ
 - 1TB のストレージで 128MB/s 以上を実現
- 事前予防保証
- 1 つの型番で発注可能
- サーバおよびストレージの性能を最大にする最適化された構成
- 色分けされたケーブルによる容易なケーブル管理

Smart アレイクラスタ



Smart アレイクラスタストレージは、リダンダントコントローラにより高いデータ可用性を提供する、SCSIを使用する2ノードクラスタ化用のエンクロージャです。Smart アレイクラスタストレージは、一般的なSCSIプロトコルをベースにしているため、ハブ、スイッチ、およびケーブルなどの追加のインフラストラクチャは必要ありません。このため、高可用性システムの初期投資を低く抑えることができます。Smart アレイクラスタストレージは、ファイバチャネル SAN ソリューションに容易に移行できます。

Smart アレイクラスタストレージには、ハイエンドストレージシステムのいくつかの高可用性機能を備えています。

- リダンダントホットプラグ対応コンポーネント
 - RAID コントローラ、キャッシュコヒーレント、バッテリーバックアップ式キャッシュを装備
 - 電源とファン
 - キャビネット前面からアクセス可能なハードドライブ。容易に交換可能
- ステータスメッセージ用の統合された管理ディスプレイ
- 2つのドライブで同時に障害が発生した場合にも保護機能を提供する ADG
- 直接接続されたストレージ
- SCSI 接続

MSA1000 用の ProLiant Cluster HA/F100、HA/F200

HA/F100 および HA/F200 クラスタは、エントリーレベルのファイバチャネルを求めている顧客向けに設計されています。HA/F100 と HA/F200 は、SCSI ベースの DL380 パッケージクラスタとは異なり、ファイバチャネルテクノロジーをベースにしています。

MSA1000 用の ProLiant Cluster HA/F100 と HA/F200 は、直接接続されたサーバ環境から、統合されたストレージ環境へ移行するための簡単かつ低コストな方法を求めている顧客向けに設計されています。また、パフォーマンスと拡張性の高いファイバチャネルストレージを求めている ProLiant クラスタの顧客にも適しています。

Enterprise Virtual Array (EVA) 用の ProLiant Cluster HA/F500

Enterprise Virtual Array (EVA) 用の ProLiant Cluster HA/F500 は、ネイティブのファイバチャネルドライブを備えたストレージシステムをベースにしたハイエンドなエンタープライズソリューションです。高いパフォーマンスと大容量、および事実上容量無制限のスナップショットなどのツールを必要とするファイバチャネル SAN 環境の顧客に最適なソリューションです。

HP ハイエンドクラスタソリューション

HP では、AlphaServer または NonStop Server 上で稼動する以下のハイエンドクラスタソリューションを提供しています。

- Tru64 UNIX TruCluster Server
- OpenVMS Cluster
- NonStop ServerNet Cluster

HP クラスターの管理

HP クラスターの管理には、Insight マネージャ 7 が最適です。これにより、1 つのツールセットを使用して、サーバやクラスターを管理することができます。また、管理者は、これらのツールを使用してクラスターコンポーネントを表示および監視できます。保留状態になっている障害や障害が発生したコンポーネントを、ポケットベルや電子メールで管理者に通知することもできます。

Insight マネージャ 7 の Cluster Monitor

Insight マネージャ 7 に組み込まれている Cluster Monitor を使用して、クラスターを表示または監視したり、クラスター固有のイベントのアラートを送信することができます。

復習問題

1. 現在のトランザクションを再起動しなければならなかったり、パフォーマンスが低下するおそれがあるが、ユーザのオンライン状態を維持することができる可用性レベルを挙げてください。

.....

2. クリティカルアプリケーションの 2 つのクラスの名前を挙げてください。

.....

.....

3. 3 つの基本的なクラスタモデルを挙げてください。

.....

.....

4. 障害に対処するための 3 つの方式を挙げてください。

.....

.....

5. ProLiant クラスタがサポートしている 3 つのプラットフォーム名を挙げてください。

.....

.....

6. クラスタ化の利点を挙げてください。

.....

.....

.....

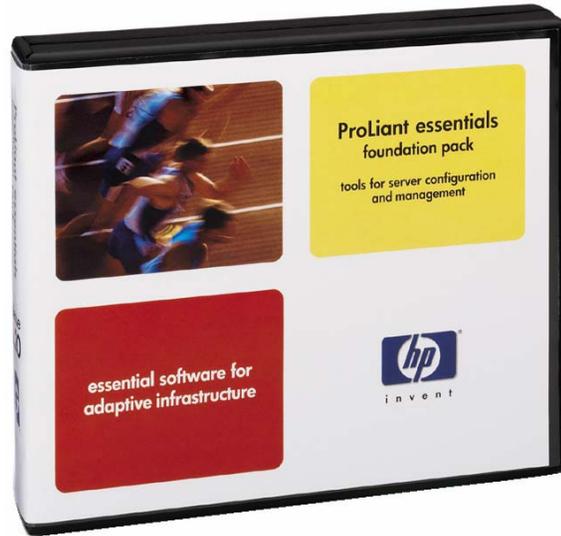
.....

目的

この章では、次の内容について学習します。

- ProLiant Essentials の構成要素
- SmartStart ユーティリティの機能
- Management CD の内容
 - Insight マネージャ 7
 - マネジメントエージェント
 - ActiveUpdate
 - Survey ユーティリティ
- Lights-Out 管理の特徴と機能

ProLiant Essentials



ProLiant Essentials Foundation パック

ProLiant サーバ用のソフトウェアツールやユーティリティは、HP が提供するソリューションにおいて非常に重要です。これらのツールやユーティリティは、ProLiant サーバを簡単、かつ高速でセットアップするだけでなく、構成、展開および管理機能を強化します。

- **Foundation パック** — すべての顧客が必要とするインストール、構成および管理機能を提供します。
- **Performance Management パック** — マネジメントソフトウェアに統合されたソリューションとして、ProLiant サーバのハードウェアに特化したパフォーマンスおよびボトルネックの検出、解析機能を提供します。
- **Rapid Deployment パック** — サーバブレードと大量の ProLiant サーバを素早く一貫性のある状態で配備します。
- **リカバリ サーバ オプション パック** — 1 台のサーバがプライマリ(アクティブ)サーバとして動作し、もう 1 台のサーバがパッシブ(初期化)モードで待機するアクティブ-スタンバイ構成を実現します。
- **内蔵 Lights-Out Advanced パック** — 内蔵 Lights-Out プロセッサをアップグレードして、グラフィカルコンソールと仮想メディア機能を追加します。
- **Workload Management パック** — 大規模な SMP サーバ上の Windows サーバアプリケーションを信頼性の高い状態で統合し、リソースの最適化を可能にする Resource Partitioning Manager (RPM) 2.0 を含んでいます。

Foundation パック



Foundation パックは、主要な ProLiant サーバに付属し、信頼性および統合性の高いサーバをセットアップするのに必要な以下の構成要素を提供します。

- SmartStart - 信頼性が高く一貫性のある状態でサーバを配備
- Insight マネージャ 7 - 簡単かつ強力な Web ベースのサーバ管理
- Insight エージェント - 以下のもので構成される
 - マネジメントエージェント - 予測的な障害管理を可能にする測定機能を提供
 - Version Control エージェント - システムソフトウェアのバージョン管理
- Survey ユーティリティ - 構成の分析およびトラブルシューティング
- Support Pack、ACU、Insight Diagnostics、ROMPaq、などのサポートソフトウェアとユーティリティ

SmartStart

SmartStart
invent

ホーム | セットアップ | メンテナンス | CD情報 | 終了

ProLiantサーバ

サーバのモデル: ProLiant ML370 G2
 ROM: P25 04/16/2002
 RAM: 512 MB
 ブート コントローラ: Compaq Smart Array 5300 Controller
 ブート コントローラのディスク容量: 17351 MB
 オペレーティングシステム: Microsoft Windows

起動
セットアップ

「すべてのデバイスを表示」

PCIデバイス情報

デバイス 0:	Compaq NC3163 Fast Ethernet NIC
デバイス 1:	Compaq Advanced System Management Controller
デバイス 2:	Compaq 64-bit-66 MHz-Dual-Channel-Wide Ultra3 SCSI Host Bus Adapter
デバイス 3:	Compaq 64-bit-66 MHz-Dual-Channel-Wide Ultra3 SCSI Host Bus Adapter
デバイス 4:	Compaq Smart Array 5300 Controller
デバイス 5:	ATI RAGE XL Video Controller
デバイス 6:	Compaq RILOE II Main Processor
デバイス 7:	Compaq RILOE II FPGA
デバイス 8:	Compaq PCI Hotplug Controller

SmartStart and Support Software CD には、SmartStart セットアッププログラムが収録されています。これは、プラットフォームの構成を最適化し、ProLiant サーバのセットアップとインストールを容易にする統合ツールです。

SmartStart は、ハードウェアを構成し、最適化されたドライバを読み込み、ソフトウェアのインストールを自動化します。また、複数の ProLiant サーバ上のオペレーティングシステムを統合して、信頼性とパフォーマンスを最適化します。統合管理機能により、信頼性が高く一貫性のある状態でサーバの配備と保守を行えます。

SmartStart の機能

SmartStart の機能は以下のとおりです。

- **セットアップ(インストール)**
 - 構成に必要な情報を、ウィザード形式の一連のグラフィカルなインタビュー画面に統合します。
 - ソフトウェアがインストールされる前に確認できるように、構成と選択したパラメータのサマリを表示します。
 - 選択したソフトウェアに適するようにハードウェアを構成し、適切な HP ドライバと多数のサードパーティ製ドライバを自動的にインストールします。
 - 適切な CD の挿入後は、ユーザ介入なしにソフトウェアのインストールを実行します。
- **ユーティリティ**
 - マネジメントエージェントを自動的にインストールおよび構成します。
 - 最適なパフォーマンスを実現できるようにソフトウェアを設定します。
- **ProLiant Support Pack (PSP)**
 - ドライバとユーティリティをインストールまたはアップグレードできます。
 - 複数のオペレーティングシステム用のソフトウェアをサポートします。
- **ブラウザベースのインタフェース**
 - hp.com ダウンロードソフトウェアの Web ページを模倣しています。
 - 実行したい機能にすばやく移動することが可能です。

SmartStart6.00 の機能

SmartStart 6.00 は、進化が加速し、より複雑になりつつあるサーバプラットフォームに対応するために登場した新しい世代の配備ツールです。SmartStart 6.00 は、内蔵ユーティリティである RBSU と ORCA を使用してハードウェアとアレイの構成を行います。

CD ベースで動作し、インテグレーションサーバはサポートされません。

リモートあるいは、ネットワークを経由した展開が必要な場合は、SmartStart Scripting Toolkit あるいは、Rapid Deployment パックを使用します。

ソフトウェア配布については、ディスクビルダに代わり、hp.com ダウンロード ソフトウェアの Web ページを模倣した、ブラウザ インタフェースに置き換えられました。

新しいインタフェースは、実行したい機能にすばやく移動できるタブを持っており、次の特徴があります。

- **ホームページ** — サーバの構成パラメータと PCI デバイス情報を表示します。ナビゲーション バーは、関連するサーバ製品の補足情報を表示します。
- **セットアップ** — オペレーティングシステムのプリインストールインタビューを開始します。
- **メンテナンス** — 展開およびサーバメンテナンスに必要なメンテナンスユーティリティへのリンクを提供します。
- **CD 情報** — CD に収録されている最新情報やリリースノートを表示します。

SmartStart 6.00 は、さらに次の機能を備えています。

- 単一サーバのセットアップを容易かつ高速化します。
 - リブート回数の減少
 - サーバプロファイルディスクが不要
 - システムパーティションが不要
 - 使いやすい ROM ベースユーティリティ
 - CD ベースのインストールのみ
 - 自動インストールのみ
- ブラウザベースの新しいユーザインタフェースによりナビゲーションが容易になります。
 - タブを使用したメニューにより、ナビゲーションが容易
 - より詳細なシステム情報
 - メンテナンスユーティリティは、ブラウザをベースとした共通のユーザインタフェースに変更されました
 - ProLiant Essentials ソフトウェアとの一貫したユーザインタフェース
- SmartStart CD からメンテナンスユーティリティに直接アクセスします。
- レプリケートインストールが必要な場合は、SmartStart Scripting Toolkit や Rapid Deployment パックなどのマルチサーバ展開ツールを使用します。
- 自動実行メニューにより、イメージの作成が容易になります。
 - メンテナンスユーティリティへのアクセス
 - SmartStart CD からのユーティリティのインストール
 - hp.com のサーバサポート Web ページへのリンク

SmartStart 5.5 と SmartStart 6.x の比較

SmartStart 5.5 までは、システムパーティションベースのセットアップユーティリティを使用する ProLiant サーバ用に設計されています。SmartStart 6.x 以降は、RBSU (ROM ベースセットアップユーティリティ) と ORCA (Option ROM Configuration for Arrays) を実装している ProLiant サーバ用に設計されています。



重要

ProLiant ML/DL のうち G2/G3 は RBSU ベースのサーバです。

RBSU と ORCA をサポートしないサーバのインストールやメンテナンスが必要な場合は、SmartStart 5.5 を使用します。

SmartStart 6.x は、ナビゲーションが容易なブラウザベースの新しいユーザインタフェースを使用した、サーバセットアッププロセスの全面的な改良が特徴です。SmartStart 6.x においては、サーバのハードウェア構成に RBSU を使用し、基本的なアレイ構成は、ORCA を使用して作成します。高度なアレイ構成を作成するには、ACU-XE を使用します。

SmartStart 6.x の自動インストールでは、サーバプロファイルディスクレットを使いません。また、システムパーティションの作成も不要になりました。

SmartStart 5.5 でサポートされるサーバ

SmartStart 5.5 でサポートされる ProLiant サーバは以下の通りです:

- ProLiant ML330/e/G2/G3
- ProLiant ML350/G2/G3
- ProLiant ML370/G2/G3
- ProLiant ML530
- ProLiant ML570
- ProLiant ML750
- ProLiant DL320
- ProLiant DL360/G2
- ProLiant DL380/G2
- ProLiant DL580/G2
- ProLiant DL760

SmartStart 6.x でサポートされるサーバ

SmartStart 6.x でサポートされる ProLiant サーバは以下の通りです:

- ProLiant ML310
- ProLiant ML330 G2
- ProLiant ML330 G3
- ProLiant ML350 G2
- ProLiant ML350 G3
- ProLiant ML370 G2
- ProLiant ML370 G3
- ProLiant ML530 G2
- ProLiant ML570 G2
- ProLiant DL320 G2
- ProLiant DL360 G2
- ProLiant DL360 G3
- ProLiant DL380 G2
- ProLiant DL380 G3
- ProLiant DL560
- ProLiant DL580 G2
- ProLiant DL740
- ProLiant DL760
- ProLiant DL760 G2

サポートしているオペレーティングシステム

SmartStart は、パッケージ製品として市販されている以下のオペレーティングシステムのインストールを自動化し、ドライバ/ユーティリティのサポートを提供します。

- 自動インストール
 - Windows 2003
 - Windows 2000
 - Windows NT Server 4.0
 - ◆ Windows NT Server Enterprise Edition 4.0
 - ◆ Windows NT Server 4.0. Terminal Server Edition
 - NetWare 6.x および 6.x SBS
 - NetWare 5.1 および 5.1 SBS
- ドライバ/ユーティリティのサポート(オペレーティングシステムは手動でインストールする必要がある)
 - Red Hat Linux 8.0 Professional
 - Red Hat Linux 7.2/7.3 Professional
 - Red Hat Linux Advanced Server 2.1
 - SuSE Linux Enterprise Server 7
 - United Linux 1.0

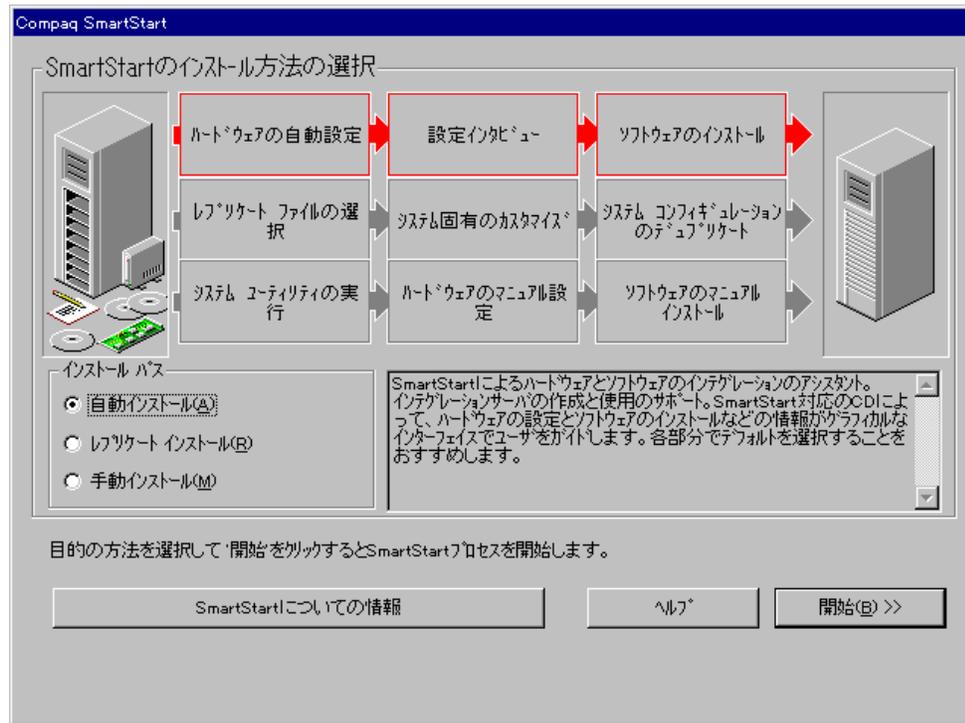
▲ 注記

SmartStart がサポートしているオペレーティングシステムの一覧は、SmartStart CD に収録されています。

ハードウェアの最小要件

- 128MB RAM (SmartStart 6.x 以降) / 32MB (SmartStart 5.50 まで)
- CD-ROM
- マウスとキーボード
- VGA モニタ

SmartStart 5.5 のインストールパス



SmartStart を使用して ProLiant サーバをセットアップする場合、以下の 3 つのインストールパスのいずれかを選択します。

- **自動インストール** — ハードウェアとソフトウェアの統合に関する SmartStart のすべてのメリットを得られます。このパスでは、ハードウェアを構成するための情報収集手順が順を追って提示され、システムソフトウェアがインストールされます。パス全体で検証機能、オンラインヘルプ、推奨デフォルト値が提供されます。いつでもサマリを表示して、インストールの設定を確認できます。また、あとで参照できるようにサマリが保存されます。

▲ 注記

自動インストールを実行するには、サーバプロファイルディスクセットが必要です。サーバプロファイルディスクセットを作成するには、メモ帳またはコマンドプロンプトの `edit` ユーティリティを使用して空の `spd.ini` ファイルを作成します。

- **レプリケートインストール** — 保存したオペレーティングシステムの構成情報を複数のサーバに複製できます。このパスでは、サポートしているソフトウェアのインストール時のパラメータが取得および保存されます。その後、構成情報はプロファイルに保存されます。このプロファイルを繰り返し使用することで、ソフトウェアのインストール時間を短縮できます。レプリケートインストールパスを使用することで、時間を節約できるとともに、エンタープライズ全体にわたってオペレーティングシステムを一貫した状態で配備できます。

**重要**

- レプリケートインストールでは、インストール先サーバは元の構成を保存したサーバと同じ状態になっている必要があります。
 - すべてのデバイスが同じスロットに装着されていなければなりません。
 - メモリが同じでなければなりません。
 - ディスクドライブの構成が同じでなければなりません。構成情報が含まれているテキストファイルは、ディスクに保存されます。
-

- **手動インストール** — システムコンフィギュレーションユーティリティを手動で実行して、ソフトウェア製造元のインストール手順に従うことができます。しかし、統合に関するすべてのメリットが得られるのは、自動インストールパスだけです。

手動インストールパスは、**SmartStart** に対応していない CD を使用してオペレーティングシステムをインストールする際に使用できます。また、HP のソフトウェア製品 CD に収録されているソフトウェアをインストールする際に、より柔軟にインストール設定を行う必要がある場合にも使用できます。

**重要**

- **SmartStart** は、新しいシステムを初期化する場合のみ使用してください。SmartStart を使用して構成済みサーバを初期化する場合は、まず **System Erase** ユーティリティを使用して以前のハードウェア構成とソフトウェア構成を完全に消去してください。
-

ユーティリティへのアクセス

SmartStart 6.x CD から、以下のユーティリティを実行できます。

- ROM ベースセットアップユーティリティ(RBSU)
- アレイコンフィギュレーションユーティリティ XE (ACU-XE)
- Array Diagnostics Utility (ADU)
- System Erase ユーティリティ
- Insight Diagnostics
- ROM アップデートユーティリティ

SmartStart 5.5 CD から、以下のユーティリティを実行できます。

- システムコンフィギュレーションユーティリティ (SCU)
- アレイコンフィギュレーションユーティリティ (ACU)
- Array Diagnostics Utility (ADU)
- システムパーティションアップグレードユーティリティ
- ディスケットビルダユーティリティ
- System Erase ユーティリティ
- Fibre Channel Diagnostics
- Fibre Channel Fault Isolation Utility
- システム ROMPaq

これらのユーティリティにアクセスするには、構成済みのサーバに挿入されている SmartStart CD にアクセスします。



注記

これらのユーティリティの大半は、他の形式のメディアからも実行できます。

システムコンフィギュレーションユーティリティ(SCU)

COMPAQ

COMPAQ SYSTEM CONFIGURATION
バージョン 2.58A

(C) 1989, 2001 コムパックコンピュータ株式会社

COMPAQ, DESKPRO, SYSTEMPRO, SYSTEMPRO/LT, LTE, FASTSTARTは、
米国Compaq Computer Corporationの商標です。

ここに登場するその他の製品名は、それぞれ各社の
商標および登録商標です。

どれかキーを押して、先へお進みください。

システムコンフィギュレーションユーティリティ(SCU)は、主に構成要素を設定するのに使用しますが、強力な診断ツールとしても機能します。システムに構成要素を追加する際の問題の発生を防止したり、問題を修正することができます。

SCUは、SmartStart、ディスクット、またはハードディスクドライブ上のシステムパーティションから実行できます。システムパーティションは、インストールプロセス中に SmartStart からインストールされ、起動時に[F10]キーを押すことでアクセスできます。このパーティションは、SCU ディスクットから作成することもできます。このパーティションは、ハードディスクドライブ上の最初のパーティションでなければなりません。

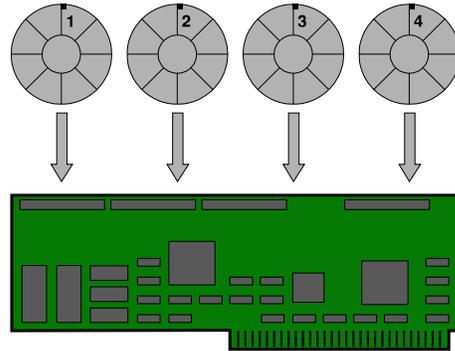
サービスの問題の解決や防止に有効なユーティリティの機能は以下のとおりです。

- サーバ上の大容量ストレージコントローラオーダの設定
- 使用可能なデバイスリソースの判断
- リソースの構成
- 使用可能なデバイスリソースの割り当て
- システムパーティションの更新
- 不揮発性 RAM(NVRAM)の消去(Advanced モードで実行可能)

▲ 注記

すべてのリソースが正しく構成されるようにするために、SCU はサーバのハードウェアの変更がすべて完了してから実行することをお勧めします。

大容量ストレージコントローラオーダの設定



コントローラオーダは、起動時にブート可能ドライブを検出するのに、システムが各ディスクコントローラにアクセスする順序を決定します。この順序は、ユニットに複数の **SCSI** コントローラとアレイコントローラが存在する場合に特に重要です。新しい **SCSI** コントローラまたはアレイコントローラをインストールした場合は、順序を手動で再構成しなければならないことがあります。手動で再構成しないと、サーバが正常に起動しないことがあります。

例

サーバに 3 つのアレイコントローラがあり、そのうちの 1 つがコントローラオーダで 1 番目に設定されているとします。また、内蔵コントローラにブート不能ドライブがいくつか接続されていて、コントローラオーダで最後に設定されているとします。この場合、**NVRAM** を消去してリセットすると、内蔵コントローラがコントローラオーダで 1 番目に設定され、起動できなくなります。

コントローラオーダを設定するには、[View or edit details] を選択し、リストを下にスクロールして、ディスクコントローラとそれに関連付けられているオーダを見つけます。

サーバ上の **IDE** コントローラ (**CD-ROM** ドライブで使用) は、[Integrated Fixed Disk Controller] として表示され、コントローラオーダで 2 番目に設定されており、それ以降のオーダに設定することはできません。このため、ブート可能ドライブを含んでいる **SCSI** コントローラが 1 番目、**IDE** コントローラが 2 番目、追加の **SCSI** コントローラまたはアレイコントローラが 3 番目以降になります。

注記

まれに、ユニットにアレイコントローラをもう 1 つインストールすると、それが誤って 1 番目になることがあります。これにより、オペレーティングシステムの起動時に起動エラーが発生することがあります。この場合は、システムコンフィギュレーションユーティリティ (**SCU**) を使用して (または **NVRAM** を消去して) システムを再起動して、オーダを正しく構成します。

使用可能なデバイスリソースの判断

ステップ 3: 設定の参照/変更 Alt+F1

全ての情報を見るにはカーソルキーを押してください。
 する場合は[Enter]を押してください。
 する場合は[F6]を押してください。

アドバンストメニュー

詳細システム情報の表示

> ホードの仕様 Windows 2000
システム仕様 7011DKF10017
使用されているリソース
利用可能なリソース

> 選択=ENTER < 取消=ESC

内蔵IDEコントローラ

標準インターフェイス.....	セカンドコントローラとして有効
ドライブ タイプ - エント 1.....	CD-ROM
ドライブ タイプ - エント 2.....	インストールされていません

COMPAQ メモリ

利用可能なシステム メモリ 4GB.....	262144 KB
------------------------	-----------

続き: PgUp/PgDn

プラグアンドプレイに対応していない ISA カードをインストールする場合、または電源投入時セルフテスト(POST)で発生したリソースの競合に関するエラーをトラッキングする場合、使用可能なデバイスリソースを判断することは重要です。

使用中のリソースと使用可能なリソースを判断するには、次の手順に従ってください。

1. [詳細の参照/変更]画面で[アドバンスト]([F7]キー)を押します。
2. [詳細システム情報の表示]を選択します。

使用済みリソースと使用可能リソースがオプションの上に表示されます。ここには、以下の情報が表示されます。

- IRQ
- DMA (Direct Memory Access)
- I/O(ベース)アドレスポート
- メモリアドレス

▲ 注記

[詳細システム情報の表示]で[システムの仕様]を選択することで、PCI スロットがバスマスタリングをサポートしているかどうかを確認できます。

使用可能なデバイスリソースの割り当て

SCU の [設定の参照/変更] 画面で、インストールされているカードの設定を変更します。これらの設定には、リソースとリソース以外のオプションがあります。

例

リモート Insight ボードには、ボード上の IP アドレスを構成したり、ユーザ、権限、およびパスワードを割り当てたりするためのオプションがあります。

インストールされている EISA、ISA、PCI デバイスのリソースは、[設定の参照/変更] 画面で構成します。プラグアンドプレイに対応していない ISA デバイスをシステムに認識させるには、それらを手動で構成する必要があります。

