

RAIDコントローラー (オンボード/オプション)

SAS/SATA/NVMe

Type-C

本ガイドについて

本ガイドは、RAID コントローラに関し、ユーザズガイドでは説明しきれない詳細な情報を記載しております。

目次

第1章 基礎知識編	4
1.1 RAID	4
1.1.1 スプリットシークによる高速化	4
1.1.2 パックによる大容量化	4
1.1.3 冗長構造による高信頼性	4
1.2 RAID コントローラ	5
1.2.1 メザニンカードタイプ	5
1.2.2 OCP カードタイプ	6
1.2.3 PCI カードタイプ	7
1.3 PCI 規格	8
1.3.1 PCI Express	8
1.3.2 PCI ボードのサイズ	8
1.4 物理デバイスのインタフェース	9
1.4.1 SAS 規格	9
1.4.2 SATA 規格	9
1.4.3 SAS/SATA 比較	9
1.4.4 SAS/SATA の奨励用途	10
1.4.5 セクタサイズ	10
第2章 機能編	11
2.1 RAID システム構築機能	12
2.1.1 ディスクアレイ (パック)	12
2.1.2 ディスクアレイ (パック) の構成ルール	12
2.1.3 RAID の種類 (RAID レベル)	13
2.2 初期化	18
2.2.1 ノーマルイニシャライズ (NI) とファストイニシャライズ (FI)	18
2.2.2 バックグラウンドイニシャライズ (BGI)	19
2.2.3 ノーマルイニシャライズの時間の目安	19
2.3 再構築 (リビルド)	20
2.3.1 ホットスペア (スタンバイ) リビルド	20
2.3.2 ホットスワップリビルド	20
2.3.3 リビルド時間目安	20
2.3.4 スペアドライブ	20
2.4 メディアチェック機能	21
2.4.1 メディアチェック機能の種類と違い	22
2.4.2 整合性チェック時間目安	22
2.5 キャッシュ機能 Flash-Backed write cache (FBWC)	23
2.5.1 無効 (WriteThrough)	23
2.5.2 有効 (WriteBack)	23
2.5.3 フラッシュバックアップユニット	23
2.6 アレイの拡張	24
2.6.1 アレイの拡張説明	25
2.6.2 アレイの拡張時間目安	26
2.7 コピーバック機能	27
2.8 スマーターコピーバック機能	28
第3章 ハードウェア編	29
3.1 RAID コントローラ製品	29
3.2 RAID コントローラの仕様	30
3.3 フラッシュバックアップユニットの仕様	30
3.4 物理デバイス選定における確認事項	30
3.5 注意事項	30
第4章 ソフトウェア編	31
4.1 RAID コントローラのソフトウェア	31

4.2 オフラインユーティリティ	31
4.3 オンラインユーティリティ	31
4.4 RAID コントローラのソフトウェアでの注意事項	32
4.5 ソフトウェアダウンロードサイト	32
第5章 運用編	33
5.1 性能比較	33
5.2 RAID レベルの比	33
5.3 RAID システムの構築	34
5.4 安定運用のために	35
5.4.1 パトロールリード	35
5.4.2 整合性チェック	35
5.4.3 RAID システム管理ユーティリティの使用について	35
5.4.4 RAID コントローラ用ドライバ、RAID システム管理ユーティリティのアップデート	36
5.4.5 RAID 構成物理デバイス台数の設定による保守運用性の向上	37

第1章 基礎知識編

1.1 RAID

サーバを構成する部品の中でハードディスクドライブは機械的な動作を伴う為に非常にデリケートです。その上 CPU やメモリなど他の構成部品と比べ桁違いに動作速度が遅いのも特徴です。RAID (Redundant Array of Independent Disks)とは複数台のハードディスクドライブを用いて、I/O 処理を分散する事で高速化し、データとそのパリティを分散して格納する事で大容量化・高信頼性を確保する技術です。

1.1.1 スプリットシークによる高速化

ハードディスクドライブは機械的な動作を伴うために、CPU やメモリに比べると桁違いに遅くなります。しかし、速度の遅い機械的な動作でも複数台のハードディスクドライブを用いて同時におこなう事(スプリットシーク)でファイルの I/O 性能を向上させることができます。

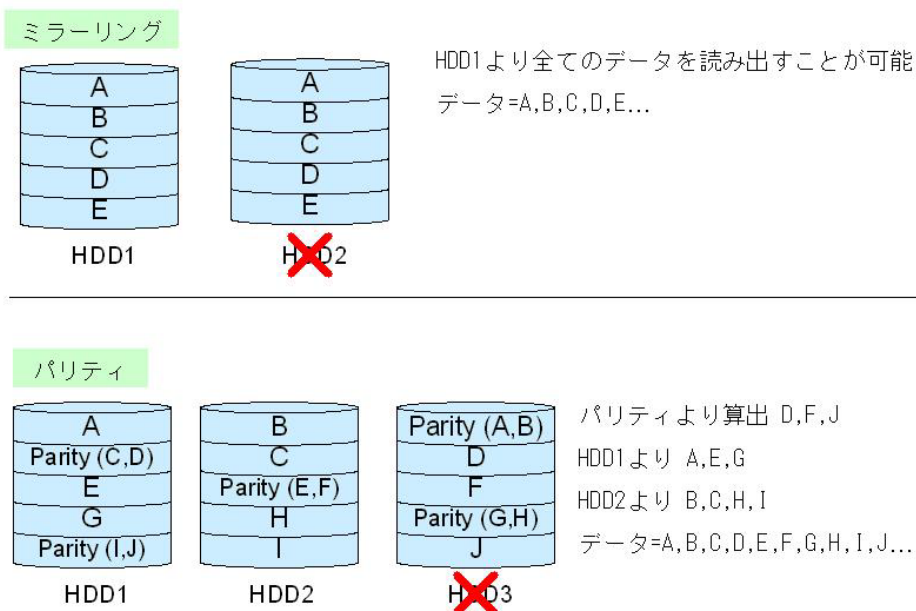
1.1.2 パックによる大容量化

複数台のハードディスクドライブを単一ドライブとして認識させる(パックする)ことで、大容量のドライブを構成することができます。50TB のハードディスクドライブを一台準備することは現時点では難しいですが、10TB のハードディスクドライブを 5 台準備することで 50TB のドライブを構成することができます。

1.1.3 冗長構造による高信頼性

格納されているデータとそのパリティを保存することで、論理ドライブに冗長性を持たせることができます。冗長構造を持つ論理ドライブを構成することで、ハードディスクドライブに障害が発生してもシステムを停止せずに復旧作業をおこなうことができます。

(例) 1 台のHDDに障害が発生した場合



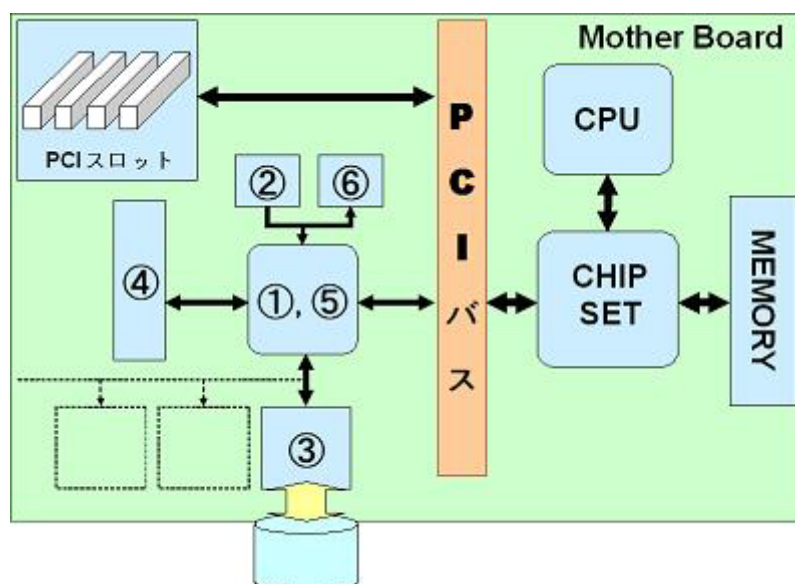
1.2 RAID コントローラ

RAID コントローラは複数のハードディスクドライブにて構成される RAID システムに対し、パリティ計算やデータ読み出し/書き込み等の処理を行う専用ハードウェアです。

- 高信頼性の RAID5、RAID6 もサポートしています。
- ほとんどの RAID 処理を専用のプロセッサで実行するため、本体装置の CPU やメインメモリ等のリソースに与える影響は小さくなります。
- RAID のシステム構成はハードディスクドライブにのみ記憶されます。そのため、故障したハードディスクドライブの状態によっては、RAID システムの回復に影響を及ぼす可能性があります。
- RAID のログ情報は本体装置に保存されます。これにより、障害発生時からの復旧や障害発生原因の解析を容易に行う事ができます。

1.2.1 メザニンカードタイプ

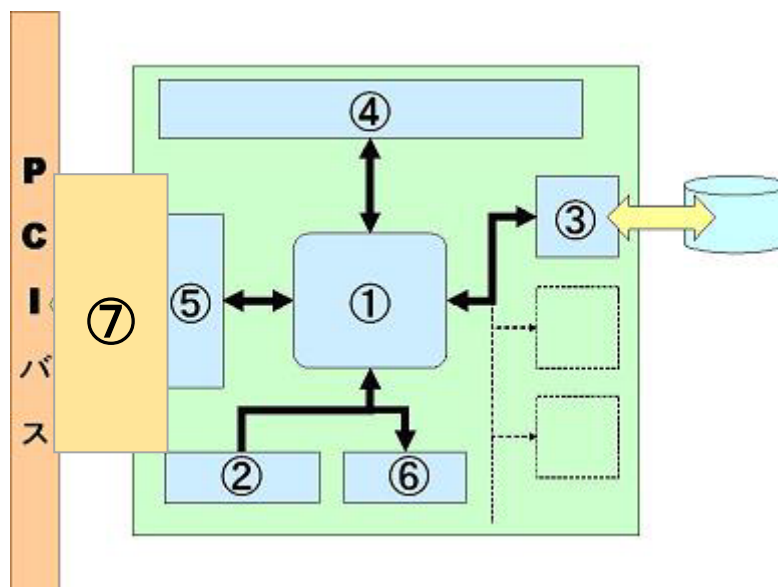
メザニンカードタイプの RAID コントローラは、PCI スロットを介さずにマザーボード上に実装しています。ほとんどの RAID 処理を専用マイクロプロセッサにて行うため、システムのパフォーマンスに影響を与えません。



