IBM TotalStorage ®

IP技術者のための ストレージネットワーク基礎知識入門と 最新動向解説

2005年12月

佐野正和　　日本アイ･ビー･エム株式会社
sanomasa@jp.ibm.com

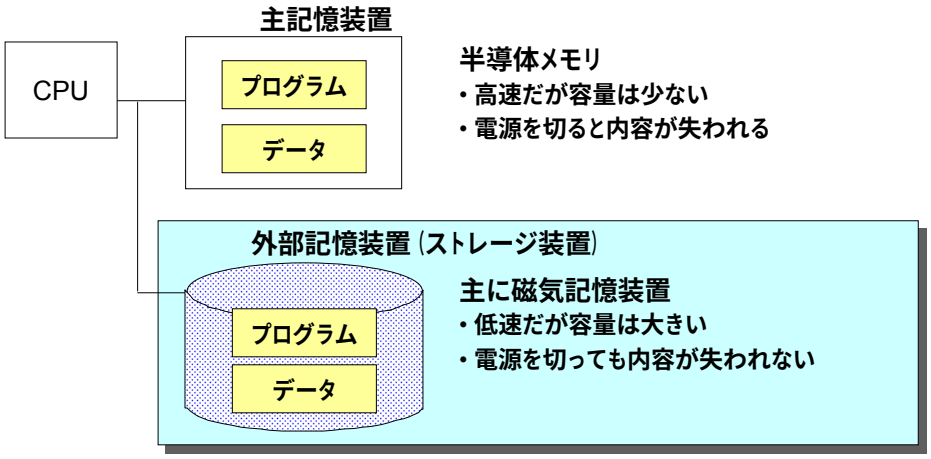
目次

第一章

基礎的なことを
そこそこ知っておきましょう

ストレージとは

コンピュータ・システムの構成



半導体メモリ
- 高速だが容量は少ない
- 電源を切ると内容が失われる

主に磁気記憶装置
- 低速だが容量は大きい
- 電源を切っても内容が失われない

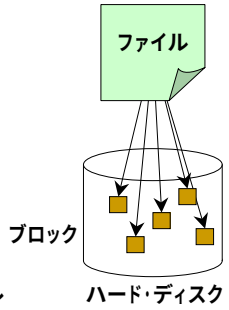
磁気記憶装置

- 特長
  - 不揮発性
    - 磁気を利用して情報を保持するので永続性がある
  - 読み書き可能
    - 外部から磁界を与える（電磁石）ことで保持する内容を変更
    - 保持している内容（磁界）を電気信号として取り出す

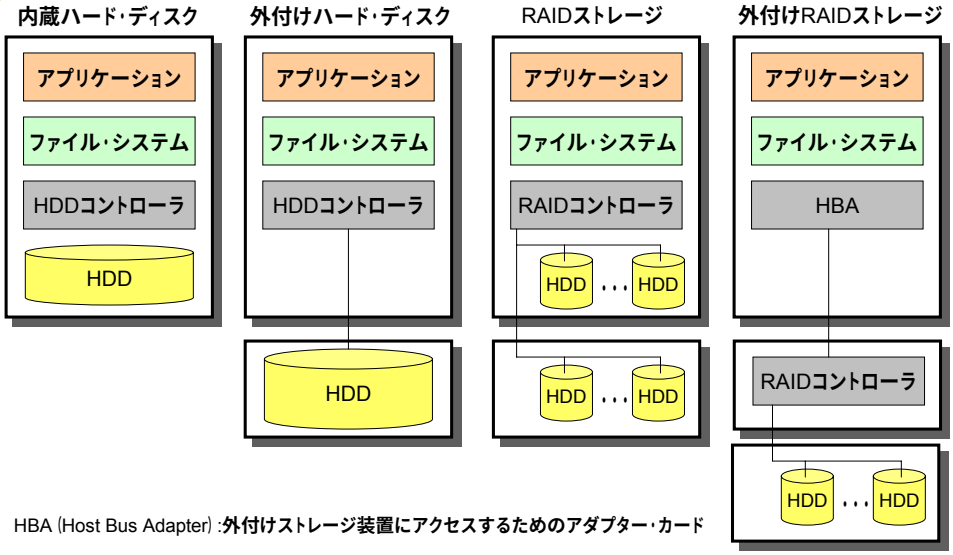
- アクセス方式
  - シーケンシャル（順次）・アクセス・メディア
    - 磁気テープ
    - データがテープの先頭から末尾まで一方向に格納
  - ランダム・アクセス・メディア
    - FDDやHDD
    - ディスク表面が2次元であることを生かしてデータを格納
      - トラック、セクター
      - 複数のディスク（プラッタ）を用いることがほとんどなので実際は3次元に配置される

ディスク装置とファイル・システム

- ディスク装置への入出力
  - ディスク記憶装置はクラスタと呼ばれる領域で管理される
    - クラスタの指定方法はディスクインタフェースなどに依存する
  - 指定したクラスタへのデータの書き込み
  - 指定したクラスタからのデータの読み出し
- ファイル・システムへの入出力
  - ファイル・システムはOSなどに依存し様々な方式が存在
  - ファイルは複数のクラスタ（ブロック、セクター）から構成される
  - ファイル名やクラスタ構成を保持するアロケーション・テーブル（i-node）を管理する



サーバーとHDDの分離

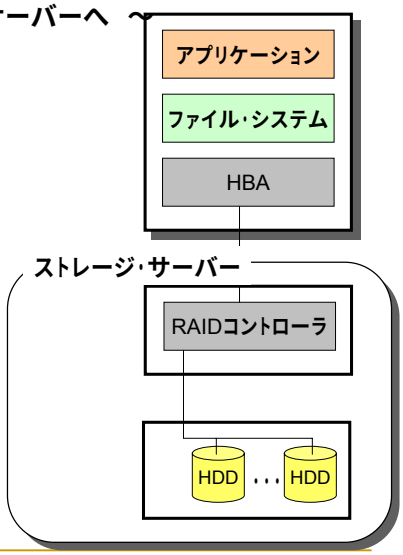


HBA（Host Bus Adapter）：外付けストレージ装置にアクセスするためのアダプター・カード

外付けストレージのインテリジェンス化

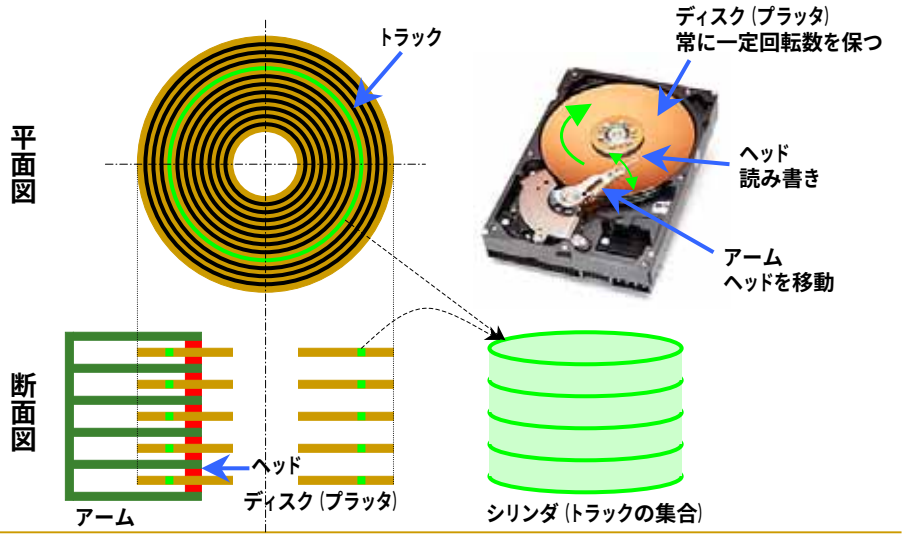
～ ストレージ・デバイスからストレージ・サーバーへ ～

- 高性能化
  - 高度なRAID制御
  - 高性能コントローラ
  - ハードウェアパリティ計算
  - 大容量キャッシュメモリ搭載
- 高可用性
  - コントローラの2重化
  - キャッシュの保護
    - バッテリ・バックアップ
    - キャッシュのミラーリング
- 接続性
  - サーバーとのさまざまな接続形態サポート
    - SCSI、ファイバーチャネル、など
  - 複数サーバーへの並列アクセス
  - サーバーアクセスの制限
    - LUNマスキング
- 付加機能
  - サーバーとは独立に各種処理ができる
    - 高速コピー
    - ミラーリング
    - 遠隔コピー



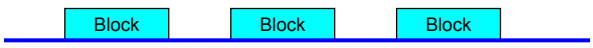
HDDの基本構造とインタフェース

HDD (Hard Disk Drive) ハードウェア構造

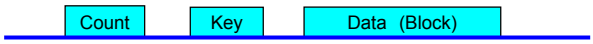


FBA方式とCKD方式

- ディスクの方式には、大きく2つのタイプがある
  - FBA方式
    - ブロック・サイズ(セクター・サイズ)を固定長にする方式
    - 現在主流となっている方式 => 通常のディスク装置

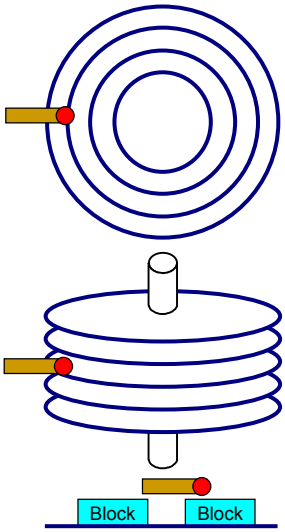


  - CKD方式
    - カウント、キー、データの順序でディスク上にデータを書き込む方式
    - ブロックのサイズは可変長な点が特徴
    - 主にホスト系のディスク装置で利用されていた
    - 現在はインタフェースとして残って入るが、実際のハードウェアとしては存在しない
      - 例：IBM 3380、IBM 3390



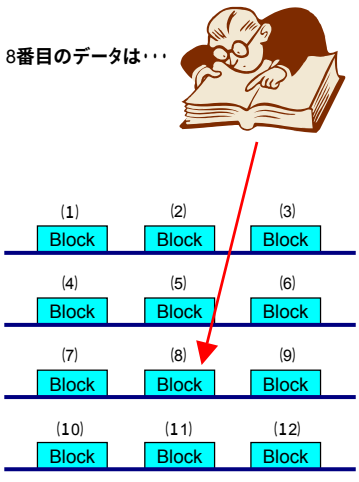
HDDのアクセス方法：CHS

- Cylinder/Head/Sector (C/H/S)
  - シリンダ番号 (Cylinder) 、ヘッド番号 (Head) 、セクター番号 (Sector) の3つのパラメータを用いてハード・ディスクにアクセスする方式
  - HDDの物理構造に密接に関連するアクセス方法
  - 各用語解説
    - トラック
      - 1回転でアクセスできる1周分の円
    - シリンダ
      - プラッタの両面を使用したり、複数のプラッタから構成される場合、複数のヘッドはアームに連動して動くのでヘッド位置を決めるとアクセスできるトラックは同一円筒状に並ぶ
    - ヘッド
      - どのプラッタのどの面をアクセスするかヘッド番号により指定
    - セクター
      - 各トラックをセクター(ブロック)と呼ばれる短い単位に分割
      - セクターがHDDにおける記録単位
        - 標準的なHDDでは各セクターは512バイトの固定サイズ



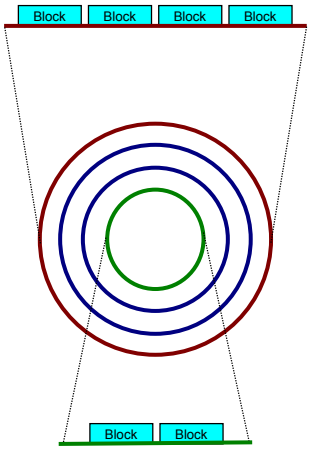
HDDのアクセス方法：LBA

- Logical Block Addressing (LBA)
  - ハード・ディスク内のすべてのセクターに通し番号を振り、その通し番号によってセクターを指定する方式
  - LBA自体にビット数の規定はないため、理論上は無限に拡張することが可能
    - IDE方式では28ビットまで、「Big Drive」方式は48ビットまで
    - SCSI方式では32ビットまで
  - IDE方式では、BIOSの規定以上の容量を持つディスクにはCHSでアクセスできないことから、現在はハード・ディスクの全セクター (ブロック) に通し番号を振るLBA方式(限界は正確には128GB)が主に使用されている



マルチ･ゾーンド･ビット･レコーディング

- ゾーン記録方式とも呼ばれる記録密度を高めるための方式
  - ディスクの最内周と最外周のシリンダ数をいくつかの領域（ゾーン）に分割
  - それぞれのゾーン毎に1トラックあたりのセクター数を定める
    - CHS方式では、最外周と最内周ではトラック長が異なる
      - 最内周トラックの記録密度が最も高く、最外周トラックの記録密度が最も低い
    - ハード･ディスクの回転数は一定なので、最内周と最外周ではデータ転送速度が異なる
      - 最外周ゾーンの転送速度が最も高く、最内周ゾーンの転送速度が最も低い
  - 近年のほとんどの大容量ドライブは8～16ゾーン程度のゾーン記録方式を採用しており、ドライブあたりの容量は従来方式と比較すると20%～50%も増加しているといわれている
- CHSアドレスは、物理的なハード･ディスクの内部構成とは異なっている