PCI が原因で構成に関するいくつかの問題が発生することがあります。PCI 仕様は EISA 仕様の改良版で、この仕様では 128 の割り込みが許可されています。ただし、オペレーティングシステムおよび関連付けられているデバイスドライバにより、マップモードまたは EISA マップモードと呼ばれるフルテーブルモードの PCI 割り込みの一部の使用が制限されます。この制限は、柔軟性の低い方法で割り込みを共有する PCI スロットによって発生し、場合によっては、共有割り込みが互換性のあるデバイスドライバ間で共有されるようにカードを物理的に再配置しなければならないこともあります。

PCI カードがオペレーティングシステムによって認識されない場合、または新しい PCI カードを追加した後システムが動作しなくなった場合は、SCU を実行して割り込みの共有が原因であるかどうかを確認してください。

アドバンスドモード

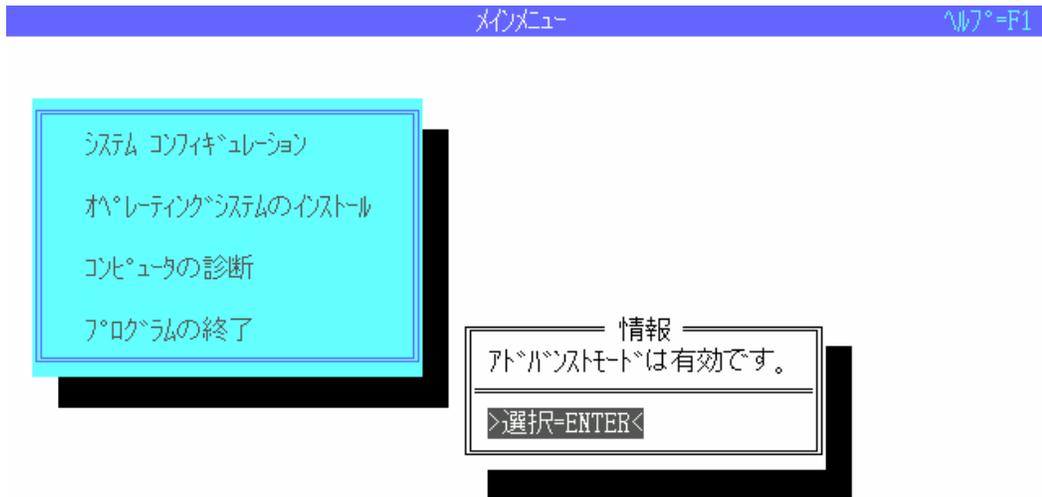
SCU のアドバンスドオプションにアクセスするには、メインメニューで [Ctrl]+[A] キーを押します。Advanced モードにアクセスすると、[Advanced Programmable Interrupt Controller (APIC)] プロセッサ設定や [不揮発性メモリの消去] オプションなどの追加オプションを使用できるようになります。

Pre-Boot Execution Environment (PXE) の有効化

PXE により、ブートディスクが物理的にシステムに存在している場合と同様に、ネットワークベースのブートイメージを使用してサーバを起動できるようになります。これにより、無人サーバのリモート構成および配備が可能になります。

PXE サポートを有効にするには、SCU を実行して [設定の参照/変更] を選択し、有効にする NIC を選択します。有効にできる NIC は 1 つだけです。

NVRAM の消去



NVRAM は、EEPROM (電氣的に消去可能なプログラマブル ROM) で、重要なエラーログ、システム情報テーブル、およびステータスフラグなどのさまざまなデータを格納します。まれに、NVRAM が破損したり、NVRAM に正しくない情報が含まれたりすることがあります。これにより、以下のようなさまざまなエラーが発生します。

- No video
- No POST
- POST エラー
- システムエラー
- ロックアップ

NVRAM を消去することで、異常なエラーや原因不明のエラーを解決することができます。

NVRAM を消去するには、SCU がアドバンスモードになっている必要があります。NVRAM を消去するには、次の手順に従ってください。

1. [システム コンフィギュレーション]メインメニューで、[Ctrl]+[A]キーを押してアドバンスモードに入り、[Enter]キーを押します。
2. [システム コンフィギュレーション]→[ハードウェアの設定]→[ハードウェア設定の参照と変更]→[設定の参照と変更]を選択します。
3. 構成要素リストを下にスクロールして、[拡張機能]を表示します。
4. [不揮発性メモリの消去]を選択し、プロンプトに対して[はい]と応答します。
5. システムを再起動します。システムの再起動後、「162-System Options Not Set」などのいくつかの POST エラーメッセージが表示されます。
6. [F10]キーを押して、SCU に入ります。
7. バックアップバイナリ(.SCI)ファイルを使用してシステムを自動構成します。

8. プライマリオペレーティングシステムを選択します。

**警告**

オペレーティングシステムを選択すると、ハードディスクドライブの変換設定が構成されます。誤ったオペレーティングシステムを選択すると、起動時にハードディスクにアクセスしようとしたときに、システムがブートしなくなります。システムが正常にブートした場合でも、誤った設定によってデータが破損することがあります。

9. [ハードウェア設定の参照と変更]→[設定の参照と変更]を選択します。
10. 必要に応じて、コントローラオーダと他の構成要素を再構成します。

**注記**

新しいサーバでは、SCU の代わりに RBSU が採用されています。このユーティリティで NVRAM を消去するには [アドバンスド オプション] メニューから [不揮発性メモリの削除] を選択します。

システムメンテナンススイッチによる NVRAM の消去

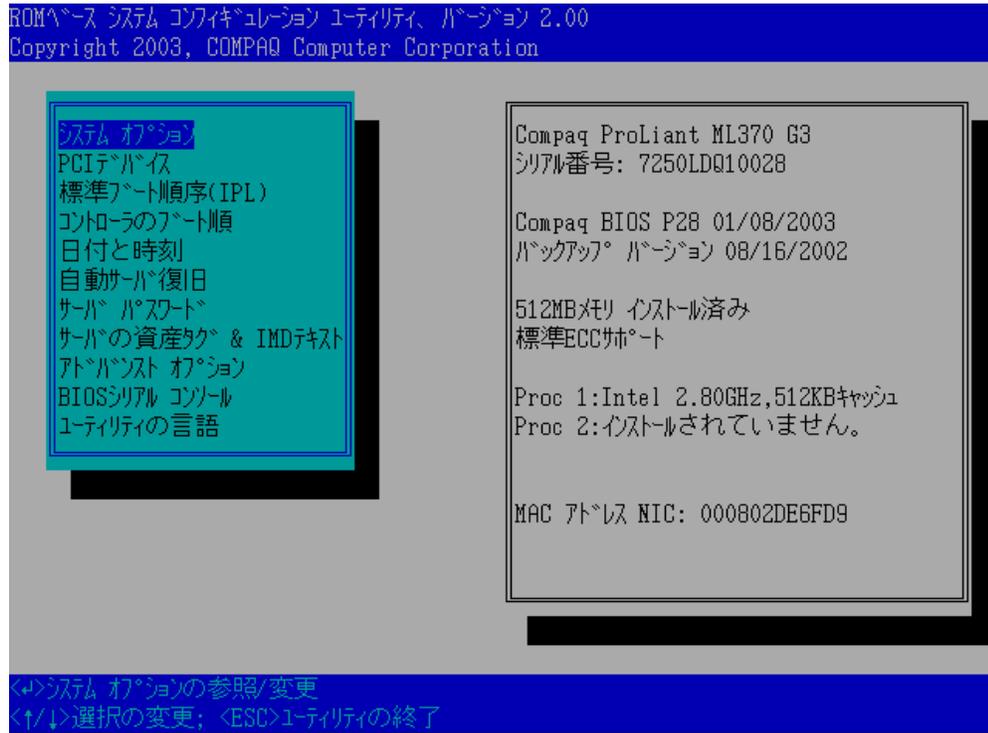
ほとんどの ProLiant サーバには、一連のシステムメンテナンススイッチ (通常は SW1-6) があり、これを設定して NVRAM を消去できます。これらのスイッチは、通常、サーバシステムまたは I/O ボード上にあります。

NVRAM を消去するには、次の手順に従ってください。

1. サーバの電源を切ります。
2. スイッチ 6 (通常は SW1-6) を切り替えます。
3. サーバの電源を入れ、POST が完了するのを待ちます。
4. オペレーティングシステムのロードが開始する前にサーバの電源を切ります。
5. スイッチ 6 (通常は SW1-6) を元の位置に戻します。
6. サーバの電源を入れます。

NVRAM が正常に消去されると、POST 中に「162-System Options Not Set」メッセージが表示されることがあります。特定のサーバでの詳細については、アクセスパネルラベルなどを参照してください。

ROM ベースセットアップユーティリティ(RBSU)



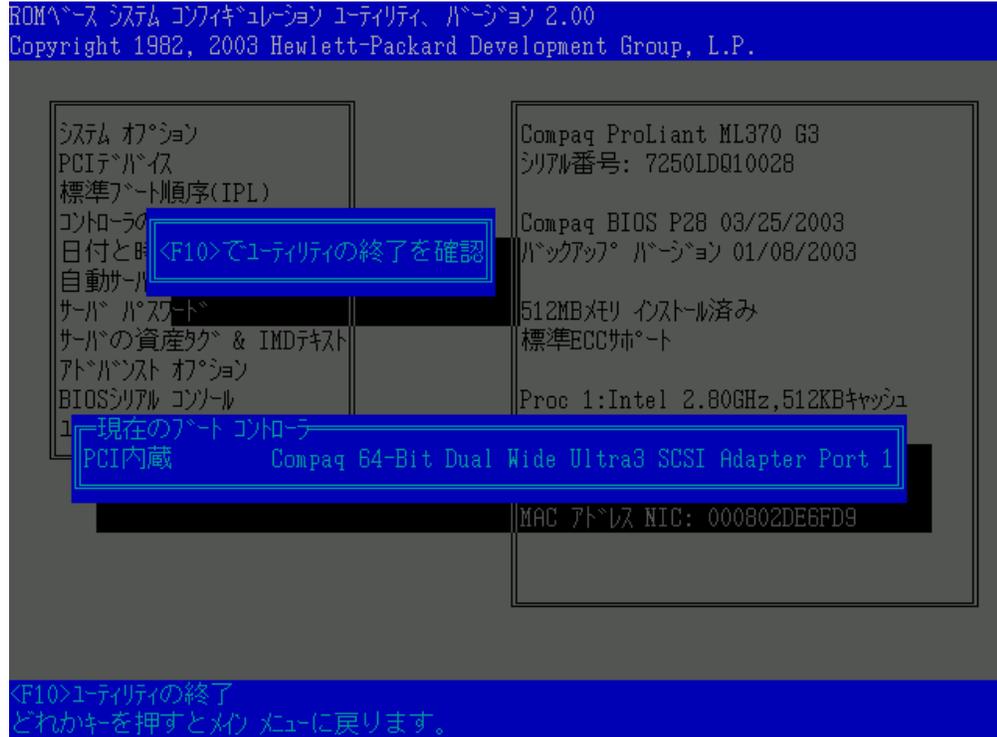
RBSUは、サーバのROMに内蔵されている更新可能な構成ユーティリティです。RBSUは、サーバハードウェアの設定を行ったり、オペレーティングシステムをインストールできるようにサーバの状態を整えたりするのに使用します。SCUと同様、RBSUを使用して、システムを初めて起動するときにサーバの構成情報を表示し、確認できます。また、サーバを構成した後でそれらを変更できます。

RBSUは、2000年後半から出荷されたProLiantサーバに実装され、SCUと同じ機能を提供しています。

以下の表に、RBSUとSCUの違いを示します。

ROM ベースセットアップユーティリティ	システムコンフィギュレーションユーティリティ
変更が行われた場合に、変更内容をNVRAMに保存する	終了するまで変更内容を保存しない
メッセージを表示せずに競合を解消する	競合解消時にメッセージを表示する
システムROMに内蔵。ディスクを使用しない	ディスクベース。システムパーティションにインストール可能
各サーバ用にカスタマイズされているため、サイズが小さく高速なユーティリティ	包括的なユーティリティ。1つのバージョンですべてのサーバをサポート
構成指向、テーブル方式	デバイス指向、ファイル方式
RBSUテーブル内の構成情報を使用した、レプリケーションユーティリティのサポート	構成バックアップを使用した場合を除いて、直接的なレプリケーションユーティリティのサポートなし
RBSUROMのフラッシュまたはROMの物理的な変更によるユーティリティの更新	新しいバージョンのソフトウェアによるユーティリティの更新
詳細オプションはメニューから選択	詳細モードには、[Ctrl]+[A]キーを押してアクセス

RBSU の起動と終了



RBSU にアクセスするには、サーバの電源投入後、システムの起動プロセス中に[F9]ファンクションキーを押します。これにより、SmartStart CD または起動ディスクを使用することなく、RBSU を起動してサーバの構成を変更できます。

サーバが未構成状態の場合は、RBSU が自動的に実行されます。サーバの構成が完了したら、RBSU を終了して、新しい設定でサーバを再起動する必要があります。メインメニューで [Escape] キーを押すと、終了確認が表示されます。現在のブートコントローラも参考情報として表示されます。[F10] キーを押して RBSU を終了すると新しい構成でサーバが起動します。

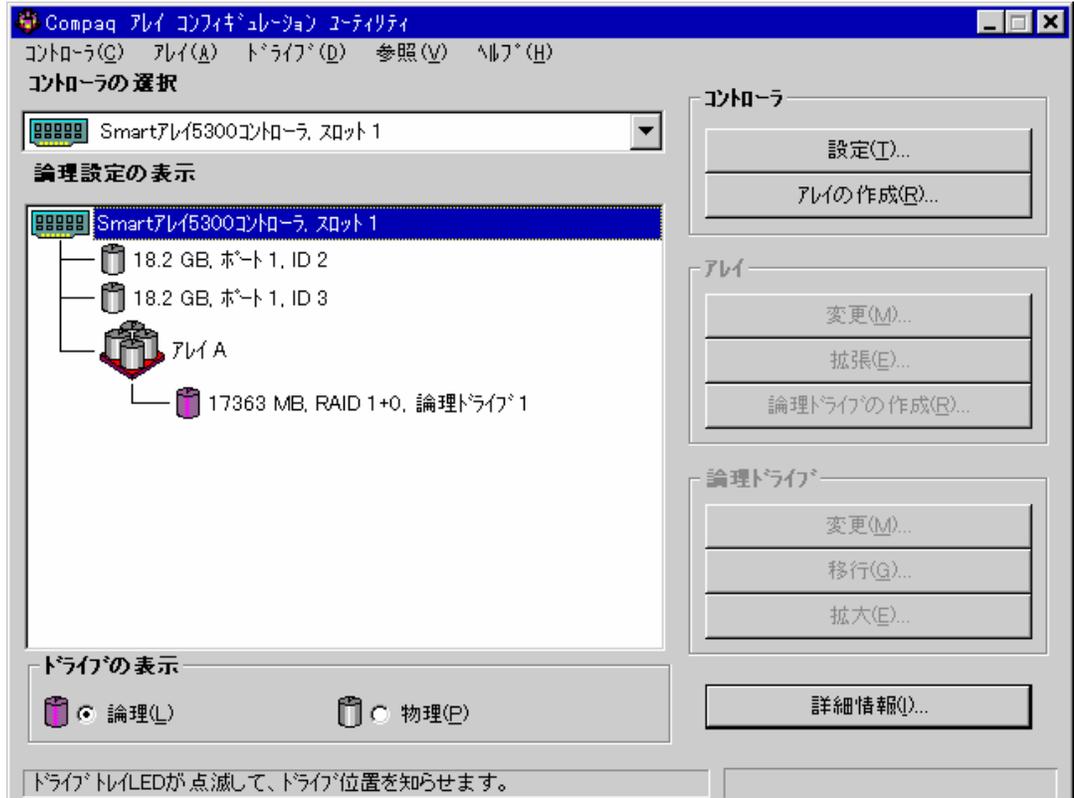
初期ブート

まだ構成されていないシステムを初めて起動するときには、オペレーティングシステムの情報を入力するように求められます。

▲ 注記

この手順をバイパスするには、サーバをブートする前に、ROMPack や署名されたディスクをディスクドライブに挿入する必要があります。これにより、ROM のアップグレードや、SmartStart Scripting Toolkit の実行ができるようになります。

アレイコンフィギュレーションユーティリティ(ACU)



ACUでは、ストレージの構成、変更、拡張、管理、監視を容易にする Smart アレイコントローラのインテリジェントな機能へのインタフェースを提供しています。

ACUを使用すれば、物理ドライブをアレイに迅速に割り当て、各コントローラ用のオンラインスペアや論理ドライブを作成することができます。ACUでは、すべての領域が割り当てられるまで、論理ドライブを構成するための手順が順を追って提示されます。論理ドライブごとにフォールトトレランスのレベルを選択できます。

ACUは、以下のものから実行できます。

- Windows 2000
- Windows NT
- NetWare
- SmartStart

独自の構成パスにより、論理ビューと物理ビューの詳細を確認し、必要に応じて特定の構成変更を行えます。アレイと論理ドライブのグラフィカルな表示により、ストレージがどのように構成されているかをビジュアルに確認することができます。

容量の拡張

容量の拡張は、論理ドライブレベルではなく、アレイレベルで行います。通常、1つのコントローラに接続されているすべてのディスクドライブは、1つのアレイにグループ化します。この構成により、RAID 冗長機能を最も効率的に使用できるようになります。ACUを使用することで、複数の物理ドライブを1つのアレイに割り当て、1つのアレイコントローラあたり最大4つのドライブをオンラインスペアとして割り当てることができます。

ACUでは、最大8つのコントローラを構成することができます。最大構成の場合、ACUの処理速度が多少低下することがあります。ACUを使用すると、構成中のコントローラの物理ドライブトレイの3つのLEDインジケータが点滅します。

注記

▲ ACUは、SMARTアレイコントローラを認識しません。また、これを構成することはできません。

アレイコンフィギュレーションデータの場所

ACUでアレイコンフィギュレーションを保存すると、これらの情報は、各ハードディスクドライブのRAID情報セクタ(RIS)に格納されます。これにより、コンピュータの構成要素を交換したり、一連のドライブをマシン間で移動したりする際にデータが失われることがなくなります。

通常、論理ボリューム構成やRAIDレベルを変更すると、データが破壊されます。

注記

▲ ACUは、構成を自動的に保存しません。必ず設定を保存してからユーティリティを終了してください。

再構築の優先順位

ドライブで障害が発生し、オンラインスペアが使用可能な場合、アレイコントローラは再構成処理をバックグラウンドで自動的に開始します。システムによって割り当てられている再構成の優先順位はACUで変更できます。

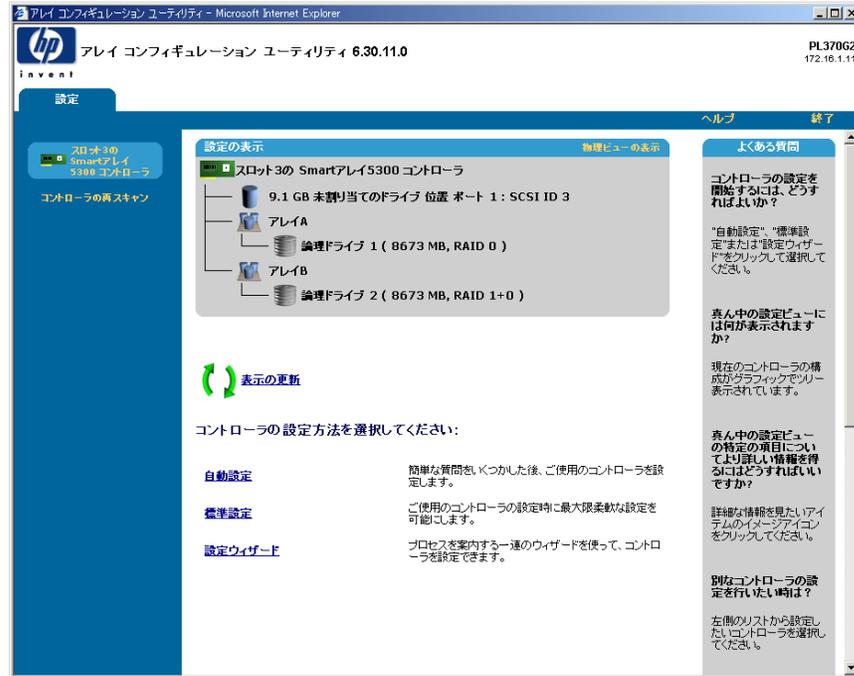
- 再構成を完全にバックグラウンドで実行し、オペレーティングシステムの要求よりも優先順位を低くする場合は、再構成の優先順位を「低」に設定します。
- オペレーティングシステムのどの要求よりも再構成の優先順位を高くする場合は、再構成の優先順位を「高」に設定します。



警告

再構成が完了する前にドライブを取り外したり、サーバの電源を切ったりすると、データが失われることがあります。

アレイコンフィギュレーションユーティリティ XE (ACU-XE)



ACU-XE では、ACUと同じ機能をブラウザを通じて利用でき、アレイをローカルおよびリモートから構成できます。

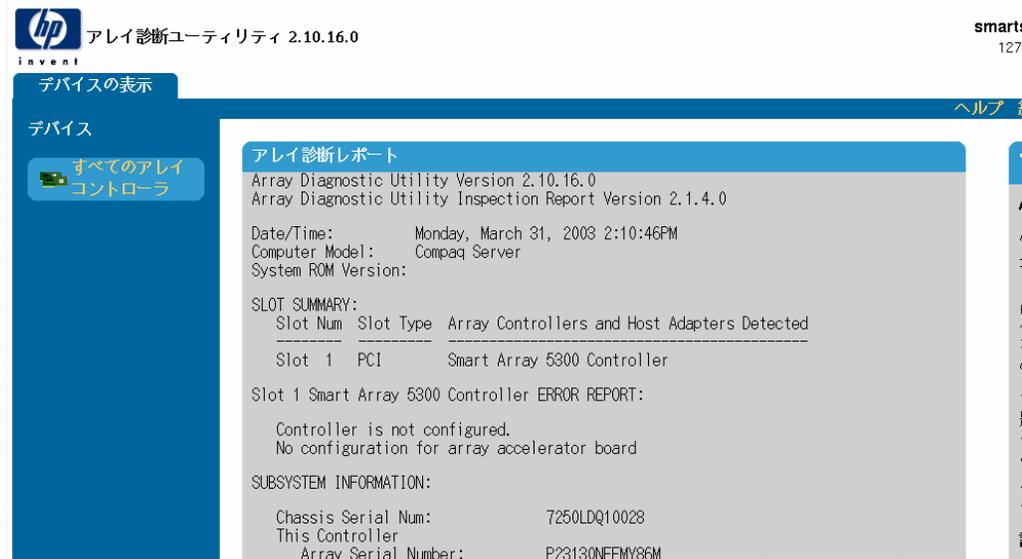
▲ 注記

ACU-XE は、Microsoft Internet Explorer 5.0 以降で動作します。

ACU-XE を表示すると、システムにインストールされているコントローラが自動的に検出されます。コントローラの種類に応じて、以下のオプションが表示されます。

- **自動設定** — ACU-XE ユーティリティにより、コントローラの最適な構成が自動的に設定されます。
- **標準設定** — ユーザは、手動でコントローラを構成できます。
- **設定ウィザード** — ユーザは、コンフィギュレーションウィザードを使用して、手動でコントローラを構成できます。

Array Diagnostics Utility (ADU)



ADU は、アレイコントローラを診断して、検出された問題のリストを作成するのに使用します。このデータをファイルに保存して分析に使用することができます。

ADUを開始すると、検出されたエラーを示すウィンドウが表示されます。エラーが検出されなかった場合は、ADUのメインダイアログボックスに以下の情報が表示されます。

- アレイコントローラ
- SCSI バス
- 物理ドライブ

システムパーティションアップグレードユーティリティ

SmartStart CD からしてシステムパーティションアップグレードユーティリティを実行することで、システムパーティションを作成または更新できます。ブートプロセス中に[F10]キーを押すことで、システムパーティションにアクセスできます。システムパーティションにある利用可能なユーティリティは、以下のとおりです。

- SCU
- オペレーティングシステムのインストール
- 診断とユーティリティ

▲ 注記

古いバージョンの ProLiant サーバでは、システムパーティションのサイズが小さいために、SmartStart を処理するのに十分な容量がないことを示すエラーメッセージが表示されることがあります。この場合は、ディスクまたは CD-ROM からユーティリティを実行します。または、データをバックアップして、ドライブを消去し、パーティションのサイズを大きくしてデータを復元します。

ディスクビルダ

ディスクビルダユーティリティ(SmartStart から実行およびブートしている場合は、Create Support Software と表示されます)を使用して、さまざまな HP ドライバとユーティリティを含むディスクを作成できます。ディスクビルダユーティリティを使用して製品のディスクを作成することで、CD から直接インストールすることができないサポートソフトウェア製品にアクセスできるようになります。

通常、Windows 環境では、ディスクビルダユーティリティは自動的に起動されます。ディスクビルダが自動的に起動されない場合は、以下の手順に従ってください。

1. CD-ROM ドライブに CD を挿入します。
2. Windows のプログラムマネージャから[スタート]、[ファイル名を指定して実行]を順に選択します。
3. 「<ドライブ>:\DSKBLDR\DSKBLDR.EXE」と入力します。
ここで、<ドライブ>は、CD-ROM ドライブのドライブ文字です。
4. プログラムが起動したら、画面の指示に従います。

System Erase ユーティリティ

System Erase ユーティリティは、SmartStart を使用してサーバを初期化する前に、ネットワークオペレーティングシステムを含む以前のすべてのハードウェア/ソフトウェア構成を消去するために使用します。System Erase ユーティリティは、SmartStart から実行できます。



警告

System Erase ユーティリティは、システムに存在する NVRAM/CMOS、ハードディスクドライブ、Smart アレイコントローラからすべての構成情報、ソフトウェアおよびデータを完全に消去し、新たに SmartStart セットアップを実行できるようにシステムを準備します。ファイバチャネルで接続されている他の構成要素は消去しません。

SmartStart を使用してシステムを再初期化する場合は、必ず System Erase ユーティリティを使用してください。インストールプロセス中にエラーが発生した場合は、System Erase ユーティリティを実行して、SmartStart インストールプロセスをやり直してください。



警告

構成済みのサーバを SmartStart で起動し、System Erase ユーティリティを実行するように求められた場合、既存のサーバ構成とデータをすべて消去する場合を除いて、System Erase ユーティリティは**実行しないでください**。System Erase ユーティリティは、すべての構成情報とデータを消去し、**すべてのハードディスクドライブを完全に消去**します。

Fibre Channel Diagnostics と Fibre Channel Fault Isolation Utility

Fibre Channel Diagnostics により、ファイバチャネルアレイコントローラの構成情報を表示して、ファイバチャネルアレイコントローラのエラーを診断できます。