①	マイクロプロセッサ(MPU)	サポートする RAID レベルに必要な処理を全て担う専用プロセッサ
②	Flash ROM	マイクロプロセッサを制御するソフトウェアを格納するメモリ
③	インタフェースコントローラ	RAID コントローラに接続する各種ハードディスクドライブに対応したインタフェースを制御するコントローラ
④	メモリ	パリティ処理やハードディスクドライブへの読み出し、書き込みに使用するキャッシュメモリ
⑤	PCI ブリッジ回路	RAID コントローラと PCI バスを接続するためのバスインタフェース
⑥	NvRAM	RAID システム構成、設定情報を記録するメモリ

1.2.2 OCP カードタイプ

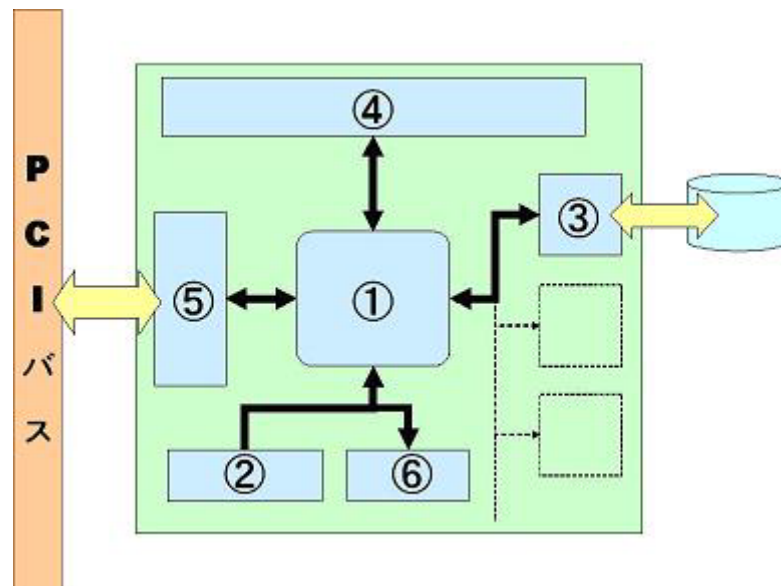
OCP カードタイプの RAID コントローラは、RAID 処理専用マイクロプロセッサを搭載しており、ほとんどの RAID 処理を RAID コントローラ単体でおこなうためシステムのパフォーマンスに影響を与えません。



①	マイクロプロセッサ(MPU)	サポートする RAID レベルに必要な処理を全て担う専用プロセッサ
②	Flash ROM	マイクロプロセッサを制御するソフトウェアを格納するメモリ
③	インタフェースコントローラ	RAID コントローラに接続する各種ハードディスクドライブに対応したインタフェースを制御するコントローラ
④	メモリ	パリティ処理やハードディスクドライブへの読み出し、書き込みに使用するキャッシュメモリ
⑤	PCI ブリッジ回路	RAID コントローラと PCI バスを接続するためのバスインタフェース
⑥	NvRAM	RAID システム構成、設定情報を記録するメモリ
⑦	OCP コネクタ	OCP コネクタ

1.2.3 PCI カードタイプ

PCI カードタイプの RAID コントローラは、RAID 処理専用マイクロプロセッサを搭載しており、ほとんどの RAID 処理を RAID コントローラ単体でおこなうためシステムのパフォーマンスに影響を与えません。



①	マイクロプロセッサ(MPU)	サポートする RAID レベルに必要な処理を全て担う専用プロセッサ
②	Flash ROM	マイクロプロセッサを制御するソフトウェアを格納するメモリ
③	インタフェースコントローラ	RAID コントローラに接続する各種ハードディスクドライブに対応したインタフェースを制御するコントローラ
④	メモリ	パリティ処理やハードディスクドライブへの読み出し、書き込みに使用するキャッシュメモリ
⑤	PCI ブリッジ回路	RAID コントローラと PCI バスを接続するためのバスインタフェース
⑥	NvRAM	RAID システム構成、設定情報を記録するメモリ

1.3 PCI 規格

PCIとは PCI SIG(PCI Special Interest Group)により策定されているバスアーキテクチャのことです。

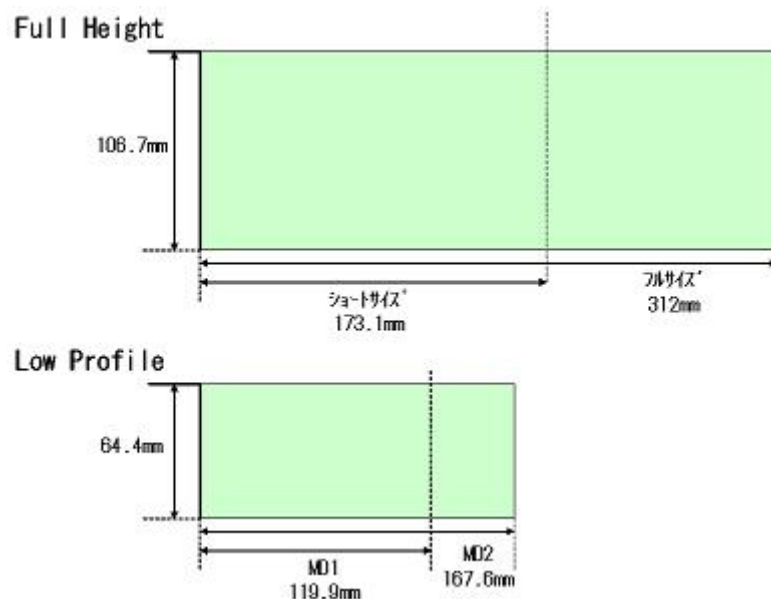
従来の PC 互換機にて最も多く使用されてきた拡張バス ISA(Industry Standard Architecture)と比べ、機能面や性能面でも優り、現時点での業界標準となっています。

1.3.1 PCI Express

2002 年に PCI-SIG によって策定された、PCI バスに代わるパソコン、サーバ向けシリアル転送インタフェースであり、3GIO を標準規格化したものです。PCI バスはパラレル転送方式を使用しており、PCI Express との間に物理レベルでの互換性はありませんが、通信プロトコルなどは共通のものが使われています。最小構成の伝送路(レーン)は片方向 2.5Gbps(双方向 5.0Gbps)の全二重通信が可能で、8 ビットのデータを送るのにクロック信号など 2 ビットを追加した 10 ビットを費やすため、実効データ転送レートは片方向 250MB/s(双方向 500MB/s)です。また、片方向 500MB/s(双方向 1Gb/s)に向上した PCI Express 2.0、片方向 1Gb/s(双方向 2Gb/s)に向上した PCI Express 3.0 も策定されています。実際の PCI Express ポートはこのレーンを複数束ねた構成になっていることが多く、1 レーンで構成された PCI Express ポートを「PCI Express x1」、2 レーンのポートを「x2」といった具合に呼称します。

1.3.2 PCI ボードのサイズ

PCI ボードとは PCI 規格に適合した拡張ボードのことを指します。PCI ボードには物理的寸法においてもボード幅およびボード長にさまざまな種類があります。



1.4 物理デバイスのインタフェース

物理デバイスの種類にはハードディスクドライブ(HDD)とソリッドステートドライブ(SSD)があり、HDD は磁性体が塗布されたアルミ合金やガラスの円盤(磁気ディスク)を複数枚かさね磁気的にデータを入出力する記憶装置、SSD は不揮発性のフラッシュメモリを用いた記憶装置です。これらを本体装置や RAID コントローラと接続するインタフェース規格として、SAS(Serial Attached SCSI)規格、および SATA(Serial ATA)規格があります。

1.4.1 SAS 規格

SCSI をシリアル化したのが SAS です。SAS のデータ転送速度は 300MB/s(3Gb/s)で Ultra320 SCSI のバス当たり 320MB/s より遅いように思えますが、SAS はポイント・ツー・ポイント接続なので、1 台のデバイスが 300MB/s を占有できます。それを考慮すれば、Ultra320 SCSI より高速と言えます。SAS は拡張性も非常に高く、SAS のホスト・コントローラとエンド・デバイス(例えばハードディスクドライブ)は、中継デバイスとなる「SAS エキスパンダ」を通じて接続できます。小規模な構成ならホスト・コントローラを中心としたスター型、大規模構成なら複数の SAS エキスパンダを利用したツリー型トポロジを採れます。デバイス間の距離は最長 8m(外部ケーブル)で、接続可能なエンド・デバイス数は、規格上 1 万 6384 台までとなっています。また、規格の拡張にともない、最大転送速度が 1200MB/s(12Gb/s, SAS3.0)と性能面が向上しています。

1.4.2 SATA 規格

IDE をシリアル化したのが SATA です。IDE とくらべ SATA はケーブルやコネクタなどの物理的な仕様が大きく変更されています。ケーブルは 7 芯のシリアル伝送用のケーブルに変更され、規格が保証する線長は IDE 規格の 45cm であったのに対し、SATA では 100cm まで保証されています。信号線の接続形態としては IDE 規格では 1 本の信号線に対し 2 台までの IDE 機器を接続する事が可能でしたが、SATA では 1 本の信号線に対し 1 台しか接続できません。その分電気的な特性は向上しており、最大転送速度は 150MB/s(1.5Gb/s)と IDE 規格よりも速く、後継の Revision では最大転送速度が 600MB/s(6Gb/s)となり、性能面が向上しています。

1.4.3 SAS/SATA 比較

特徴	SAS	SATA
Scalability (拡張性)	1 対 1 接続のため電気特性が高い。 (ノイズに強い)	
	1ch につき最大 1 台まで接続可能※1	1ch につき最大 1 台まで接続可能
Performance (能力)	1ch あたり最大転送速度 12Gb/s	1ch あたり最大転送速度 6Gb/s
	回転数 10,000～15,000rpm	回転数 5,400～7,200rpm
	複数同時処理能力が高い (コマンドキューイング機能あり)	複数同時処理能力が高い (コマンドキューイング機能がある製品に限り)
Reliability (信頼性)	リアサイン機能等の異常時のリカバリ処理能力が高く エラー発生時のステータス情報が豊富。	リカバリ処理能力が低く、 エラー発生時のステータス情報が乏しい。
	メディア記録密度が低い分、塵等の影響に強く Head マージンも高くなるためエラーが少ない。	
Maintainability (保守性)	活栓挿抜可能な標準インタフェースを HDD に持つため、システム稼動中に HDD 交換が可能。	
Cost (費用)	メディア記録密度が SATA に比べ低い GB あたりの単価が高い	メディア記録密度が SAS に比べ高い GB あたりの単価が高い