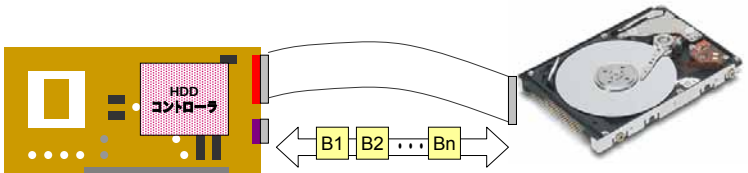


マルチ･ゾーンド･ビット･レコーディング採用ディスク例

HGST Ultrastar 146Z10

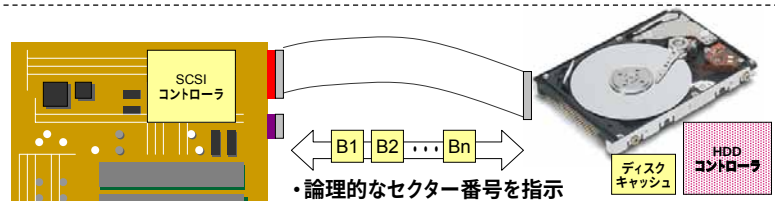
Zone	Physical Cylinders	Sectors/Track
Data Zone 0	0 - 383	864
Data Zone 1	384 - 3967	840
Data Zone 2	3968 - 5631	800
Data Zone 3	5632 - 6527	780
Data Zone 4	6528 - 8703	768
Data Zone 5	8704 - 15359	720
Data Zone 6	15360 - 18047	672
Data Zone 7	18048 - 19199	660
Data Zone 8	19200 - 21503	640
Data Zone 9	21504 - 24959	600
Data Zone 10	24960 - 27775	560
Data Zone 11	27776 - 29183	540
Data Zone 12	29184 - 30719	520
Data Zone 13	30720 - 35199	480
Data Zone 14	35200 - 36735	440

HDDの高機能化



・シリンダ、ヘッド、セクターを制御し、ブロック単位で入出力

HDD制御はコンピュータ側

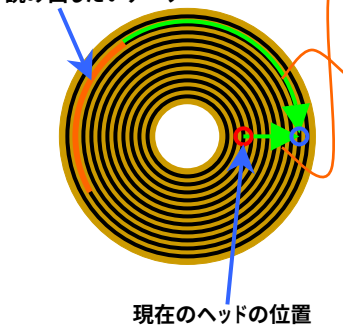


・論理的なセクター番号を指示し、ブロック単位で入出力
・複数の入出力をまとめて、コマンド&データで受け渡し

HDD制御はHDD側
読み書き動作をスケジューリング

HDDの物理的なアクセス時間

- シーク時間＋回転待ち時間＋データ転送時間



① 読み出すデータのあるトラックにヘッドを移動する
（シーク時間）

② 読み出すデータの先頭がヘッドの下に来るのを待つ
（回転待ち時間）

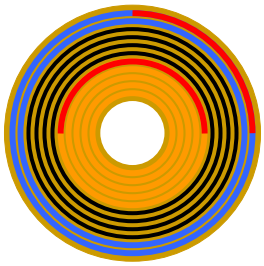
③ 読み出すデータが全てヘッドの下を通過して、ヘッドから読み出される
（データ転送時間）

HDDの外周配置と内周配置の違い

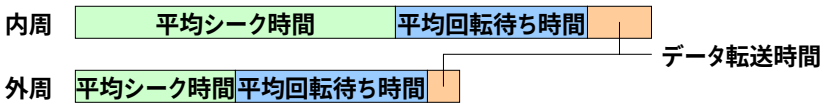
- 外周の方が高性能
  - 平均シーク時間が小さい
  - データ転送時間が小さい
  - 使用頻度の高いデータを外周付近（トラック番号の小さいトラック）に配置するのが良い

しかし、、、

- LBA方式IDE HDDやSCSI HDDでは、セクター番号とHDD内の記録位置を結びつけることは困難



同じデータ量を記録するトラック数は、
外周付近の方が内周付近よりも少ない。
同じデータ量を読み出すための回転量も、
外周付近の方が内周付近よりも少ない。



HDDインタフェース概要 (1)

- E-IDE（ATA）：Enhanced Integrated Drive Electronics
  - パソコンとハード・ディスクなどの記憶装置を接続する方式の一つ
  - Western Digital社が提唱した、IDE (ATA) 方式の拡張仕様であり、IDEでは2台までだった最大接続機器数は2系統2台ずつの合計4台まで増加し、CD-ROMドライブなどハード・ディスク以外の機器も接続できるようになった (E-IDE)
  - IDEではハード・ディスクの最大容量が528MBに制限されていたが、8.4GBまでのハード・ディスクが使えるよう改善され、データ転送速度の向上も図られた
  - アメリカ規格協会 (ANSI) によって、ハード・ディスク部分の仕様はATA-2、ハード・ディスク以外の機器の接続に関する仕様はATAPIとして、規格化された

- SCSI：Small Computer System Interface
  - パソコン本体と周辺機器の接続方法の取り決め
  - アメリカ規格協会 (ANSI) によって規格化されている。最初の規格はShugart社 (現在のSeagate Technology社) の開発したSASIをベース
  - 現在では汎用性や性能が大幅に強化された後継規格、SCSI-2やSCSI-3が普及している

HDDインタフェース概要 (2)

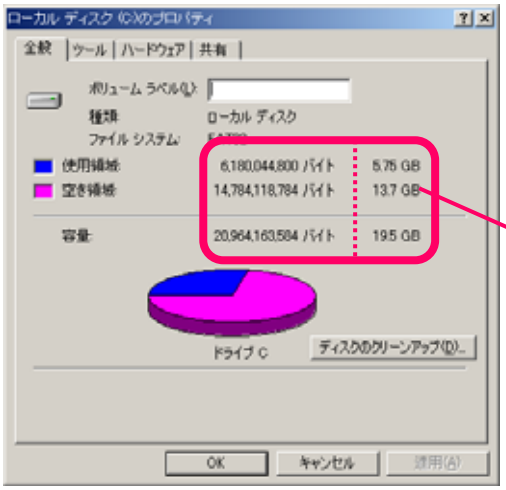
- **FCP (Fibre Channel Protocol: SCSI over Fibre Channel)**
  - Fibre Channel物理層上でSCSIコマンド&データを転送するための規格
  - SCSI-3規格で規定
  - 転送速度は100MB/s
  - SANの普及に伴い、現在、オープン系システムの代表的なインタフェースとなっている
- **SSA : Serial Storage Architecture**
  - IBM社が中心となって開発されたシリアル転送方式を採用したSCSI規格の一種
  - SCSI-3規格に含まれており、転送速度は最大160MB/s
  - 接続時の機器間の距離は最大25m、最大接続台数は96台で、ループ状の接続が可能になっている
  - ケーブルには基本的にシールド付より対線(STP)を使うが、光ファイバーケーブルを用いることで接続距離を最大2.5kmまで伸ばすこともできる
  - 外部インタフェースとしてはFibre Channelの普及が進んでいることもあって、普及率はそれほど高くない
    - 代表的な装置
      - IBM TotalStorage Enterprise Storage Server (ESS)の内部ディスク・インタフェース
      - IBM TotalStorage 7133 ディスク・サブシステム

HDDインタフェース概要 (3)

- Serial ATA
  - パソコンとハード・ディスクなどの記憶装置を接続するIDE(ATA)規格の拡張仕様の一つ
  - 2000年11月に業界団体「Serial ATA Working Group」によって仕様の策定
  - Serial ATAは、Ultra ATAなどの現在のATA仕様で採用されていたパラレル転送方式を、シリアル転送方式に変更したもの
    - シンプルなケーブルで高速な転送速度を実現することができる。従来のパラレル方式のATA諸規格との互換性も持っている。
    - 従来のパラレル方式のATA仕様で転送速度が最も高速なのはUltra ATA/133の133MB/sで、パラレル方式ではこれ以上の高速化は難しいとされる
  - Serial ATAの最初の規格「Ultra SATA/1500」は1.5Gbpsと、従来の約1.4倍の速度を実現する
  - Serial ATA仕様は今後も拡張を続け、近い将来には2倍の3Gbps、その後6Gbpsに引き上げられる予定

ディスク容量

- ディスク容量を表す単位は2通りが混在している



147.8GB=137.6GB

HDD物理容量 1K=1,000

OSの容量 1k=1,024= 2^{10}

G=$1{,}024^3$
=$(2^{10})^3$
=2^{30}

単位：K (キロ)、M (メガ)、G (ギガ)

- コンピュータの単位：K=1,024=2^{10}
- HDDの物理容量：K=1,000

分野		K (キロ)	M (メガ)	G (ギガ)
コンピュータ メモリ LAN	2進法	1,024 2^{10}	1,048,576 $(2^{10})^2=2^{20}$	1,073,741,824 $(2^{10})^3=2^{30}$
科学技術一般 HDD 通信業者	10進法	1,000 10^3	1,000,000 $(10^3)^2=10^6$	1,000,000,000 $(10^3)^3=10^9$

※ 3.5“ 1.44MBのフロッピー・ディスクの場合は特別。
M=1,024,000=1,024×1,000=$2^{10}\times10^3$

ストレージ・サーバー、ディスク・サブシステムの制御装置機能

システム製品としてのディスク装置

- ディスク製品をHDD単体ではなく、システム製品として提供
  - ストレージ・サブシステム（ストレージ・サーバー）としてメーカーは提供する
  - 単純にHDDを提供するのではなく、可用性やパフォーマンスの面で付加価値を付けた製品としてユーザーに提供
  - 一般にHDD単体による利用に比べ、使用するユーザーのメリットは大きい
  - 多くの機能は「制御装置」、または「制御機構」と呼ばれる仕組みを保持し、それらがインテリジェンスを持ってサブシステム全体の制御を行い、各種機能を提供する
- ストレージ・サーバーの制御装置が持つ代表的な機能（例）
  - 論理的なディスク・イメージの提供
  - RAID機能
  - キャッシュ
  - 高速コピー機能
  - 遠隔コピー機能
  - 電源/ファン/制御装置の二重化
  - マルチ・パス機能
  - LUNマスキング機能

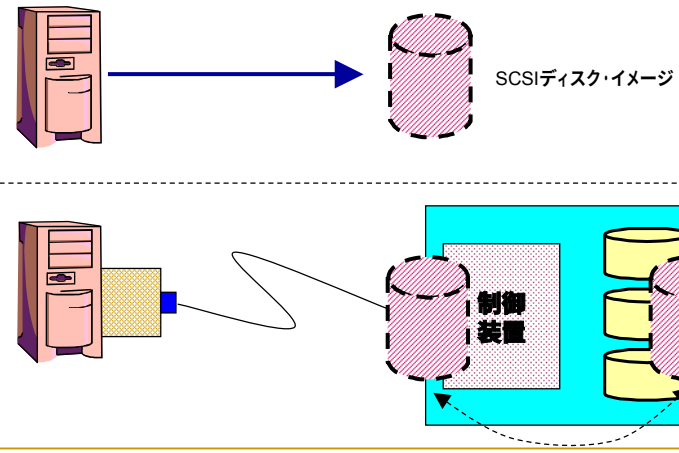
高性能、高可用性のストレージを構成するには

- 高性能
  - IOPS : Input Output per Second （トランザクション性能）
    - 1秒間に何回、データを書き込み/読み出しできるか
    - RAIDアレイ内の物理HDD数を多くする
  - MB/s, GB/s （スループット性能）
    - 1秒間に何MB(GB)、データを書き込み/読み出しできるか
    - 回転数が早いHDDを使用する、HDD数を多くする
- 高可用性
  - ホットスワップ
  - 自動リビルド
  - ホットスペア
    - RAID装置を集約した方がスペアドライブ配置によるオーバヘッドが少ない
  - 多重コントローラを使用した自動フェールオーバ

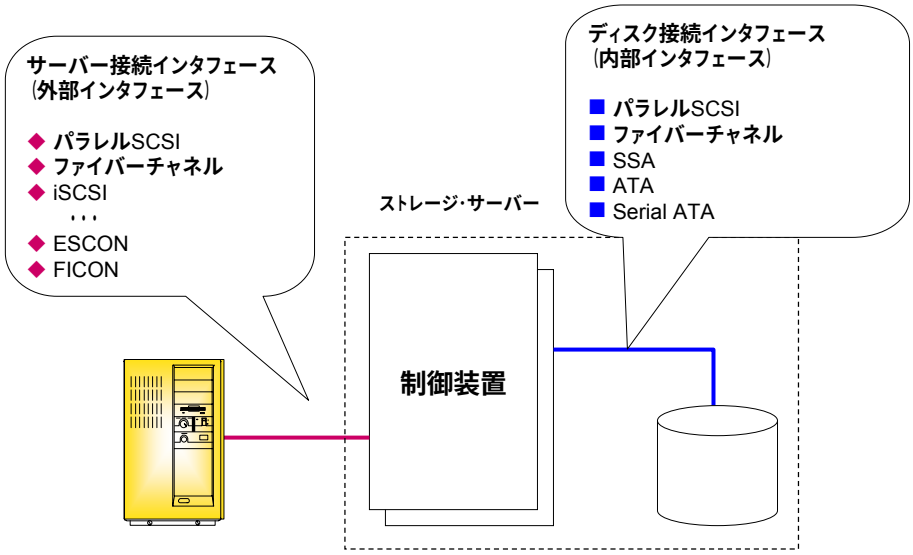
たくさんのHDDを接続でき、かつ、転送帯域が広いインタフェース
コントローラのフェールオーバに対応可能なインタフェース

サーバー、制御装置とHDDの関係

- 制御装置はHDDとサーバーとの間で活動を行い、各種機能を提供する



インタフェースの組み合わせ

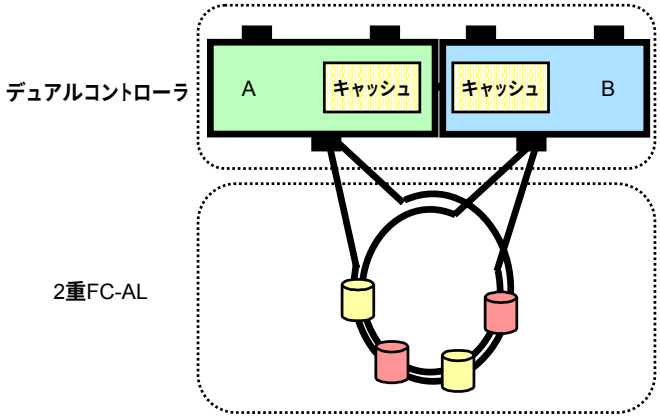


サーバーとの接続における考慮点

- 接続距離
  - SCSIは最大25m
  - ファイバーチャネルでは最大10km
    - リピータ使用により最大100km
  - IP技術を使用すれば…
- 接続可能なサーバー数
  - SCSIではチャネル当たり最大15装置
  - ファイバーチャネル
    - FC-AL構成では127装置
    - ファブリック・スイッチ構成では1600万装置
- サーバーとの接続の柔軟性
  - SCSIでは、固定されたチャネル接続
  - ファイバーチャネル
    - FC-AL構成では固定リング接続
    - FCPではスイッチを使い、ダイナミックにルートを変更させることが可能
    - ファブリックスイッチ構成では多重パス構成も可能