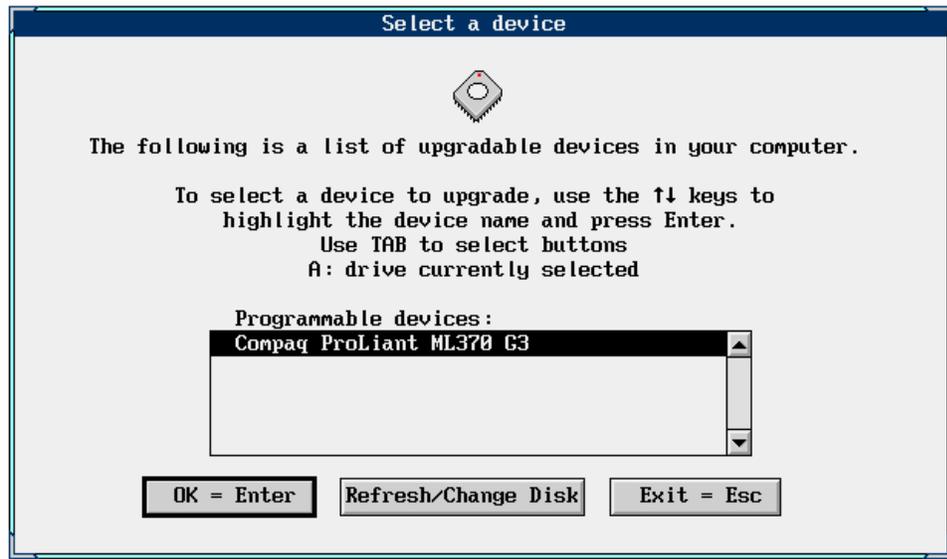
Fibre Channel Fault Isolation Utility は、ファイバチャネルアービトラリー接続されているノードの整合性を確認します。



注記

Fibre Channel Diagnostics と Fibre Channel Fault Isolation Utility は、サーバでファイバチャネルを使用している構成のみ使用できます。

ROMPaq Utility



ROMPaq のメインメニュー

HP のシステムとオプションには、アップグレード可能なファームウェアが採用されています。SmartStart CD、ROMPaq ディスケット、またはオンライン ROM フラッシュを使用して ROMPaq Utility を実行し、プログラマブルハードウェアのファームウェアを更新できます。

プログラマブルハードウェアとしては、サーバ BIOS、Smart アレイカード、およびハードディスクドライブなどがあります。

ProLiant サーバや特定のオプションコントローラではフラッシュ ROM が採用されているため、システムユーティリティやオプションの ROMPaq Utility を使用して BIOS を更新できます。



警告

ファームウェアのアップグレード中はサーバの電源を切らないでください。アップグレード中に電源が供給されなくなると、ファームウェアが破壊されることがあります。これにより、システムボードが故障し、交換しなければならなくなることがあります。



注記

システム ROMPaq は、現在のイメージのバックアップを作成しますが、オプション ROMPaq は作成しません。

ProLiant Support Pack(PSP)

PSP は、HP サーバサポートソフトウェア用のパッケージ/インストールプログラムです。PSP は、ProLiant 用に最適化されたドライバ、ユーティリティ、およびマネジメントエージェントが組み合わされた、オペレーティングシステム固有のバンドル製品です。

PSP には、Smart コンポーネントと呼ばれる、独立した自己インストール可能な複数の構成要素(ドライバ、エージェント、ユーティリティ)で構成されています。

PSP は、以下のいくつかの場所から入手できます。

- SmartStart and Support Software CD
- ActiveUpdate
- HP の Web サイト

リモート展開ユーティリティ(RDU)

RDU は、ドライバとマネジメントエージェントのアップデートをネットワーク接続サーバにリモートから配備するためのアプリケーションです。RDU は、新しい PSP をリモートからインストールしたり、以前のバージョンの PSP でインストールされたサーバサポートソフトウェアをアップグレードするのに使用します。

RDU は、サーバのソフトウェアインストール履歴にアクセスして、PSP のインストールが正常に完了したかどうかを確認することが可能なログファイル機能を備えています。

リモート展開コンソールユーティリティ

リモート展開コンソールユーティリティは、RDU と同じ機能を備えているコマンド行ユーティリティです。ユーザ介入なしでコマンドプロンプトから PSP をインストール、アップグレードできます。また、スクリプトを記述して、PSP や個別の Smart コンポーネントのローカル/リモートインストールを実行できます。

サーバの展開

ProLiant サーバではいくつかの配備オプションを提供しています。使用するツールは展開の種類によって決まります。

- SmartStart CD
 - 「Server Setup and Management」パックで提供
 - 単一サーバのセットアップ
 - すべての ProLiant ML および DL サーバに対応
 - 初心者から熟練者までのあらゆるユーザ向き
 - 有人インストール
- SmartStart Scripting Toolkit
 - Web からのダウンロードでのみ提供
 - 熟練ユーザによるセットアップと初心者ユーザによる展開用
 - ProLiant ML/DL サーバに対応
 - 無人インストール
 - RILOE 仮想フロッピーを使用してリモート操作も可能
- Rapid Deployment パック
 - BL サーバ エンクロージャおよび Rapid Deployment パックで提供
 - 中程度のスキルのユーザによるセットアップと、初心者ユーザによる展開用
 - ProLiant ML、DL、BL サーバに対応
 - 展開イベントのドラッグアンドドロップ操作

SmartStart Scripting Toolkit

SmartStart Scripting Toolkit には、カスタマイズ、予測可能な無人方式で構成および展開するための一連の DOS ベースのユーティリティが含まれています。これらのユーティリティは、必要最小限のユーザ介入で、ソースサーバの構成をターゲットサーバに複製します。



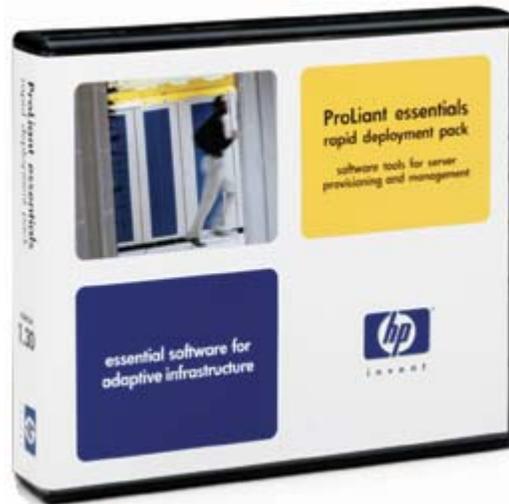
警告

Scripting Toolkit ユーティリティの使用方法を誤ると重要なデータが失われることがあります。データが失われる危険性があるため、Toolkit ユーティリティは熟練ユーザだけが使用してください。Scripting Toolkit を使用する前には、障害が発生した場合でもミッションクリティカルなシステムが稼動し続けるよう、必要なあらゆる予防措置を講じる必要があります。

Scripting Toolkit ユーティリティを使用してターゲットシステムを構成するには、構成済みのソースサーバから一連の構成スクリプトを作成します。その後、未構成のターゲットサーバ用にスクリプトファイルを編集し、そのファイルと Scripting Toolkit ユーティリティをネットワーク上の共有フォルダまたはサーバ構成用ブートディスクケットにコピーします。

サーバ構成用ブートディスクケットには、新たにインストールしたターゲットサーバ上で実行するサーババッチファイルを入れることもできます。これにより、サーバのハードウェアの構成とセットアップ、およびサーバのオペレーティングシステムのインストールを無人で行うことができます。

Rapid Deployment パック(RDP)



Rapid Deployment パック(RDP)は、HPとAltirisによって共同開発された以下のツールで構成されています。

- Altiris eXpress Deployment Solution for Servers
 - スクリプティングまたはイメージングによって展開
 - ◆ 組み込みのスクリプト世代管理機能と編集機能を実装
 - ◆ 既存のスクリプトファイルを実行可能
 - PXE サービスとイメージツールを実装
 - ◆ ヘッドレス運用(現地での操作が不要)を実現するネットワーク経由でのブート
- ProLiant インテグレーションモジュール for Altiris eXpress
 - ProLiant サーバ用に最適化されている
- RDP の試用版は、HP の Web から入手可能

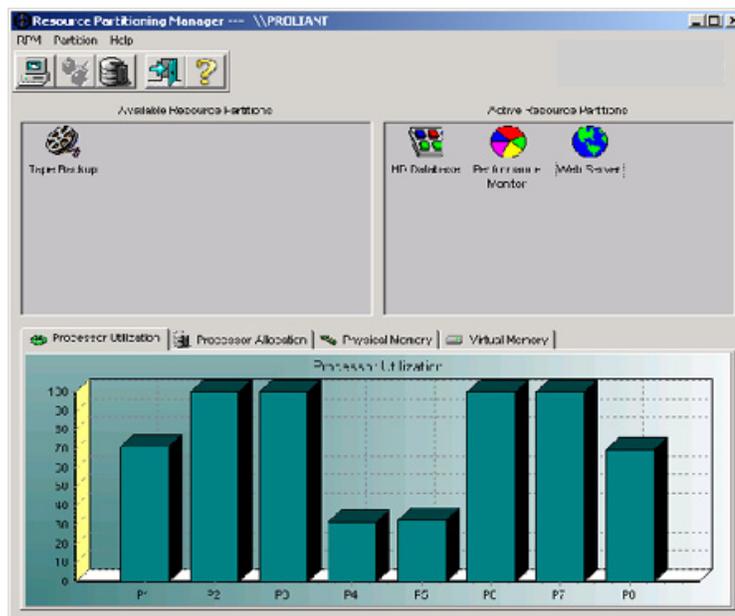
RDP は、有償ソフトウェアです。ハードウェアのコストを低く抑え、サーバの配備および管理用の費用対効果の高いツールを提供できるようにこの方式を採用しています。

10 日間の無償試用期間が設定された RDP が HP の Web から入手可能です。この試用版では、製品版のすべての機能を使用できますが、最初に使用した日から 10 日間で試用期限が切れます。

Workload Management パック

ProLiant Essentials Workload Management パックは、Resource Partitioning Manager を備えています。

Resource Partitioning Manager



Resource Partitioning Manager は、グラフィカルユーザインタフェース方式のツールで、Microsoft Windows 2000 Server オペレーティングシステムの機能を拡張して、個別のアプリケーション、サービス、および他のプロセスが使用可能なシステムリソースのサイズと物理的な場所を制御できるようにします。

Resource Partitioning Manager ユーザは、複数のリソースパーティションを作成したり、各パーティションのリソース数を一定の数に制限したり、プロセッサとメモリの動的再割り当てを可能にするルールを作成できます。

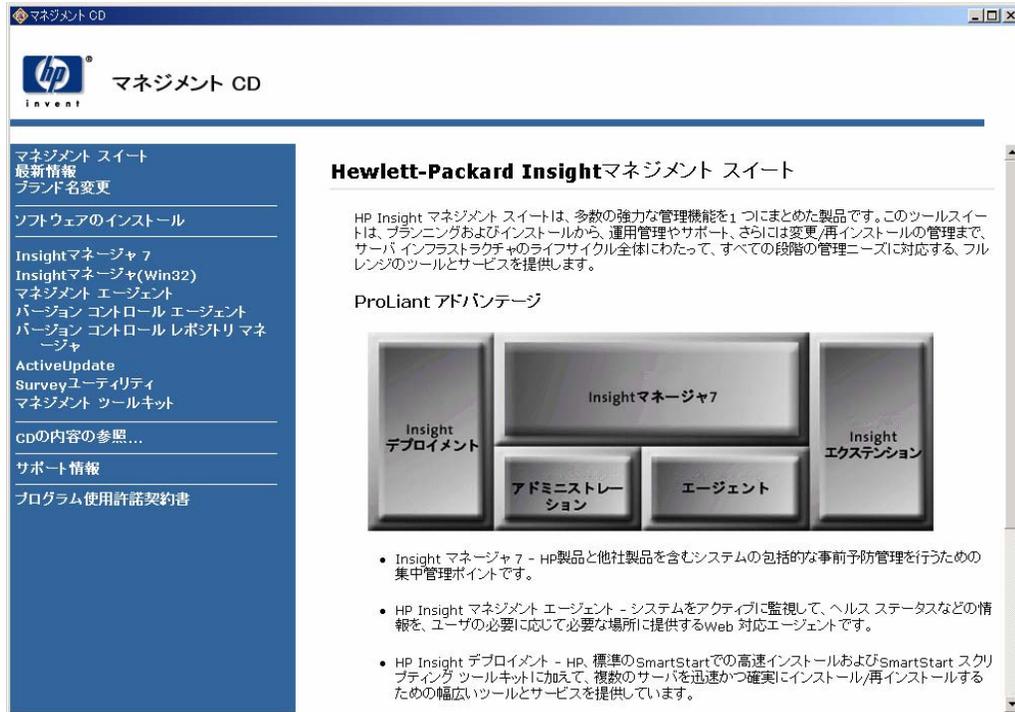
リソースパーティション

リソースパーティションとは、Windows 2000 ジョブオブジェクトに割り当てられているプロセッサリソースとメモリリソースの集まりです。Resource Partitioning Manager は、リソースパーティションを作成および管理します。これらの境界は、オブジェクトのリソースの必要性に応じて動的に再割り当てすることができます。

ジョブオブジェクト

ジョブオブジェクトとは Windows 2000 の機能の1つで、プロセスをグループ化することができます。これらのジョブオブジェクトは、ユーザ定義のアプリケーション、サービス、および他のプロセスを保持するコンテナと考えることができます。

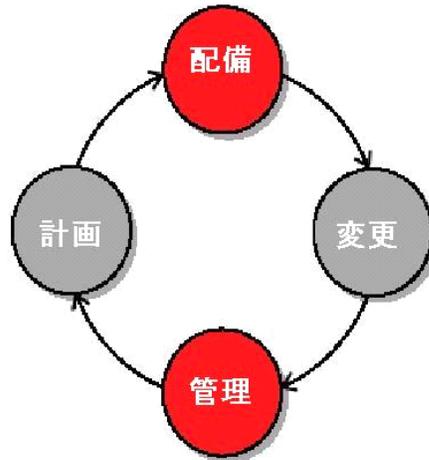
Management CD



Management CD は、サーバの統合、モニタリング、およびパフォーマンス最適化のためのさまざまなサービスおよびソフトウェアツールを提供します。Management CD 6.x には、以下のコンポーネントが含まれています。

- Insight マネージャ 7
- マネジメントエージェント
- ActiveUpdate
- Survey ユーティリティ
- マネジメントツールキット

Insight Management Suite

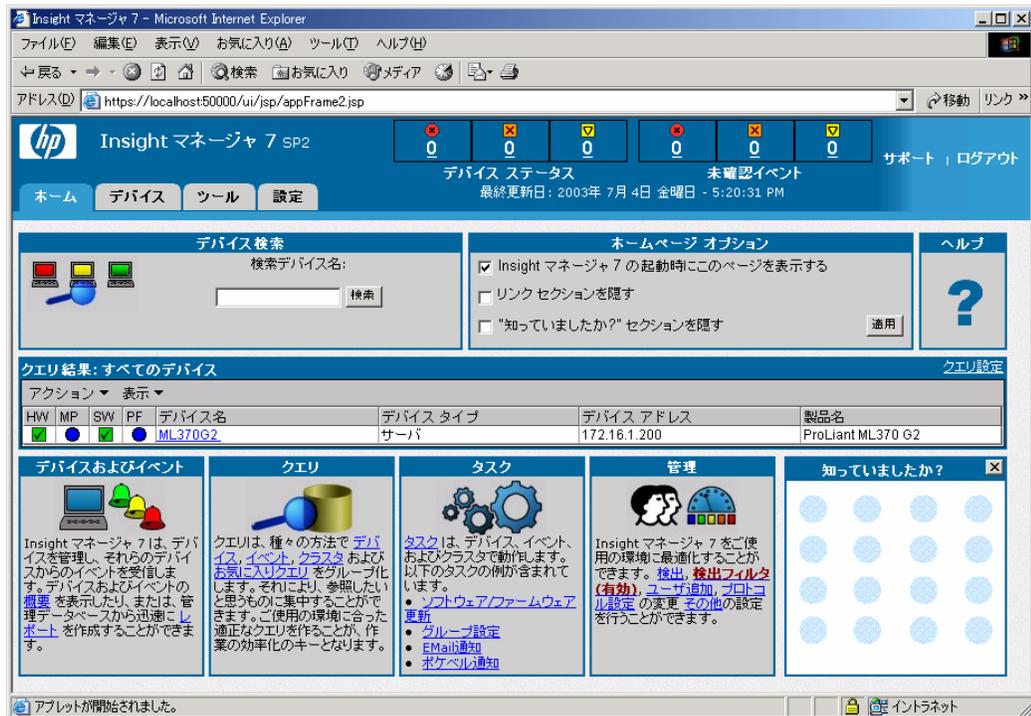


ProLiant サーバの ライフサイクル管理

Insight Management Suite は、以下の製品で構成されています。

- **Insight 配備ツール** — 1 台から数千台までのサーバのセットアップを単純化します。一貫性およびサポート性の高いインストールが可能です。
- **Insight マネージャ 7** — スイートの核となる製品で、アダプティブインフラストラクチャのすべての属性をサポートしています。オプションで用意されている新しいバージョン管理機能により、サーバグループ全体を自動更新できます。
- **Insight エージェント** — サーバ全体の状態をトラッキングし、障害が発生する前にアラートを送信します。
- **Insight 管理ツール** — リモート Insight Lights-Out オプションボードや内蔵 Lights-Out マイクロプロセッサなどのツールにより、実際にサイトに出向いた場合と同様の制御を行えます。これにより、移動時間やリソースなどのコストを大幅に削減できます。
- **Insight 拡張機能** — 基本的な管理機能をベースに、異機種環境をサポートするためのマネジメントプラグインや、Microsoft Cluster Service 用のクラスタ障害予測管理を提供するクラスタリングツールなど、ユーザ特有のインフラストラクチャを機能強化します。

Insight マネージャ 7



Insight マネージャ 7 は、Web ベースのエンタープライズ管理アプリケーションで、分散型システムを管理するための、自動化され費用対効果の高い、障害予測機能を備えたソリューションです。メインページからクリックするだけで、システムの共通ビューを表示できます。以下の HP 製品を管理します。

- サーバ
- デスクトップ
- クラスタ
- 以下に準拠するサードパーティ製デバイス
 - HTTP
 - SNMP (Simple Network Management Protocol)
MIB (Management Information Base) -2
 - DMI (Desktop Management Interface)

Insight マネージャ 7 は、インストール直後からシステムの探索とレポートの作成を実行できる、標準的なタスクとクエリーを備えています。

機能

Insight マネージャ 7 は、IT 環境全体の管理対象デバイスから、デバイスのステータスや資産構成などの重要な管理情報を集計します。また、これを使用すれば、1つの場所から、HP のサーバ、デスクトップ、ワークステーション、ポータブルコンピュータにインストールされている以下の HP 管理アプリケーションにアクセスできるようになります。

- マネジメントエージェント
- リモート Insight Lights-Out Edition (RILOE) ボード
- Version Control エージェント

IT 管理者は、Insight マネージャ 7 のクエリーおよびタスクシステムを使用して、デバイスグループの監視や操作ができます。パフォーマンス、資産、および構成に関する大量のデータは、Microsoft SQL Server または Microsoft Data Engine (MSDE) データベースに格納され、標準の ODBC (Object Database Connector) 準拠のレポートツールを使用してアクセスできます。

Insight マネージャ 7 は、以下の機能を備えています。

- **自動検出と識別** — Insight マネージャ 7 は、HP のサーバ、デスクトップ、ワークステーション、ポータブルコンピュータに加え、ネットワークに接続されているその他の SNMP および DMI 対応のデバイスを自動的に検出および識別するように構成できます。デフォルトでは、自動検出は 1 日 1 回実行されるように設定されます。
- **デスクトップインストールのサポート** — Insight マネージャ 7 は、Windows 2000 Professional が稼動している DeskPro および Evo デスクトップ、または環境のワークステーションにインストールできます。また、Microsoft Windows 2000 Server または Windows 2000 Advanced Server 環境の ProLiant サーバにもインストールできます。
- **言語サポート** — Insight マネージャ 7 は、英語版、フランス語版、ドイツ語版、スペイン語版、および日本語版の Microsoft Windows 2000 Professional、Windows NT Server 4.0、Windows 2000 Server、Windows 2000 Advanced Server 環境にインストールできます。英語、フランス語、ドイツ語、スペイン語、および日本語のデータベースもサポートしています。
- **セキュリティ** — Insight マネージャ 7 は、マネジメント情報への安全なアクセスを可能にします。ブラウザと管理サーバ間のすべての接続は、業界標準の Secure Sockets Layer (SSL) テクノロジーで暗号化されます。

- **シングルログイン** — Insight マネージャ 7 は、管理対象デバイスで実行されている HP 管理アプリケーションへユーザを自動的に認証します。これにより、インテグレートド マネジメントログの消去など、管理者レベルの権限を必要とするタスクを実行する際に、HP マネジメントエージェントを再認証する必要がなくなります。
- **Insight マネージャ 7 ホームページ** — Insight マネージャ 7 ホームページでは、以下の 3 つの主な機能を提供しています。
 - Insight マネージャ 7 の最も使用頻度の高い機能と、それらの機能の詳細な説明へのリンクを提供しています。
 - Insight マネージャ 7 を初めて使用したときに表示されるデバイスやイベントのセットを選択できます。このデバイスまたはイベントクエリーは、ユーザごとにカスタマイズできます。
 - 大量のデバイスリストから探し出すことなく、デバイスの検索を実行して特定のデバイスを見つけることができます。
- **クエリーとタスク** — Insight マネージャ 7 のクエリーとタスクにより、ネットワークに接続されている HP デバイスやサードパーティ製デバイスをグループ管理することができます。クエリーとは、ユーザが定義した条件に基づいたデバイスグループまたはイベントグループです (たとえば、すべてのサーバまたはすべての重要なイベントなど)。

タスクとは、ソフトウェアの配備や SNMP ステータスのポーリングなど、管理対象デバイスのグループに対して実行する操作です。すべてのタスクがクエリーに基づいているため、自己更新されます。
- **複数のシステムのバージョン管理とシステムソフトウェアの更新** — Insight マネージャ 7、Version Control エージェント、Version Control Repository (VCR) により、バージョン管理と HP システムソフトウェアの更新のための新しいアーキテクチャが実現されます。ソフトウェアメンテナンスアーキテクチャにより、管理対象環境の一部または全体にわたるバージョン管理の基準となるソフトウェアのセット (BIOS、ドライバ、エージェント) を選択できます。

注記

▲ PSP は、ProLiant サーバで最適なパフォーマンスと信頼性を発揮するように設計およびテストされたドライバ、ユーティリティ、およびマネジメントエージェントのセットです。

- **グループ構成** — 管理者は、複数のデバイスグループを同時に管理できます。指定した構成情報を基準システムから取り出して編集し、複数のターゲットシステムにコピーすることができます。
- **電子メールおよびポケットベルによる通知** — Insight マネージャ 7 は、指定したイベントの受信やデバイスのステータスの変更に基づいて、電子メールおよびポケットベルの両方で通知を送信できます。これにより、管理コンソールを常時モニタリングする必要がなくなり、IT 担当者の生産性を大幅に向上できます。
- **Cluster Monitor** — Cluster Monitor は、Microsoft Cluster Service (MSCS)、OpenVMS、Tru64 UNIX、SCO UnixWare 7、NonStop、NetWare 6.0、Oracle など各種クラスタの拡張管理機能を提供します。
- **マネジメントエージェントの統合** — Insight マネージャ 7 は、定義済みの HTTP ポートで実行されているマネジメントエージェントや他の Web ベースの管理アプリケーションを自動的に検出します。マネジメントエージェントによって報告されたハードウェアのステータスがデバイスリストに表示されます。
- **リモート Insight の統合** — Insight マネージャ 7 は、SNMP と HTTP を使用して、RILOE ボードと iLO (内蔵 Lights-Out) を検出および識別します。

マネジメントエージェント

Management Agents for Servers は、デバイス上で動作し、パラメータを収集および測定して、デバイスのステータスを詳細にモニタリングします。これらのパラメータは、特定のイベントの発生数を数えたり、重要な機能の状態をモニタリングしたりすることでサブシステムの現在の状態を示します。

Web 対応のマネジメントエージェントには、ブラウザから直接アクセスすることも、Insight マネージャ 7 やサードパーティ製のシステム管理アプリケーションなど管理アプリケーションからアクセスすることもできます。

ActiveUpdate



hp ActiveUpdate

User name:

Password:

Log On

Remember my username and password [\(What's this?\)](#)

New User? [Click here](#) POWERED BY BACKWEB®

ActiveUpdate は Web ベースのクライアントアプリケーションで、IT 管理者が HP に直接接続された状態を保ち、最新のソフトウェアアップデートの配信とそれに関する通知を得られるようにします。

ActiveUpdate を通じて配信される内容は、以下のとおりです。

- 管理アプリケーションのアップデート
- PSP
- デバイスドライバ
- システム ROM
- BIOS イメージ
- ソフトウェアユーティリティ

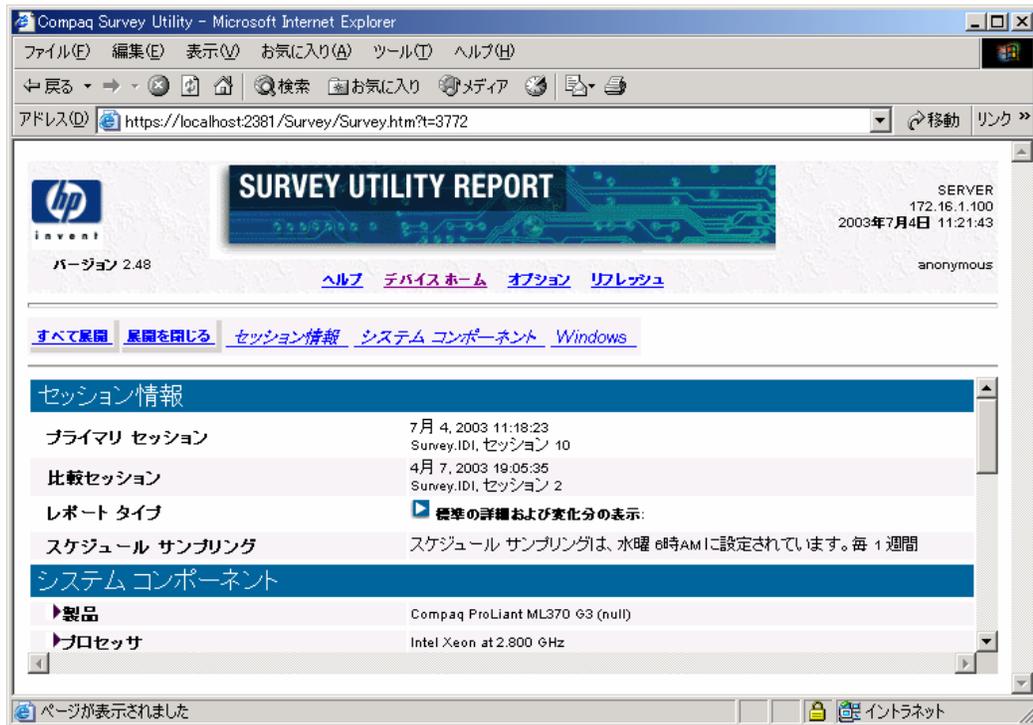
ActiveUpdate はカスタマイズ可能なプロファイルとソフトウェアフィルタを使用して、ユーザが各自の環境に関するソフトウェアアップデートだけを受け取るようにします。サブスクリプションページのトピックツリー構造により、ユーザは使用しているハードウェアプラットフォームだけでなく、実行しているオペレーティングシステムもドリルダウンして選択できます。