※1: SAS エキスパンダを利用することで、1ch に複数台接続可能

1.4.4 SAS/SATA の奨励用途

物理デバイスのインタフェース	奨励用途	主な理由
SAS	データベースサーバなどのハイエンド環境	高い性能、および信頼性を有する
SATA	ワークグループサーバなどの SOHO 環境	価格メリットを活かし、用途に応じサーバを使い分けるシステム展開が可能

1.4.5 セクタサイズ

HDD のセクタサイズには、媒体上に書き込まれるデータ単位、I/F 上でやり取りされるデータ通信単位により、512 native(512n)、4K native(4Kn)、512 emulation(512e)の 3 種類に分けられます。本タイプの RAID コントローラは 4Kn をサポートしていません。

		媒体上の記録単位(Byte)	
		512	4096
データ通信単位 (Byte)	512	512n	512e
	4096	—	4Kn

4Kn/512e はより大容量の HDD を実現するための技術で、媒体上に記録するデータの単位を従来の 512n HDD より大きくしたものです。4Kn と 512e の違いは、ホスト・コントローラとのデータ通信の単位です。4Kn は対応するホスト・コントローラに接続することで、媒体上に記録されるデータ単位と同じ 4096Byte 単位で通信を行います。512e は従来のホスト・コントローラに接続可能とするため、媒体上に記録されたデータとは異なる 512Byte 単位に分割してデータ通信を行います。全ての通信が分割されるため、512e では見かけ上のメディアエラーの発生数(4Kn の 8 倍)、メディアエラーによる HDD 交換基準(160 回)は 4Kn と比べて多くなりますが、品質/性能への影響の程度は 4Kn と変わりません。

第2章 機能編

本章では RAID コントローラの機能を説明します。
各製品のサポート RAID レベル、キャッシュ容量、ポート数の情報を下記に記載します。

機能対応表

	対応 RAID レベル								キャッシュ	スぺアディスク	アレイの拡張	ポート数
	RAID 0	RAID 1	RAID 5	RAID 6	RAID 10	RAID 50	RAID 60	JBOD				
N8103-240	○	○	○	○	○	○	○	○(*)	4GB	○	○	内蔵 16 ポート
N8103-248	○	○	×	×	○	×	×	○(*)	なし	○	○	内蔵 16 ポート
N8103-249	○	○	○	○	○	○	○	○(*)	4GB	○	○	内蔵 8 ポート
N8103-250	○	○	○	○	○	○	○	○(*)	8GB	○	○	内蔵 16 ポート
N8103-251	○	○	×	×	○	×	×	○(*)	なし	○	○	内蔵 16 ポート
N8103-252	○	○	○	○	○	○	○	○(*)	8GB	○	○	内蔵 16 ポート

(*)JBOD は N8103-189/190/191/192/193/194/195/201/196/237/238/242/243/244/245/246 の HBA パススルー機能と同等です。

2.1 RAID システム構築機能

2.1.1 ディスクアレイ(パック)

ディスクアレイ(パック)とは複数の物理デバイスのグループを表し、論理ドライブを設定するための基となります。設定可能なディスクアレイ数は、本体装置やディスク増設ユニットに搭載した物理デバイスの台数、ディスクアレイの種類(RAID レベル)、および RAID コントローラの最大作成可能ディスクアレイ数により異なります。

2.1.2 ディスクアレイ(パック)の構成ルール

ディスクアレイ(パック)は以下のルールに則り構成する必要があります。

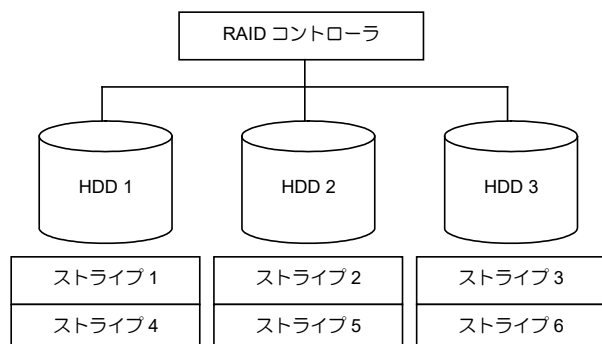
[ルール]

- 同一 RAID コントローラ配下の物理デバイスを使用して、(RAID レベルを問わず)複数のディスクアレイを組むことが可能です。
- RAID コントローラをまたいだ物理デバイスを使用して、ディスクアレイを組むことはできません。

2.1.3 RAID の種類(RAID レベル)

RAID0 (ストライピング)

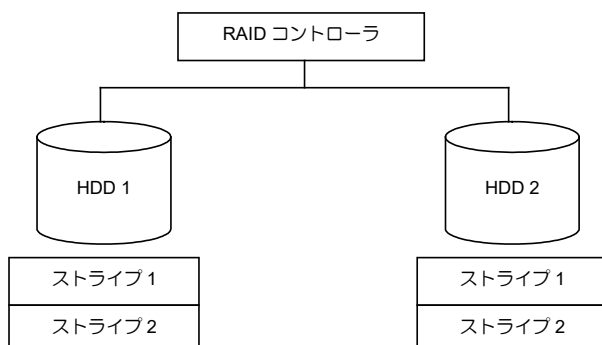
複数台の物理デバイスを単一ドライブに見立て、アクセスを分散する事で高速化、大容量化を実現します。



RAID0 の特徴	
冗長性	無し
特徴	全物理デバイス容量をデータディスクとして使用可能
	RAID レベルの中で最も高速
	冗長構造ではないため物理デバイスが故障(Dead)するとデータをロストしてしまう
使用に適した AP	クリティカルでないデータに対して高い性能を必要とする AP
ドライブ数	1 台以上

RAID1 (ミラーリング)

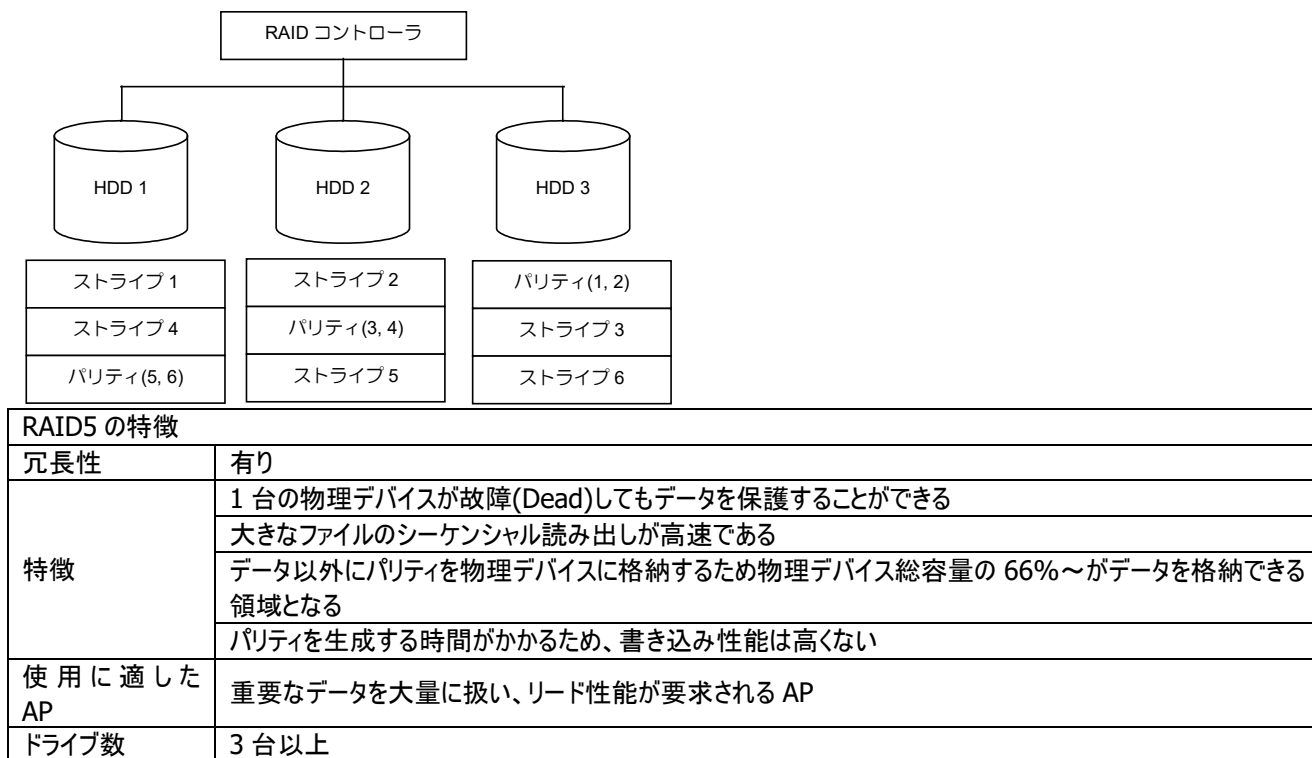
2 台 1 組の物理デバイスに対し常に同じデータを格納する事でデータを二重化し高信頼性を実現します。



RAID1 の特徴	
冗長性	有り
特徴	1 台の物理デバイスが故障(Dead)しても、もう片方の複製物理デバイスより復旧をおこなう
	2 台の物理デバイスのみで冗長性のある RAID システムを構築できるため、必要な総コストは最も低くなる
	書き込み性能は理論値で単一物理デバイスへの書き込みと比べ 1/2 になる
	データを書き込める容量は物理デバイスの総容量の 1/2 になる
使用に適した AP	論理ドライブ、重要なファイルを格納するドライブ
ドライブ数	2 台