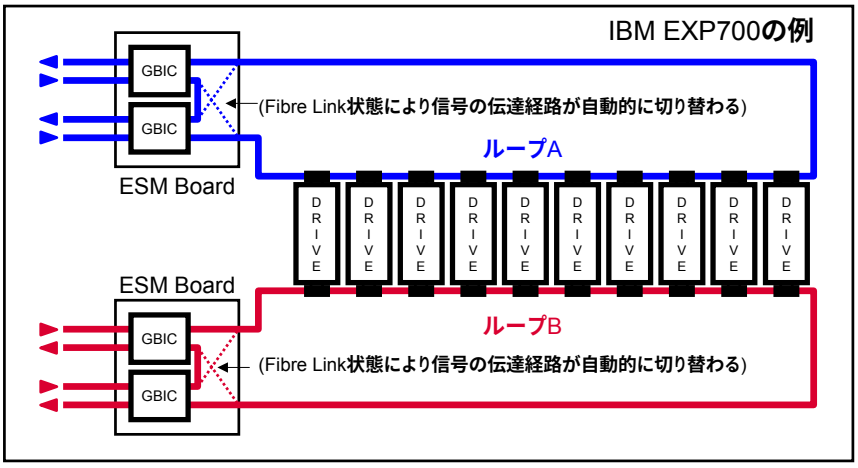
デュアル・コントローラ

- コントローラを2重化し、コントローラ故障時でも処理を継続可能
  - コントローラはホット・スワップ可能



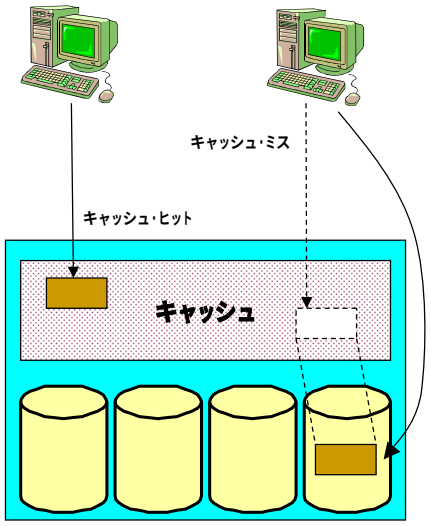
内部接続の高可用性：2重ループFC-AL構成

- ファイバーチャネルディスク拡張ユニットの内部は2重ループ構成になっている
- 各ドライブは両方のループからアクセス可能



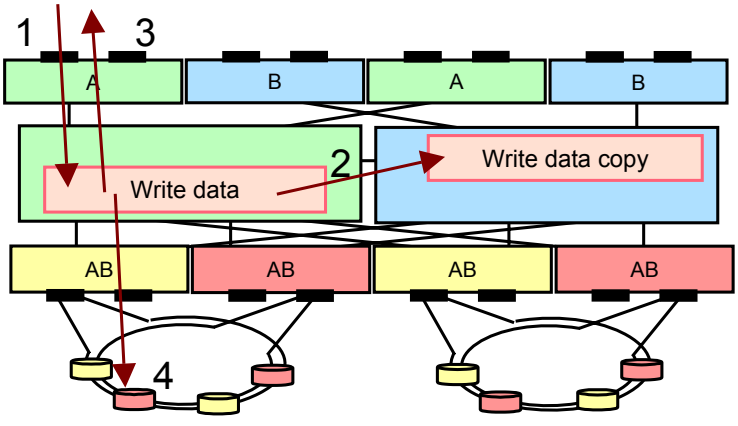
キャッシュ

- 半導体メモリーを制御装置に装備し、HDDの機械動作速度より高速に読み書きを行わせる事を可能にする機構
- キャッシュの種類
  - 読込みキャッシュ (Read only cache)
    - 一般に「キャッシュ」と言った場合はこれを指す
  - 書込みキャッシュ(Read/Write cache)
    - 通常読み込みキャッシュの機能も持つ
    - 書込みを高速化するためのキャッシュ
    - 通常二重化、不揮発性などのデータ保護機能が必要
- キャッシュ・ヒット
  - 目的とするデータがキャッシュ上にあった場合
  - 機械的動作が不要なため、高速な入出力処理が可能となる
- キャッシュ・ミス
  - 目的とするデータがキャッシュ上に無かった場合
  - HDDに直接読み書き動作をしなければならない

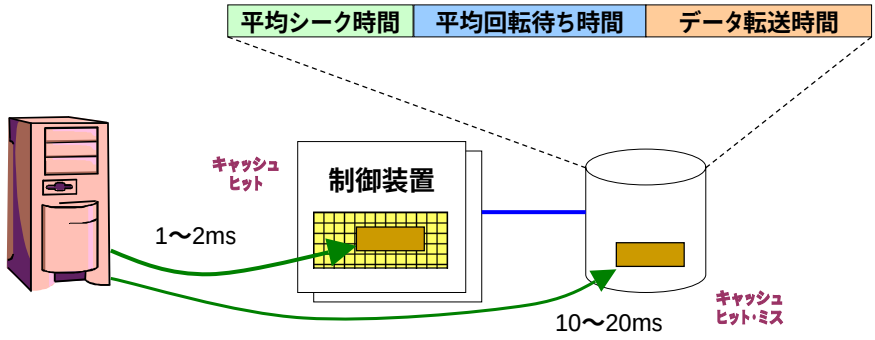


キャッシュ制御の例

書込みデータをミラーする実現例



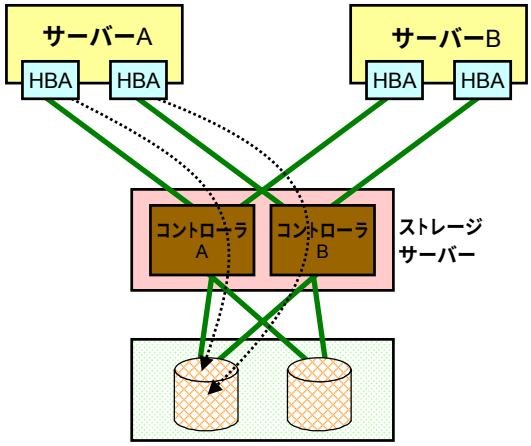
キャッシュ使用時のデータ・アクセス時間



- キャッシュ・ヒット率を考慮したレスポンス・タイム
  - 前提
    - ヒット時のレスポンスを1ms、ヒット・ミス時のレスポンスを10msと仮定
    - ヒット率を80%と仮定
  - レスポンス・タイムを計算すると・・・・
    - 1ms ＊ 80% ＋ 10ms ＊ 20% ＝ 2.8ms

マルチ・パス・アクセス

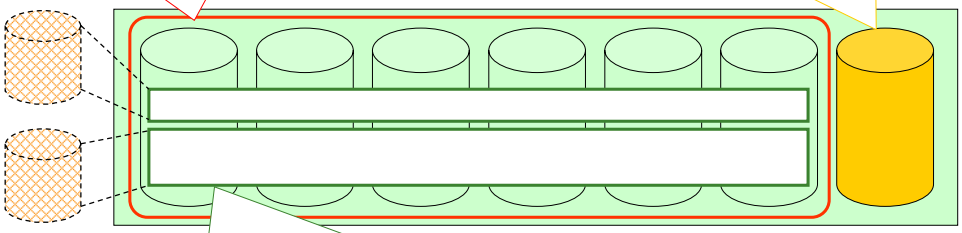
- サーバーからマルチ・パスによるアクセスをサポート
  - 可用性の向上
    - 自動フェールオーバ
  - パフォーマンスの向上
    - 負荷分散



論理ドライブ/LUN (論理装置番号)

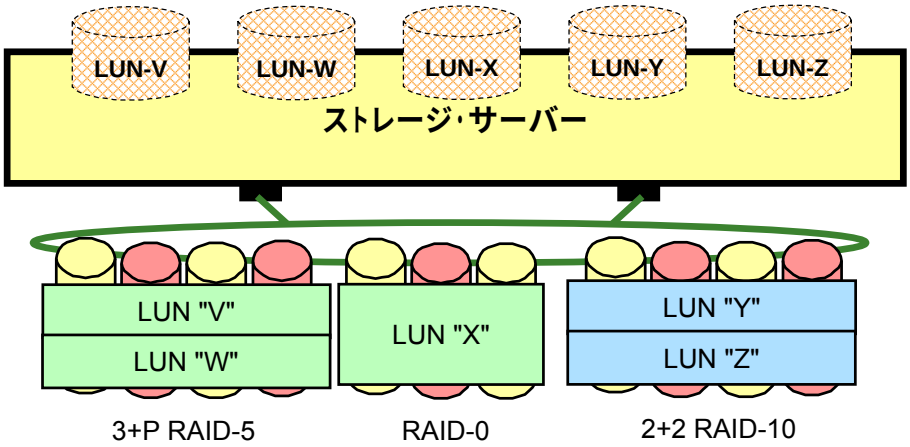
RAIDアレイ (RAID Array)
複数の物理HDDをまとめたもの

ホットスペアドライブ (Hot Spare Drive)
障害に備えてスタンバイしているスペアディスク



論理ドライブ (Logical Drive) /論理装置番号 (LUN: Logical Unit Number)
RAIDレベルを指定して作成
OSからは1台のドライブに見えます

LUNの構成例



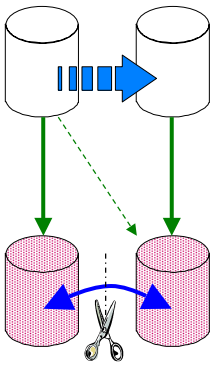
機種にもよるが、各種RAIDアレイ内にLUNを構成することも可能

高速コピー機能

- 高速コピー機能
  - サーバーを介することなく瞬時にデータの複製をストレージ・サーバーだけで作成することができる機能
    - 参考：ソフトウェアのファイル・システム・レベルで行う製品も存在する
  - スナップ・ショットとも呼ばれる
  - 用途
    - バックアップの取得
    - テスト・データの作成
    - データの並列処理
- 通常、同一ストレージ・サーバー内のボリューム(LUN)の複製を行う
  - ハードウェアは「ファイル」を認識できないためボリューム（SCSIディスクのイメージ）でコピーを作成する
- 大別すると3つの方式がある
  - どの方式も瞬時にコピーできる機能を提供するが、バックグラウンドの作業のやり方が異なる
    - スプリット・ミラー方式
    - バックグラウンド・コピー方式
    - ポインター・コピー方式

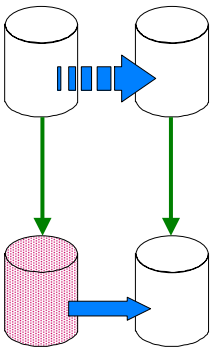
高速コピー機能の3つの方式

■ スプリット・ミラー方式



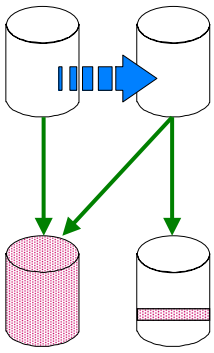
- 通常はミラーリングしている
- 高速コピー起動後にミラーを分割
- その後は個別に利用

■ バックグラウンド・コピー方式



- 高速コピー起動後、ポインターのみをコピーし、同じ内容に見せる
- 実データは後からバックグラウンドでコピーを行う

■ ポインター・コピー方式

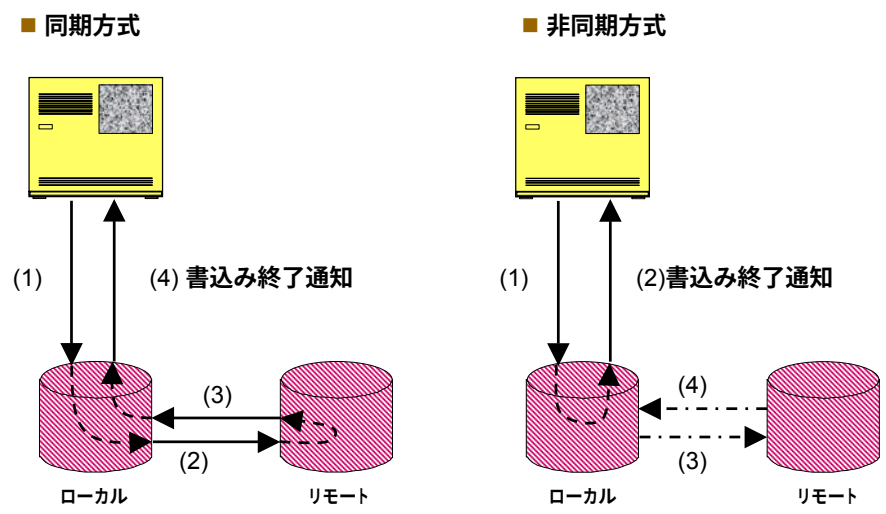


- 高速コピー起動後、ポインターのみをコピーし、同じ内容に見せる
- 実データはコピーしない
- 変更分のみ別エリアに保管

遠隔コピー機能

- 遠隔地間でのデータ複製機能
  - 実装方式の違い
    - ハードウェア方式
      - 一般に各ストレージ・サーバーの固有の機能のため、コピー元とコピー先は同一メーカ、同一機種である必要がある
      - サーバー資源を消費しない
    - ソフトウェア方式
      - ストレージ・サーバーの機種制限が無いというメリットがある
      - サーバー資源を消費し、システム・パフォーマンスへの影響を考慮する必要がある
  - 転送方式の違い
    - 同期方式
      - ローカルのディスクに書かれ、更に遠隔地のディスクへの書込みも確認された段階で書き込み終了とする方式
      - データの整合性が取りやすいが、パフォーマンスへの影響が大きい
    - 非同期方式
      - ローカルのディスクに書かれたことをもって書き込み終了とする方式
      - 遠隔地への転送は、非同期に転送される
      - パフォーマンスへの影響は小さいが、回復手順が複雑になる傾向がある