ActiveUpdate の利点は以下のとおりです。

- 新しいアップデートを自動的にダウンロードおよび保存できるため、時間を節約可能
- カスタマイズされた情報を配信する
- ソフトウェアアップデートに関するわかりやすい説明を提供する
- 単一アクセスポイントの提供により、HP のサーバ、デスクトップ、ポータブルコンピュータ、およびワークステーション用の最新のソフトウェアアップデートへのアクセスが単純化される

Web サイトでは、ダウンロードする SoftPak ファイルのモデル、オペレーティングシステム、および言語を選択できます。ダウンロードを受け取るには、サブスクリプションを送信する必要があります。

Survey ユーティリティ



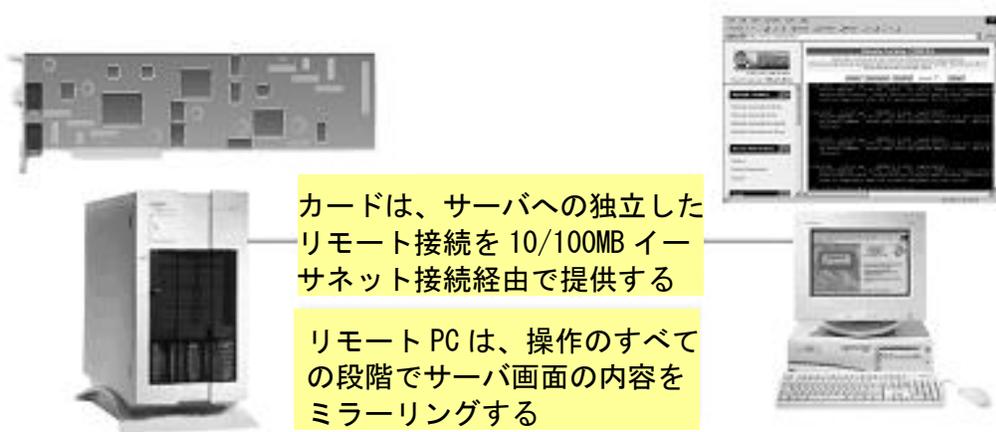
Survey ユーティリティは、ProLiant および Prosignia サーバで動作する情報収集エージェントです。構成分析とトラブルシューティングプロセスを単純化することで、サーバの可用性を最大化できるように設計されています。

Survey ユーティリティは、以下の環境のハードウェアおよびオペレーティングシステムに関する重要な情報を収集します。

- Windows 2000
- Windows NT
- NetWare
- Linux

このユーティリティは、Web ベースとコマンド行方式の両方に対応しており、重要なサーバの構成を迅速かつ容易に認識することができます。

リモート Insight Lights-Out テクノロジ



HP は、リモートシステム管理の必要性を認識しており、ProLiant サーバをリモート PC から管理できる RILOE ボードを提供しています。RILOE ボードは、コンピュータ内のコンピュータとして機能し、サーバがダウンしているときでもサーバにアクセスできるようになります。RILOE ボードには、以下の構成要素があります。

- プロセッサ
- メモリ
- ファームウェア
- オペレーティングシステムとブラウザ
- ビデオカード
- ネットワーク

RILOE オプションは、仮想のオンサイト管理機能を提供します。リモートサイトの担当者呼び出して、キーボード操作を実行してもらい、サーバからの応答に関する説明を聞き取る代わりに、直接サーバの動作を確認しサーバを管理することができます。

RILOE ボードは、管理 PC への SNMP トラップの配信を通じてアラート通知もサポートしています。

RILOE の機能

RILOE テクノロジーの機能は以下のとおりです。

- **ハードウェアベースのグラフィカルリモートコンソール** — 管理 PC の標準ブラウザを、ホストサーバの仮想デスクトップとして使用できます。
- **サーバ障害アラート** — サーバの電源が失われたことを検出します。または、オペレーティングシステムが応答しなくなった後、Automatic Server Recovery (ASR) 回路によってサーバがリセットされたことを検出します。どちらの場合も、アラートを管理アカウントに送信できます。
- **リセットと障害シーケンスの再生** — サーバの起動および停止シーケンスを再生できます。これには、システムおよびオペレーティングシステムのエラーメッセージと、NetWare の Abend 画面や Windows 2000 のブルースクリーンなどの致命的なエラー画面も含まれます。
 - **現在のリセット** — サーバの前回のリセットを再生します。
 - **以前のリセット** — 以前のリセット(前々回のリセット)を再生します。
 - **障害** — 前回のサーバリセットで表示された情報を再生します。これには、サーバがダウンしてその後リセットされる前にオペレーティングシステムによって表示されたエラー情報も含まれます。
- **リモートリセット** — ホストサーバが応答していない場合に、リモートコンソールからコールドリセットを実行してオンラインに戻すことができます。このタイプのリセットは、サーバのオペレーティングシステムの猶予付きシャットダウンはしませんが、オペレーティングシステムが応答不能な場合に役立ちます。
- **Insight マネージャ 7 との統合** — ハードウェアベースの非同期管理を提供します。これには、Insight マネジメントエージェントへのアクセスが含まれます。
- **iPAQ との統合** — iPAQ H3600 シリーズハンドヘルドデバイスからの RILOE へ接続できます。
- **RILOE ボード** — ホストのオペレーティングシステムの状態にかかわらず、サーバのキーボード、ビデオ、およびマウス (KVM) 機能を提供します。ボードは単一ボードコンピュータで、サーバのビデオ入力の完全なグラフィカル表示を提供します。これにより、リモート管理 PC は、Web ブラウザを通じて仮想デスクトップとして機能することができます。
- **内蔵 Lights-Out (iLO)** — システムボードに組み込まれたインテリジェントなプロセッサとファームウェアで構成され、Advanced System Management ASIC (アプリケーション固有の集積回路) と RILOE ボードの機能が組み合わされています。iLO は、標準および高レベルの Lights-Out 機能を提供します。

内蔵 Lights-Out (iLO)



iLO は、主要なサーバのシステムボードに組み込まれている ASIC です。iLO は、インテリジェントなプロセッサとファームウェアで構成され、標準および高レベルの Lights-Out 機能を提供します。

iLO Standard は、基本的なシステムボード管理機能、診断、および重要な Lights-Out 機能を提供します。iLO Advanced は、高度な機能を提供し、オプションの iLO Advanced パックでライセンスを取得することができます。

iLO Standard

iLO Standard の機能は以下のとおりです。

- **仮想テキストリモートコンソール** — テキストモード画面用に、組み込みのハードウェアリモートコンソール機能を提供します。オペレーティングシステムに依存しないコンソールは、起動およびシャットダウン操作などのリモートホストサーバのアクティビティを表示するテキストモードをサポートしています。
- **仮想電源ボタン** — サポートされているブラウザインタフェースを使用して iLO にアクセスし、リモートサーバの電源ボタンを操作できます。これにより、リモートサーバの電源の入切、またはサーバの再起動を行うことができます。オペレーティングシステムで障害が発生した場合など、ボタンを一瞬押しただけではサーバの電源を切ることができない場合に備えて、ボタンを押し続ける操作も可能です。
- **ユニット ID LED の制御** — ユニット ID LED は、サーバの正面と背面にあり、iLO によって監視および制御されます。iLO の機能は以下のとおりです。
 - ユニット ID LED のステータスを表示します。
 - ユニット ID LED を有効/無効にします。
 - ユニット ID LED を点滅させて、中断してはいけない処理 (iLO ファームウェアのフラッシュやアクティブ Remote Console セッションなど) を iLO が実行していることを示します。

- **簡単なセットアップ/使用**
 - Domain Name Service (DNS)/Dynamic Host Configuration Protocol (DHCP) による自動構成
 - RBSU
 - Microsoft Internet Explorer 5.0 以降による高いアクセス性
- **常時オン/常時稼働** — iLO プロセッサの電源は、サーバの補助電源から供給されるため、サーバの状態にかかわらずプロセッサには電源が常に供給されます。
 - サーバの電源を切ってもプロセッサにアクセスできます。
 - オペレーティングシステムのエラーによってサーバがダウンした場合、サーバのモニタリングを続行できます。

注記

この機能を使用するためには、サーバが AC 電源に接続されている必要があります。

- **システムボード管理** — iLO により、次のことを行えます。
 - POST エラーメッセージの発生を監視できます。
 - インテグレートド マネジメント ログと iLO イベントログによって記録されたイベントを使用してシステムを診断できます。
- **ヘッドレスサーバ運用** — iLO の機能により、ローカルキーボード、ポインティングデバイス、モニタ、または KVM スイッチを使用せずに、サーバをシームレスに制御できます。
- **Insight マネージャとの統合** — Insight マネージャ 7 との統合により、効率的な集中管理を行えます。

iLO Advanced

オプションの iLO Advanced パックのライセンスを取得することで、高度な機能を使用できるようになります。iLO Advanced パックの機能は以下のとおりです。

- **仮想グラフィカルリモートコンソール** — iLO は、サポートしているブラウザを仮想デスクトップとして使用できるようになる、組み込みのハードウェアグラフィカルリモートコンソール機能を提供しています。これにより、ホストサーバのディスプレイ、キーボード、およびマウスを完全に制御できるようになります。オペレーティングシステムに依存しないコンソールは、起動およびシャットダウン操作などのリモートホストサーバのアクティビティを表示するグラフィックモードをサポートしています。また、標準のブラウザで機能するため、リモートサーバやクライアントシステムで追加のソフトウェアは不要です。
- **仮想フロッピードライブ** — ブートするリモートホストサーバを簡単に指定して、ネットワーク上の任意の場所にある標準メディアを使用することができます。これにより、リモートサーバが設置されている場所に出向いてディスクを挿入および使用する手間を省くことができるため、時間を節約して作業効率を高めることができます。
ROMPack のアップグレードをリモートサーバに適用して、リモートサーバ上のオペレーティングシステムをネットワークドライブから配備したり、SmartStart Scripting Toolkit を使用して展開したりできます。または、障害が発生したオペレーティングシステムの災害復旧を実行できます。

Universal Serial Bus (USB) ディスケットドライブデバイスをサポートしているオペレーティングシステムでは、内蔵 Lights-Out Virtual Floppy を使用してサーバをブートすることができます。これは、オペレーティングシステムが稼働中またはロード中にメディアデバイスとして使用することもできます。

iLO Advanced 機能の有効化

iLO Advanced 機能のライセンスは、内蔵 Lights-Out Advanced パック (製品番号 263825-B21) を注文することで取得できます。

iLO Advanced パックではライセンスキーを提供していて、これを iLO に入力することで高度な機能を使用できるようになります。

復習問題

1. ProLiant Essentials の構成要素の名前を挙げてください。
.....
.....
.....
.....
2. マネジメントエージェントの機能を挙げてください。
.....
.....
3. リモート Insight Lights-Out テクノロジーの 4 つの機能を挙げてください。
.....
.....
.....
.....
4. Insight マネージャ 7 の 4 つの機能を挙げてください。
.....
.....
.....
.....
5. バージョン管理の基準となるソフトウェアのセットは、Insight マネージャ 7 でどのように定められていますか?
.....
.....

6. Management 6.x CD に含まれている構成要素を挙げてください。

.....
.....
.....
.....
.....

7. RILOE が管理 PC にアラート通知を配信する方法を挙げてください。

.....

8. RBSU (ROM ベースセットアップユーティリティ) はディスクを使用します。

- 正
- 誤

目的

この章では、次の内容について学習します。

- HP 製品の機能と特徴
 - ラック
 - コンソールマネジメントコントローラ (CMC)
 - UPS (無停電電源装置)
 - サーバコンソールスイッチ

HP ラック/電源テクノロジー

HP の革新的なラック/電源テクノロジーは、エンタープライズ環境におけるパフォーマンスと価値の標準となっています。HP のエンタープライズクラスのラック/電源製品は、個々のワークステーションから分散エンタープライズ環境までのコンピュータシステムを保護、管理するさまざまなパワーマネジメント製品を提供しています。

HP は、最新のテクノロジーが採用された、以下のさまざまなラック/電源製品を提供しています。

- エンタープライズクラスのラック
- ラックマウントモニタとキーボード
- ラックマウントファンキット
- 無人リモートモニタリング
- CMC および環境センサ
- パワーディストリビューションユニット (PDU)
- サーバコンソール スイッチ
- PCI カード KVM スイッチ
- UPS (無停電電源装置)
- パワーマネジメントソフトウェア
- rack builder online

ラック 10000 シリーズ



HP ラック 10000 シリーズは、ProLiant、Alpha、サーバおよびワークステーション、ネットワーク、ストレージ製品などのすべての HP エンタープライズハードウェアを設置するための統合化された共通プラットフォームを提供します。

ラック 10000 シリーズは、奥行き 1000mm のラックで、幅 600mm の 22U、36U、42U および 47U と幅 800mm の 42U があります。

ラック 10000 シリーズでは、数多くのケーブル管理機能が採用されています。

- 取り外し可能なアクセスパネルのある分割式リアドア
- ラック天板のケーブル開口部
- ラック間のケーブル接続を可能にするマウンティングレールの切り欠きと、ケーブルを束ねるためのストラップ
- 最大積載重量の増加
- グラファイトメタリックカラー

▲ 注記

U は、ラックおよびラックマウント型構成要素の寸法単位で、1U は 1.75 インチです。

オプション

ラック 10000 シリーズ専用のオプションがいくつか用意されています。以下に示す他のラックアクセサリも使用できます。

- KVM スイッチ
- モニタ
- キーボード
- ラックマウント型キーボードとモニタ
- コンソール マネジメント コントローラ (CMC)
- PDU
- UPS (無停電電源装置)

ラック接地キット

オプションのラック接地キットは、HP ラック 9000 シリーズおよび 10000 シリーズの両方で使用できます。このキットにより、ラックキャビネット内で動作する電子部品によって放出される電磁ノイズを軽減できます。

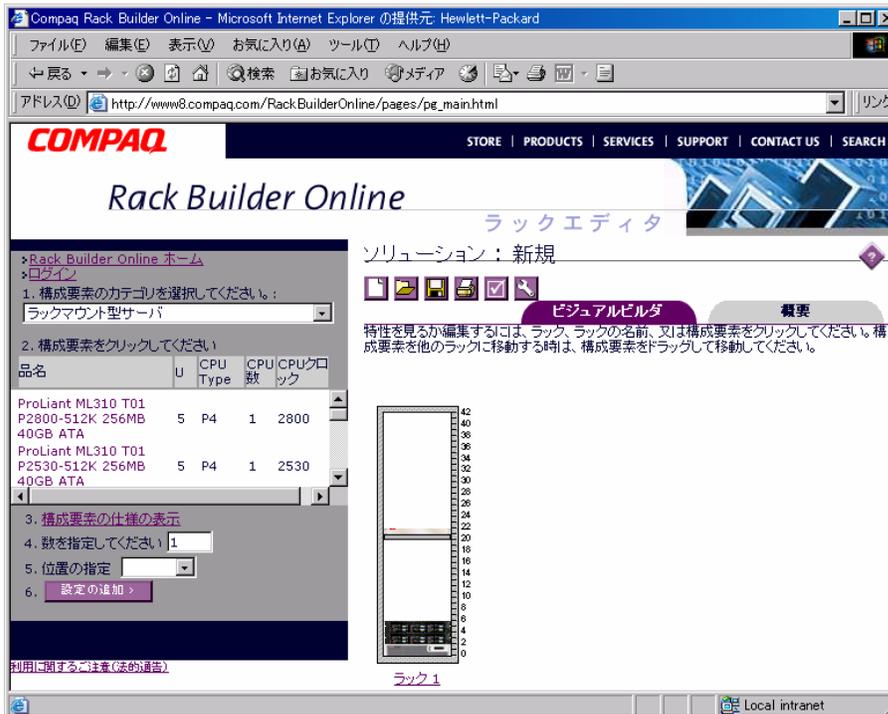
ラック接地キットは、ラックのすべての導電構造部品を導電ケーブルで接続します。キットは、ケーブルの接地ポイントに取り付け、ラック上の AC 電源分配装置上の接地ポイントに接続します。

ルーフマウント型ファンキット

ルーフマウント型ファンキットは、ラック内の通気性を高めることで自然対流による冷却性能を向上させます。このオプションは、10000 シリーズのラックの上部に取り付けます。低電圧モデルと高電圧モデルを取り揃えています。

- ラックファンキット 100V 257413-B21
- ラックファンキット 200V 257414-B21

rack builder online



rack builder online は、ラック環境の構築および管理作業を容易に行える、Web ベースの構成ツールです。これは、顧客がラックマウント型製品と HP ラックを計画/構成するのに役立ちます。rack builder online の機能は以下のとおりです。

- HP ラック 4000、7000、9000 および 10000 シリーズをサポート
- 販売完了した製品は「EOL (End of Life)」と表示
- 入力ワット数と漏電電流の計算
- 各構成要素の電源数の選択
- 漏電電流に加え、機能要件や規制要件を適用して、パワーマネジメント製品を正しく構成
- サーバコンソールスイッチとケーブルを構成
- すべてのラックオプションとアクセサリを構成
- アシストモードとビルダーモードを提供

追加機能

rack builder online では、以下の 2 つの方法で構成を作成することができます。

- **アシストモード** — ウィザード方式でラックマウント型構成要素の最適化された構成を決定することができます。
- **ビルダーモード** — アシスタンスなしで構成要素を追加および削除することができます。

ラックソリューションを印刷または表示することが可能な各種レポートが用意されています。

rack builder online では、以下の有用なレポートを作成できます。

- ラックレイアウトのグラフィック表示
- 以下のことに関する詳細レポート
 - 特定の部品番号
 - 必要なラック数
 - 必要なパワーマネジメント製品

これらのレポートは、新しいインストールをすばやく効率的に実装するのに役立ちます。

rack builder online の機能は、以下のとおりです。

- 構成を保存するための集中化された Web リポジトリ
- インターネット経由で瞬時にアクセス可能
- アプリケーションおよび製品データベースのオンラインヘルプ
- ラック情報の編集、印刷、およびローカルドライブまたは HP サーバへの保存
- ドラッグアンドドロップ機能に対応した構成のグラフィック表示
- 改良された電源容量決定アルゴリズム
- 製品の発表と同時に最新の製品データに直接アクセス可能
- 新しい HP 製品と従来の HP 製品の完全なデータベース
- 他社製ラック構成要素の作成、保存、および構成
- 2 つのユーザモード

HP ラックの保守に関する注意事項

HP ラックを取り扱う際には、以下の保守に関する注意事項に従う必要があります。

- ラックを単体でインストールする場合は、構成要素をインストールする前にラックが水平かつ安定した状態になっていることを確認してください。スタンドアロンラックの背面には固定脚を取り付ける必要はありません。けがのリスクを減らすには、すべてのスタンドアロンラックに固定脚を取り付けてください。
- ラックの四隅には水平調整脚があります。ラックに構成要素をインストールする前に、脚を回して床につけ、ロックナットで固定する必要があります。キャストが損傷しないように、ラックの全重量がキャストにではなく、水平脚に載るようにしてください。キャストは、ラックを所定の場所に移動するためだけのものです。
- 複数のラックをインストールする場合は、ラック接続キットでそれらを結合して安定性を高めてください。ラックを目的の場所に置き、ラック接続キットをインストールしてから、ラックに構成要素をインストールしてください。
- 複数の構成要素をインストールする場合は、まず最も重い構成要素をラックの一番下にインストールしてください。UPS は、必ず一番下にインストールする必要があります。
- 複数の構成要素をラックから引き出した状態にしないでください。複数の構成要素をラックから引き出した状態にすると、ラックが不安定になり、傾くことがあります。
- ラックマウント型 ProLiant サーバは、フロントドアから外気を吸入して、内部の熱気を後部から排出します。このため、ラックのフロントドアの前方には、外気をキャビネット内に吸入できる適度な隙間が必要で、リアドアの後方には、熱気を排出できる適度な隙間が必要です。
- 構成要素がインストールされていない部分には、ブランクプレートを取り付ける必要があります。ブランクプレートは、良好なエアフローを得るのに重要です。顧客のラックシステムを保守しているときに、ブランクプレートが付いていないことに気が付いた場合は、それらを取り付けるように依頼してください。

コンソールマネジメントコントローラ(CMC)

HP ラックマウント CMC は、1 台のラック内、または特定の用途における 2 台構成のラック内の環境を監視および制御します。これには、温度、湿度、煙、振動、セキュリティ、およびラックファン用のセンサとアラーム、データ交換用のネットワークポートとシリアル通信ポートがあります。

HP ラックマウント CMC は、監視対象のラックシステムの温度、湿度、煙、衝撃および振動、セキュリティを監視するためのセンサ入力を受け付けます。ユーザは、インテリジェント ラック マネージャを使用して許容値とそれを超えた場合の動作を設定します。ラックの状態がそれらの値を超えた場合、インテリジェント ラック マネージャは電子メールやポケベルでユーザに通知を送信し、定義済みの動作を実行します。

監視項目は以下のとおりです。

- CMC の入力電圧。ラックの主電源など、100～240 VAC の任意の供給電源を監視できます。ラックファンがインストールされている場合は、監視される電圧はラックファンの供給電圧になります。
- 温度。
- ラック環境内の湿度レベル。ユーザが設定した値を超えた場合に通知が送信されます。
- 煙の粒子が特定の大きさ(ミクロン単位)を通過した場合、アラートがユーザに送信されます。
- 衝撃と振動(顧客が調整可能)。
- 侵入。

応答を送信したりイベントアラームをトリガする状態パラメータは、ユーザが指定します。たとえば、温度が上昇してユーザが設定した値を超えた場合に、その次の状態に到達したときにファンの電源が入り、アラーム音が鳴り、サービスエンジニアのポケットベルにアラートが送信されるように設定できます。

CMC の機能は、以下のとおりです。

- ホストコンピュータとのデータ交換用のイーサネット通信ポート
- 1U のフォームファクタ
- CMC 管理用のインテリジェントラックマネージャ Lite ソフトウェア
- アナログ入力:4 つ
- 温度センサ:2 つ
- 湿度センサ:1 つ
- AC 電源センサ:1 つ。100～240 VAC 電源を接続および監視可能。ラックファンがインストールされている場合は、ラックファンの電圧を監視可能
- 衝撃/振動の検知
- セキュリティ
- 煙探知
- 電子メールおよびポケベルによりサービスエンジニアにアラートを送信
- ドアの開閉と侵入検知(4 つのセンサ)。ラックキャビネットの不正な操作を監視
- ユーザがプログラミング可能なアラーム出力:2 つ。ストロボライト、アラーム音、またはアラーム状態を示すその他のサポートされているデバイスを接続可能

CMC には、ホストコンピュータとのデータ交換用の 2 つの通信ポートがあります。

- ネットワーク経由で通信するためのネットワーク通信ポート
- CMC からホストサーバへの直接接続用のシリアル通信ポート

インテリジェント ラックマネージャ Lite

インテリジェント ラックマネージャ Lite ソフトウェアにより、柔軟かつ包括的に HP ラックマウント CMC を制御して、HP コンピュータシステムの信頼性を最大限引き出すことができます。このソフトウェアは、ネットワーク接続されているすべての CMC に関する情報を自動的に組み合わせるため、これらのデバイスをリモートから監視、構成、管理、および制御することができます。

タワーUPS モデル



1000/1500/2200 VA タワーモデル

タワーUPS モデルは、ラックマウント型以外のサーバ、ストレージ、コンピュータ機器、およびネットワーク環境の高度な電源保護機能を提供します。各モデルは、変動が激しく不安定な電源による損傷からコンピュータ機器とデータを保護できるように設計されています。

モデル番号	VA 定格	ワット
T700J	700	500
T1000 XR	1000	700
T1500 XR	1500	1050
T2200 XR	2200	1600

T1000 XR、T1500 XR、T2200 XR には、オプションの Extended Runtime Module (ERM) があります。

タワーUPS モデルの主な機能は、以下のとおりです。

- 最大 2 つの ERM のサポートにより、システム稼働率を向上可能
- 定格電力の増大
- 機能強化されたフロントパネルディスプレイ
- 機能強化されたバッテリー管理とホットスワップ対応バッテリー

これらの UPS デバイスには HP パワーマネジメントソフトウェアが付属しています。また、ホストコンピュータとのデータ交換用の通信ポート、個別のロードセグメントの制御、リモートシャットダウンなどの機能があります。

タワーUPS モデルの特徴は、以下のとおりです。

- **ERM によるシステム稼働率の向上** — 最大 2 つのモジュールをサポートしているため、稼働時間を伸ばすことができます。
- **拡張バッテリーマネジメント (EBM) による投資保護** — バッテリーの寿命が 2 倍になり、バッテリーの充電時間が最適化されます。また、バッテリーが寿命に達する前にそれが通知されます。EBM を使用すれば、TCO を削減できるとともに、業界最先端の保護を実現できます。
- **HP パワーマネジメントソフトウェアによるインテリジェントな管理性** — HP UPS デバイスをローカルまたはリモートから監視および制御し、ネットワークの電源状態に関する詳細情報をシステム管理者に提供します。パワーマネジメントソフトウェアにより、起動とシャットダウンの時間を設定したり、UPS 診断テストを実行したり、停電時に優先順位に従ってシャットダウンを実行できます。
- **個別に制御可能なロードセグメント** — システム管理者は、複数のサーバを優先順位に従ってシャットダウンできるため、最も重要なサーバの稼働時間を延ばすことができます。また、この機能により、週末や休日など使用率が低いときのシャットダウンと起動をスケジューリングして、電力の消費を抑えることができます。

ラックマウント UPS モデル



R12000 XR



R1500J XR

ラックマウント型 UPS デバイスは、停電が発生した際に、さまざまなラックマウント型コンピュータシステムの稼働時間を最大限延ばすことが可能な高性能 UPS で、ラック占有スペースも必要最小限に抑えることができます。

これらの UPS は、Insight マネージャ 7 の HP パワーマネジメントソフトウェアを使用してローカルまたはリモートから管理できます。パワーマネジメントソフトウェアは、HP UPS ユニートを監視し、UPS 診断を実行します。

各モデルでは、ロードセグメントを個別に制御できるため、複数のサーバを優先順位に従ってシャットダウンすることができます。

ラックマウント型 UPS モデルの主な機能は、以下のとおりです。

- HP パワーマネジメントソフトウェア。インテリジェントな管理性を提供し、シャットダウンの優先順位を指定したり、装置の起動スケジュールを設定したり、ロードセグメントを個別に制御できる柔軟性を提供します。
- EBM。バッテリーの寿命が 2 倍になり、充電時間が最適化され、バッテリーで障害が発生する前に警告が通知されます。
- UPS オプションカード。高い柔軟性、拡張性、管理性を提供します。
- ホットスワップ対応バッテリー。重要な機器の電源を切ることなくバッテリーを交換することができるため、保守性が向上します。
- 3 年間保証。

- 事前予防保証。HP の 3 年間保証を拡張し、バッテリーが実際に故障する前にも適用されます。事前予防保証は、すべての HP UPS デバイスに付属されていて、バッテリーが故障する可能性があることがパワーマネジメントソフトウェアから通知された場合は、保証規定に基づいてバッテリーを無償で交換します。