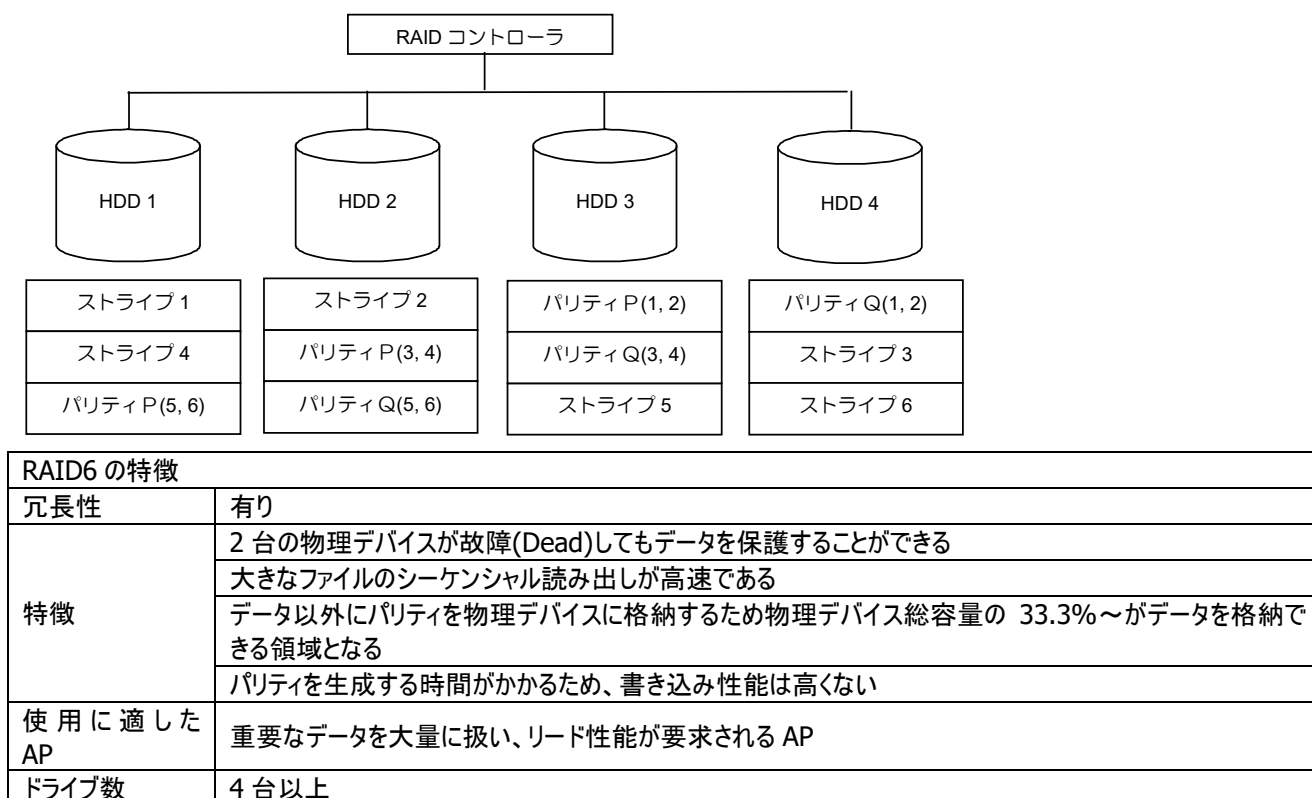
RAID5 (パリティ付きストライピング)

複数台の物理デバイスを単一ドライブに見立て、アクセスを分散します。また、保存するデータのパリティを生成し各物理デバイスに保存します。これにより高速化、大容量化および高信頼性を実現します。

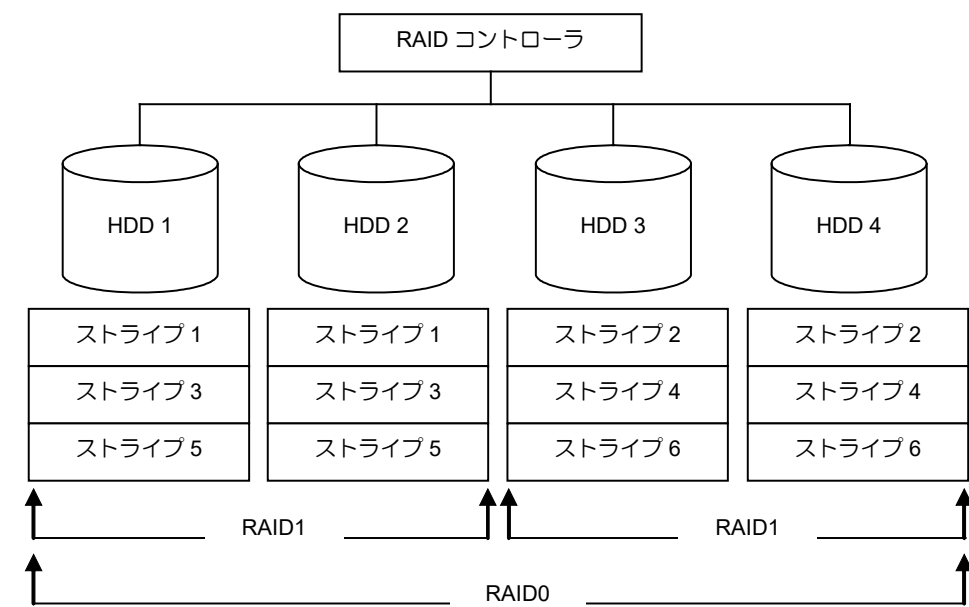


RAID6 (二重化パリティ付きストライピング)

複数台の物理デバイスを単一ドライブに見立て、アクセスを分散します。また、保存するデータのパリティを生成し、各物理デバイスに二重化して保存します。これにより高速化、大容量化および高信頼性を実現します。

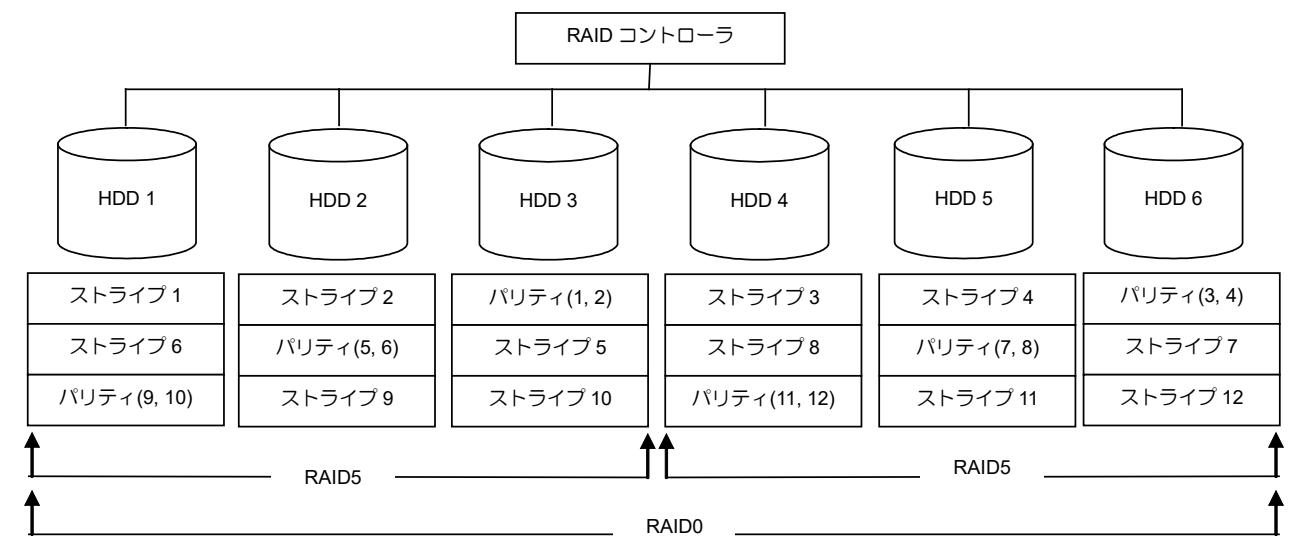


RAID10



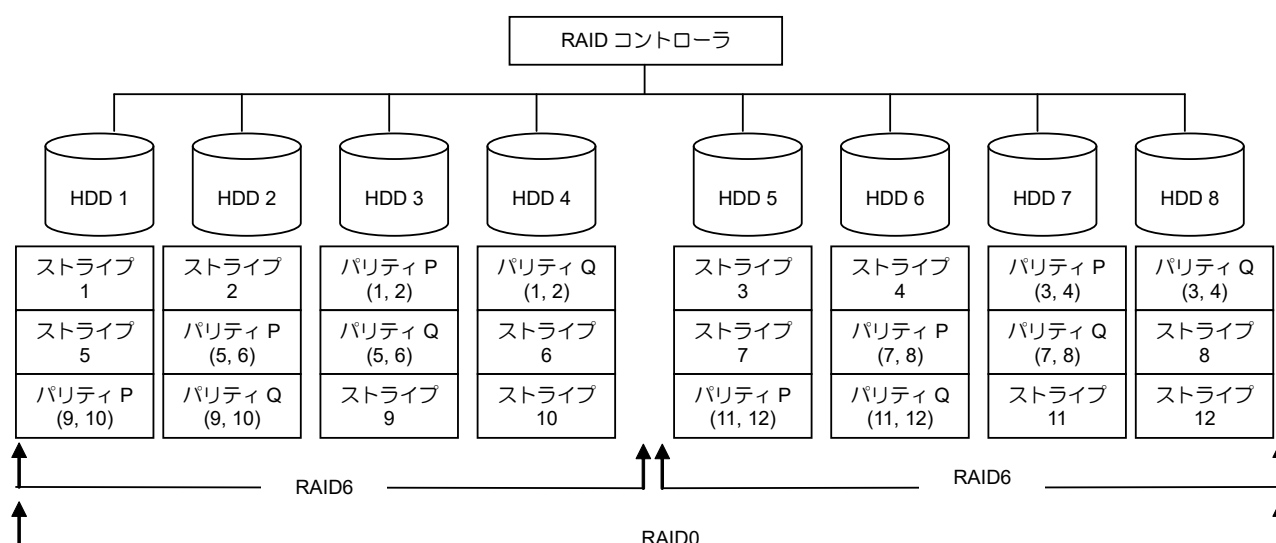
RAID10 の特徴	
冗長性	有り
特徴	RAID1 のスパン
	1～2 台の物理デバイスが故障(Dead)してもデータを保護することができる(物理デバイス 2 台故障(Dead)の場合は組み合わせによる)
	書き込み性能は RAID1 を多少上回る
	データを書き込める容量は物理デバイスの総容量の 1/2 になる
使用に適した AP	論理ドライブ、重要なファイルを格納するドライブ
ドライブ数	4 台以上の偶数デバイス

RAID50



RAID50 の特徴	
冗長性	有り
特徴	RAID5 のスパン
	1～2 台の物理デバイスが故障(Dead)してもデータを保護することができる(物理デバイス 2 台故障(Dead)の場合は組み合わせによる)
	書き込み性能は RAID5 を多少上回る 大きなファイルのシーケンシャル読み出しが高速である
	データ以外にパリティを物理デバイスに格納するため物理デバイス総容量の 66%～がデータを格納できる領域となる
使用に適した AP	重要なデータを大量に扱い、リード性能が要求される AP
ドライブ数	6 台以上