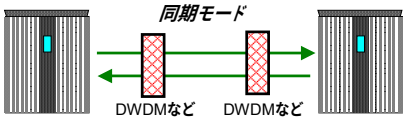
遠隔コピー：同期方式と非同期方式



遠隔コピー機能の実装例

IBM DS8000 Metro Mirror

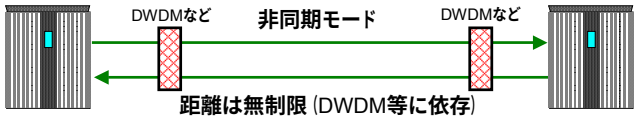
- 同期モードによる
- 最大103Kmまでの距離をサポート
  - 103Km以上が必要な場合はRPQ



距離は標準で103Kmまで

IBM DS8000 Global Mirror

- 非同期モードによる
- 最大距離の制限は無し
  - 実際の制限はDWDM等の機能に依存



距離は無制限 (DWDM等に依存)

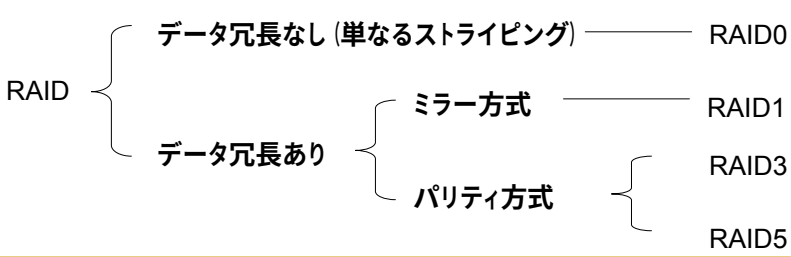
RAID

RAIDとは

- RAID=Redundant Array of Independent Disks
  - 本来は、Redundant Array of Inexpensive Disksの略であり、低価格であるが信頼性の低いHDDを組み合わせて高信頼化を実現することが目的
    - 各メーカーは一般に安いディスクは使っていないと言う意味で「Independent」を使う
  - HDDは稼働部が多いため、故障率の高いコンポーネント
  - HDDが同時に複数個故障する確率は低い
    - 単一HDD故障に対応できる仕組みができれば、可用性を向上できる
  - 筐体全体での故障やオペレーションミスによるデータ損失には対応できない
    - 外部装置（テープ装置など）へのバックアップは必要
- 一つのHDDでは不可能なことを、複数のHDDで実現する技術
  - 容量の向上
    - 複数台のHDDを1台のHDDとして取り扱う
  - 性能の向上
    - データを複数のHDDに分割、並列入出力することで性能を向上
  - 信頼性の向上
    - ドライブ間で分割した情報を重複して記憶することで、あるドライブにエラーが発生しても別のドライブから正確な情報を復元可能

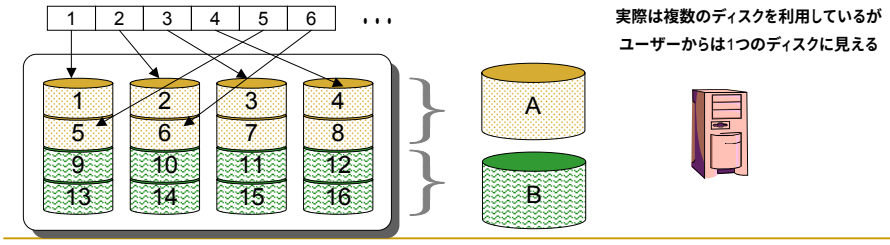
RAIDの分類

- 基本はRAID0～RAID5の6方式
  - RAID 0（ストライピングとも言う）
  - RAID 1（ミラーリングとも言う）
  - RAID 2（ECC適用方式）
  - RAID 3（パリティ保護 + ストライピング）
  - RAID 4（固定パリティ + データ単位でストライピング、）
  - RAID 5（ローテート・パリティ + データ単位でストライピング、）
  - RAID 10, RAID 1+0（RAID 1 とRAID 0 の組み合わせ）など



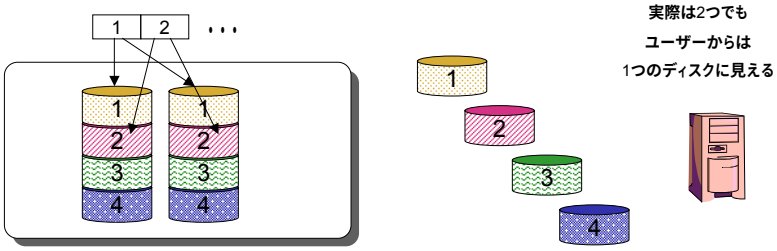
RAID 0 （ストライピング）

- 複数のHDDにデータをブロック単位で分散させて（ストライピング）記録させる方式
- 利点
  - 複数のHDDを並列に動作させるため高速に読み書きできる
    - 複数のHDDに対してコマンドとデータを送り、見かけ上シーク時間や回転待ち時間をなくす
    - HDDの数に比例して入出力のスループット性能が向上
  - 1台のHDDでは実現できない大容量のHDDを実現
- 欠点
  - 複数のHDDをデータの書き込みに使用するので、1台でもHDDが故障すると全データが読み書きできなくなる
  - 厳密には「RAID」とは呼びにくい（便宜上RAIDという用語が慣習として使われている）



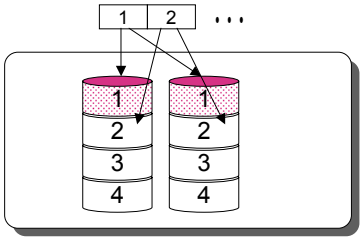
RAID 1 （ミラーリング）

- 2台のHDDに同じデータを記録し（ミラーリング）、常に同じ状態に保つ
- 利点
  - 1台が故障しても同じ内容のデータを記録したもう1台が残り、処理を継続することができる
- 欠点
  - 2台のHDDに同じデータを書き込むため多少オーバヘッドがある
  - 利用可能な容量は実装容量の半分となる

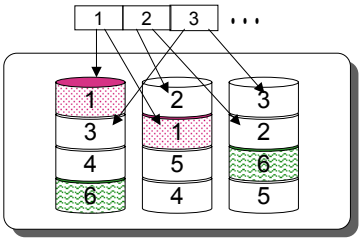


RAID 1E

- RAID 1の拡張版で、3台以上の物理ドライブを使用したミラーリングである
- 利点
  - ドライブ追加によって性能が向上する場合がある
  - 1台のHDDでは実現できない大容量のHDDを実現
- 欠点
  - 論理ドライブの容量はRAID1と同じく物理ドライブ容量の50%である



RAID1: 通常のミラーリング

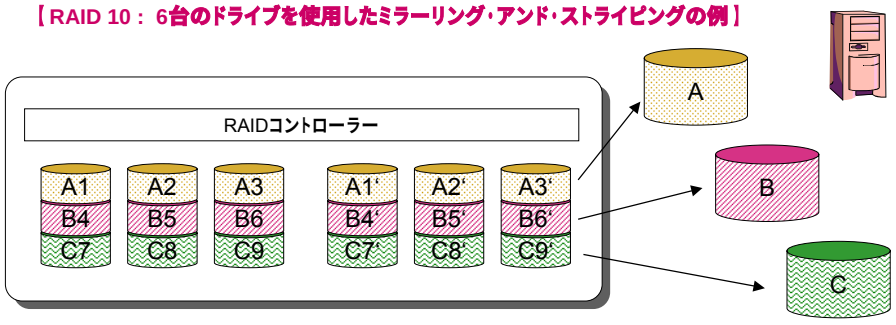


RAID1E: 3台のドライブを使用したミラーリング

RAID 1＋0（RAID10）

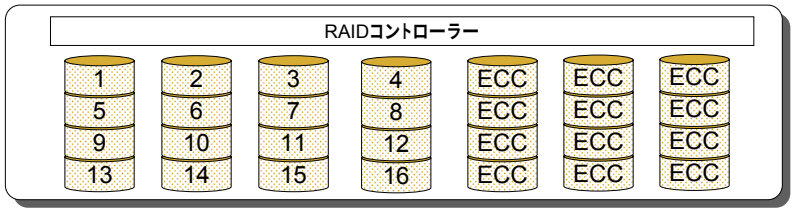
- RAID 1とRAID 0を組み合わせた技術である
  - 複数台の物理ドライブを使用したミラーリング＋ストライピングである
  - 論理ドライブの容量はRAID 1と同じく物理ドライブ容量の50%である
  - ドライブ追加によって性能が向上する場合がある

【RAID 10： 6台のドライブを使用したミラーリング・アンド・ストライピングの例】

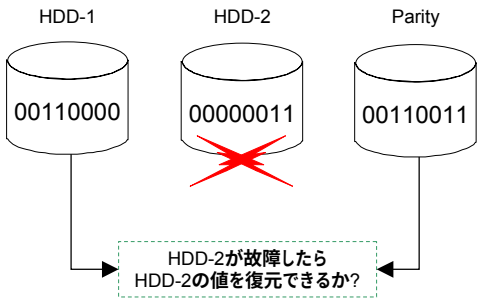


RAID 2

- メモリーなどで利用されているECCの手法をディスクに取り入れた方式
- 現在、実現された製品はない（はずである）
  - 製造回路が複雑になり、パフォーマンス及びコスト的なメリットも得にくいため



パリティを使ったデータ冗長化



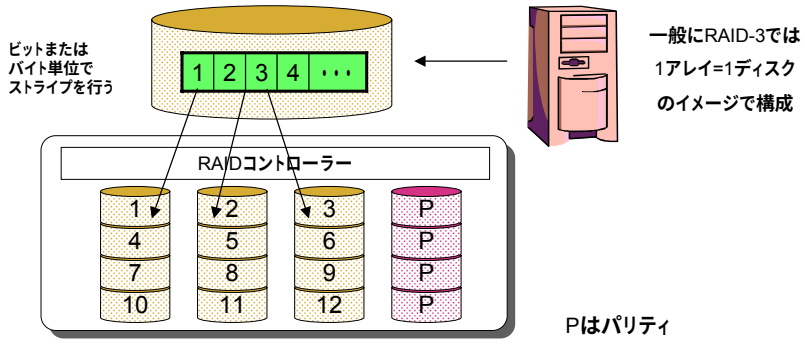
x	y	x XOR y
0	0	0
0	1	1
1	0	1
1	1	0

HDD-1	00110000	(0x30)
XOR（排他的論理和）		
HDD-2	00000011	(0x03)
パリティ	00110011	(0x33)

HDD-1	00110000	(0x30)
XOR（排他的論理和）		
パリティ	00110011	(0x33)
HDD-2	00000011	(0x03)

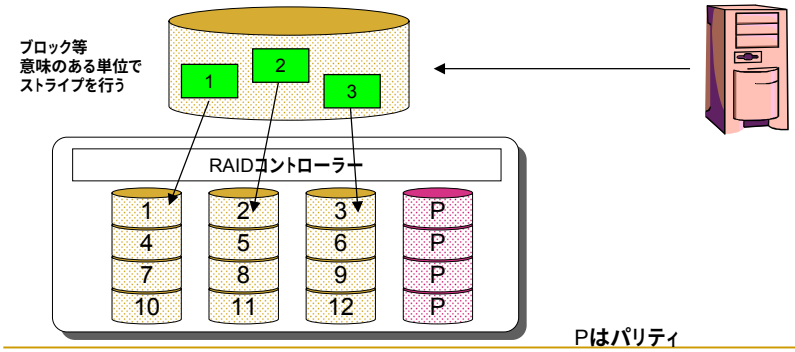
RAID 3

- 複数あるディスクのうち1台をパリティの記録に割り当て、他のディスクにデータを分散して記録する方式。
  - 意味のないデータの単位（例えば10バイト）に分割してストライプを行う
- どれか1台が故障しても交換してデータを復旧することができる
- 複数のディスクにはデータを分散して同時並行で記録するため、高速化もはかられる
  - 特に科学技術計算のアレイの読込み、書出しなどに向く



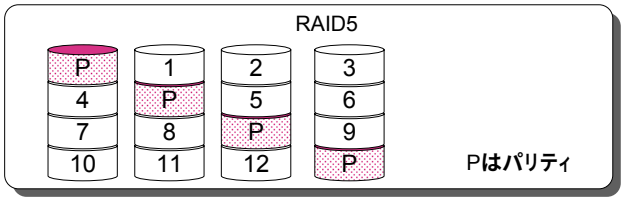
RAID 4

- 意味のある単位（ブロック、セクターなど）でストライプを実施
  - パリティーは固定
  - RAID 3 の不得手なランダム・アクセスに対応可能
- パリティ・アクセスにボトルネックが発生してしまうため、実装された製品例は少ない



RAID 5

- RAID 4の改良版、パリティーをローテートする点が特徴
- アレイの全てのドライブを越えてデータとパリティをストライプする
  - 意味のあるデータの単位（例えばブロック）に分割してストライプを行う
  - 並列アクセス（トランザクション・タイプのアクセス）に向いている
- 最も実用的なRAID方式と一般に考えられている
- 実際に使用できるディスク容量はディスク1台分（パリティデータ記憶域用）だけ少なくなりなる
- 少なくとも3台以上の物理HDDが必要
- アレイ中の1台の物理ドライブに障害が発生しても残りの物理ドライブでサービスを継続
  - ホットスペアドライブ、あるいは、交換したドライブを使ってRAID5を再構成可能