ラックマウント型 UPS モデルに固有の機能を以下の表に示します。

機能	R1500J XR	R3000J XR (低電圧モデル)	R6000J	R12000XR
形式	ラックマウント型	ラックマウント型	ラックマウント型	ラックマウント型
サイズ	2U	2U	6U	10U
電力定格 (W)	1340	2250	6000	12000
VA 定格	1500	2400	6000	12000
EBM	有	有	有	有
オプションの ERM	1 つまたは 2 つ	1 つ	1 つまたは 2 つ	1 つまたは 2 つ
その他の機能	UPS ネットワーク デバイスをリモートから 監視可能	ホットスワップ対応: <ul style="list-style-type: none"> ▪ バッテリー ▪ 電子モジュール 	電力密度: 1kW/1U	「N+x」並列リタ ンダント

すべての UPS の保守に関する注意事項

- あらゆる HP UPS で、TEST/ALARM/RESET ボタンを 3 秒間押し続けることで自己診断テストを実行できます。また、このボタンを 2 秒間押し続けることで、アラーム音をリセットできます。
- ホットスワップ対応バッテリーは、フロントパネルから操作できます。バッテリーの交換作業は、必ずサービスエンジニアが実施してください。
- オプションのマルチサーバ UPS カードを追加できます。これにより、シリアルポート経由で最大 3 台のサーバと直接通信できるようになります。スケーラブルカードオプションにより、3 台の HP UPS が 1 つの仮想 UPS として動作するようになります。
- 700、1000、2000、および 3000 シリーズでは、ホットスワップ対応バッテリーとアドオンカードに加え、ユニット全体も交換できます。3000XR、6000 および 12000XR には、交換可能な電子モジュールがあります。
- スリープモード機能はデフォルトで有効になっていて、UPS ポートで 5%～10%の最小ロードが検出されなかった場合に、UPS は約 5 分間でシャットダウンします。このシャットダウンは、ソフトウェアのシャットダウン設定とは独立に機能します。スリープモードは、バッテリーの電力を節約する目的で用意されています。スリープモード機能が最小ロードで有効にならないようにするには、UPS のユーザガイドの手順に従って、これを無効にしてください。
- すべての HP UPS で、内蔵リレーが buck-to-boost 状態または boost-to-buck 状態から切り替わる時にカチッという音がします。これは正常な動作です。この音が繰り返し鳴る場合は、UPS のユーザガイドの説明に従って自己診断テストを実行してください。問題が検出された場合は、アラーム音が鳴り、LED が点灯します。自己診断テストにパスした場合は、電気技術者に AC 電源の入力電圧の変動を調べてもらってください。
- UPS を介して同じ AC 電源に大量のハードウェアを接続すると、ハードウェアが損傷したり、感電する危険性があります。ラックベースのソリューションは、必ず安全基準と電気法規に従って構成してください。
- UPS にレーザープリンタは接続しないでください。レーザープリンタは、電源投入時に過度の電力を消費します。



警告

UPS または UPS の電源コードは絶対に改造しないでください。改造した場合は、UPS 保証の対象外になります。

パワーディストリビューションユニット(PDU)



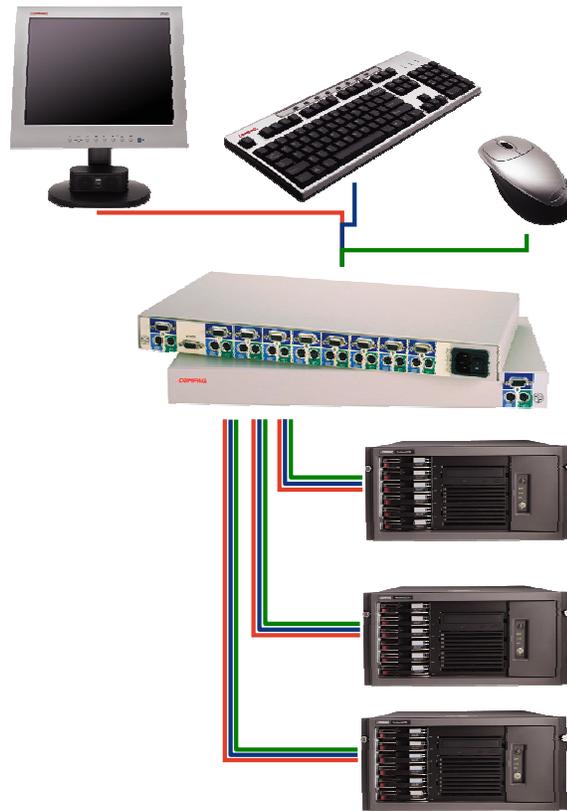
PDU は、1 つの電源から複数のオブジェクトに電力を供給します。ラックマウント環境の場合、PDU は、サーバ、ストレージ、および他の周辺機器に電力を分配します。

PDU システムにより、コンピュータキャビネット内の構成要素への配電に関する問題に対処することができます。つまり、キャビネット内に引き込む電源ケーブルの数を減らしたり、一連のサーキットブレーカによってある程度の電源保護機能を提供できます。

新しい PDU は従来のもものよりも小さいため、顧客の要望に応じて 0U または 1U 構成が可能です。PDU には、ラックキャビネット内で配電するための出力ソケットが最大 32 個あります。

- 0U 24Amp 低電圧 PDU
- 0U 24Amp 高電圧 PDU

サーバコンソールスイッチ



コンソールスイッチボックスは、複数のコンピュータサーバへ直接接続するためのデバイスで、これらのシステムで 1 組の KVM を論理的に切り替えることができます。異なるオペレーティングシステム環境の 64 台のコンピュータを 1 台の画面で操作できます。



注記

コンソールを切り替えるには、キーボードの [Print Scrn] ボタンを押してオンスクリーンメニューを呼び出します。

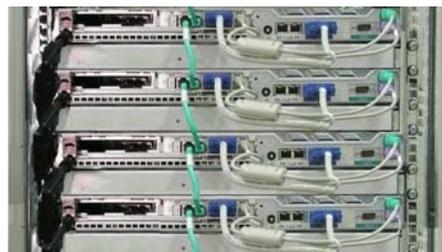
コンソールスイッチボックスの標準機能は、以下のとおりです。

機能	説明
形状	1U 構成 (側面または後方に取り付けることで 0U 構成も可能)
接続	2、4、または 8 ポート
対応台数	8 ポートのスイッチボックスを接続 (デ이지ーチェーン) することで、最大 64 台のサーバを管理可能
サーバへのアクセス	オンスクリーンディスプレイとキーボードホットキー
オンスクリーンディスプレイ (OSD)	電源投入時のテストデータや構成メニューなどのシステム関連情報がモニタに表示される
プログラマブルスキャン	システム内のすべてのコンピュータまたはそのいくつかを順番にスキャンして、システムのパフォーマンスを評価可能。また、スキャンするコンピュータや接続時間を指定可能
コンフィギュレーション NVRAM	電源が供給されなくなった場合でも、構成情報は次回更新するまで維持される
ハードウェアリセットスイッチ	キーボードおよびマウス、あるいはそのいずれかが応答しなくなったときに、システムの電源を切ることなくコンソールを再起動するのに使用
無人再起動	停電後、スイッチボックスに接続されている各サーバは、設定されているとおりに自動的に再起動する。スイッチは、再起動が正常に完了したかどうかを確認するための応答を生成する
ケーブルのロック	外部デバイスとの確実な接続が可能
色分けされたコネクタ	正しいインストールが可能
スイッチのファームウェアの更新	スイッチのアプリケーションコードはフラッシュ ROM に格納されるため、すばやく簡単な更新が可能
[Print Scrn]	KVM にアクセスするためのキー
[Reset] ボタン	デバイスの設定を復元し、ユニットをリセットする

コンソールスイッチボックスを使用する前に、以下のことを考慮してください。

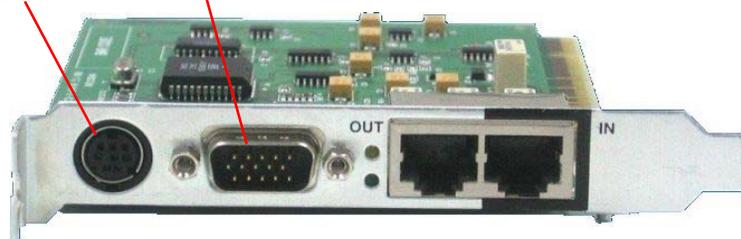
- **[Print Scrn]キーの使用方法** — [Print Scrn]は、画面の内容をプリンタに送信するだけでなく、オンスクリーンディスプレイ(OSD)を表示するのにも使用します。[Print Scrn]を1回押すと、OSDが表示されます。2回押すと、選択したコンピュータの画面がキャプチャされます。
- **ポート/コンピュータの状態のチェック方法** — [Print Scrn]を押して、OSDを表示します。[Port]または[Name]列にそれらの状態が表示されます。コンピュータが接続されていて実行されている場合は、緑色の円(O)またはプラス記号(+)が表示されます。スレーブスイッチが接続されていて実行されている場合は、×またはアスタリスク(*)が表示されます。
- **OSD ウィンドウの閉じ方** — 選択ウィンドウを閉じるには、[Esc]キーを押します。
- **ユニットのリセット方法** — キーボードまたはマウスがロックアップした場合は、背面パネルの[Reset]ボタンを押してスイッチをリセットします。[Reset]を押すことで、コンピュータの電源を入れ直すことなくデバイスの設定を復元できることがあります。

PCI カード KVM スイッチ



ProLiant DL360のPCIスロットに装着されたPCIカードKVM。

キーボード/マウス用
コネクタ ビデオ用コネクタ



PCI カード KVM スイッチは、カード形式の KVM ソリューションで、ラック内のケーブル数を削減できます。このソリューションにより、カードからの距離が 110m 以内のサーバを 64 台までサポートできます。

KVM は、CAT5 ケーブルで接続されたサーバとカードとともにローカルループになります。この新しいシステムと既存のスイッチボックスを併用することもできます。

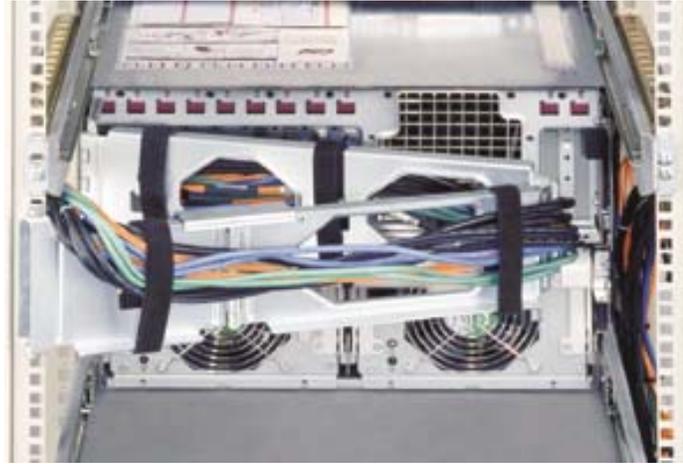
ラックアクセサリ

HP は、HP 製ラックの構成に役立つさまざまなラックアクセサリを提供しています。

ラックアクセサリ	説明
TFT5600 RKM	TFT5600 ラックマウント型キーボード/モニタは、TFT フラットパネルモニタとキーボードが統合された製品で、わずか 1U のラックスペースに 15 インチの表示領域とプログラマブルホットキーを提供します。フラットパネルモニタとキーボードが一体化された革新的なデザインが採用されています。 ラックマウント型の設計でスペースの消費が少ないので、HP 製スイッチボックスを追加できます。TFT 技術の採用により最高のパフォーマンスと信頼性が実現されているとともに、CRT 製品よりもシャープでちらつきのない表示と広い表示領域を提供します。
ラックマウント型フラットパネルモニタ	格納式の 2U シェルフに収納できる省スペース型の 15 インチ TFT フラットパネルモニタです。ラック スペースを 7U 以上も節約できるので、HP 製サーバ、ストレージ、ネットワークング、または電源保護製品を追加して、ラックを最大限に利用することができます。
モニタシェルフキット	この 1U 固定シェルフは、ラック内の任意の高さに設置して、モニタなどの構成要素を収納することができます。モニタシェルフキットの最大積載重量は 68kg で、15 インチモニタ用および 17 インチモニタ用のベゼルが同梱されています。
キーボードシェルフキット	この省スペース型 1U シェルフは、トラックボール付き内蔵キーボードをラックに設置するのに便利です。また、キーボードを保護し、すっきりと収納することができます。このシェルフの後ろに HP 製スイッチボックスを設置して 1U のラックスペースを節約することもできます。HP 製ラック内の任意の高さに設置し、スライドさせてラックに出し入れすることができます。
トラックボール内蔵キーボード	トラックボールが内蔵されたこのキーボードは、1U キーボードシェルフキットにすっきりと収納できます。このキーボードには、9 つのホットキーにそれぞれ最大 7 回のキー入力をプログラムすることができるウィザードキーも付いています。これらのホットキーにより、ラックマウント型デバイス間をすばやく移動することができます。
ラックスライドシェルフキット	このスライド シェルフを使用すると、一部のラックマウント型製品を HP 製ラック内の任意の高さに設置することができます。設置した機器はスライドさせてラックに出し入れできるので、アクセス、保守、およびアップグレードを簡単に行えます。ケーブルマネジメントアームとハンドルが付属しています。
ラックレールキット	このキットは、19~31 インチに奥行きを調節可能な 2 本のレールで構成されています。このキットにより、HP 製のラックマウント型サーバ、ストレージ ボックス、UPS、およびネットワークング製品を他社製のラックに設置することができます。また、ほとんどの他社製ラックマウント型製品を HP 製ラックに設置することもできます。
ブランクパネルキット	ブランクパネルは、ラック内の未使用スペースの開口部をふさいで適切なエアフローを保つための板金プレートです。
ラックサイドパネルキット	36U、42U および 47U ラック用のサイド パネルです。ラックを連結したラック グループ 1 つに対して、このキットが 1 つ必要になります。
ラック固定脚セット	36U、42U および 47U ラックをスタンドアロンで使用する場合は、固定脚が必要です。
ラック接続キット	ラック接続キットは複数のラックを接続し、ラックによるフロアスペースの占有面積を最小限にします。
キーボード/モニタ延長ケーブル	2m の延長ケーブルです。キーボード、モニタ、およびマウスをより柔軟に配置できるようになります。

ラックアクセサリ	説明
コンソールスイッチボックス	この 1U のプログラマブルスキャニングスイッチボックスは、1 組のキーボード、モニタ、マウスを、4 台または 8 台の ProLiant サーバ間で切り替えて使用することができます。このスイッチボックスを使用すると、すべてのケーブルをラックキャビネットの背面に通し、オンスクリーンディスプレイによってステータス情報を表示させることができます。スイッチボックスをカスケード接続すれば、1 つのコンソールから 16 台または 64 台までのサーバをサポートでき、最高の利便性を得ることができます。
スイッチボックスコネクタキット	このキットを使用すると、キーボード、モニタ、マウス、および電源コード用の 2/4/8 ポートスイッチボックスに前面からアクセスできるようになります。
スイッチボックス接続ケーブル	この 1~12m のケーブルを使用すると、サーバをスイッチボックスに柔軟に接続できます。
ラックファンキット	このキットをラックの最上部に設置することによって、熱せられた空気をラックの背面から最上部を通して排出することができます。
ケーブルマネジメントキット	このキットは、ラックに標準装備されているケーブル整理用の D 型リングに加えて使用できる 10 個組の D 型リングです。
コンソール マネジメント コントローラ (CMC)	このコントローラは、リモートから、またはオンサイトでラック環境を監視および制御するモニタリングデバイスです。このキットは、温度、侵入、衝撃、および湿度を監視します。

ケーブルマネジメントアーム



ケーブルマネジメントアームにより、サーバに接続されているすべてのケーブルを整理整頓して、ケーブル外れを防止することができます。

このアーム上にケーブルを載せて固定することで、サーバ移動時のケーブル外れを防止できます。サーバとラックに取り付けます。

パワーマネジメントソフトウェア

HP パワーマネジメントソフトウェアを使用して、ローカルまたはリモートの Windows 2000、Windows NT、NetWare 環境から UPS デバイスを監視および制御できます。

HP パワーマネジメントソフトウェアは Insight マネージャとともに機能して、ネットワーク内のすべての UPS を柔軟かつ包括的に制御することで、システムの電源の信頼性を最大限高めめます。

HP パワーマネジメントソフトウェアは、すべてのネットワークコンピュータの電源環境を監視、管理、および制御するための機能を提供します。HP パワーマネジメントソフトウェアは、以下の場合に使用します。

- 個別の UPS のロードセグメントを管理して、接続されている装置の電源を個別に管理する
- 装置のシャットダウンの優先順位を指定して、ロードセグメントごとに接続されている装置を再起動する
- 手動またはユーザが指定したスケジュールに従って、HP UPS と接続されている装置をシャットダウンして起動する
- Insight マネージャのカスタマイズされたアラート生成機能にアクセスする

マルチサーバ UPS カード

マルチサーバ UPS カードは、HP パワーマネジメントソフトウェアによりインテリジェントに管理されます。HP マルチサーバ UPS カードを使用すれば、各ロードセグメントは、個別のシャットダウンスケジュールとロードセグメント構成が設定された個別の UPS のように機能します。

マルチサーバ UPS カードにより、システム管理者は複数のプラットフォームを 1 つの HP UPS で管理できるようになります。ネットワークがダウンした場合でも、マルチサーバ UPS カードは、優先順位に従ったシャットダウンを実行できます。

復習問題

1. ラック内の通気性を向上させる製品を挙げてください。
.....
2. UPS をリモートからシャットダウンするのに必要なソフトウェアを挙げてください。
.....
3. HP UPS でカチッという音が繰り返し鳴っている場合の対処方法を挙げてください。
.....
4. 各用語とその説明を対応付けてください。
 - a. 水平脚 誤ってケーブルが外れるのを防止する
 - b. ラック接続キット スタンドアロンラックに必要
 - c. ケーブルマネジメントアーム ラック構成ツール
 - d. ラック固定脚 ラックの全重量を支える
 - e. rack builder online 複数のラック構成で必要

目的

この章では、次の内容について学習します。

- ドキュメント、オンラインなどのサポートサービス
 - ドキュメント
 - ◆ シリアル番号
 - ◆ パーツ番号
 - ◆ Maintenance and Service Guides (MSG)
 - ◆ アクセスパネルラベル
 - オンライン
 - ◆ HP サポート Web サイト
 - ◆ 標準保証

ドキュメント、オンラインなどのサポートサービス

HP 製品に記載されている製品リファレンス情報は、以下のとおりです。

- シリアル番号
- パーツ番号

ProLiant サーバのサービスドキュメントは、以下のとおりです。

- アクセスパネルラベル
- Maintenance and Service Guides (MSG)

HP がオンラインで提供しているサービスは、以下のとおりです。

- ソフトウェアとドライバ
- Reference Library
- ActiveAnswers

シリアル番号

HP の 12 桁または 10 桁のシリアル番号は、製品固有の ID で、以下の情報を示しています。

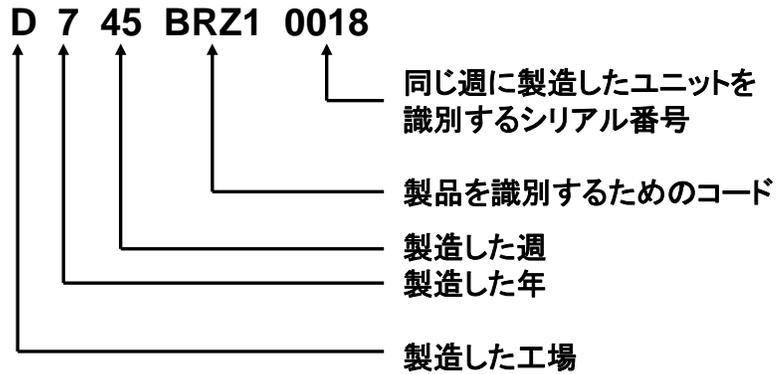
- 製造年月日 (年と週または月で表示)
- モデル構成

ProLiant サーバでは、現在 3 種類のシリアル番号形式が使用されています。

コンフィギュレーションコード

コンフィギュレーションコードは、製品シリアル番号の 5～8 番目の値です。

シリアル番号の詳細(形式 1)



- S = サイトコード
- Y = 製造年コード
- WW = 製造週コード
- CCCC = コンフィギュレーションコード
- ZZZZ = シーケンシャル番号

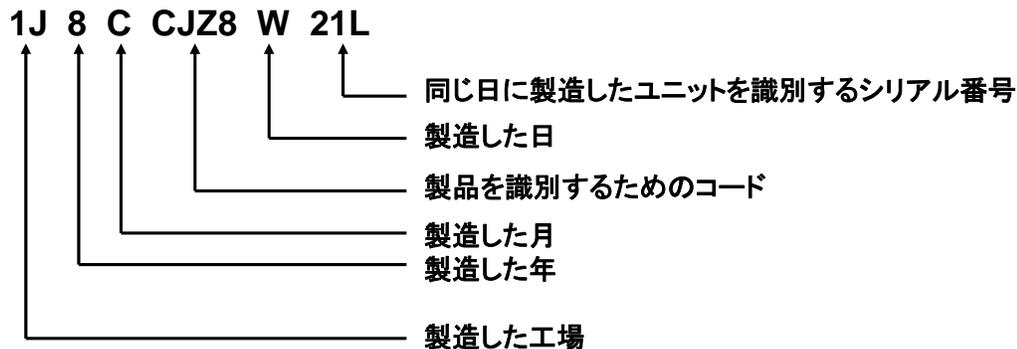
例

シリアル番号 8811BBN51234 は、コンフィギュレーションコード BBN5 の製品であり、サイト 8 で 1998 年の第 11 週に製造され、その週の 1234 番目の製品であることを示しています。

注記

▲ コンピュータが保証期間外であることをシリアル番号が示している場合でも、常にコンピュータの購入日をチェックして保証の状態を確認してください。たとえば、システムのシリアル番号の製造日が保証期間外であることを示している場合でも、コンピュータがパーツ保証の対象となる場合があります。

シリアル番号の詳細(形式 2)

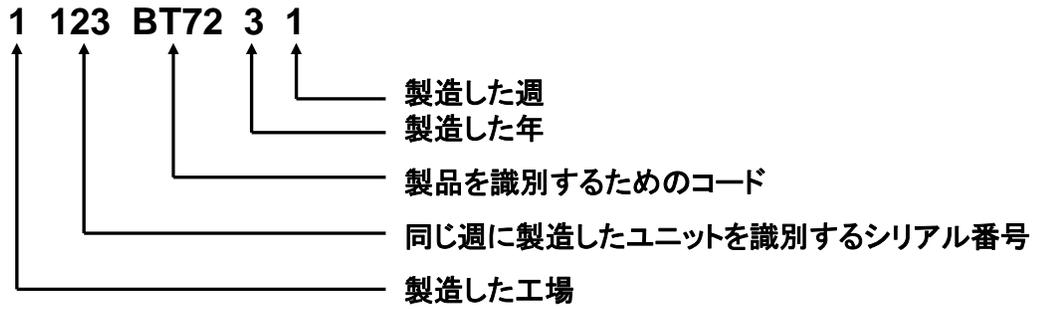


- SS = サイトコード
- Y = 製造年コード
- M = 製造月コード
 - 1月 = 1、2月 = 2、3月 = 3
 - 10月 = A、11月 = B、12月 = C
- CCCC = コンフィギュレーションコード
- D = 製造日コード
 - 使用される文字は、1~9、A~Z
 - I、O、Q、および0(ゼロ)は未使用
- ZZZ = シーケンシャル番号を表す Base 32 文字
 - 0、1、2、3...9、A、B、C...Z
 - 毎日0にリセットされる

例

シリアル番号 4J1CBT71134Y は、コンフィギュレーションコード BT71 の製品であり、サイト 4J で 2001 年の 12 月 1 日に製造され、その日の 34Y 番目の製品であることを示しています。

シリアル番号の詳細(形式 3)



- S = サイトコード
- ZZZ = シーケンシャル番号
- CCCC = コンフィギュレーションコード
- Y = 製造年コード
- W = 製造週コード

例

シリアル番号 1123BT7231 は、コンフィギュレーションコード BT72 の製品であり、サイト 1 で 2003 年の第 1 週の 123 番目に製造された製品であることを示しています。

パーツ番号

HP では、3 種類のパーツ番号を使用しています。すべて同じ形式(#####-AAA。#は数字、A はアルファベットを表す)ですが、値の意味は異なっています。これらの値に互換性はありません。

- **アセンブリパーツ番号** — HP 製品を組み立てるために製造工程で使用されます。同じアセンブリ番号の部品は同じ部品です。部品が変更された場合(チップセットが変更されたり、構成要素の製造元が変更された場合など)は、新しいアセンブリパーツ番号が割り当てられます。通常、最後の 3 桁は部品のリビジョンを表しています。

- **スペアパーツ番号** — 既存の HP 製品の障害が発生した構成要素の交換部品を注文する際に使用します。提供されるスペア部品は、通常、1 つの構成要素です。

同じスペアパーツ番号の部品は、機能的に同等であり、互換性があります。スペアパーツ番号が同じ部品でもアセンブリ番号が異なることがあります。スペアパーツは、フィールド交換可能ユニット(FRU)とも呼ばれます。

- **オプションパーツ番号** — 新しい部品を注文する際に使用します。提供されるオプションパーツには、通常、HP 製品に新しいパーツをインストールするのに必要なすべてのものが含まれています(ケーブル、アダプタ、ユーザガイドなど)。

Maintenance and Service Guides (MSG)

各 ProLiant サーバファミリに個別の Maintenance and Service Guides (MSG) が発行されます。MSG は、トラブルシューティングガイドおよび HP コンピュータを保守する際の参照ツールとして使用できるように用意されています。

MSG は、いくつかの章で構成されており、それぞれ以下の情報を提供しています。

- **Illustrated Parts Catalog** — そのマシン固有のスペア部品カタログとすべてのスペア部品番号のリストがイラスト形式で掲載されています。
- **Service Preliminaries** — サービス開始前の警告と注意、必要な機器に関する情報、および保証情報が掲載されています。
- **Removal and Replacement Procedures** — 特定のコンピュータファミリのフィールドサブアセンブリの取り外しおよび交換方法について説明しています。
- **Switch and Jumper Settings** — スイッチとジャンパの設定方法の詳細について説明しています。各ボードの固有の設定について説明しています。
- **Power-On Self-Test (POST)** — 電源投入時に自動的に実行される内蔵システム診断プログラムについて説明しています。
- **Error Messages and Codes** — POST および HP Diagnostics Error Codes の一覧と、各エラーコードが示している問題の解決方法について説明しています。また、LED の一覧とその意味についても説明しています。
- **LED Indicators** — グラフと表により、システムインターロック ステータスインジケータ、ホットプラグ対応 I/O ファン、PCI ホットプラグ LED インジケータ、ホットプラグ対応電源、およびホットプラグ対応 SCSI ハードディスクドライブ上の LED の場所と機能について説明しています。
- **Specifications** — ガイドの対象となっている特定の HP コンピュータの動作およびパフォーマンス仕様について説明しています。
- **Index** — ガイドで特定の情報を見つけるのに役立ちます。

INTERNET MSG は、Reference Library (リファレンスライブラリ) に製品ごとに掲載されています。
http://www.compaq.com/support/techpubs/maintenance_guides/