RAID60



RAID60 の特徴	
冗長性	有り
特徴	RAID6 のスパン
	2～4 台の物理デバイスが故障(Dead)してもデータを保護することができる(物理デバイス 4 台故障(Dead)の場合は組み合わせによる)
	大きなファイルのシーケンシャル読み出しが高速である
	データ以外にパリティを物理デバイスに格納するため物理デバイス総容量の 33.3%～がデータを格納できる領域となる
	パリティを生成する時間がかかるため、書き込み性能は高くない
使用に適した AP	重要なデータを大量に扱い、リード性能が要求される AP
ドライブ数	8 台以上

JBOD

論理ドライブを構築することなく、物理デバイスを直接 OS からあつかえる機能です。RAID コントローラに物理的に接続されていますが、監視は行いません。RAID コントローラのキャッシュを使用しないため、ディスクアクセス性能は物理デバイスの性能になります。

JBOD の特徴	
冗長性	無し
特徴	全物理デバイス容量をデータディスクとして使用可能
	RAID コントローラによる(IML も)監視は行わない
	冗長構造ではないため物理デバイスが故障(Dead)するとデータをロストしてしまう
	ディスクアクセス性能は物理デバイスの性能
	JBOD ドライブにはリビルド/整合性チェック/コピーバック/パトロールリードは実行できない
使用に適した AP	単体ディスクとしてディスクを認識させることを必要とする AP
ドライブ数	1 台

2.2 初期化

初期化機能とは論理ドライブを構築している物理デバイスに対し、初期化処理を行う機能です。

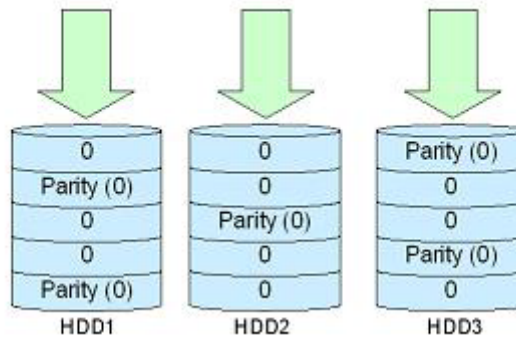
前項 2.1 に説明したように、RAID コントローラは複数の物理デバイスを論理的に結合することで論理ドライブを構築することができます。しかし、論理ドライブを構築している物理デバイスのすべてが新品であるなどの場合、物理デバイス内のデータが消去されているとは限りません。そのため、初期化機能を使用して論理ドライブを構築している物理デバイスに対し初期化処理を行います。初期化は、ノーマルイニシャライズ(NI)、ファストイニシャライズ(FI)、バックグラウンドイニシャライズ(BGI)の3種類に大別されます。

2.2.1 ノーマルイニシャライズ(NI)とファストイニシャライズ(FI)

① ノーマルイニシャライズ(NI)

ノーマルイニシャライズは論理ドライブを構築している物理デバイスの全領域に対し、0 データを書き込みます。物理デバイス内の情報は全て 0 クリアされるため、物理デバイス内にもともと保存されていた無効なデータを全て削除することができます。オール 0 データが記録されるため、パリティ情報の整合性も整った状態になります。

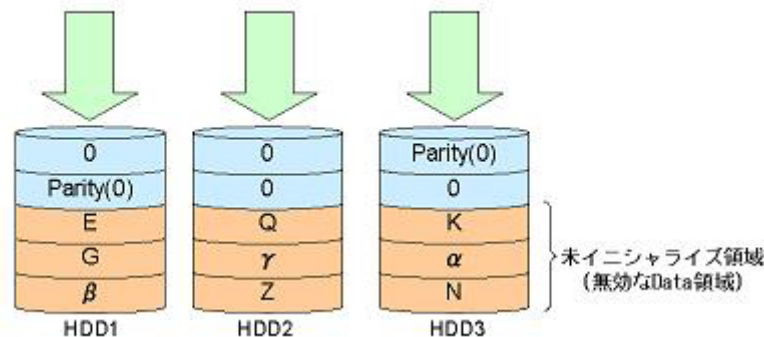
ディスクアレイ全領域に対し、“0”Data書き込み



② ファストイニシャライズ(FI)

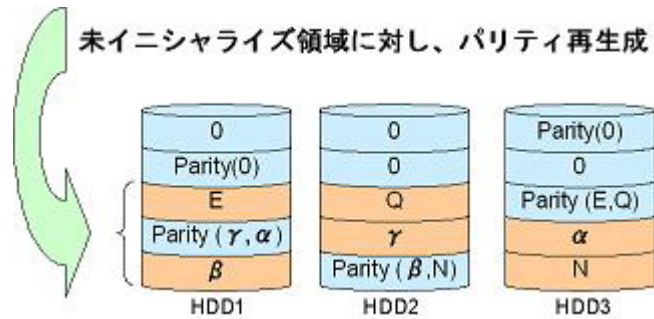
ファストイニシャライズは論理ドライブを構築している物理デバイスの先頭部分のみに 0 データを書き込みます。OS のインストール情報や、パーティション情報をクリアすることができます。ノーマルイニシャライズより早く終了するため、次の作業へすぐに移行することができます。ただし、未初期化領域が発生するため論理ドライブ全領域の整合性は整っていません。

初めの数ブロックに対し、“0”Data書き込み



2.2.2 バックグラウンドイニシャライズ(BGI)

ファストイニシャライズを実行した場合、および、ノーマルイニシャライズを中断した場合、初期化を実行していない場合、論理ドライブには未初期化領域が存在する事になります。この未初期化領域に対し、バックグラウンドでパリティ合わせを行う機能がバックグラウンドイニシャライズです。



2.2.3 ノーマルイニシャライズの時間の目安

完了までに必要な時間目安については、付録 B を参照してください。

2.3 再構築(リビルド)

再構築(リビルド)は、論理ドライブを構築している物理デバイスが故障(Dead)した場合、障害が発生した物理デバイスを正常な物理デバイスと交換することで、元の正常な論理ドライブを再構築する機能です。

RAID1 や RAID5、RAID6、RAID10、RAID50、RAID60 といった冗長性のある論理ドライブに対して実行することができます。

2.3.1 ホットスペア(スタンバイ)リビルド

冗長性のある RAID システムにて、ホットスペアをあらかじめ RAID システムに組み込み、物理デバイスの障害発生時に自動的にホットスペアを用いて実行されるリビルドをホットスペア(スタンバイ)リビルドといいます。

2.3.2 ホットスワップリビルド

冗長性のある RAID システムにて、システム稼働中でも電源を落とすことなく、障害が発生した物理デバイスを交換する、この機能をホットスワップと呼びます。そしてホットスワップにて交換された物理デバイスに対して自動的に実行されるリビルドをホットスワップリビルドといいます。

2.3.3 リビルド時間目安

完了までに必要な時間目安については、付録 C を参照してください。

2.3.4 スペアドライブ

・専用スペア (Dedicated Spare)

専用スペアは各アレイ専用に設定するスペアドライブです。

RAID1、5、6、10、50、60 などのフォールトトレラントな論理ドライブがサポートされています。

専用のスペアドライブは、アレイ内のドライブに障害が発生したときにアクティブになります。

コピーバック機能に対応しています。

・グローバルスペア (Global Spare)

グローバルスペアドライブは、次の条件が満たされた場合に、アレイ内の障害の発生したドライブを置き換えます。

- ・ ドライブの種類が同じである。
- ・ グローバルスペアドライブの容量が、障害ドライブの容量以上である。

フォールトトレラント論理ドライブ内でドライブに障害が発生すると、グローバルスペアドライブがアクティブになります。

RAID0 論理ドライブの場合、メンバードライブが予測障害を報告すると、グローバルスペアがアクティブになります。

コピーバック機能に対応しています。

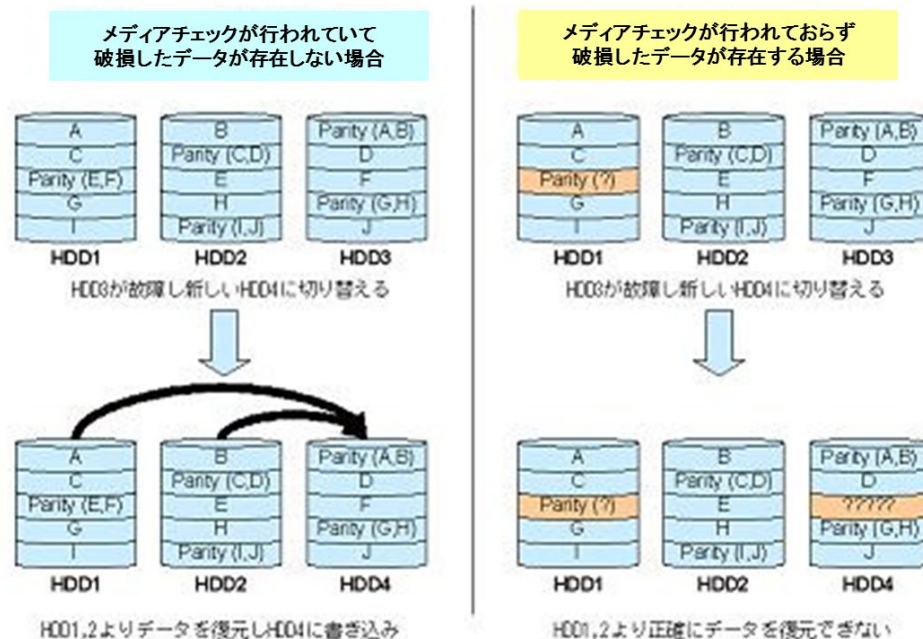
2.4 メディアチェック機能

メディアチェック機能とは RAID コントローラ配下に接続された物理デバイスの全領域を読み取り動作することにより、物理デバイス上で読み込み不能なセクタがあった場合にエラーの修正を行い、信頼性向上を図ることを目的に作られた RAID コントローラの機能です。メディアチェック機能にはパトロールリード機能と整合性チェック機能があり、違いについては「2.4.1 メディアチェック機能の種類」の項で説明します。メディアチェックを実施することにより、次の効果が期待できます。

① データ復旧時の障害を未然に防ぐ

メディアチェックが定期的におこなわれることで、全領域のリードエラーを訂正します。複数台エラーの場合はデータを復旧することができません。したがって、縮退状態が発生した際にリードエラーが発生する領域が存在しないようにする事は大切です。

例：RAID6 HDD×3台の構成にて1台のHDDにエラーが発生した場合



② データの書き込まれていない領域をチェックする

メディアチェックは論理ドライブを構成する物理デバイス(パトロールリードの場合はスタンバイのディスクも対象(*1))の全ての領域に対して、正常であるかをチェックします。これにより物理デバイスの異常を早期に発見することができます。