一般論としてのRAIDレベル比較

一般的なファイルサーバーの活動の統計結果： 読み取り80%、書き込み10%、検索10%

RAIDレベル	データ冗長	ドライブ容量の使用率	ランダム読取性能	ランダム書込性能	順次読取性能	順次書込性能	コスト
RAID 0	なし	100%	△	△	◎	◎	高
RAID 1	あり	50%	○	○	○	○	高
RAID 1E	あり	50%	○	○	○	○	中
RAID 3	あり	67% - 94%	△	△	◎	◎	低
RAID 5	あり	67% - 94%	◎	△	◎	◎	低

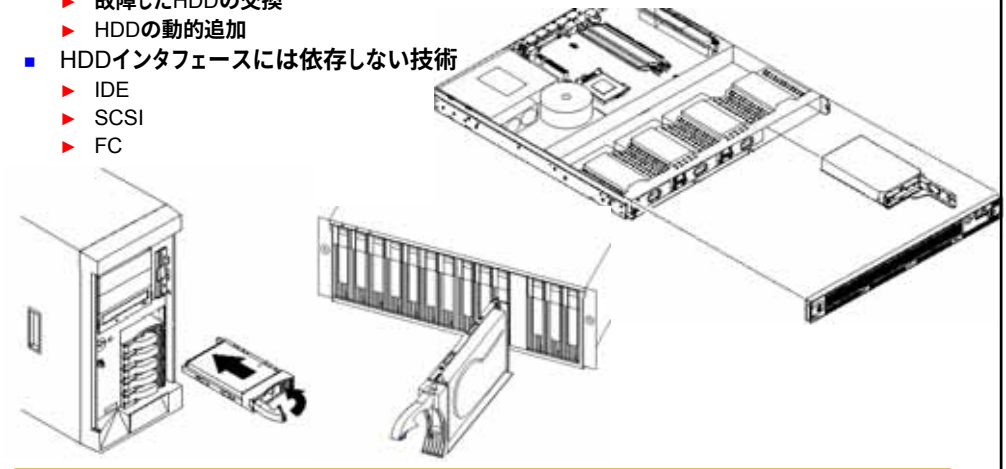
これら比較はあくまで参考と考えてください。実際の各社製品においては、
各種仕組みの実装により、この表に当てはまらないケースもある

性能良 ←→ 性能悪

◎ ＞ ○ ＞ △

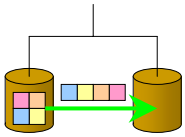
ホット・スワップ機構

- 通電中にドライブを取り外したり、取り付けたりできる仕組み
  - 故障したHDDの交換
  - HDDの動的追加
- HDDインタフェースには依存しない技術
  - IDE
  - SCSI
  - FC

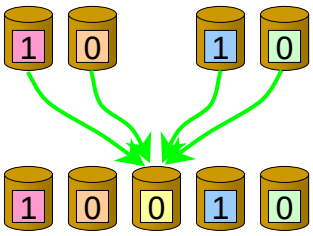


RAIDのリビルド

- RAID1、RAID5を構成しているHDDが故障した場合、その他のHDDが壊れる前に、そのHDDを交換すれば、元の信頼性を取り戻すことができる。
- 交換後、本来そのHDDにあるべきデータを再構築することをRebuild（リビルド）と言う



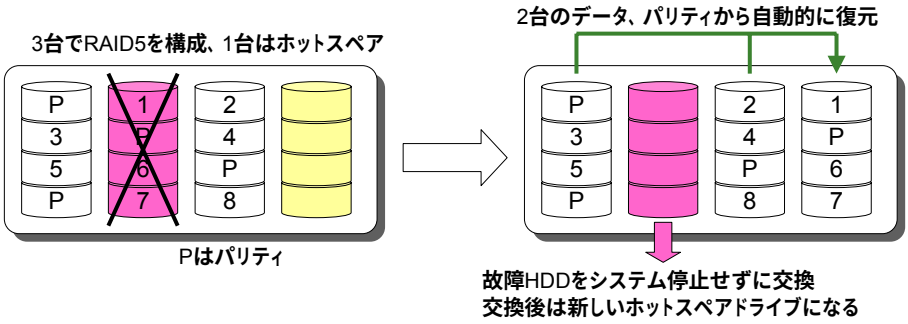
RAID1の場合、
データのコピー



RAID5の場合、他の全ての
HDDのデータを用いて、Parity
生成と同じ容量でデータを再生

ホット・スペアー・ドライブ

- ホット・スペアー・ドライブ（通常は使用されずスタンバイしている）をあらかじめ確保しておけば、HDD故障時にはオペレータが介在することなく、自動的に故障したHDDに代わりホット・スペアー・ドライブを使用してデータの復元が行われる
  - ホット・スペアー・ドライブの構成、装備数は一般に任意にカスタマイズ可能（機種にも依存）
  - RAID 0ではリビルドができないため、スペアーを準備する意味は無い
  - ホット・スペアー・ドライブの機能が無い装置も、当然世の中には存在する

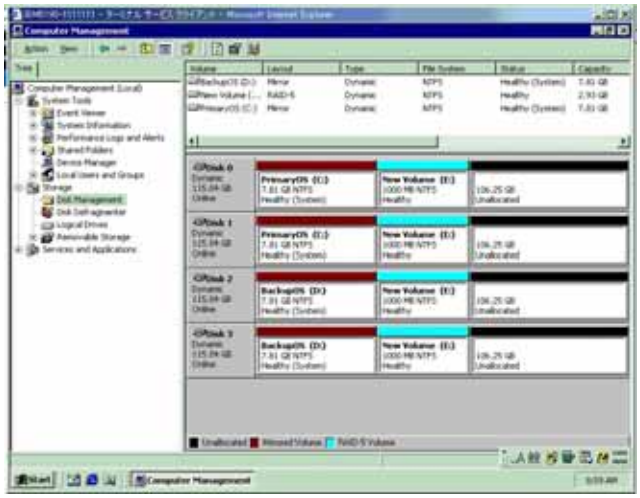


RAIDの実現方法

- サーバー直結ストレージの場合
  - ソフトウェアRAID
    - Windows 2000 Serverなどに実装されている方式で、ソフトウェア処理により、RAID1やRAID5を実現
  - ハードウェアRAID
    - RAIDアダプタ・カードなどを使用してアダプタ上のプロセッサによりRAID機能を実現
    - OSからは通常SCSIアダプタとして認識される
    - 専用のRAID制御ツールによりRAIDを構成する
- SAN接続ストレージ・サーバーの場合
  - ストレージ・サーバー内でRAID機能を実現
  - ハードウェアとソフトウェアの組み合わせ
    - 近年の高機能ストレージ・サーバーでは、ストレージ装置内での仮想化技術が取り入れられている

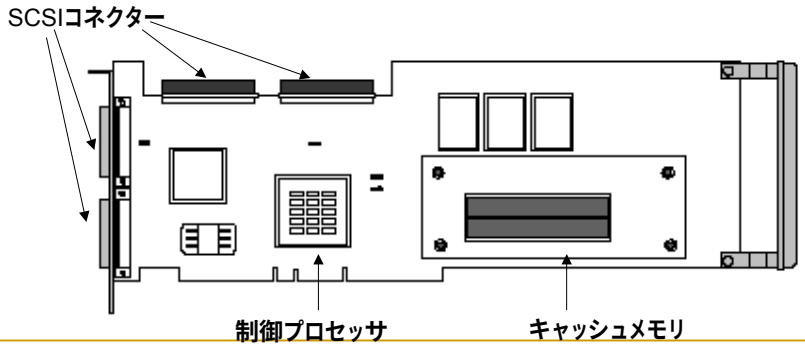
ソフトウェアRAIDの例

- Windows 2000サーバーなどでは、Windows NTFSのダイナミックディスク機能を使用し、ソフトウェアRAIDを実現



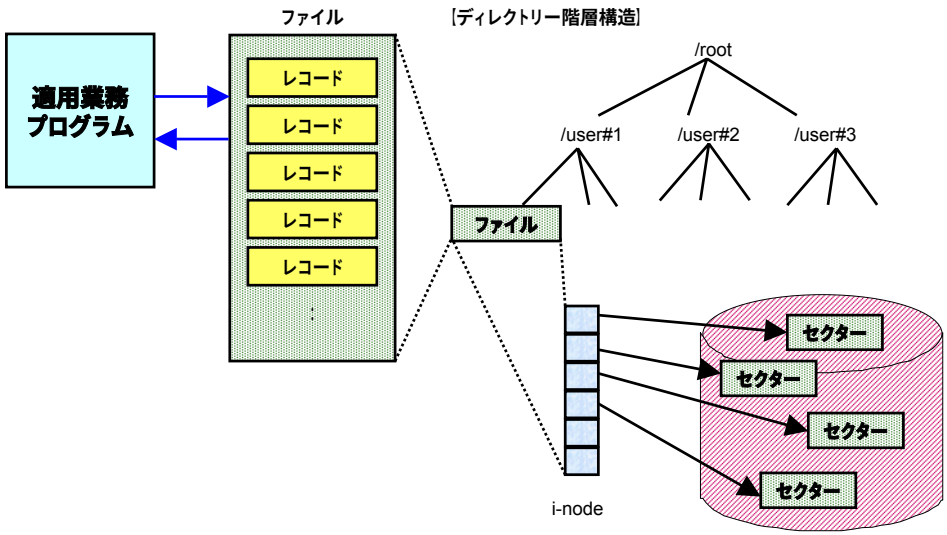
RAIDコントローラー・カードの例

- ハードウェアRAID PCIカードを搭載し、RAIDをハードウェアにより実現
  - RAIDアレイ構成情報の保持
  - パリティ計算
  - RAID自動復元、ホット・スペアリングなど

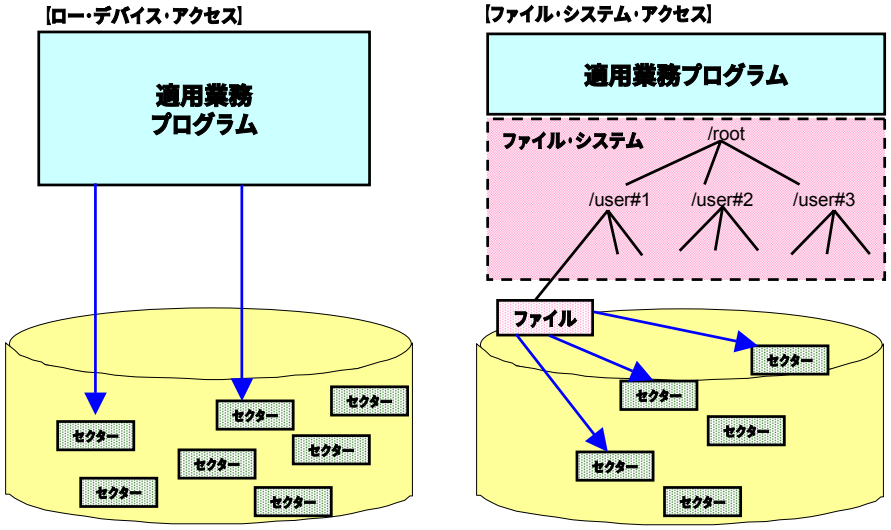


ファイルとファイル・システム

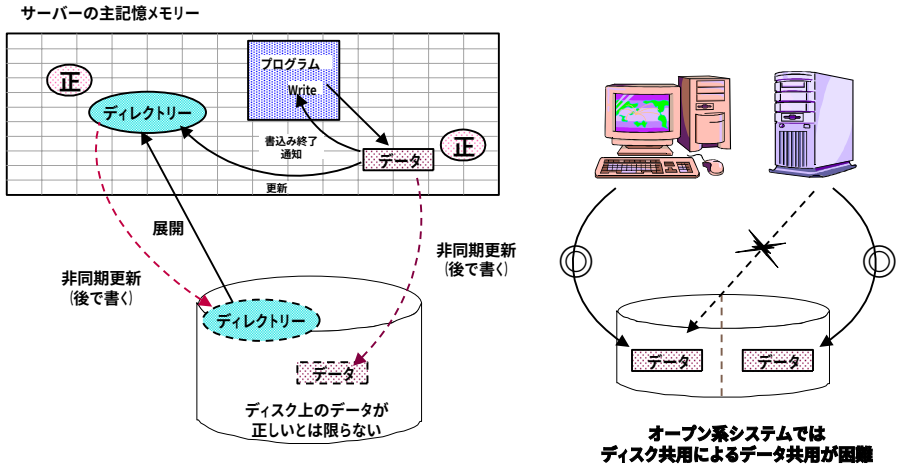
ファイルとファイル・システム



ロー・デバイス・アクセスとファイル・システム・アクセス



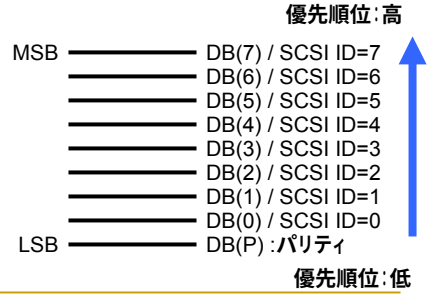
オープン系ファイル・システムのI/O処理の特徴



SCSI 規格

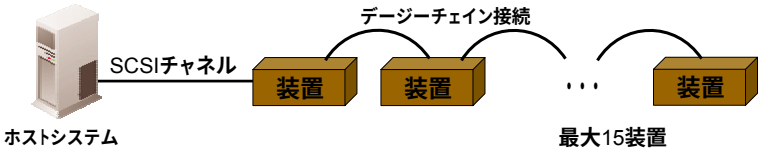
SCSI (Small Computer System Interface)

- SCSIとは、コンピュータと補助記憶装置を接続するための規格
  - スカジィ
  - HDDを接続する標準的なインタフェース
  - 対象は、HDDのみではなく、フロッピーディスク、テープ装置、CD、DVD、スキャナ、プリンタ、コンピュータ・システムなど、幅広い
- バス型
- パラレル・インタフェース
  - データ巾：8+1（パリティ）
  - データ線とアドレス線（SCSI ID指定）が共通
  - バス使用権の優先順位が固定
- ANSI NCITS（旧X3委員会） T10部会
  - SCSI-2 : ANSI X3.131-1994
  - SCSI-3 : ANSI X3.270-1996他