アクセスラベル

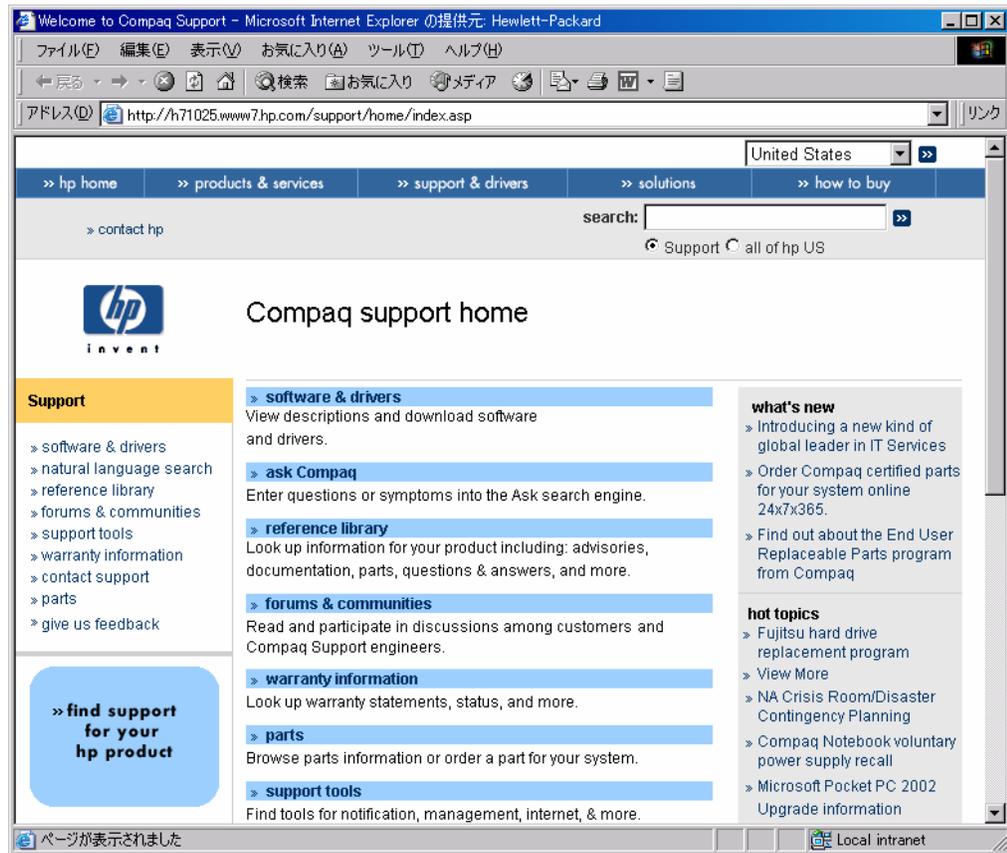
ProLiant サーバには、1 つまたは複数のアクセスパネルラベルがあり、ここにはサーバ構成やオプションのインストールに関する情報が記載されています。サーバの電源を切ってカバーを取り外すことで、サービスエンジニアは十分なサービス情報を得ることができます。

アクセスパネルラベルには、以下の情報が記載されています。

- ディップスイッチの場所と状態
- ジャンパの場所と状態
- LED の場所と意味
- I/O ボードの情報
- オプションのインストール情報
- モジュールへのアクセス方法
- プロセッサボードの構成
- ケーブル情報
- メモリのインストール方法

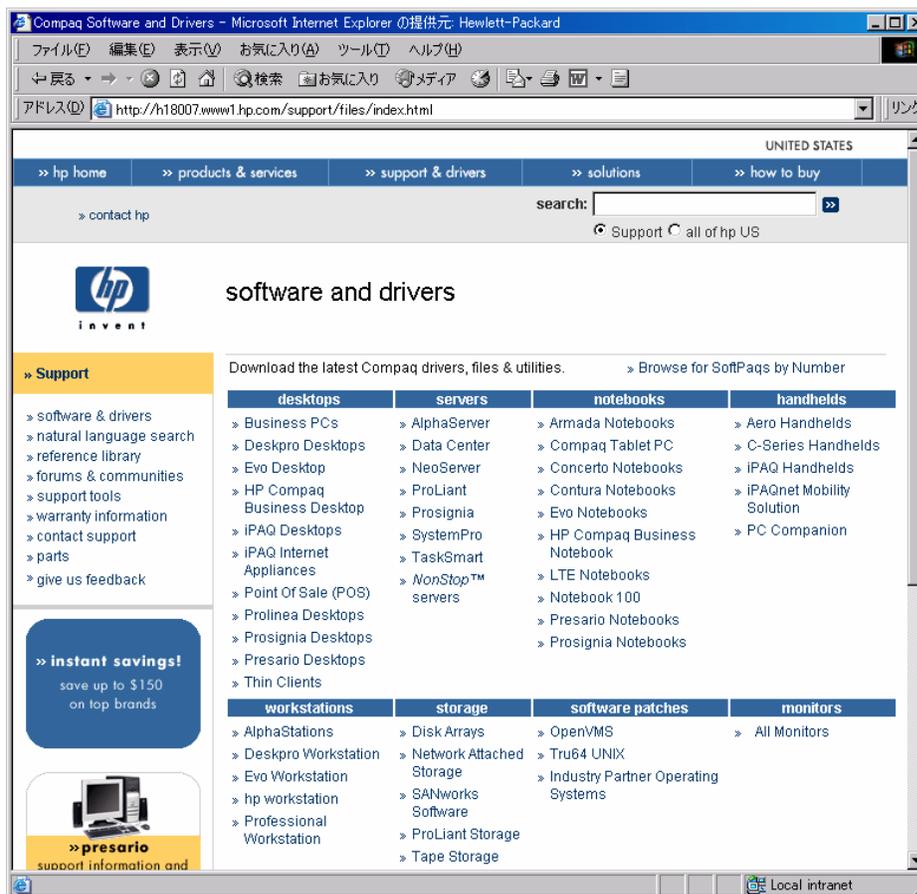
通常これらのラベルには、警告や注意情報は記載されていません。

HP サポートリソース Web サイト



HP サポート Web サイト (<http://h71025.www7.hp.com/support/home/index.asp>) にアクセスするか、HP の Web サイトで [Support] を選択します。

ソフトウェアとドライバ



HP 製品が最適なパフォーマンスを維持するのに必要なすべてのドライバ、ユーティリティ、サポートソフトウェア、ROMPaq は、<http://h18007.www1.hp.com/support/files/> から入手できます。

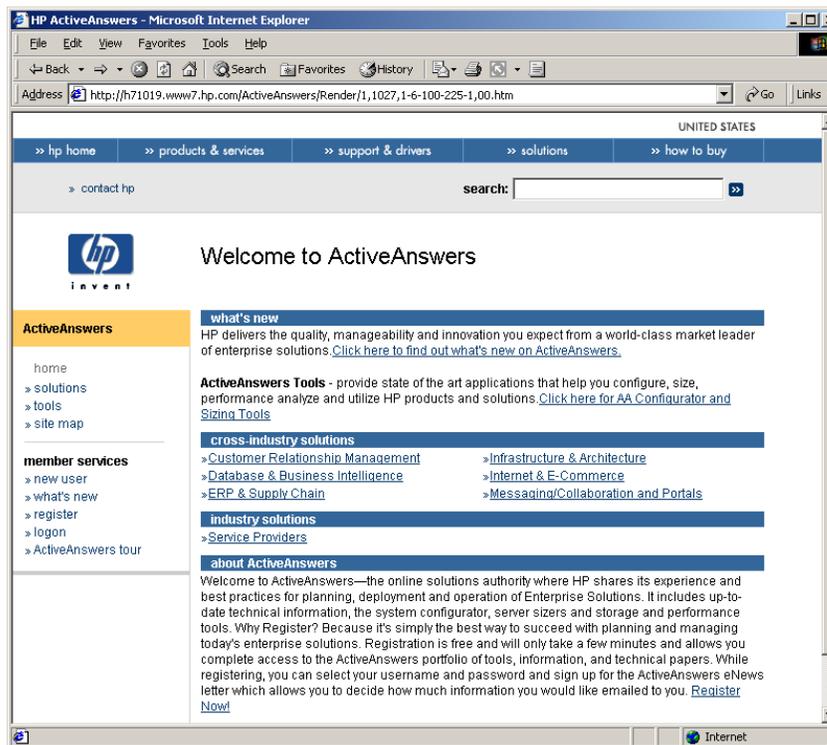
製品プラットフォームを選択し、コンピュータファミリ、モデル、言語、およびオペレーティングシステムを検索して正しい SoftPaq を見つけます。各 SoftPaq の説明は Readme ファイルにあり、このファイルは SoftPaq とともに自動的にダウンロードされます。

Reference Library

[Support Home]ページの[Reference Library]では、以下のサポートドキュメントが製品別に用意されています。

サポートしている製品	説明
Customer Advisories (アドバイス集)	<p>以下の内容に関するアドバイスを掲載しています。</p> <ul style="list-style-type: none"> ■ コントローラとアダプタ ■ ハードウェア ■ オペレーティングシステムとユーティリティ ■ インターネットソリューション ■ 通信とネットワーク ■ クラスタ ■ システム管理
Drivers, SoftPaq, Software, and Utilities (ドライバ、SoftPaq、ソフトウェア、ユーティリティ)	<p>ソフトウェア情報を製品ファミリー、モデル、およびオペレーティングシステム別に、またはオペレーティングシステムとソフトウェアカテゴリ別に参照できます。</p> <ul style="list-style-type: none"> ■ Support Pack ■ ディスプレイ ■ マネジメントエージェント ■ 管理アプリケーションとユーティリティ ■ ネットワーク ■ ストレージ ■ システム ROMPaq/BIOS ■ ユーティリティ
Manuals(マニュアル)	<p>製品ドキュメントは以下のとおりです。</p> <ul style="list-style-type: none"> ■ MSG ■ Setup and Installation Guide ■ User and Reference Guides ■ Option Related Guides
Other(その他)	<p>製品別に用意されているその他の情報は、以下のとおりです。</p> <ul style="list-style-type: none"> ■ TechNotes ■ サービス ■ ホワイトペーパー

ActiveAnswers



ユーザがビジネスソリューションを計画、配備、運営するのに役立つツール、e サービス、情報を提供します。

INTERNET 初めてアクセスした場合には、登録が必要です。

<http://h71019.www7.hp.com/ActiveAnswers/Render/1,1027,1-6-100-225-1,00.htm>

登録することで、すべての機能にアクセスして、ActiveAnswers コミュニティを十分に活用できるようになります。System Configurator を使用すれば、ソリューションと対応する部品表を作成することができます。Application Sizer により、ソリューション環境のサイズと範囲を判断することができます。

Solutions List を使用するには、調査するカテゴリを選択します。選択肢は以下のとおりです。

- Business Intelligence
- NonStop eBusiness
- Data Management and Data Storage
- Enterprise Applications
- Enterprise Management
- Internet and E-Commerce
- Messaging and Collaboration
- Software

HP の標準保証

HP では、さまざまな保証プログラムを提供しています。

- **ProLiant サーバ** — 通常、3-3-3 年間保証が付いています。この保証には、3 年間の稼働保証 (オンサイトまたはセンドバック)、3 年間のパーツ保証、および翌営業日での対応が含まれます。
- **事前予防保証** — 保証の一部で、Insight マネージャソフトウェアによって障害が予測されると認識されたパーツ (プロセッサ、メモリ、ハードドライブを含む) を修理または交換します。

復習問題

1. HP で使用されているパーツ番号の 3 つのタイプを挙げてください。

.....
.....
.....

2. HP が提供している保証プログラムを挙げてください。

.....
.....

第 2 章のトピック

RISC 命令と CISC 命令

当初、命令はプロセッサ内部に組み込まれていました。このため、新しい命令を使用するには、プロセッサを再設計しなければならないという問題が発生しました。また、プロセッサを変更すると、命令が変更されるという逆の問題も生まれました。

その後、命令はキャッシュに保存されるようになりました。これにより、マイクロコードとも呼ばれる命令の変更や新規作成が容易になりました。ほとんどの場合、複数の組み込み命令が実行していたことを 1 つの命令に実行させました。この結果、ソフトウェアプログラムは小さくなり、メモリ消費量が減少しました。

ただし、これらの複雑な命令により、プロセッサの作業は増大しました。場合によっては、1 つの複雑な命令を処理するほうが、複数の単純な命令を処理するよりも時間がかかることがテストにより判明しました。また、ほとんどの場合、プログラムは同じ単純な命令を繰り返し使用し、複雑な命令はあまり使用しないことも判明しました。

あるエンジニアグループは、複数の単純な命令で構成された命令セットを作成し、すべての命令を同じ長さにしてみることにしました (マイクロコード命令は可変長であるため、デコードがより困難でした)。この結果、プロセッサが処理しなければならない命令数は増えましたが、処理速度は向上しました。

この技術は RISC (Reduced Instruction Set Computing) と呼ばれるようになり、これに対して元のマイクロコード技術は、CISC (Complex Instruction Set Computing) と呼ばれるようになりました。

Motorola プロセッサと Alpha プロセッサは、RISC ベースのプロセッサです。Pentium Pro をはじめとする Intel プロセッサのほとんどは、CISC ベースのプロセッサでした。近年、Intel プロセッサには、RISC プロセッサを高速化するために開発された数多くのテクノロジーが導入されているため、RISC と CISC の境界線があいまいになってきました。

命令セットと拡張

あるプロセッサは、ある特定の命令グループまたは命令セットだけを実行するように設計されています。たとえば、x86 プロセッサは、x86 命令セットを使用します。

Intel は、プロセッサのパフォーマンスを向上できるように、Pentium プロセッサの各世代で x86 命令セットに追加命令を導入しました。これらの追加命令は、拡張命令セットと呼ばれています。

MMX や SSE (Streaming SIMD Extensions) などの拡張命令セットは、クライアントとして使用される PC において効果的で、SSE2 (Streaming SIMD Extensions 2) はサーバで効果を発揮します。

MMX

1997 年に Intel は、Pentium II プロセッサに MMX 命令セットを導入しました。マルチメディアおよび通信プログラムをより効率的に実行できるように、MMX 命令セットとして 57 個の新しい命令が追加されました。

マトリックス内のデータを乗算する場合など、これらの多くのプログラムでは、1 つの命令をデータの複数部分で繰り返し実行する必要がありました。この手法は、SIMD (Single Instruction Multiple Data) コンピューティングモデルと呼ばれています。

SIMD モデルを使用するプログラムをサポートできるように、MMX 命令セットでは、命令を繰り返し実行する代わりに、1 つのマイクロインストラクションが複数のデータ項目上で同時に動作できるようになっています。

MMX 拡張は、2D グラフィックスに効果的です。現在では、オーディオを操作したりプロシージャルテクスチャ (プログラムによるテクスチャの自動生成) といった利用法が主流になっています。

SSE

1999 年に Intel は、Pentium III プロセッサに SSE (Streaming SIMD Extension) 命令セットを導入しました (当初これは、Katmai New Instructions または KNI と呼ばれていました)。SSE 命令セットとして、56 個の新しい命令が追加されました。

この命令セットにより、リアルタイム 3D アプリケーションなど、ビジュアルパフォーマンスが重要なプログラムのパフォーマンスが向上しました。また、MPEG-1 や MPEG-2 などの新しいビデオ形式を処理したり、音声認識の信頼度を上げるための命令も含んでいました。

SSE2

2000 年に Intel は、SSE2 (Streaming SIMD Extension 2) を Pentium 4 プロセッサに導入しました。この命令セットとして、144 個の新しい命令が追加されました。

SSE2 命令セットの長所の 1 つは、1 つの命令で処理可能な値のサイズが 64 ビットから 128 ビットに増えたことです。

1 つの命令で処理できる値が大きいと、タスクを完了するために必要な命令数は少なくて済むため、プロセッサはタスクをより高速に処理できるようになります。これは、アプリケーションの速度がユーザの印象を直接左右するビデオストリーミングなどのマルチメディアアプリケーションで特に有効です。

パイプラインハザードの例

ここでは、コンピュータが次のタスクを与えられた場合について考えます。

例

「x」が「animal」（動物）の場合は、
「I am an animal.」（私は動物です）と表示する
それ以外の場合は
「I am a plant.」（私は植物です）と表示する

プロセッサがこのコマンドを処理するのに必要な命令は以下のとおりです。

1. 「x」が「animal」であるかどうか判断する。
2. 「x」が「animal」でない場合は、ELSE にジャンプする。
3. THEN: 「I am an animal.」と表示する。
4. NEXT にジャンプする。
5. ELSE: 「I am a plant.」と表示する。
6. NEXT: 次の命令に進む。

分岐の問題

	クロック サイクル 1	クロック サイクル 2	クロック サイクル 3	クロック サイクル 4	クロック サイクル 5	クロック サイクル 6	クロック サイクル 7	クロック サイクル 8	クロック サイクル 9	クロック サイクル 10
1. 「x」が「animal」であるかどうか判断する。	命令を取得する	命令をデコードする - pt1	命令をデコードする - pt2	命令を実行する - pt 1	命令を実行する - pt 2	命令を実行する - pt 3	データを取得する - pt 1	データを取得する - pt 2	命令を書き出す - pt 1	命令を書き出す - pt 2
2. 「x」が「animal」でない場合は、ELSEにジャンプする。		命令を取得する	命令をデコードする - pt1	命令をデコードする - pt2	命令を実行する - pt 1	命令を実行する - pt 2	命令を実行する - pt 3	データを取得する - pt 1	データを取得する - pt 2	命令を書き出す - pt 1
3. THEN:「I am an animal.」と表示する。			命令を取得する	命令をデコードする - pt1	命令をデコードする - pt2	命令を実行する - pt 1	命令を実行する - pt 2	命令を実行する - pt 3	データを取得する - pt 1	データを取得する - pt 2
4. NEXTにジャンプする。				命令を取得する	命令をデコードする - pt1	命令をデコードする - pt2	命令を実行する - pt 1	命令を実行する - pt 2	命令を実行する - pt 3	データを取得する - pt 1
5. ELSE:「I am a plant.」と表示する。					命令を取得する	命令をデコードする - pt1	命令をデコードする - pt2	命令を実行する - pt 1	命令を実行する - pt 2	命令を実行する - pt 3
6. NEXT: 次の命令に進む。						命令を取得する	命令をデコードする - pt1	命令をデコードする - pt2	命令を実行する - pt 1	命令を実行する - pt 2

「x」が「tomato」の場合、「x」が「animal」または「plant」であるかどうかを判断する最初の命令は、クロックサイクル 6 まで実行が完了しません。その時点で、命令 2、3、4 はすべてパイプラインに入っていますが、「x」が「tomato」である場合、命令 3 と 4 は処理すべきではないため、問題になります。

これを分岐の問題といいます。命令 1 には 2 つの分岐があり、一つは「x」が「animal」の場合と、もう一つは「plant」の場合です。この場合、「x」が「animal」でなくてもパイプラインは「animal」分岐(命令 3 と 4)を処理します。この処理は、クロックサイクル、パイプライン領域、レジスタおよびキャッシュ領域などの貴重なリソースを浪費します。

依存の問題

別の問題が命令 2 で発生します。命令 1 の結果はサイクル 10 が完了するまで不明ですが、命令 2 はサイクル 5 を実行するのにその情報を必要とします。これが依存の問題です。

パイプラインハザードの影響

パイプラインハザードが発生すると、ハザードが排除されるまでパイプラインはブロックされません。このとき、パイプラインはストール(失速)します。サイクル 10 の後、パイプライン内のすべての命令は、命令 3 と 4 がパイプラインから排除されるまでストール(失速)します。その後、プロセッサは命令 5 を再度取得して最初からやり直します。

VLIW

1980年代半ばに、プロセッサの処理を高速化できるように、VLIW (Very Long Instruction Word) プロセッサが開発されました。VLIW システムでは、VLIW コンパイラがいくつかの短い RISC ベースの命令を組み合わせて 1 つの長い命令を作成しました。コンパイラは、どの命令を組み合わせたことができるかを判断する際に、依存と分岐を考慮しました。プロセッサでは、長い命令が分割され、複数の実行ユニットへ並行にストリーミングされました。

VLIW システムの特徴の一つとして、ハードウェアではなくソフトウェアが依存と分岐に関してコードを分析することが挙げられます。これにより、プロセッサを小さくすることができます。プロセッサが小さければ、チップ上のトランジスタの数が少なくなるため、製造コストが安くなり実行時の温度も低下します。ただし、このような結果を得るためには、コンパイラが非常に高度でなければならず、コンパイラのプログラマに依存します。通常、ソフトウェアで依存と分岐を判断するほうが、ハードウェアで判断する場合よりも時間がかかります。

ProLiant サーバで採用されているプロセッサ

HP は、以下の Intel プロセッサを ProLiant サーバで採用しています。

- Pentium III
- Pentium III Xeon
- Pentium 4
- Xeon
- Xeon MP

これらのプロセッサの特徴と、それらが搭載されているサーバモデルを以下の表に示します。

Pentium III プロセッサ

機能	Pentium III 0.13 μ プロセスベース	Pentium III 512KB L2 キャッシュ	超低電圧版 Pentium III
プロセッサアーキテクチャ	32ビット	32ビット	32ビット
速度	1.13A、1.2GHz	1.13、1.26、1.40GHz	700、800、900MHz
サポートしているシステム バス速度	133MHz	133MHz	100MHz
L1 キャッシュ	16KB 命令キャッシュ 16KB データキャッシュ	16KB 命令キャッシュ 16KB データキャッシュ	16KB 命令キャッシュ 16KB データキャッシュ (ライトバック)
内蔵 L2 キャッシュ	256KB	512KB	512KB
バックサイドバス速度	フル	フル	フル
データバス	64	64	64
マルチプロセッシング	2	2	不可
パッケージング	FC-PGA2	FC-PGA2	FC-BGA
拡張命令セット	MMX、SSE	MMX、SSE	MMX、SSE
マイクロアーキテクチャ	P6	P6	P6
サーバ	DL320、 ML330	ML330 G2、ML350 G2、 ML370 G2、 DL320、DL360 G2、 DL380 G2、 BL20p	BL10e

Pentium III Xeon プロセッサ

機能	Pentium III Xeon 256KB L2 キャッシュ	Pentium III Xeon	Pentium III Xeon
プロセッサアーキテクチャ	32ビット	32ビット	32ビット
速度	600MHz~1GHz	700MHz	900MHz
サポートしているシステム バス速度	133MHz	100MHz	100MHz
L1 キャッシュ	16KB 命令キャッシュ 16KB データキャッシュ	16KB 命令キャッシュ 16KB データキャッシュ	16KB 命令キャッシュ 16KB データキャッシュ
内蔵 L2 キャッシュ	256KB	1MB、2MB	2MB
バックサイドバス速度	フル	フル	フル
データバス	64	64	64
マルチプロセッシング	2	システムバスあたり 4 個の プロセッサ	システムバスあたり 4 個の プロセッサ
パッケージング	SC330.1	SC330.1	SC330.1
拡張命令セット	MMX、SSE	MMX、SSE	MMX、SSE
マイクロアーキテクチャ	P6	P6	P6
サーバ	ML530	ML570、ML750、 DL580、DL760	ML570、ML750、 DL580、DL760

Pentium 4、Xeon、Xeon MP プロセッサ

機能	Pentium 4 512KB L2 キャッシュ	Xeon	Xeon MP
プロセッサアーキテクチャ	32ビット	32ビット	32ビット
速度	2.2、2.26、2.53、2.8GHz	2.0、2.4、2.8、3.0GHz	1.4、1.5、1.9、2.0GHz
サポートしているシステム バス速度	400MHz	400/533MHz	400/533MHz
L1 キャッシュ	12KB 実行トレース 8KB データキャッシュ	12KB 実行トレース 8KB データキャッシュ	12KB 実行トレース 8KB データキャッシュ
内蔵 L2 キャッシュ	512KB	512KB	512KB
バックサイドバス速度	フル	フル	フル
データバス	64	64	64
マルチプロセッシング	不可	2	システムバスあたり 4 個の プロセッサ
パッケージング	FC-PGA2	μ PGA	μ PGA
拡張命令セット	MMX、SSE2	MMX、SSE2	MMX、SSE2
マイクロアーキテクチャ	NetBurst	NetBurst	NetBurst
サーバ	tc2120、 ML310、 DL320 G2	ML330 G3、ML350 G3、 ML370 G3、ML530 G2、 DL360 G3、DL380 G3、 BL20p G2	ML570 G2、 DL560、DL580 G2、 DL740、DL760 G2、 BL40p

Windows ベースのサーバと論理プロセッサ

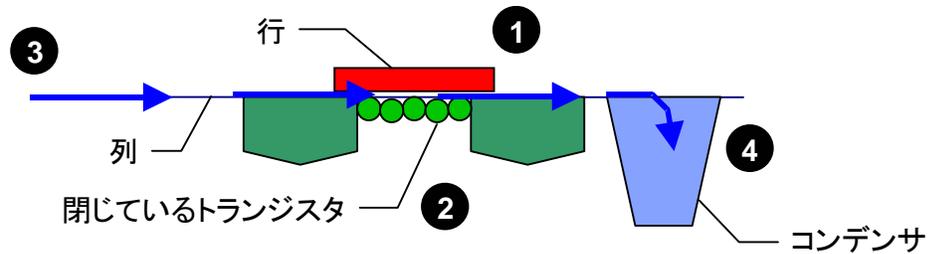
Windows ベースのサーバは、論理プロセッサを利用できますが、利用可能な数はオペレーティングシステムによって異なります。

- **Windows 2000 Server** — Windows 2000 Server には、4 プロセッサライセンス制限があり、システム BIOS から得たプロセッサ数とライセンス制限が照合されます。BIOS が 8 個のプロセッサが存在することを示した場合、4 プロセッサライセンスでは最初の 4 個だけが使用されます。Windows 2000 Server は、残りの 4 個の論理プロセッサは使用しません。
- **Windows 2000 Advanced Server** — Windows 2000 Advanced Server には、8 プロセッサライセンス制限があり、システム BIOS から得たプロセッサ数とライセンス制限が照合されます。BIOS が 8 個のプロセッサが存在することを示した場合、サーバは 8 個の論理プロセッサをすべて使用します。
- **Windows Server 2003** — Windows Server 2003 は、物理プロセッサと論理プロセッサの両方を認識するように設計されています。物理プロセッサだけに基づいてプロセッサライセンス制限を判断するので、論理プロセッサを十分に活用することができます。

例: Windows Server 2003, Enterprise Edition には 8 プロセッサライセンス制限があります。4 ウェイシステムでは、8 個の論理プロセッサをすべて使用しますが、8 物理プロセッサ制限のうち 4 個しか使用しません。

第 3 章のトピック

セルのデータ格納方法

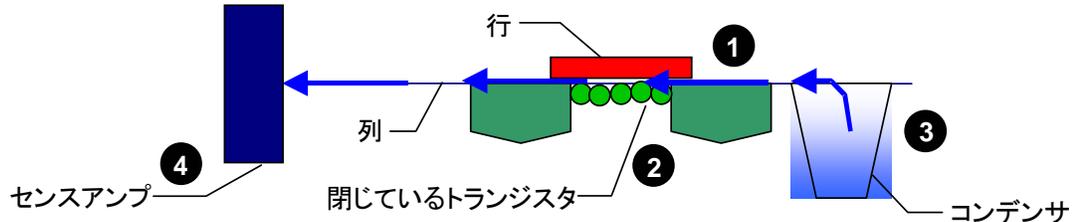


メモリセルの側面図 - データの格納

セルは、以下の手順でデータを格納します。

1. 電子パルスが行を流れます。
2. このパルスは、その行のすべてのトランジスタを閉じます。
3. 別の電子パルスが特定の列を流れます。
4. この 2 番目のパルスは、トランジスタを通過してコンデンサを帯電します。