*1 メディアチェック機能の種類と違いは「2.4.1 メディアチェック機能の種類」の項を参照してください。

③ 物理デバイスの機械的なコンディションを整える

物理デバイスの全ての領域にチェックをおこなうことにより物理デバイスの磁気ヘッドを適度に動かすことにつながります。機械的な部分が大部分を占める物理デバイスにとって、内部の機械を定期的に動かすことは非常に大切なことです。

2.4.1 メディアチェック機能の種類と違い

メディアチェック機能にはパトロールリード機能と整合性チェック機能があります。
パトロールリード機能と整合性チェック機能の違いを以下に示します。

	パトロールリード	整合性チェック
起動条件	工場出荷時より自動実行	工場出荷時より自動実行
検査範囲	アレイ構成を組んだ物理デバイスおよびスタンバイに設定した物理デバイス（実装しただけの物理デバイスには実施されません）	冗長なアレイ構成を組んだ物理デバイスのみ（RAID1,5,6,10,50,60 を構成した物理デバイス）
性能インパクト	なし	あり
動作	ディスク単体毎にダミーリードを実施	パリティとデータの整合性チェックを実施 *1

*1 RAID1 ではミラーリングを行っている双方の物理デバイスを比較します(データの不一致を検出した場合はあらかじめ決められた物理デバイス上のデータを他方の物理デバイスに上書きすることでデータの整合性を整えることができます)。RAID5、および、RAID6 ではデータからパリティを計算し、格納済みのパリティと比較します(このパリティの不一致を検出した場合は、パリティの再生成をおこなうことでデータの整合性を整えることができます)。

パトロールリード、整合性チェックともに毎週土曜日の AM3 時に開始します。実行間隔はともに 168h(1 週間)です。実際には整合性チェック優先なので、整合性チェック完了後にパトロールリードが実施されます。

2.4.2 整合性チェック時間目安

完了までに必要な時間目安については、付録 D を参照してください。

2.5 キャッシュ機能 Flash-Backed write cache(FBWC)

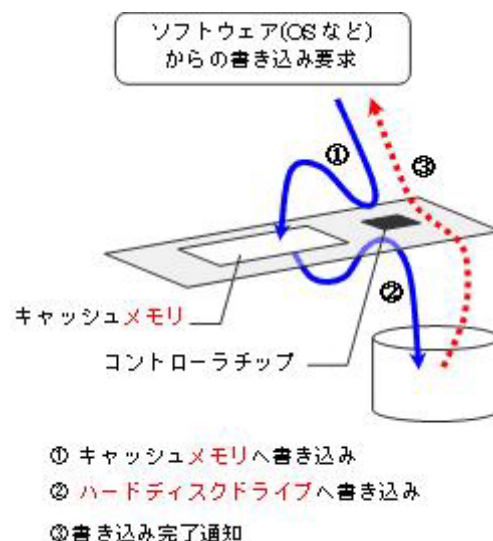
本製品は、ライト時にRAIDコントローラのキャッシュメモリを使い、ライト性能を大幅に向上させることができます。また、停電が発生した場合にキャッシュメモリ内のデータをバックアップすることができます。

2.5.1 無効 (WriteThrough)

どのソフトウェアから書き込み要求がきた場合に、RAIDコントローラ上のキャッシュメモリと物理デバイスの両方へ書き込みを行う方式。

ソフトウェアは、物理デバイスへの書き込み処理が終了するのを待ってから次の処理に移るため、一般的にアクセス性能は劣ります。

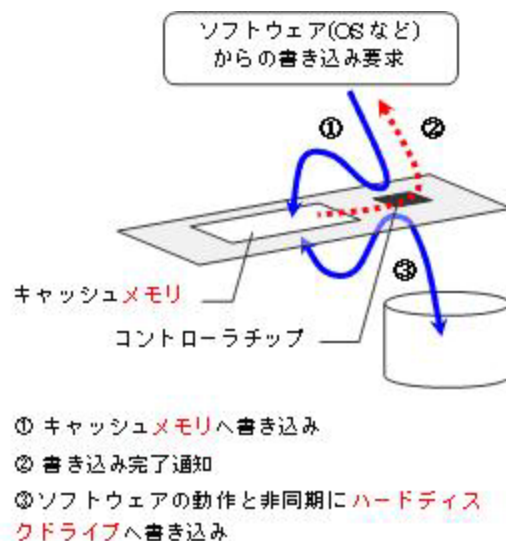
しかし、ソフトウェアからの書き込み要求が即時に物理デバイスに反映されるため、電源瞬断などの不慮の事故が発生してもデータを損失する危険性が少ないという利点があります。



2.5.2 有効 (WriteBack)

OSなどのソフトウェアから書き込み要求がきた場合に、RAIDコントローラ上のキャッシュメモリへのみ書き込みを行い、物理デバイスへの書き込みはキャッシュメモリ上のデータを元にRAIDコントローラが非同期に行う方式。

キャッシュメモリにデータが書き込まれた時点でソフトウェア側に完了通知が発行されるため、物理デバイスへの書き込み処理が完了するのを待たずにソフトウェア側は次の処理を継続することができます。一般的にアクセス性能が向上しますが、電源瞬断などの不慮の事故が発生した際にキャッシュメモリの内容が物理デバイス上に反映されない場合があります、データ損失の危険性があります。



2.5.3 フラッシュバックアップユニット

本体装置にフラッシュバックアップユニット(FBU)を接続することで、停電やサーバ障害時にFBUの電力を使用し、キャッシュデータをフラッシュメモリに転送します。フラッシュメモリは不揮発性メモリのためメモリ保持に電力が不要です。この機能により、キャッシュ有効で運用しているシステムにおいて、電源瞬断などの不慮の事故によるデータ損失を防ぐことができます。

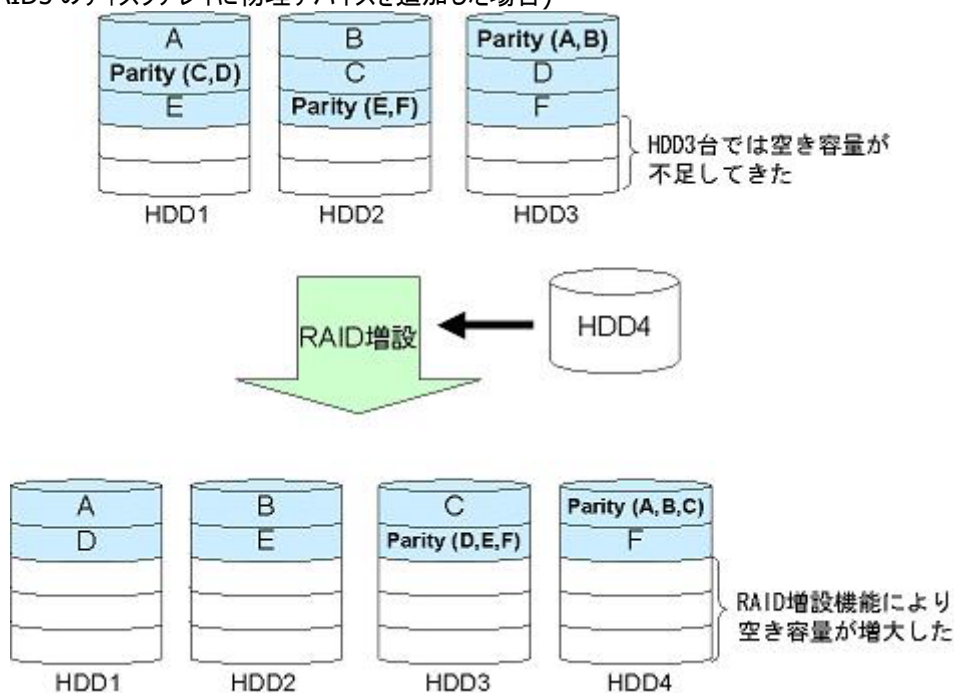
2.6 アレイの拡張

既に設定済みのディスクアレイ容量を拡大するために、最終ディスクアレイに物理デバイスを追加して 1 つのディスクアレイにまとめる機能(スパン構成の場合、増設機能を実行することはできません。)

現在割り当てられていない既存のドライブを追加することにより、既存のアレイの容量を増やします。追加するドライブは、以下の基準を満たしている必要があります。

- ・ 割り当てられていないドライブである必要があります。
- ・ アレイに含まれる既存のドライブと同じタイプ(SAS HDD、SAS SSD、SATA HDD、SATA SSD など)である必要があります。
- ・ アレイに含まれる最小のドライブ以上の容量を持っている必要があります。

(例: RAID5 のディスクアレイに物理デバイスを追加した場合)