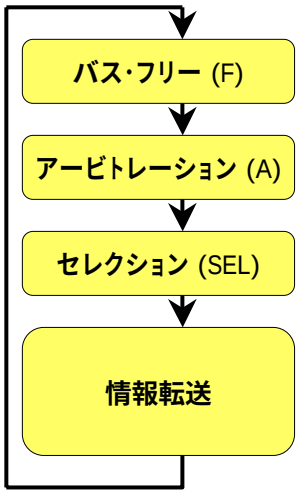


SCSIデバイスと接続形態

- イニシエータ･デバイス
  - コマンドの発行
  - ホスト (HBA)
- ターゲット･デバイス
  - コマンドに対する応答
  - HDD
- イニシエータとターゲットの間で1:1の伝送
  - 他のデバイスはバスが開放されるのを待つ



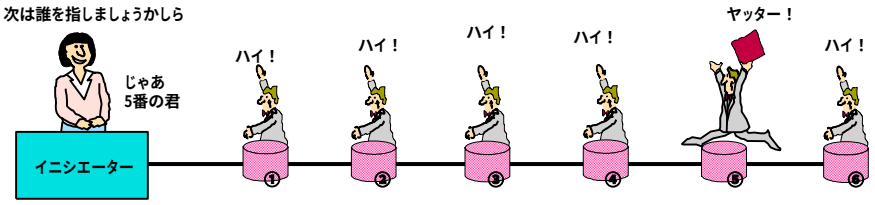
SCSIデバイスの動作概要



イニシエータ	ターゲット
バス要求を発行	
バス権を獲得	
ターゲットを指定	
接続を確立	
コマンドを発行	
	コマンドを実行 (Read/Write)
	ステータスを返送
	バスを開放

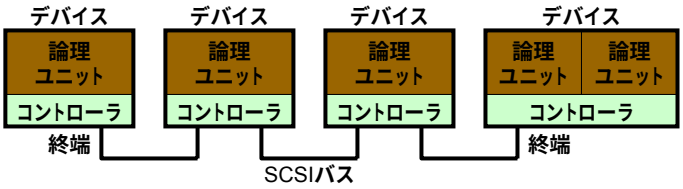
アービトレーション

- 次のサービス提供のための、選定作業
  - イニシエーターは次にサービスを提供するデバイスを選定するために、アービトレーションを行う
  - デバイス全てに対し、サービス要求があるかどうかを尋ね、優先順位に従って次にバスの使用権を得るデバイスを選定する
    - SCSIではデバイス番号によって優先順位の高さが決まる



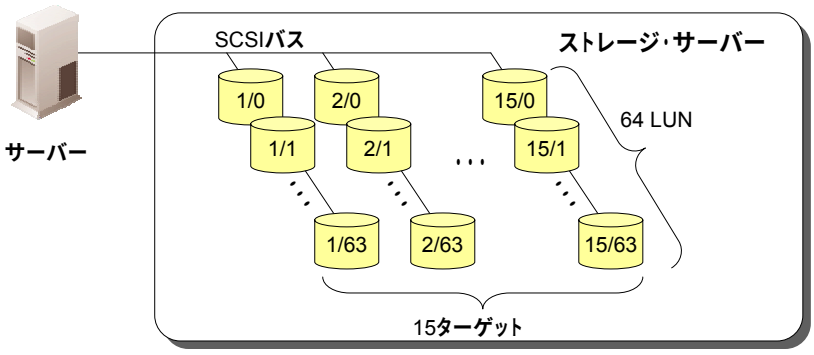
SCSIデバイスの接続

- SCSIデバイス＝コントローラ + 論理ユニット
  - コントローラ : インタフェース
  - 論理ユニット(Logical Unit、LUN) : ユーザーから見て装置そのもの
  - 8コントローラ/バス
  - 8論理ユニット/コントローラ
  - デイジーチェイン接続



LUN (Logical Unit Number)

- ストレージ・サーバー内の各ディスクはLUNとしてサーバーから認識される
  - LUNは物理SCSI HDDイメージで認識される
    - ターゲット数とLUN数の最大値はサーバー側のOSに依存
    - ポートあたり最大960ディスク (15ターゲット×64 LUN)

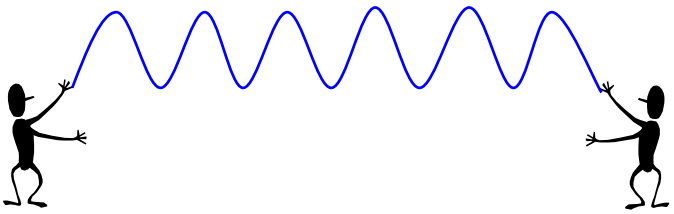


SCSIの進化：SCSI-2

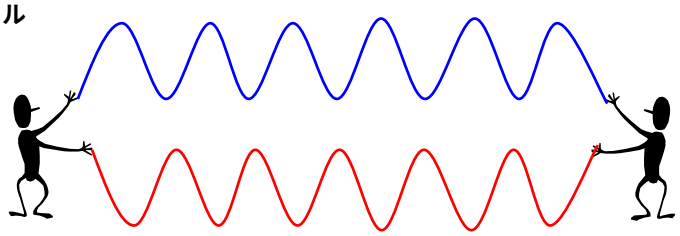
- コンピュータ本体とハード・ディスクなどの記憶装置の接続に用いられるSCSI規格の一つで、SCSI-1を改良した第二世代の規格群
  - 1994年にアメリカ規格協会(ANSI)が標準化
- 2種類の規格
  - 「Fast SCSI」(Fast Narrow SCSI)
    - 8ビット幅
    - 10MB/s
    - 最大接続台数は8台
  - 「Fast Wide SCSI」
    - 16ビット幅
    - 20MB/s
    - 最大接続台数は16台
  - 最大ケーブル長はどちらもシングルエンド駆動で3m、ディファレンシャル駆動で12m(LVD)～25m(HVD)である。
- SCSI-2からは、高速なデータ転送の妨げとなる回線終端での信号の反射を抑えるため、終端に取り付けるターミネータはアクティブ型が必須となった

シングル・エンドとディファレンシャル

■ シングル・エンド

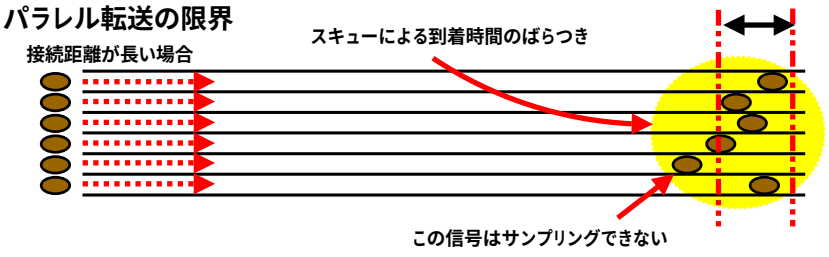


■ディファレンシャル



パラレル・インターフェースの限界：スキュー

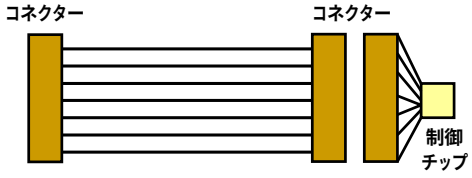
- 高速化が進んでいるが以下の欠点がある
  - 短い接続距離
  - 限られた装置数
  - 制限された拡張性
  - 単一経路のアクセス



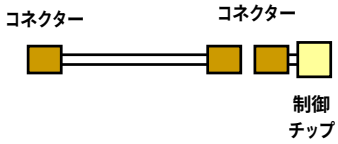
パラレル・インターフェースの限界：微細化とコスト削減

- 高度に微細化加工が進み、パラレルでは配線が困難
  - 大きさが限られる
  - 制御チップの微細化がコスト削減にもつながっている
  - コネクターからチップへの配線が困難
  - シリアルであれば2本で済む

《パラレル転送方式》

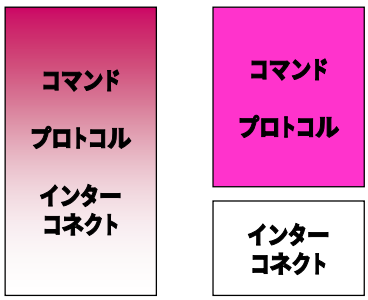


《シリアル転送方式》



SCSIの進化：SCSI-3

- コンピュータ本体とハード・ディスクなどの記憶装置の接続に用いられるSCSI規格の一つで、SCSI-2を改良・拡張した第三世代の規格群
- パラレル転送方式を用いていた従来のSCSI規格群(パラレルSCSI)の延長に当たるUltra SCSI等の仕様に加え、Fibre Channel、SSA、IEEE 1394など、シリアル転送方式を採用したシリアルSCSI規格群を新たに制定
- SCSI-3アーキテクチャモデル
  - コマンド・レベル
    - デバイス個別のコマンド
    - 共通コマンド (SCSI-3 Primary Command : SPC)
  - プロトコル・レベル
    - デバイスを接続するプロトコル
  - インターコネクト・レベル
    - 物理媒体
    - シリアル・インタフェースの追加



パラレルSCSI-3規格

- Ultra SCSI規格
  - 「Ultra SCSI」
    - 8ビット幅、20MB/s
  - 「Wide Ultra SCSI」
    - 16ビット幅、40MB/s
  - 駆動方式は従来のシングルエンド駆動のままだったため、最大接続長が8台接続時で1.5mまで、4台接続時で3mまでと短くなってしまった
- Ultra2 SCSI規格
  - 「Ultra2 SCSI」「Wide Ultra2 SCSI」では、駆動方式をディファレンシャル方式に改め、それぞれ40MB/s、80MB/sという高い転送レートを維持しながら、最大接続長を12m(LVD)～25m(HVD)に伸ばすことに成功
- Ultra3 SCSI規格
  - 転送速度を160MB/sに高めた「Ultra3 SCSI」(Ultra160 SCSI)や、320MB/sの「Ultra320 SCSI」などの仕様が策定されている

SCSI規格：LVDとHVD

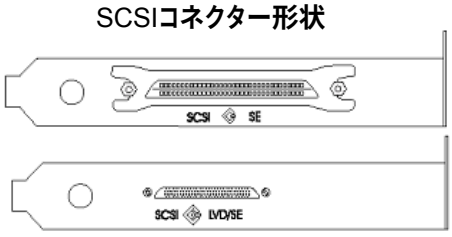
SCSIインタフェース	SCSIバス転送速度	説明
HVD (High Voltage Differential)	40MB/sec	＋と－の2本の信号線でデータを転送する方式である 信号線への供給電圧は5Vを使用します
LVD (Low Voltage Differential)	80MB/sec	＋と－の2本の信号線でデータを転送する方式である 信号線への供給電圧は3.3Vを使用します
SE (Single-Ended)		1本の信号線でデータを転送する方式である

注） テープドライブ自体の転送速度はドライブによって決まります
注） 同一SCSIバス上に、LVD/SEおよびHVD SCSIアダプタ、テープドライブ、ターミネータを混在しないでください。これらの機器が破壊されるおそれがある。

高密度/ハーフピッチ68ピン
HD68（High Density 68 pin）

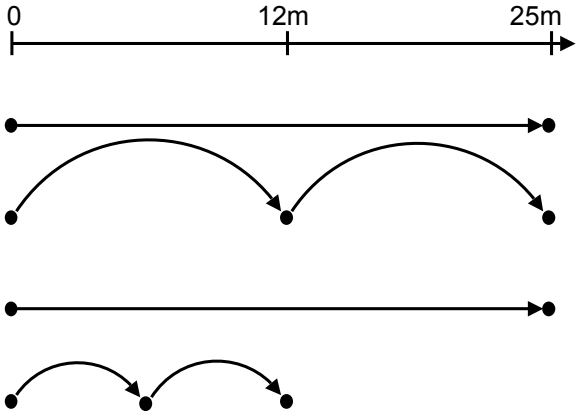
0.8mm VHDCI

SCSIコネクター形状



SCSIケーブル長の制限事項

- HVDの場合
  - Point-to-point接続
    - 全長25mまで
  - マルチドロップ接続
    - 全長25mまで
- LVDの場合
  - Point-to-point接続
    - 全長25mまで
  - マルチドロップ接続
    - 全長12mまで



SCSI規格のまとめ

	Pケーブル (68)	Qケーブル (68)	データ巾 兼 SCSI ID数	転送モード					
				非同期	同期	Fast	Ultra	Ultra2	Ultra3
Narrow	○	-	8	1.5MB/s	5MB/s	10MB/s	20MB/s	40MB/s	80MB/s
Wide	○	-	16	-	10MB/s	20MB/s	40MB/s	80MB/s	160MB/s

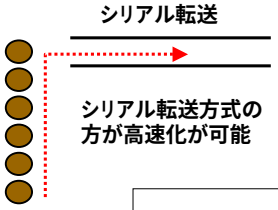
※ コマンドやステータスは常に非同期転送モードで伝送される

ファイバーチャネル
SCSI over Fibre Channel

シリアルSCSIの登場

パラレルSCSI			
速度に限界	距離に限界	デバイス数が少ない	ホストが1台 (DAS)

SCSI-3



シリアル転送方式の方が高速化が可能

	シリアルSCSI			
	高速化	長距離化	デバイス数が多い	平等化 ネットワーク化
Fibre Channel	200 MB/s	10 km	2^{24}	SAN

Fibre Channel規格

- 高速シリアル・インターフェイス・テクノロジ
  - 送信と受信の2本の転送路を使用
- いくつかのオープンな標準から構成
  - メディアと物理インターフェイス (光ファイバーと銅線)
  - データ転送、リンク・サービス、信号プロトコル
  - 上位レベル・プロトコルのマッピング(異なるコマンド・セット: SCSI, FICON, HIPPI, IP, IPI-3, ATM, 他)
- FCP上のSCSIプロトコルは、従来のパラレルバス上のSCSIプロトコルと同一