セルのデータ読み取り方法



メモリセルの側面図 - データの読み取り

セルは、以下の手順でデータを読み取ります。

1. 電子パルスが行を流れます。
2. このパルスは、その行のすべてのトランジスタを閉じます。
3. コンデンサが帯電されている場合、閉じているトランジスタと列を通してコンデンサから電荷が流れ出します。
4. 列の終端にはセンスアンプがあり、行のコンデンサが電荷を流したかどうかを検出します。

完了時にコンデンサが電荷を保持しなくなるため、この動作は破壊読み出しと呼ばれます。この状態を直すために、センスアンプは列を通じて電荷を戻し、コンデンサを再帯電します。この間は他のデータの読み取りまたは書き込みを行うことはできません。

メモリコントローラのデータビットの格納方法

メモリコントローラは、以下の手順でデータビットをメモリモジュールに書き込みます。

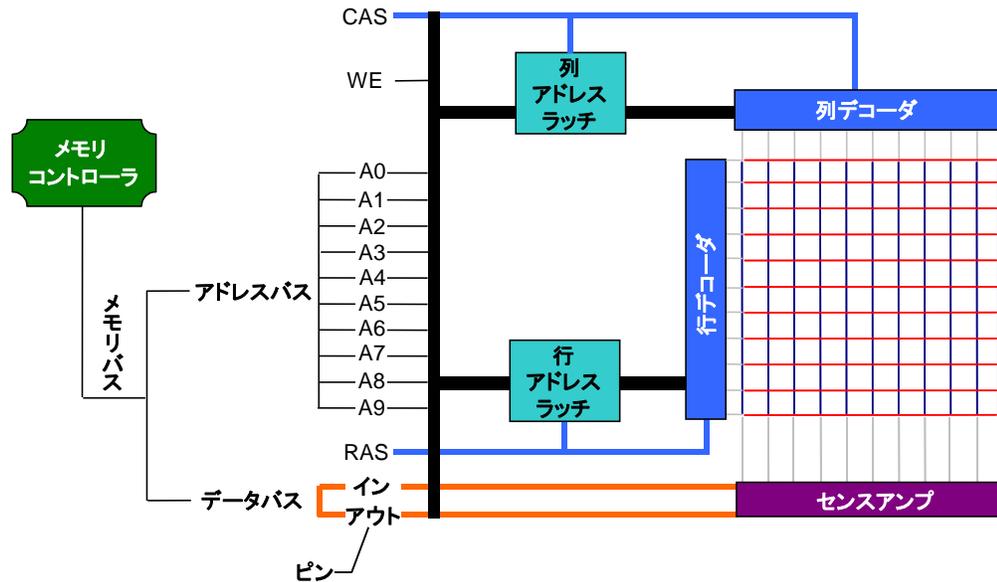
1. メモリコントローラは、メモリモジュールへのアドレスバスを通じて行アドレスを送信します。アドレスは、アドレス専用のピンを通じてメモリモジュールに入ります。
2. メモリコントローラは RAS 電圧を下げて、行アドレスを行アドレスラッチに入れます。アドレスは、書き込み処理が完了するまでそこに残ります。
3. 行アドレスデコーダは、ラッチ上のアドレスをデコードし、どの行を使用するかを判断します。その後、選択した行に沿って電子パルスを送信します。これにより、その行のすべてのトランジスタが閉じます。
4. その行に格納されているすべてのデータがコンデンサからセンスアンプに向かって流れます。
5. メモリコントローラは、列アドレスをアドレスバス上のメモリモジュールに送信します。
6. メモリコントローラは、WE ピン上の電圧を下げて、操作が読み取り操作ではなく書き込み操作であることを示します。
7. データは、データバスを通じて Data In ピンとセンスアンプに送信されます。
8. メモリコントローラは CAS 電圧を下げて、列アドレスを列アドレスラッチに入れます。
9. 列アドレスデコーダは、ラッチ上のアドレスをデコードし、どの列を使用するかを判断します。その後、選択した列に沿って電子パルスを送信します。これにより、行と列アドレスの交点にあるコンデンサが帯電されます。
10. RAS と CAS の電圧が上がり、サイクルの終わりが示されます。
11. センスアンプは、書き込み操作前にデータを保持していた選択した行のすべてのコンデンサを再帯電します。

メモリコントローラのデータビットの読み取り方法

データビットの読み取り処理は、データの書き込み処理と似ています。

1. メモリコントローラは、メモリモジュールへのアドレスバスを通じて、必要なデータの行アドレスを送信します。
2. メモリコントローラは RAS 電圧を下げて、行アドレスを行アドレスラッチに入れます。
3. 行アドレスデコーダは、ラッチ上のアドレスをデコードし、どの行を使用するかを判断します。その後、選択した行に沿って電子パルスを送信します。これにより、その行のすべてのトランジスタが閉じます。
4. その行に格納されているすべてのデータがセンスアンプに流れます。
5. メモリコントローラは、列アドレスをアドレスバス上のメモリモジュールに送信します。
6. メモリコントローラは、WE ピン上の電圧を上げて、操作が書き込み操作ではなく読み取り操作であることを示します。
7. メモリコントローラは CAS 電圧を下げて、列アドレスを列アドレスラッチに入れます。
8. 列アドレスデコーダは、ラッチ上のアドレスをデコードし、どの列を使用するかを判断します。
9. 行と列の交点にあるセルの電荷がセンスアンプと Data Out ピンに放電されます。
10. メモリコントローラは、RAS および CAS 回路の電圧を上げてサイクルの終わりを示します。
11. センスアンプは、書き込み操作前にデータを保持していた行のすべてのコンデンサを再帯電します。

データのリフレッシュ



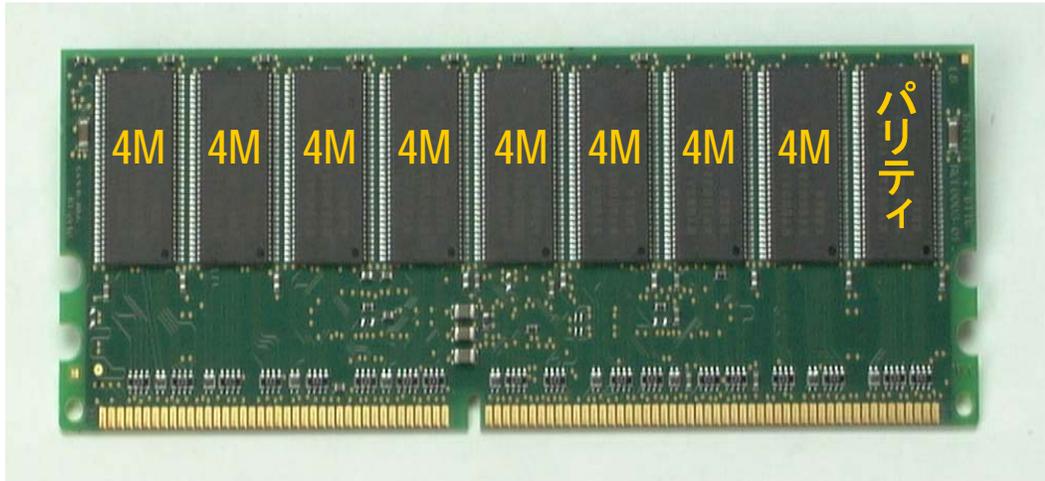
メモリコントローラは、すべての行のセルを1秒間に数千回読み取る

DRAM では、コンデンサは電荷を永久に保持することができません。放置した場合、コンデンサ内の電荷は、数ミリ秒以内に完全に放電されます。

コンデンサが帯電された状態を保てるように、メモリコントローラは各グリッド内の各行を1秒間に数千回読み取ります。行を読み取ることで、各コンデンサの電荷がセンスアンプに送られます。その後センスアンプは、受け取った電荷を増幅して適切なコンデンサに戻します。

メモリコントローラのジョブの1つに、読み取りおよび書き込みサイクル前後に実行するリフレッシュサイクルのスケジュール設定があります。メモリコントローラは、各行を読み取ってリフレッシュを実行するため、メモリマトリクスの行と列の数が同数であることはほとんどありません。通常、列よりも行が少なくなっています。行が少ないことは、リフレッシュ処理が短時間で終了することを意味します。

メモリチップサイズ



メモリチップはセルの数で表され、メモリモジュールはメモリチップのサイズと数で表される

メモリチップの数は、それに含まれているセルの数で表されます。たとえば、メモリチップの行と列がそれぞれ 2048 の場合、セルの数は 4,194,304 になります。このメモリチップは、4 メガビット (4Mb) のメモリチップと呼ばれます。131,072 列、32 行のチップも、4Mb チップになります。

4Mb は 4MB (4 メガバイト) とは異なります。8 ビットが 1 バイトであるため、4MB は 32Mb と等しくなります。

メモリモジュールは、メモリチップのサイズと数で表されます。メモリチップが 4Mb チップの場合、上図のモジュールは 4Mb x 8 モジュールになります。つまり各チップにはセルが約 400 万あり、チップが 8 個あります。

メモリモジュールによっては、512Mb のメモリチップが搭載されているものや、複数の Data In と Data Out ピンがあるものもあります。これは、各メモリチップが 1 ビットではなく 8 ビット提供できるようにするためです。チップの内部回路は、1 つのチップ上の 8 つの異なる部分が、それぞれが 1 ビットを提供できるように設計されています。たとえば、メモリチップが 512Mb チップで、8 つの Data In と Data Out ピンがある場合、上図のモジュールは、512Mb x 64 モジュールになります。この場合、データバスはすべてのデータを処理できるように 64 ビット幅でなければなりません。

CAS レイテンシ

CASレイテンシ(CL)は、システムクロックのサイクルタイム(t_{CLK})に対する t_{CAC} の割合を最も近い整数に丸めた値です。

$$CL \geq t_{CAC}/t_{CLK}$$

システムクロックの速度にかかわらず、列のプリチャージにはある一定の時間がかかります。

例

100MHzシステムで、 t_{CAC} が 20nsで t_{CLK} が 10nsの場合は、 $20\text{ns}/10\text{ns} = 2$ になるため、CLは 2 になります。

同じ DIMM を 133MHz システムで使用した場合、CAS レイテンシの値は大きくなります。

例

133MHz システムでは、 $20\text{ns}/7.5 = 2.66$ となるため、CL = 3 になります (最も近い整数に丸められます)。

100MHz データバスよりも 133MHz データバスの方が、データが到着するまでに多くのサイクル数が必要です。

SDRAM には、CL2 または CL3 として分類されているものがありますが、これは CAS レイテンシを表しています。また、3-2-2 などの x-y-z の形式で分類されている SDRAM も存在しますが、これは Fast Page Memory RAM の分類とは異なります。SDRAM の場合、最初の値は CAS レイテンシ、2 番目の値は RAS-to-CAS 遅延、3 番目の値はクロックサイクルで表した RAS プリチャージ時間です。

SDRAM は、メガヘルツでも分類されていて、これは動作可能なシステムバス速度を表しています。たとえば、PC100 DIMM は、100MHz データバスで最高の性能を発揮し、PC133 DIMM は 133MHz データバスで最高の性能を発揮します。

第 4 章のトピック

AGP

Pentium II 以降のシステムボード用に開発された AGP (Accelerated Graphics Port) により、グラフィックコントローラは、PCI バスではなく、特別なポイントツーポイントバスに配置されるようになりました。AGP では、システムボード上に AGP グラフィックカード用のスロットがあります。

AGP インタフェース仕様は、PCI 仕様をベースにしていますが、PCI ビデオ帯域幅の制限に対処できるように開発されています。AGP チャネルは、32 ビットで構成されており、66MHz で動作します。AGP は、バス速度を 2 倍にするだけでなく、特殊なシグナリングを採用する 2X モードを定義しており、同じクロック速度でポートから 2 倍のデータ量を送ることができます。すなわち、クロック信号の立ち上がりエッジと立ち下がりエッジの両方で情報を送信します。

AGP は、ビデオチップセットとシステムプロセッサの間で動作する専用の高速インタフェースを構成します。主な長所は以下のとおりです。

- ポートの実装が容易になります。
- 高速化が容易になります。
- ビデオ専用の高度な設計が可能になります。

AGP は、2 つのデバイス (プロセッサとビデオカード) のみに関係しており、拡張性もないので、バスでなくポートと見なされます。AGP のもう 1 つの利点は、ビデオサブシステムをコンピュータから分離することです。AGP スロットは PCI スロットを解放するだけでなく、PCI バスからビデオカードを削除するので、その他の PCI デバイスはその分の帯域幅を有効に利用できます。

AGP1X によって、グラフィックスを処理する PCI バスが解放されるとともに、データ転送速度が向上し、グラフィックスに使用できるメモリが増大するという効果があります。

バス	転送速度
PCI	133MB/s
AGP1X	266MB/s
AGP2X	533MB/s
AGP4X	1.07GB/s

サポートするオペレーティング システム

現在 AGP をサポートしているオペレーティングシステムは以下のとおりです。

- Windows 95 OEM Service Release (OSR) 2.1
- Windows 98
- Windows NT 4.0 Service Pack 3
- Windows 2000
- Windows Server 2003

AGP バス上のデバイスを 1 つずつ確認して初期化し、さらにメモリ マッピングを管理するコードの提供により、AGP は Windows 98、Windows 2000 および Windows Server 2003 で完全にサポートされています。アプリケーションでこの機能を使用するには、DirectDraw 5.0 API を使用する必要があります。

ダイレクト メモリ アクセス

ダイレクト メモリ アクセス (DMA) は、プロセッサをバイパスしたドライブ/メモリ間のデータ転送を可能にする方式です。DMA を使用しない場合、プロセッサはあるクロックサイクルでデバイスからデータを読み取り、次のクロックサイクルで他のデバイスに書き込みます。DMA は、1 つのクロックサイクルでデータ転送を管理し、これによりプロセッサの負荷が軽減されます。

マルチタスクのオペレーティングシステムでは、DMA によりプロセッサはディスク転送中も解放されたままになり、有用なタスクを実行できるようになります。

DMA 転送は、最初に設定する必要があり、これには DMA コントローラをプログラミングするためのプロセッサ時間が必要となります。データを小さなパケットで配信する場合は、DMA 方式によりプロセッサに大きな負荷がかかることがあります。それに対して、バスマスタはバスの解放だけを処理する必要があり、これはバスアービトラータによって実行されます。

DMA チップは、保留中のデータ転送を引き継ぐことができるようにプログラミングされる必要があります。転送が開始される前に、以下の情報が DMA チップに送信されます。

- ターゲットアドレス
- 転送されるデータの量
- 転送方法と速度

DMA のタイプは以下のとおりです。

- DMA
- バスマスタリング DMA

通常、DMA は、ユニットのシステムボード上の DMA コントローラを使用して複雑なアービトラーションタスク、システムバスの操作、データ転送を実行します。バスマスタリング DMA を使用すると、すべてのタスクがインタフェースコントローラ上のロジックによって実行されます。

第 6 章のトピック

ATA 規格

ATA の各バージョンは、それ以前のバージョンと下位互換性があります。新しいバージョンの ATA は、古いバージョンをベースに構築されています。いくつかの例外を除き、以前のバージョンの拡張とみなすことができます。

ATA 規格を以下に示します。

- ATA-1 — ANSI 文書 X3.221-1994
- ATA-2 — ANSI 文書 X3.279-1996
- ATA-3 — ANSI 文書 X3.298-1997
- ATA-4 (Ultra-ATA/33) — ANSI 文書 NCITS 317-1998
- ATA-5 (Ultra-ATA/66) — ANSI 文書 NCITS 340-2000
- ATA-6 (Ultra-ATA/100) — ANSI NCITS プロジェクト番号 1410
- ATA-7 (ATA/133) (Ultra-ATA/133) — ANSI NCITS プロジェクト番号 1532
- シリアル ATA (Ultra SATA/1500) — Serial ATA Final 1.0 Specification

ATA-1

ATA-1 では、オリジナルの ATA インタフェースが定義されています。これは ISA バスをベースにしたディスクドライブとホストシステム間の内蔵バスインタフェースです。

ATA-2

拡張 IDE (EIDE) は、ATA-2 の拡張で、ATA-2 規格と ATAPI 規格の両方をベースに構築されています。PIO (Programmed Input/Output) と DMA (ダイレクトメモリアクセス) モードなどの機能が含まれていますが、これらは基本的にパフォーマンス拡張機能としての位置づけです。

ATA-3

ATA-3 では、PIO モード 4 の信頼性が向上しています。ATA-3 では、単純なパスワードベースのセキュリティスキーム、より高度な電源管理機能、および S.M.A.R.T. (Self-Monitoring Analysis and Reporting Technology) を提供しています。ATA-3 は、ATA-2、ATAPI、ATA デバイスとも下位互換性があります。

ATA-4 (Ultra-ATA/33)

ATA-4 は、DMA データ転送を使用した最大 33MB/s のバースト転送速度を持つ高性能なバスマスタリングを提供します。Ultra-ATA の実装は、通常 Ultra-DMA/33 または UDMA/33 と呼ばれます。

Ultra-ATA ドライブシステムを実装するには、Ultra-ATA ドライブ、コントローラ、およびそれをサポートしている BIOS が必要です。

ATA-5 (Ultra-ATA/66)

Ultra-ATA 規格は、DMA データ転送を使用した最大 66MB/s のバースト転送速度を持つ高性能なバスマスタリングをサポートしています。Ultra-ATA の実装は、通常 Ultra-DMA/66 または UDMA/66 と呼ばれます。

Ultra-ATA/66 ドライブを使用するには、コントローラ、ケーブル、およびそれをサポートしている BIOS が必要です。

信号ケーブルには、Ultra-ATA/66 または Ultra-ATA/100 用に特別に設計された 40 ピン/80 線ケーブルを使用します。これは以前の ATA 規格と完全な下位互換性がありますが、標準の 40 ピンケーブルでは UDMA のパフォーマンスは提供しません。オペレーティングシステムが DMA に対応していて、DMA モードをアクティブにする必要があります。

ATA-6 (Ultra-ATA/100)

Ultra-ATA/100 ドライブ (Ultra-DMA/100 または UDMA/100 と呼ばれる) は、100MB/s のバースト転送速度に対応しています。当初 Intel、Quantum、Seagate によって開発され、その後設計変更が加えられています。

Ultra-ATA/100 は、既存の EIDE/UDMA ハードディスクドライブ、リムーバブルメディアドライブ、および CD-ROM/R/RW ドライブと下位互換性があります。検出メカニズムにより、ホストは 80 線ケーブルを検出して、Ultra-ATA/66/100 転送速度を有効にするかどうかを判断します。

Ultra-ATA/100 では、Ultra-ATA/66 と同じ 80 線ケーブルが必要です。Ultra-ATA/33 では 40 線ケーブルが必要です。

ATA-7 (Ultra-ATA/133)

Maxtor によって開発された Ultra-ATA/133 インタフェースにより、ATA インタフェース速度は 133MB/s に向上し、インタフェース速度と PCI バス転送速度が同じになりました。

- Ultra-ATA/33 のダブルエッジクロッキング技術と巡回冗長検査、および Ultra-ATA/66 で導入された 80 線ケーブルをベースにしています。
- すべてのパラレル ATA デバイスと下位互換性があります。現在 Ultra-ATA/100 用に使用されているものと同じ 80 線、40 ピンケーブルを使用できます。

ATA の動作モードと速度

ATA にはさまざまな規格があるため、ATA テクノロジーが動作するモードと速度も複数あります。ATA は以下のテクノロジーをサポートしています。

- PIO
- DMA
- Enhanced DMA
- Ultra DMA

PIO (Programmed I/O: プログラム I/O)

PIO は、データ転送に割り込み方式を使用します。この方式は低速ですが、単純であるためハードウェアコストは低く抑えることができます。

以下の順序でイベントが発生します。

1. プロセッサがデータ要求を行う。
2. ディスクドライブがデータを取得する。
3. デバイスはプロセッサへの割り込みを生成する。
4. プロセッサは、割り込みに応答し、データ転送が開始される。

PIO モードは、最大データ転送速度に基づいて分類されています。モードには 0~4 があり、値が大きいほど高速です。モードと速度は以下のとおりです。

- モード 0 — 3.3MB/s
- モード 1 — 5.2MB/s
- モード 2 — 8.3MB/s
- モード 3 — 11.1MB/s
- モード 4 — 16.6MB/s

DMA (Direct Memory Access: ダイレクトメモリアクセス) モード

Intel では、データ要求のバーストサイクルをサポートするように DMA 技術を拡張して、データ転送速度を向上させました。1996 年に、Intel と Quantum は、ATA ハードディスクドライブインタフェースを大幅に拡張し、PC のハードディスクドライブの以前の規格を実効値で 2 倍にするプロトコルを開発しました。

Enhanced DMA

Enhanced DMA は、IDE コントローラ上で DMA バスマスタリング転送を実現するために使用する方式です。Enhanced DMA の速度は PIO モード 4 の速度と同じですが、Enhanced DMA の利点は、バス上でのバースト転送をサポートしていることと、DMA であることによりプロセッサ使用率が減少することです。Enhanced DMA のモードと速度は以下のとおりです。

- Multi-word DMA モード 0 — 4.17MB/s
- Multi-word DMA モード 1 — 13.3MB/s
- Multi-word DMA モード 2 — 16.6MB/s

Ultra DMA

Ultra DMA ハードディスクドライブとシステムインタフェースは、すべての EIDE および ATA モードと完全に下位互換性があります。転送速度は、チェーン内の最も遅いリンクに調整されます。インタフェースは、マスタ/スレーブおよびケーブルセレクト構成の両方をサポートしています。Enhanced DMA と異なる点は、Ultra DMA 規格では、ATA バスに対する CRC (巡回冗長検査) が追加されていることと、転送速度が向上していることです。Ultra DMA は、新しい ATA-4 仕様に含まれています。

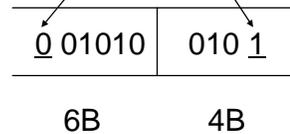
Ultra DMA は、BIOS 内および IDE コントローラによってサポートされています。Ultra DMA は、単にドライブとコントローラ間の異なるインタフェースであるため、特殊なドライバは存在しません。ソフトウェアは同じで、標準の DMA ドライバを使用することができます。

以前のユニットとは異なり、システム BIOS がドライブ速度を判断し、システムをセットアップします。以前 BIOS は、いくつかのドライバによって処理されていました。BIOS は、特定のデバイス構成での統合に関する問題を検出し、それに応じて IDE バス速度を調整します。必要に応じて、インタフェースを強制的に低速モードに設定することもできます。UDMA のモードと速度は以下のとおりです。

- モード 0 — 16.6MB/s
- モード 1 — 25MB/s
- モード 2 — 33.3MB/s

CRC(Cyclic Redundancy Check: 巡回冗長検査)

1と0のバランスをとるために
ビットが追加される



伝送キャラクタ x4 = 伝送ワード(40ビット)

CRCフィールドは、フレーム内のエラーを検出します。巡回冗長検査値と呼ばれる特殊な伝送ワード(4バイトフィールド)が計算され、データフィールドの後ろのフレーム/フィールドの終わりを示す区切り記号の前のフィールドに追加されます。値は、フレームヘッダとデータフィールドの整合性を検査します。

CRCフィールド値の計算に使用するアルゴリズムは、FDDI規格で使用されているものと同じです。

例

01010	010
5B	3B

8ビットワード

Ultra DMA の保守の問題点

Ultra DMA モードで、標準の IDE/EIDE 40 線ケーブルを使用する場合はケーブル長が重要になります。ドライブが 1 つの構成に、35cm を超えるケーブルを追加するとデータ整合性の問題が発生します。ユニットでは、35cm 未満の標準の 40 ピンケーブルが使用されていません。ハードディスクドライブが 1 つの構成では、Ultra DMA モード 2 がサポートされています。

標準の 40 ピンケーブルを使用して 2 番目のドライブを追加すると、BIOS は Ultra DMA をモード 1 に設定します。2 番目のドライブを追加する場合は、80 ピンケーブルを使用することをお勧めします。これにより、BIOS はこの構成を検出し、Ultra DMA をモード 2 に設定します。

注記

▲ 最適なパフォーマンスを得るためにも、ハードディスクドライブはプライマリコントローラに接続してください。また、CD-ROM ドライブ、テープドライブ、LS-120 などはセカンダリコントローラに接続してください。

ATA 変換

ATA は、いくつかのコンポーネントを使用してデータとアドレスを変換します。

Standard CHS (標準シリンダ/ヘッド/セクタ)

以前のコンピュータで使用されていた Standard CHS (標準シリンダ/ヘッド/セクタ) 方式では、ATA/IDE インタフェースとの互換性の問題により、1024 シリンダ、16 ヘッド、64 セクタに制限されていました。このインタフェースが、ハードディスクドライブを BIOS と Int13h インタフェース (BIOS とオペレーティングシステムをリンクする) にリンクします。このため、実際のドライブ容量が 528MB を超えていても、システムで使用可能なフォーマット済みドライブの最大容量は 528MB でした。

Extended CHS (拡張シリンダ/ヘッド/セクタ)

Extended CHS は、BIOS/Int13h インタフェースの CHS パラメータを ATA/IDE インタフェースの CHS パラメータに変換することで 528MB の制限を撤廃します。

例

ドライブが 1500 シリンダ、16 ヘッドの場合、変換 BIOS は、ドライブが 750 シリンダ、32 ヘッドであるようにプログラムに示します。

LBA (Logical Block Addressing: 論理ブロックアドレッシング)

LBA 方式の登場により、標準のアドレッシング方式はこれに変更されました。同じレジスタを使用して、シリンダ、ヘッド、およびセクタの情報が渡されますが、SCSI で使用されているアドレスなどの情報は LBA として完全に再定義されています。

LBA は、セクタアドレスに直線的にアドレッシングする方法で、LBA 0 であるシリンダ 0 のヘッド 0 のセクタ 1 から、ドライブ上の最後の物理セクタ (標準の 540MB ドライブの場合は、LBA 1,065,456) までアドレッシングします。

LBA により、プロセッサのオーバーヘッドが減少します。これは、通常はドライブを呼び出す前に BIOS が CHS アドレッシングに変換する必要があるオペレーティングシステムの要求を、LBA 呼び出しで行うためです。LBA により、この変換処理が不要になります。

ドライブ 0 とドライブ 1 の関係

1 つのポートに 2 つのハードディスクドライブが接続されている場合は、階層関係が存在しません。

- ドライブ 0 - プライマリドライブ
- ドライブ 1 - セカンダリドライブ

両方のドライブが同時に動作することはできません。

ケーブルセレクト

現在、ほとんどの ATA 実装ではケーブルセレクトテクノロジーが採用されています。これは、IDE ハードディスクドライブをプライマリ/セカンダリデバイスとして指定する機能です。この機能を使用するには、ピン 28 をケーブルセレクトに使用するケーブルセレクト用の ATA/IDE ケーブルを使用し、最初のドライブにドライブ 0 として接続する必要があります。そして、各ドライブのケーブルセレクトジャンパをケーブルセレクトモードに設定します。新しい ATA ハードディスクドライブのジャンパは、ケーブルセレクトモードにプリセットされています。

注記

▲ ドライブを 1 つだけインストールする場合は、リボンケーブルの終端のドライブコネクタに接続する必要があります。

ケーブルセレクト対応のハードディスクドライブと ATA ケーブルは、下位互換性があります。ケーブルセレクトに対応していないコンピュータでケーブルセレクトモードのハードディスクドライブを使用する場合は、ドライブのジャンパをドライブセレクトモードに切り替える必要があります。