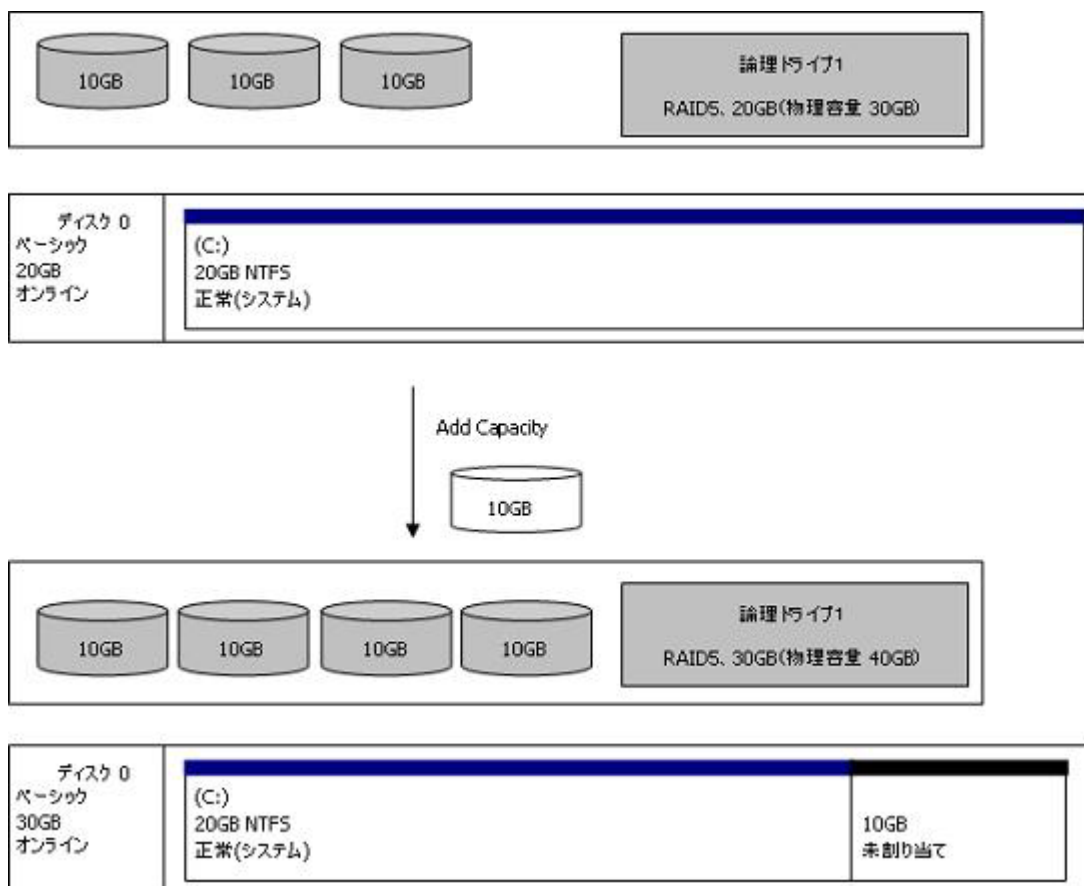


2.6.1 アレイの拡張説明

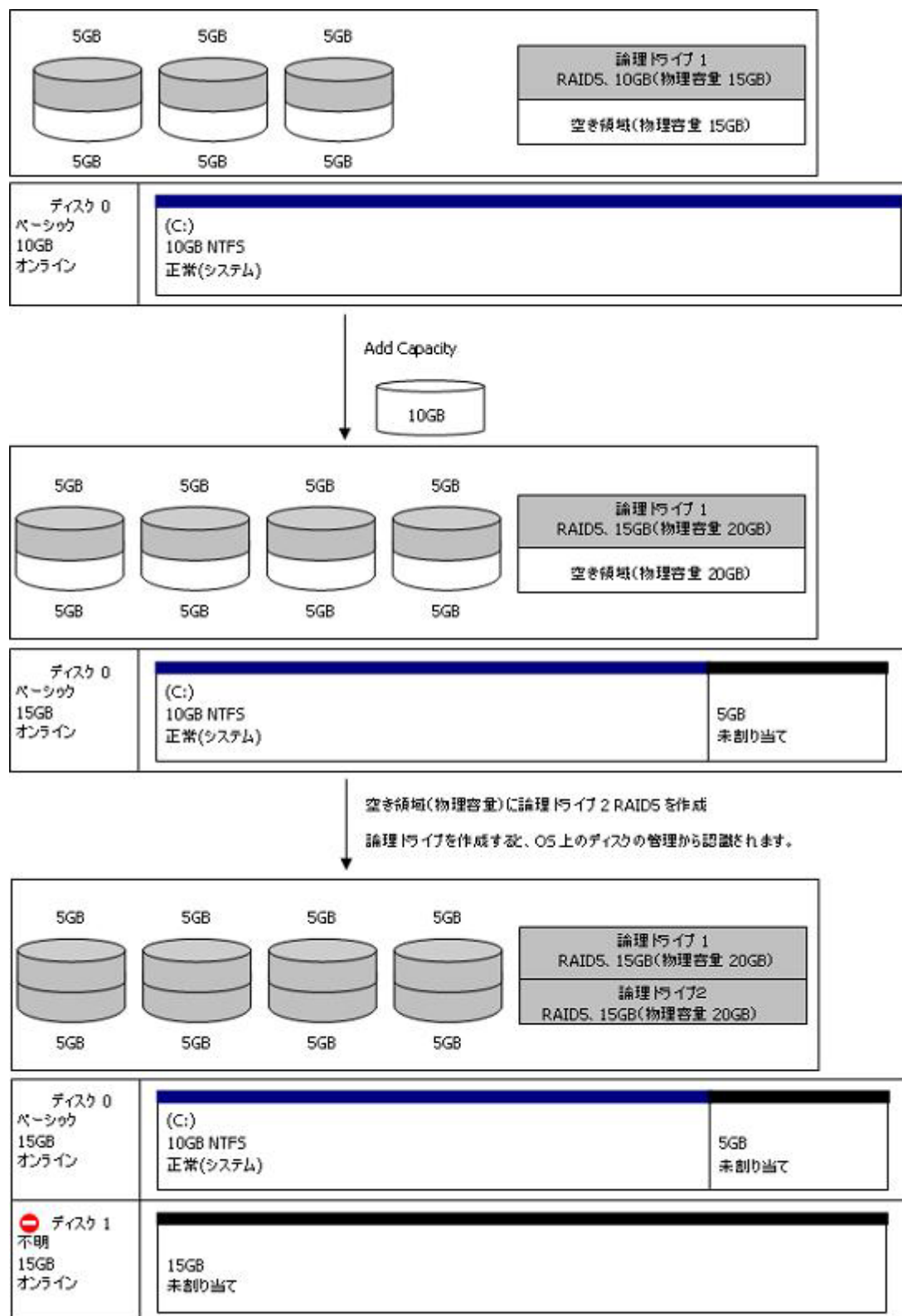
BIOS ユーティリティでのみアレイの拡張が実行可能です。

本機能はディスクアレイ容量を拡大し、そのディスクアレイに属する論理ドライブ(Logical Drive または Virtual Disk、Virtual Drive)容量を拡大することができます。OS 上では既存の物理デバイスの容量が増えたようになり、空き容量を使用して新たにパーティションを作成することで利用可能となります。

(例1)空き領域がない場合(パック内の領域全てを1つの論理ドライブとして定義している場合)



(例 2)空き領域が存在する場合。



2.6.2 アレイの拡張時間目安

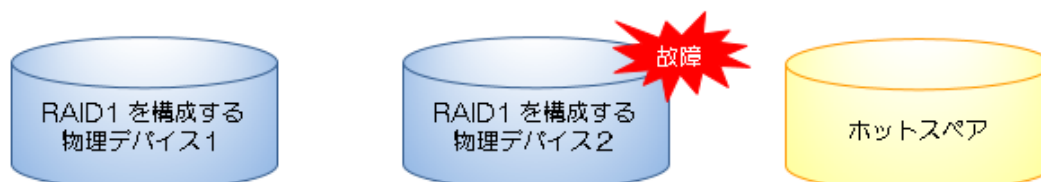
完了までに必要な時間目安については、付録 E を参照してください。

2.7 コピーバック機能

ホットスペアに対するリビルド完了後、故障した物理デバイスを交換すると、リビルドが実施された物理デバイスのデータを交換した物理デバイスにコピーする処理(コピーバック)が実施され、障害が発生する前の構成に戻ります。

(例：RAID1 構成のケース)

(1) RAID1 を構成する物理デバイスで故障が発生



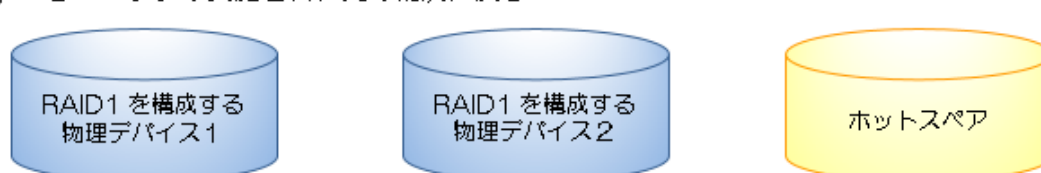
(2) ホットスペアに対してリビルドが実施される



(3) 故障した物理デバイスを交換する



(4) コピーバックが実施され、元の構成に戻る



注意：

- * コピーバック中は、本体装置のシャットダウンやリブートを実施しないでください。
- * 故障した物理デバイスを取り外してから新しい物理デバイスを取り付けるまでに、90 秒以上の間隔をあけてください。

2.8 スマーターコピーバック機能

論理ドライブを構成する物理デバイスに S.M.A.R.T.エラーが発生した場合に、この物理デバイスを置き換えることができるホットスペアが存在する構成であれば、S.M.A.R.T.エラーが発生した物理デバイスからホットスペアにデータをコピーする処理が実施され、データを退避します。コピー完了後、退避先の物理デバイスは論理ドライブに組み込まれ、S.M.A.R.T.エラーが発生した物理デバイスは「故障」となります。故障した物理デバイスを交換した後は、「2-7. コピーバック」の機能により、障害が発生する前の構成に戻ります。スマーターコピーバックはデフォルトで有効です。

(例：RAID1 構成のケース)

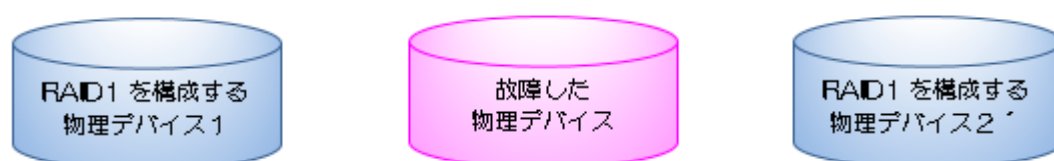
(1) RAID1 を構成する物理デバイスで S.M.A.R.T.エラーが発生



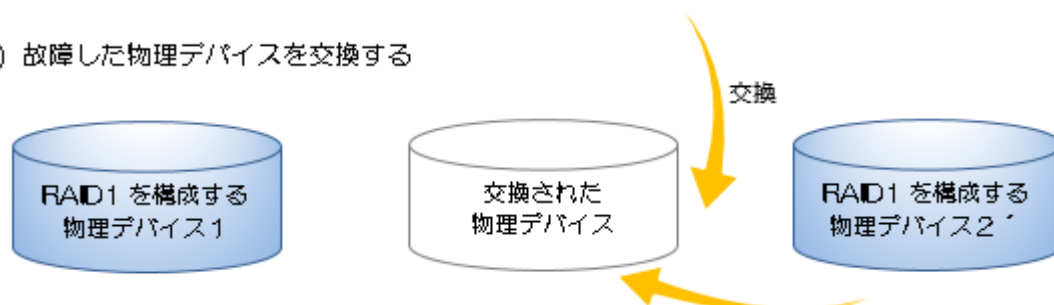
(2) ホットスペアに対してコピーが実施される



(3) コピー完了後、S.M.A.R.T.エラーが発生した物理デバイスは故障となる



(4) 故障した物理デバイスを交換する



(5) コピーバックが実施され、元の構成に戻る



- 注意：
- * スマーターコピーバック中は、本体装置のシャットダウンやリブートを実施しないでください。
 - * 故障した物理デバイスを取り外してから新しい物理デバイスを取り付けるまでに、90 秒以上の間隔をあけてください。
 - * JBOD ドライブで S.M.A.R.T.エラーが発生してもスマーターコピーバックは実行されません。
 - * スマーターコピーバックはグローバルスペアのみ対応です。