FCP とは

- ストレージエリアネットワーク (SAN) を可能にする基礎の技術
  - 高速データ転送
    - 100-200MB/sec (全二重)
    - 200-400MB/sec (全二重)
  - 最長10kmの接続距離
    - 銅線：30m
    - ショートウェーブファイバーケーブル：500m
    - ロングウェーブファー場ケーブル：10km
      - リピーターを用いることで最大100 Kmまで延長可能
  - 大規模で拡張性のある構成のサポート
    - Point-to-point
    - Arbitrated Loop
    - ファブリックスイッチ
  - コントローラあたりのデバイス数の増加
    - FC-ALで最大127台まで
    - ファブリックスイッチで最大1600万台まで
  - ネットワーク可能
    - パラレルSCSIはサーバーとストレージを接続する専用のパラレルバス
  - ホット・プラグ可能

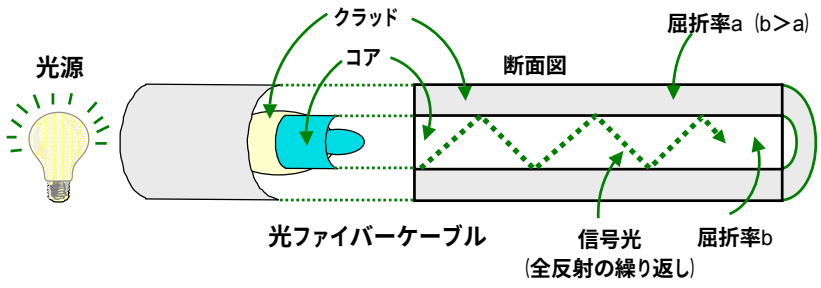
ファイバーチャネル規格：FC-0

ケーブル、コネクター、トランスミッター、レシーバーなどの媒体の物理的な特性を規定
- 光メディア、または、銅メディア上で動作
- 銅メディアはファイバーチャネルディスクドライブで主に使用されている

FC-4	プロトコル・マッピング層
FC-3	共通サービス
FC-2	フレーミング・プロトコル/フロー制御
FC-1	エンコード/デコード
FC-0	物理層

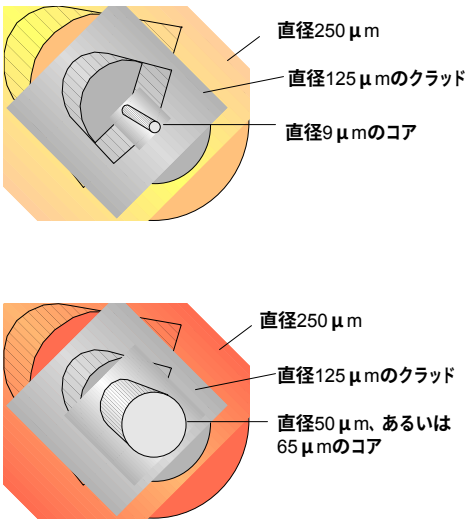
1Gbps（1.0625Gb/s）
　エントリ製品向け
2Gbps（2.125Gb/s）
　現時点での主流
10Gbpsの登場が待たれている

光ファイバーの構造



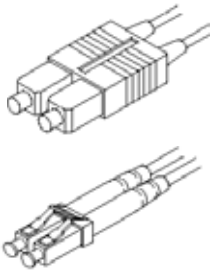
ファイバーの種類

- シングル・モード・ファイバー (SMF)
  - 直径10μm以下の1つのモードのみを伝送するファイバー
  - ケーブル中を進んでいくレーザー光は1つのモードしか存在しないので分散を起こすことなく光パルスを高速に伝えられ長距離伝送ができる
  - ファイバーの材料には純度の高い石英が使用されており折り曲げに弱く高い加工技術も必要
  - LXレーザ
- マルチ・モード・ファイバー (MMF)
  - 直径50μm、あるいは、62.5μmの異なるモードが混在するファイバー
  - 伝送距離も短く低速伝送
  - 材料としてプラスチックを使っているので安価でかつ折り曲げにも強く加工しやすい
  - LED、SXレーザ



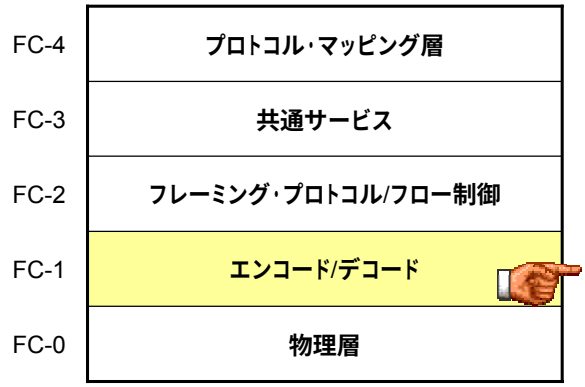
ファイバーケーブル・コネクター

- SCコネクター
  - Gigabit EthernetやFibre Channelで使用される標準的なコネクター
    - IEC 60874-14規格
  - 光ファイバー1芯毎に1つのコネクターが必要
    - 送信用 (TX)、受信用 (RX) の計2本のケーブルを接続
- LCコネクター
  - Gigabit EthernetやFibre Channelで使用される小型コネクター
  - PUSH/PULL着脱可能
- MT-RJコネクター
  - Gigabit Ethernetで使用される小型コネクター
  - PUSH/PULL着脱可能
  - 2芯ケーブルを使用するため1つのコネクターで送信・受信可能



ファイバーチャネル規格：FC-1

エンコーディング・スキームを定義
　データ転送のための同期を取るために使用される

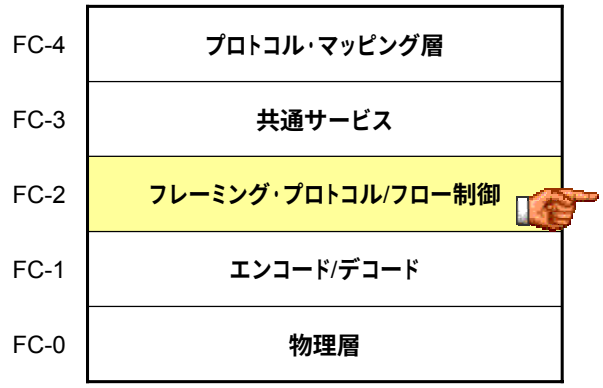


8b/10b符号化アルゴリズム
1バイトが10ビットで表現される

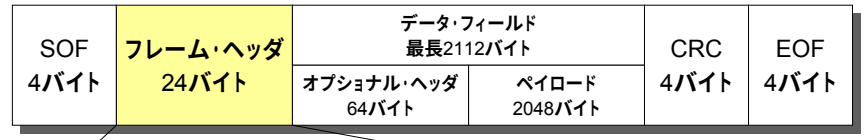
1Gbps　→　100MB/s
2Bbps　→　200MB/s

ファイバーチャネル規格：FC-2

フレーム構造とバイト順を含む信号プロトコルを定義
　データは、フレーム単位で転送される
　フレームは最大2112バイトの可変長



ファイバーチャネルのデータ構造



フレーム・フォーマット

Word	Bits 31-24	Bits 23-16	Bits 15-8	Bits 7-0
0		Destination ID (Address)		
1		Source ID (Address)		
2		フレーム制御		
3	シーケンスID		シーケン・スカウント	
4	Originator Exchange ID		Responder Exchange ID	
5				

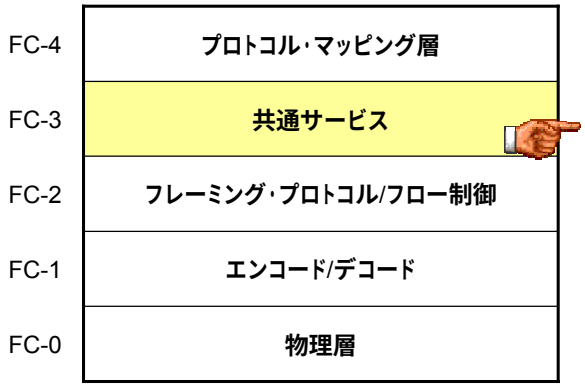
SOF Start Of Frame
EOF End Of Frame

サービス・クラス

- フロー制御はサービス・クラスにより規定される
- 現在最も利用されているのは以下の3種類
  - クラス1
    - 確認応答ありのコネクション型サービス
    - チャネルプロトコルの機能
    - 送信されるフレーム毎にACKフレームが送り返される
    - 通信するデバイスが帯域幅の全体を使用
  - クラス2
    - 確認応答ありのコネクションレスサービス
    - フレームはスイッチに伝送され、スイッチの都合の良いときに送信
    - 利用可能な帯域幅をデバイスで共有
  - クラス3
    - 確認応答なしのコネクションレスサービス
    - SAN上で最も頻繁に使用されるサービスクラス
    - トラフィックが少ないときは帯域幅をフルに使用、トラフィックが多い時は帯域幅を共有

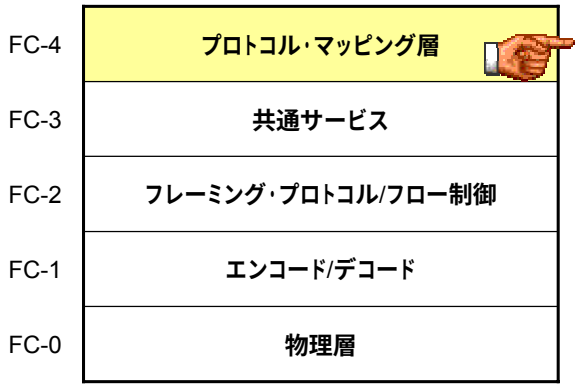
ファイバーチャネル規格：FC-3

ファイバーチャネルの一般サービス層
　ネームサーバーなど
　現時点では使用されていない



Fibre Channel上位層プロトコル

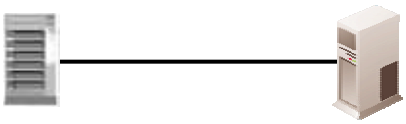
ファイバーチャネルのULPマッピング層
　SCSI、IP、HIPPI、IPI-3、ATMなどのULP (Upper Level Protocol) をファイバーチャネルを通じて伝送するための規格



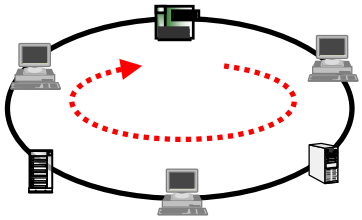
SCSI-FCP
　SCSIフレームをファイバーチャネルプロトコルにカプセル化する規格

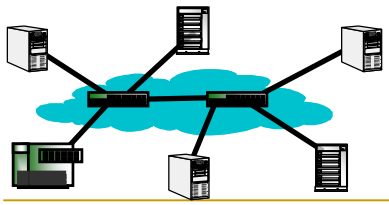
Fibre Channelの3つの接続方式

Point to Point



Arbitrated Loop

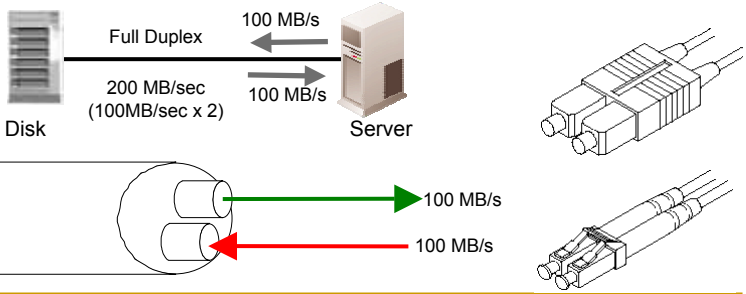




Switched Fabric

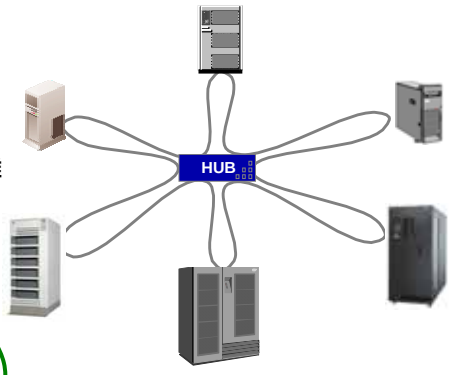
Point to Point

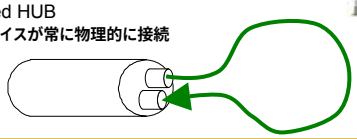
- デバイスをスイッチに直結するために主に使用される
  - 一般にイニシエータデバイスとターゲットデバイスがPoint-to-point型で接続されることはほとんどない
- アドレス指定なし
  - 送信先は常に相手側
- 初期化ルーチンは非常に簡単



FC-AL (Fibre Channel Arbitrated Loop)

- スイッチなしで127台までのデバイスを接続可能
- ループ上の全てのデバイスが1本のファイバーチャネルの帯域を共有
  - リング状にファイバーを1方向へ接続
- 8ビットのAL_PA (Arbitrated Loop Physical Address) によってデバイスを識別
- ハブとケーブルによる接続
- 性能は以下の要素に依存
  - ノード数
  - 接続距離
  - 負荷
- 2重ループを構成することで高可用性を実現
- ホットプラグ対応だが、初期化プロセスは非常に複雑
  - ループ内の全てのデバイスにAL_PAを割り当て
  - ループへのノードの追加、除去はループリセットが必要
- 2種類のHUB装置が提供されている
  - Managed HUB
    - デバイスから有効なデータが送られるまで接続しない
  - Unmanaged HUB
    - 全デバイスが常に物理的に接続