ドライブセレクト

ドライブセレクトテクノロジーは、ケーブル上の位置に従ってではなく、ハードディスクドライブ上のジャンパ設定によって、ATA ハードディスクドライブがマスタまたはスレーブであるかを指定します。1 つのドライブがインストールされている状態でスレーブドライブをインストールする場合、ドライブによっては既存のドライブのジャンパをシングルからマスタに変更しなければならないことがあります。シングルとマスタのジャンパ設定が同じであるドライブもあります。

第 1 章 – サーバテクノロジー

1. 汎用サーバはどのような目的で使用されますか。(p. 1 – 2)
汎用サーバは、中小企業および大企業で、部門内での使用やクラスターリングなど、さまざまな目的で使用されます。
2. 各用語とその説明を対応付けてください。(p. 1 – 6)
a..... MLサーバ — システム内部の拡張性が最大化されたサーバ
b..... BLサーバ — 消費電力を抑えた、モジュール式の超高密度ブレードサーバ
c..... DLサーバ — ラックマウント環境用に集積度が最適化されたサーバ
3. すべてのサーバコンポーネント間の通信とデータ交換を制御し、サーバの全般的な機能を決定するのは、サーバサブシステムのどの部分ですか。(p. 1 – 16)
チップセット
4. 高速なサブシステム(プロセッサとメモリ)の効率的な動作を可能にするのは、システム内の次のどのコンポーネントですか(4 つ選択してください)。(p. 1 – 26)
a..... コントローラ
c..... メモリバッファ
d..... パラレルバス
f..... 並列読み取り/書き込みプロセス
5. サーバ性能を向上させる 3 つの方法を挙げてください。(p. 1 – 29)
速度を上げる
並列処理を行う
レイテンシを減らす
6. システムが使用する 3 つのクロックを挙げてください。(p. 1 – 27)
システムクロック
バスクロック
リアルタイムクロック
7. HP サーバの設計に影響を及ぼす 3 つの要因を挙げてください。(p. 1 – 2)
変化する顧客のニーズ
新しいテクノロジー
新しい標準

第 2 章- プロセッサとマルチプロセッシング

1. 命令とは何ですか。(p. 2-3)
コンピュータプログラムからプロセッサに渡される指示(加算、減算、比較など)
2. プロセッサの各コンポーネントとその機能を対応付けてください。(p. 2-4~5)
 - a.プリフェッチユニット - プロセッサがその後必要とする命令やオペランドを置いておく場所
 - b.....デコードユニット - 命令をその構成要素に分解するコンポーネント
 - c.....実行ユニット - 加算や減算など、実際のデータ処理を行うコンポーネント
 - d.....制御ユニット - 実行ユニットのスケジューラとして機能するコンポーネント
 - e.レジスタ - 制御ユニットと実行ユニットによって、一時的にデータを格納するために使用される少数のメモ
 - f.....L1 キャッシュ
 - g.分岐ターゲットバッファ - 分岐予測を行うために、最近辿った分岐を格納するレジスタ
 - h.バスインタフェースユニット - アドレスバスおよびデータバスへのアクセスを制御するコンポーネント
3. プロセッサによる入力の処理方法を示す次の手順を、正しい順序に並べ直してください。(p. 2-6)
 - c.....命令をフェッチする
 - e.....命令をデコードする
 - a.....命令を実行する
 - d.....データを転送する
 - b.....データを書き込む
4. 各テクノロジーとその説明を対応付けてください。
 - a.....パイプライン - 1つの命令の完了を待たずに別の命令を開始するプロセッサ (p. 2-7)
 - c.ハイパーパイプライン - 1つの命令を実行するために使用するステップの数を増やしたプロセッサ (p. 2-
 - b.....スーパースケーラ - 1つのクロックサイクルで複数の命令を実行できるプロセッサ (p. 2-8)
 - f.EPIC - コンパイラが実行ユニットに対して、どの命令を並列に処理できるかを指示するプロセッサ (p. 2-
 - d.分岐予測 - ある分岐命令を初めて実行する際に、その命令のアドレスと正しい分岐のアドレスを分岐ター
 - e.....順序外実行 - 別の命令に依存しない命令を先に処理できるプロセッサテクノロジー (p. 2-11)

5. 次の図に示されているのは、どのタイプのパッケージですか。(p. 2 – 17)
c..... FC-PGA
6. 次の図に示されているのは、どのタイプのパッケージですか。(p. 2 – 17)
c..... FC-PGA2
7. 回路線の幅に関する次の説明のうち、正しいものはどれですか(該当するものをすべて選択してください)。(p. 2 – 19)
a.....回路線が細いほど、1枚のウェーハにエッチングできるチップの数が多くなる
b.....回路線が細いほど、プロセッサの発熱量が少なくなる
c.....回路線が細いほど、プロセッサの速度が向上する
d.....回路線が細いほど、プロセッサの動作に必要な電力が少なくなる
8. マルチプロセッサ サーバとは何ですか。(p. 2 – 20)
複数のプロセッサを搭載したサーバ
9. 各テクノロジーとその説明を対応付けてください。
a.アシンメトリック(非対称型)マルチプロセッシング - 特定のタスクを特定のプロセッサに割り当てる(p. 2 – 21)
b.シンメトリック(対称型)マルチプロセッシング - 次のタスクを、次に使用可能なプロセッサで実行する(p. 2 – 22)
c.疎結合 - 各プロセッサにメモリが割り当てられ、独立したコンピュータのように機能する(p. 2 – 23)
d.....密結合 - すべてのプロセッサがメモリを共有する(p. 2 – 22)
10. プロセッサを混用する場合、各プロセッサのコア周波数は最も速いプロセッサの周波数に合わせて設定する必要があります。(p. 2 – 28)
誤
11. キャッシュサイズの異なる Xeon プロセッサを混用する場合、同じキャッシュサイズを持つプロセッサ同士をペアでインストールする必要があります。(p. 2 – 28)
正
12. 最も高いステッピング番号を持つプロセッサを、ブートストラップ プロセッサとしてインストールする必要があります。(p. 2 – 28)
誤

第 3 章 – メモリとキャッシュ

1. データを格納するためにコンデンサを使用するのはどのテクノロジーですか。(p. 3 – 3)

DRAM

2. どちらのテクノロジーの方が高速ですか。(p. 3 – 4)

SRAM

3. キャッシュで使用されているのはどのテクノロジーですか。(p. 3 – 4)

SRAM

4. データをグリッド状に格納するのはどのテクノロジーですか。(p. 3 – 6, 3 – 25)

両方

5. DIMM で使用されているのはどのテクノロジーですか。(p. 3 – 7)

DRAM

6. セル、CAS、メモリコントローラ、RAS の各用語を使って、次の文を完成させてください。(p. 3 – 9)

データを読み書きするために、メモリコントローラがアドレスピン経由で行アドレスをRAS に送信し、列アドレスをCASに送信します。データはアドレスの交点にあるセルに送られます。

7. 各 DRAM テクノロジーとその説明を対応付けてください。(p. 3 – 12)

a.FPM — RASが有効化されると(電圧が下げられると)、メモリコントローラは、順に電圧を下げることに

b.EDO — メモリコントローラは、データがデータアウトピン上にある間に次の列を有効化することにより、レ

c.SDRAM — このDRAMテクノロジーはメモリバスと同じクロックを使用するため、特別なタイミングメカニズム

8. DDR RAM で初めて導入された革新的技術とは何ですか。(p. 3 – 13)

a..各クロックサイクルの立ち上がりエッジと立ち下がりエッジの両方でデータを転送する

9. 1 回のメモリアクセスで取得できるデータ量を 64 ビットから 128 ビットへと 2 倍に増やすメモリテクノロジーは、次のどれですか。(p. 3 – 14)

d.....デュアルインタリーブ メモリ

10. α 粒子が DIMM に当たると、メモリセル内で電気障害が発生します。これはハードエラーとソフトエラーのどちらですか。(p. 3 – 15)

ソフトエラー

11. 欠陥のあるデバイスを修理しないと再発するのは、どのタイプのメモリエラーですか。(p. 3 – 15)

ハードエラー

12. 各フォールトトレランス テクノロジとその説明を対応付けてください。
- a.パリティ — メモリコントローラがバイトをメモリに書き込む際に、バイト内の 1 の数に基づいて各バイトに 1
 - b.ECC — このテクノロジはチェックサムを使用してエラーを分析し、どのバイトが壊れているかを判断し、修
 - c.アドバンスドECC — このテクノロジは、1 つのDRAMチップ内で発生したマルチビットエラーを訂正します
 - d.オンラインスペア メモリ — 問題のあるDIMMを含むメモリバンクが自動的にスペア (予備) のDIMM(バンク)に
 - e.ホットプラグ ミラーメモリ — メモリコントローラが、2 つのメモリボード上の同様に構成されたDIMM(バンク)に
 - f.ホットプラグRAIDメモリ — 4 つのメモリコントローラがそれぞれ 4 つのDIMMの内の 1 つに 1 ブロックのラ
13. サーバのキャッシュの利点とは何ですか。(p. 3 – 25)
- a.....メモリよりも速く、プロセッサからのデータ要求を満たす
14. L1 キャッシュは、L2 キャッシュや L3 キャッシュよりも多くのデータを格納できます。
(p. 3 – 26)
- 誤
15. L2 キャッシュをプロセッサに接続するのはどのバスですか。(p. 3 – 26)
- b.....バックサイドバス
16. マザーボード上のキャッシュは、プロセッサの最高速度と同じ速度で動作します。
(p. 3 – 26)
- 誤
17. キャッシュ内の各行を何と言いますか。(p. 3 – 27)
- キャッシュライン
18. Tag RAM の機能とは何ですか。(p. 3 – 27)
- 各キャッシュライン内のデータに対するメモリアドレスを格納する

19. 各キャッシュ実装形式をその定義と対応付けてください。
- a. ルックアサイド — キャッシュとメモリの両方がメモリ要求を受け取ります。キャッシュヒットがあれば、キャッシュ
 - b. ルックスルー — キャッシュヒットがあった場合、要求はシステムバスに送られません。(p. 3 – 30)
 - c. フルアソシアティブ — メインメモリからのデータをどのキャッシュラインにも格納できます。(p. 3 – 33)
 - d. ダイレクトマッピング — 各キャッシュラインにメモリアドレスのグループが割り当てられます。(p. 3 – 33)
 - e. セットアソシアティブ — キャッシュラインの特定のグループにメモリアドレスのグループが割り当てられます。
 - f. ライトスルー — データを再度使用する前に、システムはすべてのメモリレベルにデータを書き込む必要が
 - g. ライトバック — キャッシュラインに添付されたビットに、データがまだメモリに書き込まれていないことを示
 - h. バスヌーピング — キャッシュコントローラが、バスマスタによるメモリ要求がないか、システムバスのトラフ
 - i. バスナフティング — バスマスタがメモリにデータを書き込む際に、キャッシュコントローラが書き込まれよ

第 4 章 – バスアーキテクチャ

1. バス幅とは何ですか。(p. 4-8)

b.....バス内の電気回線の数

2. バスの最大転送レートを計算するには、どの式を使いますか。(p. 4-10)

a.....(速度 x 幅)/バスサイクル毎クロックサイクル

3. システムパフォーマンスを向上させるための 4 つの基本的な対策を教えてください。
(p. 4-11)

バスサイクルの速度とデバイス(特にシステムメモリ)がプロセッサと通信する速度を高速化する。

データバスの幅を拡張する。

バーストサイクルなどのバスサイクルを変更する。

デュアル インディペンデントバスまたはマルチプロセッシングなどの並行プロセスを追加する。

4. バスマスタは、バス上の他のデバイスと直接通信するデバイスです。(p. 4-16)

正

5. PCIバスは32ビットまたは64ビットのバスで、一度に4または8バイトのデータを転送します。その時のクロックレートは33MHzまたは66MHzで、最大バーストスループットの範囲は、133~533MB/sです。(p. 4-19)

6. ブリッジ PCI バスアーキテクチャのシステムパフォーマンスは、PCI バスが追加された分だけ向上します。(p. 4-23)

誤

7. PCI-X バスセグメントは 66MHz で動作する 4 つの拡張スロットをサポートできます。(p. 4-30)

正

8. PCI-X はクロックサイクル全体でデコードロジックを発生させ、タイミング上の制約を緩和します。(p. 4-35)

正

9. 各用語とその説明を対応付けてください。

a.....PCI Express — パケット化プロトコルとロード/保存アーキテクチャを定義する
(p. 4-45)

b.InfiniBand — 高帯域幅デバイスがスイッチファブリックアーキテクチャを使い、サーバをリモートストレージ

c.USB — システムを再起動または再構成することなく、周辺デバイスのホットプラグを可能にする(p. 4-4)

第 5 章 – システムアーキテクチャ

1. ハイリーパラレル システムアーキテクチャで、性能を高めるために使用されるコンポーネントは何ですか。(p. 5 – 8)
並列の 2 つの PCI バスと 2 つのメモリコントローラ
2. バスマスタリングから生じたボトルネックを、HP はどのように解消しましたか。(p. 5 – 6)
トライバスアービトラータ
3. デュアルメモリコントローラ的设计を使用すると、メモリ帯域幅が最高 1.6GB/s になります。(p. 5 – 14)
正
4. ProFusion チップセットには、メモリバスがいくつありますか。(p. 5 – 14)
2
5. ProFusion クロスバースイッチには、どのようなコンポーネントが結合されていますか。(p. 5 – 15)
独立した 2 つのプロセッサバス
キャッシュコヒーレンシフィルタ
独立した 1 つの I/O バス
2 つのメモリポート
6. F8 チップセットを使用するとき、バス速度を表すための尺度は何ですか。(p. 5 – 17)
MT/s – メガトランザクション/秒

第 6 章 – ディスクストレージ

1. 5つの SCSI 標準とは、SCSI-0、SCSI-2、SCSI-5、SCSI-6、SCSI-7 です。(p. 6 – 11)
誤
2. シングルエンド SCSI とディファレンシャル SCSI インタフェースの主な違いは何ですか。(p. 6 – 37, 38)
シングルエンドデバイスのケーブル長が 6m以下であるのに比べ、ディファレンシャル SCSIは、ノイズ耐性が高いので、25mという長いSCSIケーブルを使うことができる。
3. パッシブターミネーションとアクティブターミネーションの主な違いは何ですか。(p. 6 – 41)
パッシブターミネーション – 古く、信頼性が低い
アクティブターミネーション – 高速SCSIバスの最低必要条件
4. Wide Ultra SCSI-3 の最大バースト転送レートは? (p. 6 – 33)
40MB/s
5. Wide Ultra2 SCSI-3 の最大バースト転送レートは? (p. 6 – 33)
80MB/s
6. 初期化フェーズで、コントローラが接続先の各デバイスと転送速度について合意するプロセスを表すのは、以下のどの用語ですか。(p. 6 – 30)
c.....ネゴシエーション
7. タグ付きコマンドキューイング (TCQ) の目的は何ですか。(p. 6 – 31)
TCQを使用すると、ドライブが多くのコマンドを受け取り、SCSIバスを使わずにそれらのコマンドを実行できる
8. ターミネータは、SCSI バスの両端に必要です。(p. 6 – 41)
正
9. Wide Ultra2 ハードドライブを Fast-Wide SCSI コントローラに接続できますか。(p. 6 – 33)
できる
10. Wide Ultra3 SCSI-3 の 3つの新機能は何ですか。(p. 6 – 16)
ダブルトランジションクロッキング
CRC(巡回冗長検査)
ドメインバリデーション

第 7 章 – アレイ テクノロジ

1. 非アレイストレージシステムと比べた場合、RAID ストレージの 4 つの主な利点は何ですか。(p. 7-2)
 - パフォーマンスの向上
 - データの可用性と信頼性
 - ストレージ容量の動的拡張
 - ストレージ管理が容易
2. RAID 実装の 2 つの主なタイプは何ですか。(p. 7-6)
 - ソフトウェア
 - ハードウェア
3. RAID レベルの中で、読み取り、書き込み操作が最も高速なのはどれですか。(p. 7-11)
 - RAID 0
4. POST の実行中、Smart アレイコントローラに接続されたすべてのハードドライブについて SCSI ID が表示されます。(p. 7-37)
 - 誤
5. ホットプラグ可能なハードドライブでサポートされるのはどの機能ですか。(p. 7-53)
 - a..... フォールトトレランス アレイでの障害ドライブの交換
 - b..... ドライブとアレイの追加
 - c..... アレイの拡張
6. Smart アレイ 532 の特徴に当てはまるのはどれですか。(p. 7-48)
 - c..... オンラインでの RAID レベルの移行のサポート
7. 次のアレイコントローラのうち、SCSI チャンネルが 3 つ以上あるのはどれですか。(p. 7-31, 32)
 - c..... Smart アレイ 5304
8. RAID 1 は、完全なフォールトトレランスを提供する安価なソリューションだと考えられています。(p. 7-14)
 - 誤
9. オンボードバッテリーバックアップキャッシュを搭載した Smart アレイ 5300 コントローラでは、サーバ障害の場合にデータが保護されます。このバッテリーによってデータが保持される時間はどれくらいですか。(p. 7-55)
 - 4 日間
10. RAID ADG は、同時に何個のドライブ障害までデータを失わずにサポートできますか。(p. 7-23)
 - 2 個

第 8 章 – ファイバチャネル

1. ファイバチャネルの 5 つの利点を挙げてください。(p. 8 – 3)
 - パフォーマンス
 - 可用性
 - スケーラビリティ
 - 柔軟性
 - ノード間の距離
2. 各用語とその説明を対応付けてください。(p. 8 – 5)
 - a.....ノード — ホストやアダプタなどのデバイス
 - b..... ノードID — デバイスに割り当てられている一意の識別子
 - c.....リンク — ファイバチャネルポートの 1 対の光ファイバ
 - d.....トポロジ — 2 つ以上のポートを接続する相互接続構造
3. ファイバチャネルは、業界標準の相互接続と高性能シリアル I/O のプロトコルであり、サーバ業界において、これまでにない水準の信頼性、スループット、距離についての柔軟性を実現します。(p. 8 – 2)
 - 正
4. 次のうち、ストレージシステムソリューションで使われる 3 つのファイバチャネルトポロジはどれですか。(p. 8 – 30)
 - d.....ポイントツーポイント、ファブリックアービトラレーテッドループ、ストレージファブリック
5. GBIC (Gigabit Interface Converter) とは何ですか。(p. 8 – 8)
 - 電気インパルスと光ファイバ媒体で使われる光信号の間でトランザクションを実行します。GBICには、光信号を発するデバイスがあり、それが光ファイバケーブルで伝送されます。
6. ファイバチャネルハブを使うのは、どのファイバチャネルトポロジですか。(p. 8 – 14)
 - FC-AL
7. ファイバチャネルスイッチの E_port にはどのような機能がありますか。(p. 8 – 30)
 - E_portは、スイッチ間拡張ポートであり、もう 1 つのスイッチのE_portに接続してファブリックを拡張する場合に使います。
8. 2 種類のファイバチャネルケーブルは何ですか。(p. 8 – 21)
 - 銅ファイバ
 - 光ファイバ
9. 次のうち、減衰についての最も適切な記述はどれですか。(p. 8 – 28)
 - b..... 信号を伝送する間に生じるパワーの損失
10. ファイバチャネル環境で、スイッチのネットワークは何という名前と呼ばれますか。(p. 8 – 33)
 - ファブリック

第 9 章 – 高度なストレージテクノロジー

1. テープストレージバックアップテクノロジーの主な目的を挙げてください。(p. 9 – 29)

長期間にわたるデータの保管と破滅的な障害からの復旧のために、セカンダリ媒体上にデータをアーカイブし、ドライブ障害、ファイル削除、自然災害、盗難などから保護すること。

2. データ損失の原因となるのは、次のどれですか。(p. 9 – 30)

d..... 上記のすべて

3. HP ストレージ製品が使用する 5 つのテープテクノロジーを挙げてください。
(p. 9 – 41)

DAT (Digital Audio Tape)

AIT (Advanced Intelligent Tape)

DLT (Digital Linear Tape)

SDLT (Super Digital Linear Tape)

ultirum

4. エンタープライズ バックアップ ソリューション (EBS) は、アプリケーションソフトウェアと業界標準のハードウェアを統合して、さまざまな環境のデータのバックアップとリストア専用の完全なソリューションを提供します。(p. 9 – 46)

正

5. 各用語とその説明を対応付けてください。(p. 9 – 2)

a.DAS — 複数のクライアント用の専用ストレージがあり、サーバとストレージの関係は 1 対 1 である

b.....NAS — 既存の LAN に直接接続される独立したインテリジェントデバイス

c.SAN — オープンなオペレーティングシステム上で顧客のアプリケーションを実行するオープンシステム

6. Enterprise Network Storage Architecture (ENSA) の目的を挙げてください。
(p. 9 – 17)

ENSA は、ストレージをエンタープライズユーティリティサービスに変換するパスを定義します。

7. コピーおよびバックアップされる情報の量を表す 4 つの語句を挙げてください。
(p. 9 – 34)

フル

部分

増分

差分

8. テープ媒体の実際の平均寿命を判断する要因を挙げてください。(p. 9 – 38)
 - ヘッドとテープパスの設計
 - ドライブが1つの完全バックアップで必要とするパス数またはトラック数
 - 湿度と周囲温度
 - テープ先頭のヘッダーを読み取りおよび更新する際の追加のパス数
 - テープストリーミング
9. 災害復旧計画の4つの要素を挙げてください。(p. 9 – 47)
 - バックアップテープローテーション方式
 - ハードウェア保護計画
 - 災害期間中にビジネス運用を維持するための計画
 - データをオフサイトで保管するための計画
10. スナップショットとクローン化は、運用データをオフラインバックアップ用にコピーするために使用します。(p. 9 – 44)
 - 正
11. HPの標準SANトポロジで使用されている4つの設計を挙げてください。(p. 9 – 12)
 - カスケード
 - メッシュ
 - リング
 - バックボーンSANファブリック
12. VersaStorテクノロジーがサポートしている3つのタイプのストレージバーチャライゼーションを挙げてください。(p. 9 – 20)
 - OpenView Storage Virtual Replicator
 - Enterprise Virtual Array (EVA)
 - CASA (OpenView continuous access storage appliance)
13. 管理者が設定した属性に基づいて容量のプールから仮想ディスクを自動的に作成するテクノロジーを挙げてください。(p. 9 – 26)
 - ENSAextended
14. オフサイトに保管されているデイリー増分バックアップ、ウィークリーフルバックアップ、およびマンスリーフルバックアップを使用するテープローテーション方式の名前を挙げてください。(p. 9 – 36)
 - Grandfather-Father-Son (GFS)

第 10 章 – 高可用性

1. 現在のトランザクションを再起動しなければならなかったり、パフォーマンスが低下するおそれがあるが、ユーザのオンライン状態を維持することができる可用性レベルを上げてください。(p. 10 – 4)

AL3

2. クリティカルアプリケーションの 2 つのクラスの名前を上げてください。(p. 10 – 3)

ミッションクリティカル

ビジネスクリティカル

3. 3 つの基本的なクラスタモデルを上げてください。(p. 10 – 20)

シェアードナッシング (Shared Nothing)

シェアードディスク (Shared Disk)

シェアードエブリッシング (Shared Everything)

4. フルスペクトラム障害管理に含まれる 3 つのコンポーネントを上げてください。(p. 10 – 9)

障害予防

フォールトトレランス

ラピッドリカバリ

5. ProLiant クラスタがサポートしている 3 つのプラットフォーム名を上げてください。(p. 10 – 22)

次のいずれか 3 つ: Windows 2000 Cluster Service、LifeKeeper for Linux、Oracle Parallel Server for Oracle 8i、Oracle 9i Real Application Clusters

6. クラスタ化の利点を上げてください。(p. 10 – 15)

リソースの高可用性

ビジネスの成長に対応したスケーラビリティ

集中管理

負荷分散

クラスタ対応アプリケーション

第 11 章 – ProLiant サーバのユーティリティ

1. ProLiant Essentials の構成要素の名前を挙げてください。(p. 11 – 2)
Foundation パック
Rapid Deployment パック
内蔵 Lights-Out Advanced パック
Workload Management パック
2. マネジメントエージェントの機能を挙げてください。(p. 11 – 40)
デバイス上で動作し、パラメータを収集および測定して、デバイスのステータスを詳細にモニタリングします。
3. リモート Insight Lights-Out テクノロジーの 4 つの機能を挙げてください。(p. 11 – 45)
次のうちいずれか 4 つ
ハードウェアベースのグラフィカルリモートコンソール
サーバ障害アラート
リセットと障害シーケンスの再生
リモートリセット
Insight マネージャ 7 との統合
iPAQ との統合
RILOE ボード
内蔵 Lights-Out (iLO)

4. **Insight マネージャ 7** の 4 つの機能を挙げてください。(p. 11 – 38~40)
次のうちいずれか 4 つ
自動検出と識別
デスクトップインストールのサポート
言語サポート
セキュリティ
シングルログイン
Insight マネージャ 7 ホームページ
クエリーとタスク
複数のシステムのバージョン管理とシステムソフトウェアの更新
グループ構成
電子メールおよびポケットベルによる通知
Cluster Monitor
マネジメントエージェントの統合
リモートInsightの統合
5. バージョン管理の基準となるソフトウェアのセットは、**Insight マネージャ 7** でどのように定められていますか?(p. 11 – 39)
ソフトウェアメンテナンスアーキテクチャにより、管理対象環境の一部または全体にわたるバージョン管理の基準となるソフトウェアのセット(BIOS、ドライバ、エージェント)を選択できます。
6. **Management 6.x CD** に含まれているコンポーネントを挙げてください。(p. 11 – 35)
Insight マネージャ 7
マネジメントエージェント
ActiveUpdate
Surveyユーティリティ
マネジメントツールキット
7. **RILOE** が管理 PC にアラート通知を配信する方法を挙げてください。(p. 11 – 44)
管理PCへのSNMPトラップの配信を通じてアラート通知をサポートしています。
8. **RBSU (ROM ベースセットアップユーティリティ)** はディスクを使用します。(p. 11 – 20)
誤

第 12 章 – ラック/電源テクノロジー

1. ラック内の通気性を向上させる製品を挙げてください。(p. 12 – 4)
ルーフマウント型ファンキット
2. UPS をリモートからシャットダウンするのに必要なソフトウェアを挙げてください。(p. 12 – 23)
パワーマネジメントソフトウェア
3. HP UPS でカチッという音が繰り返し鳴っている場合の対処方法を挙げてください。(p. 12 – 14)
UPS のユーザガイドに従って自己診断テストを実行する。
4. 各用語とその説明を対応付けてください。
 - a. 水平脚 — ラックの全重量を支える (p. 12 – 7)
 - b. ラック接続キット — 複数のラック構成で必要 (p. 12 – 7)
 - c. ケーブルマネジメントアーム — 誤ってケーブルが外れるのを防止する (p. 12 – 22)
 - d. ラック固定脚 — スタンドアロンラックに必要 (p. 12 – 7)
 - e. rack builder online — ラック構成ツール (p. 12 – 5)

第 13 章 – サポートサービス

1. HP で使用されているパーツ番号の 3 つのタイプを挙げてください。(p. 13 – 7)
アセンブリパーツ番号
スペアパーツ番号
オプションパーツ番号
2. HP が提供している保証プログラムを挙げてください。(p. 13 – 14)
ProLiantサーバ
事前予防保証