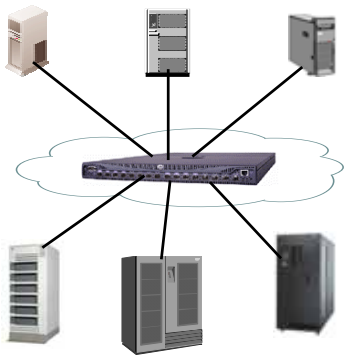




リングは常に1方向へ送信

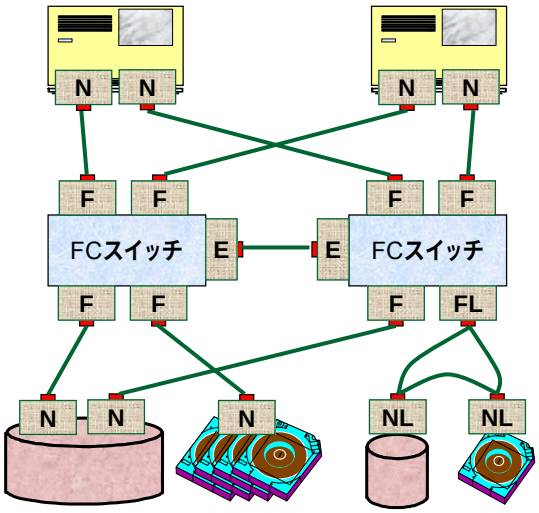
Fibre Channel Switched Fabric

- 高いデータ転送バンド幅
  - ノード当たり100-200 MB/secのバンド幅
  - 双方向転送
- ドメイン内に最大1600万ノードが接続可能
- 接続待ちは経路と負荷状態に依存
- リンクの接続距離は最長10km
- スケーラブルでフレキシブルな再構成
- Fabricの管理が必要



FC で利用されるポートの名称と種類

- F-port
  - スイッチにあり、ストレージ装置、またはサーバーのHBAとの接続に利用
- E-port
  - スイッチにあり、スイッチ間接続に利用
- N-port
  - サーバーのHBAやストレージ装置にあるポート
- NL-port
  - サーバーのHBAやストレージ装置にあるポートで、ループ・トポロジーで利用する
- FL-port
  - スイッチにあり、ループ・トポロジーで利用する



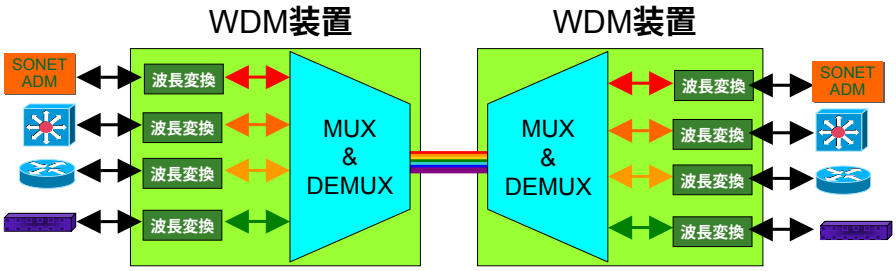
Fibre Channel HBA

- HBA (Host Bus Adapter)
  - サーバーをファイバーチャネル接続するために必須のアダプタカード
  - サーバーの各種バス毎に提供される
    - 現在主流はPCIやPCI-Xバス規格
  - 各OS用のデバイスドライバが必要
  - OSからはSCSIボードとして認識される
  - プロトコル処理などほとんどの処理はカード上で実行されるのでCPU使用率が低い
  - 送信ポートと受信ポートの1対
    - 多重ポート対応HBAも提供されているが、HBAレベルの冗長性を確保するには個別のHBAを用意する必要がある

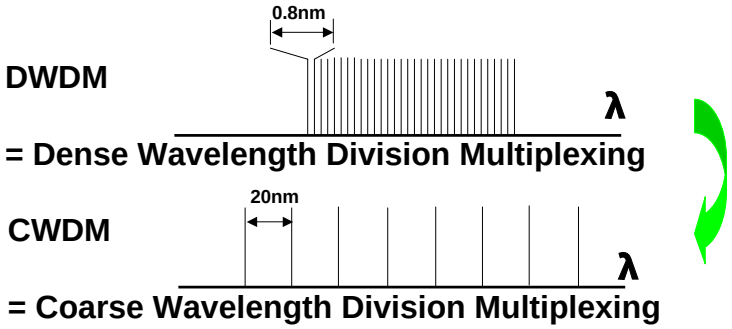
- 製品例：QLogic QLA2300シリーズ
  - 64bit 66/133MHz PCI-X対応 (標準32/64bit 33/66MHz PCI互換)
  - ファイバーチャネル転送速度最大400MB/s(Full Duplex)
  - ファイバーチャネルビットレート (1Gb/2Gb) 自動切換
  - FC-ALプライベート及びパブリックループ/Point-to-Point/スイッチファブリック接続サポート
  - FLポート及びFポートによるファブリックログイン サポート

リピータ/WDM装置

- ファイバーチャネル･リピータは、ファイバーチャネルの接続距離制限である10kmを越えるために使用される
- WDM (Wavelength Division Multiplexing)
  - 複数の波長の光信号を合波して1本の光ファイバーに挿入
  - 1本の光ファイバーを伝播してきた複数の波長の光信号を各波長ごとに分波



CWDMとDWDM



	CWDM	DWDM
多重数/伝送容量	4～16λ／4～16Gbps	64～128λ／1Tbps～
波長間隔/設置環境	波長間隔が広い(10～60nm) 光学系部品が低コスト	波長間隔が狭い(0.8nm) 光学系部品が高コスト
伝送距離/用途	10～50kmの中･短距離向け 都市内の拠点間通信	数100kmの長距離向け 都市間や国内基幹の通信

テープ装置の基礎

テープ装置選択の基準

- 信頼性
  - 信頼性の低いテープ装置を使用したのではテープバックアップの意味がありません
- 容量
  - バックアップ対象のストレージ容量に見合ったテープ容量を確保
    - 1巻あたりの容量
    - テープ・ライブラリー装置を使用し、複数テープを使用したバックアップ
- 読み書き速度
  - バックアップ・ウィンドウ内にバックアップが終了するのか
  - 転送速度は
  - 並列化（複数ドライブの使用）は可能か

代表的なテープ規格

- LTO
  - Linier Tape Open
  - Seagate、HP、IBMが作成した規格
  - オープン･システムにおける業界の標準になりつつある
  - LTO-G1、LTO-G2の2種類がある
- DLT / Super DLT
  - LTOが出るまで、オープン･システムにおける事実上の業界標準テープ 規格
  - Quantam社がドライブを一括製造
  - DLT、Super DLTなどがある
- その他
  - AIT, PetaSite、4mm/8mm DAT、QIC

テープ記録方式の比較

ヘリカル･スキャン(Herical Scan)

- 回転ヘッド
- トラックはテープに対して傾いている
- 高密度化が容易 (メディアのサイズが小さい傾向)
- Start/Stopのパフォーマンスが良くない
- テープの走行方向は終始一方向
- メディア、ヘッドが比較的劣化しやすい

→ 8mm, 4mm DAT DTF AIT PetaSite など

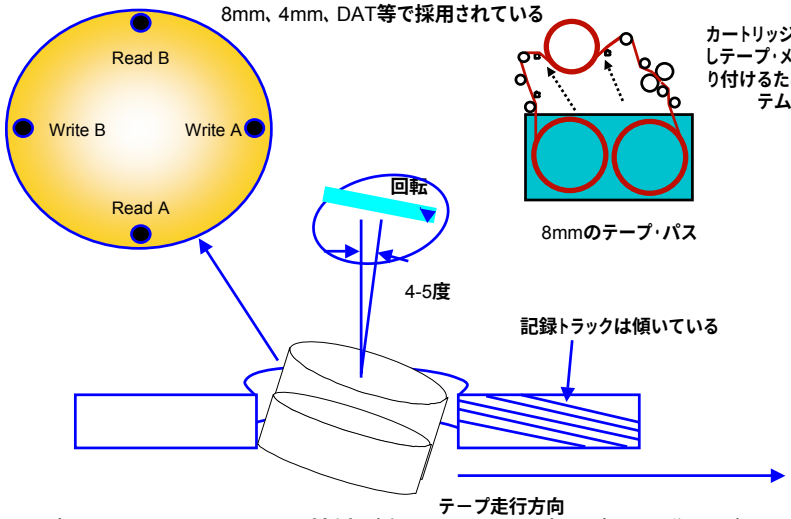
ロンギチューディナル(Longitudinal)

- 固定ヘッド (垂直移動のみ)
- トラックはテープの走行方向にに対して水平
- 高密度 (メディアのサイズが比較的大きい)
- Start/Stopのパフォーマンスが良い
- (多くの場合) テープの走行方向は双方向化可能
- メディア、ヘッドが比較的劣化しにくい

→ DLT IBM 3480 IBM 3490 IBM 3590 IBM 3592 TS1120 LTO など

ヘリカル･スキャン方式

8mm、4mm、DAT等で採用されている

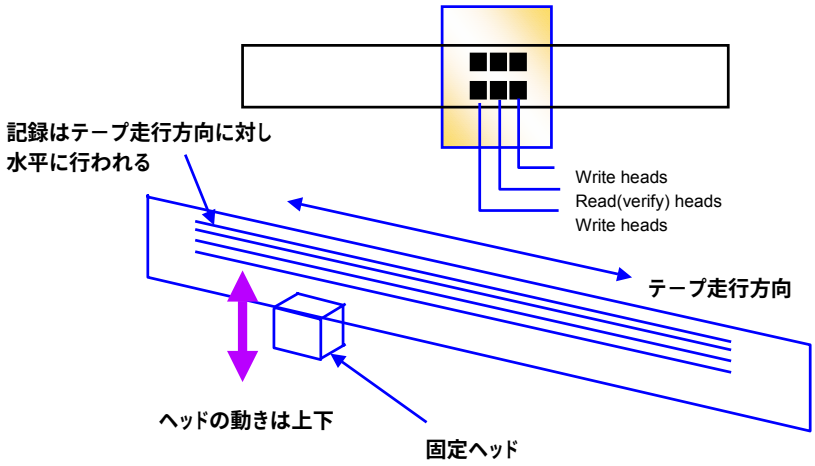


カートリッジからテープを引き出しテープ･メディアをヘッドにこすり付けるためローディング･システムが複雑になる

8mmのテープ･パス

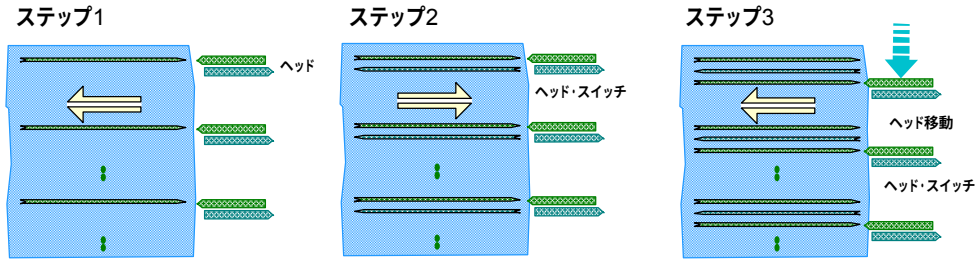
メディアとヘッドやローラーの接触が多いため、テープ･メディアの傷みが早い

ロンギチューディナル方式 (LTOで採用)



メディアとヘッドやローラーの接触が少ない、耐久性の高いテープ装置が開発可能

マルチトラックロンギチューディナル記録方式

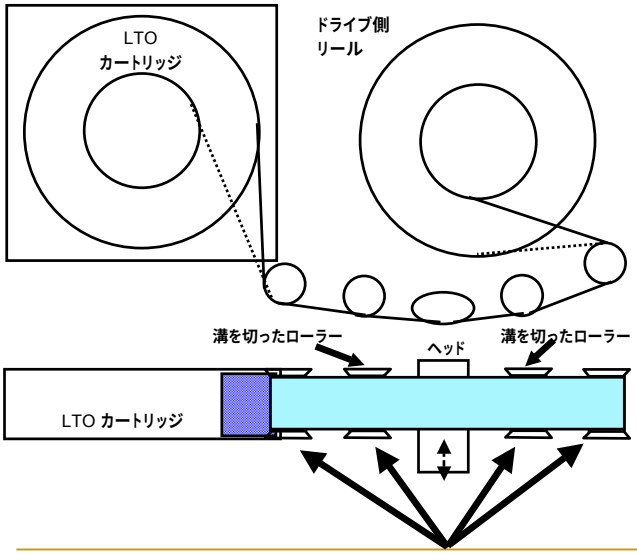


- 1時点では、一方向にのみ16トラックでライト
- 両方向にリード／ライト可能
- ヘッドの移動によりトラックを移動させる
- IBM 3480 ： 18トラック (200MB/巻)
- IBM 3590-E ： 256トラック (30GB/巻)
- IBM LTO Ultrium： 384トラック (100GB/巻)
- IBM LTO Ultrium 2： 512トラック (200GB/巻)
- IBM LTO Ultrium 3： 704トラック (400GB/巻)

特長

- 高信頼性
  - シンプルなテープ・パス
  - ヘッド、メディアのダメージを軽減
- 高速なストップ・スタート処理
- 記録容量を増加

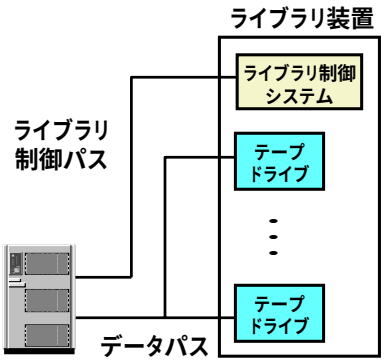
LTOの先進技術：Surface Control Guiding



- テープ表面を利用してテープ走行のずれを修正
- 旧来の方法 (Edge Guiding) と比較して、テープ・エッジへの負荷やダメージを大幅に削減。
- テープ走行の高速化と高密度化に大きく貢献

テープ･ライブラリー装置の概要

- 複数のテープドライブとライブラリ制御システム接続から構成



テープ装置の接続インタフェース

- SCSI LVD
  - サーバーのローカルバックアップやバックアップサーバー直付け
  - IAサーバーで主に使用されているSCSIインタフェース
- SCSI HVD
  - サーバーのローカルバックアップやバックアップサーバー直付け
  - UNIXサーバーで主に使用されているSCSIインタフェース
- ファイバーチャネル
  - 主にSAN接続
  - 主にテープ･ライブラリー装置向け

ちょっと休憩