第3章 ハードウェア編

3.1 RAID コントローラ製品

前述のように RAID は複数台の物理デバイスを用いて高速化、大容量化、高信頼性を実現するための技術です。RAID コントローラとは RAID の持つ、優れた特性を発揮するための専用ハードウェアです。RAID コントローラは、実装形態や、搭載されている集積回路の規模、接続可能なインターフェース等により大別する事ができます。下記に RAID コントローラの製品を示します。

① メザニンカードタイプ製品

N コード	インタフェース	コネクタ/ポート数	最大転送速度	PCI バス形式	タイプ
N8103-240	SAS/SATA	2 コネクタ/16port	SATA:6Gb/s SAS:12Gb/s (1port あたり)	PCI Express(x8)	Intelligent

② OCP カードタイプ製品

N コード	インタフェース	コネクタ/ポート数	最大転送速度	PCI バス形式	タイプ
N8103-248	SAS/SATA/NVMe	2 コネクタ/16port	SATA:6Gb/s SAS:12Gb/s NVMe:16Gb/s (1port あたり)	PCI Express(x8)	Intelligent
N8103-249		1 コネクタ/8port			
N8103-250		2 コネクタ/16port			

③ PCI カードタイプ製品

N コード	インタフェース	コネクタ/ポート数	最大転送速度	PCI バス形式	タイプ
N8103-251	SAS/SATA/NVMe	2 コネクタ/16port	SATA:6Gb/s SAS:12Gb/s NVMe:16Gb/s (1port あたり)	PCI Express(x8)	Intelligent
N8103-252					

3.2 RAID コントローラの仕様

RAID コントローラの仕様については、付録 A を参照してください。

3.3 フラッシュバックアップユニットの仕様

本体装置によって使用できるフラッシュバックアップユニット(FBU)が異なります。接続対象について詳しくはシステム構成ガイドを参照してください。

N コード	タイプ	データ 保持時間	充電 時間	定期リフレッシュ 推奨実行間隔 (リフレッシュ時間)	製品 寿命	制御ケーブル長
N8103-218	FBU	不揮発性メモリに保持するため、 長期の保持が可能	数分	不要	5 年	145mm

3.4 物理デバイス選定における確認事項

- RAID システムに使用する物理デバイスは、新しい物理デバイスあるいは完全にフォーマットされた物理デバイスを使用してください。
- ディスクアレイを構成するときは、同容量/同回転数/同規格の物理デバイスにて構成してください。
- 容量の異なる物理デバイスを同一ディスクアレイに混在させた場合、ディスクアレイ内の物理デバイス 1 台あたりの容量は最も小さい物理デバイス容量に揃えられます。
- サポートする本体装置、搭載ルールについては、システム構成ガイドを参照してください。
<http://www.nec.co.jp/products/express/systemguide/100guide.shtml>

3.5 注意事項

- 本体装置内蔵物理デバイスに OS をインストールする時は、OS をインストールする物理デバイス以外は接続せず、OS インストール後に接続してください。

第4章 ソフトウェア編

4.1 RAID コントローラのソフトウェア

RAID コントローラを用いて RAID システムを構築するには、各 RAID コントローラに対応したソフトウェアを使用する必要があります。RAID コントローラを制御するソフトウェアは、オフラインユーティリティとオンラインユーティリティに大別されます。

4.2 オフラインユーティリティ

オフラインユーティリティは RAID コントローラあるいは本体装置の BIOS ROM 内に格納されております。オペレーティングシステムを起動せずに RAID コントローラの操作を行うことができます。

N コード/名称	オフラインユーティリティ名	略称
N8103-240/248/249/250/251/252	システムユーティリティ	なし

*各ユーティリティの操作方法については本体装置及び RAID コントローラのユーザーズガイドを参照してください。

4.3 オンラインユーティリティ

Windows/Linux

名称	オンラインユーティリティ名	説明書
RAID システム管理ユーティリティ	LSI Storage Authority StorCLI	MegaRAID ユーティリティユーザーズガイド
RAID 通報サービス	RAID Report Service	Smart Storage Administrator ユーザーズガイド

VMware ESXi

名称	オンラインユーティリティ名	説明書
RAID システム管理ユーティリティ	StorCLI	MegaRAID ユーティリティユーザーズガイド

RAID 監視通報方式について、Windows/Linux 環境では RAID 通報サービス、VMware ESXi 環境では iLO SNMP Trap による通報をご利用ください。

VMware ESXi 環境について、詳しくは、下記の Web サイトを参照してください。

・NEC サポートポータル

<https://www.support.nec.co.jp/View.aspx?id=3140108419>

4.4 RAID コントローラのソフトウェアでの注意事項

[オンラインユーティリティ]

- RAID コントローラを使用する場合は、必ず RAID システム管理ユーティリティをシステムにインストールしてください。RAID システム管理ユーティリティをインストールしていない場合、RAID システムの障害検出ができません。また、RAID システム管理ユーティリティを使用する場合は、RAID システム管理ユーティリティの説明書を事前に読んでください。以下に示す注意事項以外にもシステムを運用するために留意すべき注意事項が記載されています。説明書は StarterPack (S8.10-010.04 以降) または NEC サポートポータルサイト (URL <https://www.support.nec.co.jp/>) より入手できます。(本体装置ユーザズガイド掲載 Web サイト参照ください)
- RAID システム管理ユーティリティを使用する場合は、管理者権限のあるユーザでログオンしてください。管理者権限を持たないユーザでログオンした場合は、RAID システム管理ユーティリティが動作しない、または操作できない場合があります。

4.5 ソフトウェアダウンロードサイト

【N8103-240専用】MegaRAIDユーティリティ(LSI Storage Authority および StorCLI) (Windows版)
<https://www.support.nec.co.jp/View.aspx?NoClear=on&id=9010110424>

【N8103-240専用】MegaRAIDユーティリティ(StorCLI) (VMware ESXi 7.0版)
<https://www.support.nec.co.jp/View.aspx?NoClear=on&id=9010110425>

第5章 運用編

5.1 性能比較

物理デバイスのタイプが性能に影響を与えます。

SSD はハードディスクよりも性能が高くなり、SAS デバイスは SATA デバイスよりも性能が高くなります。

また、一般的に RAID コントローラはインテリジェントタイプがノンインテリジェントタイプよりも性能が高くなります。

5.2 RAID レベルの比

RAID コントローラを用いて RAID システムを構築する前に、用途に応じた RAID レベルを選択する必要があります。耐障害性、アクセス速度および容量効率を考慮の上で最適な RAID レベルを選択してください。

①耐障害性

耐障害性は冗長構造を持つ RAID レベルを選択することで向上させることができます。RAID1 は構成に必要な物理デバイスの数が少なく、冗長構造も単純であるため耐障害性は高くなります。RAID5 は制御する物理デバイスの数が多くパリティ計算など、冗長構造が RAID1 よりも複雑であることから、耐障害性は RAID1 よりも低くなります。RAID6 は、2重のパリティを使用するため耐障害性は RAID5 より高くなります。

→「2.1.3 RAID の種類(RAID レベル)」

RAID1 > RAID6 > RAID5 RAID0 は耐障害性無し

[注意事項]

- 物理デバイスのマルチデッドによるシステム障害の発生を低減させる観点から、各アレイの物理デバイス搭載数は 8 台以下を目安とした RAID 構成を推奨します。
- 大容量物理デバイスにて RAID を構築する場合、障害復旧時に長時間のリビルドが必要です。その間冗長性が失われますので、より信頼性を高めるためにも物理デバイス 2 台の障害に対応する RAID6 あるいは RAID60 でのご利用を推奨します。

②アクセス速度

アクセス速度は物理デバイスを複数台接続しスプリットシークを行うことで向上させることができます。RAID0 はこの機能を用いたアクセス向上が期待される RAID レベルです。RAID5 と比べてパリティ情報の書き込みが必要ない分アクセス速度が速くなります。RAID1 と RAID5 との比較においてもライト動作においてパリティ計算+パリティライトが必要であるため RAID1 の方がアクセス速度は勝っています。RAID6 は、リード性能は RAID5 と同等ですが、ライト性能はパリティを2重化しているため低くなります。

→「2.1.3 RAID の種類(RAID レベル)」

RAID0 > RAID1 > RAID5 > RAID6

③容量効率

容量効率は物理デバイス総容量に占める冗長データ容量を少なくすることで向上させることができます。RAID1 と RAID5 は冗長構造を持ちます。RAID1 は2台の物理デバイスを用いて構築可能であり、小規模なシステム向けですが、容量効率は 50%に固定されます。RAID5 は3台以上の

を用いて構築可能であり、中規模～大規模なシステム向けです。容量効率も 66%～と物理デバイスを多く接続すればするほど容量効率が上がる特徴を持っています。RAID6 は常に物理デバイス2台分の容量がパリティとして使用されるため、RAID5 と比べると容量効率は低くなり 33%～になります。

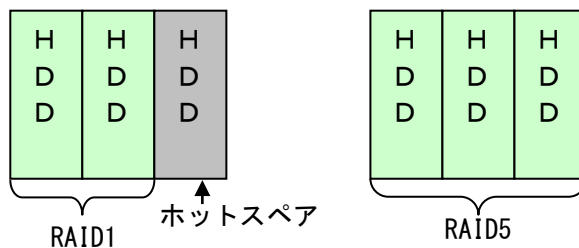
→「2.1.3 RAID の種類(RAID レベル)」

RAID0 > RAID5 > RAID6 > RAID1

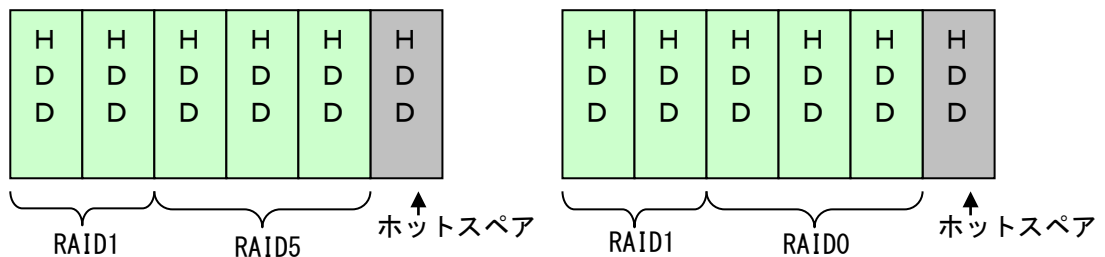
5.3 RAID システムの構築

上記、RAID コントローラの選択および RAID レベルの選択をふまえ、RAID システムの構築例を下記に示します。

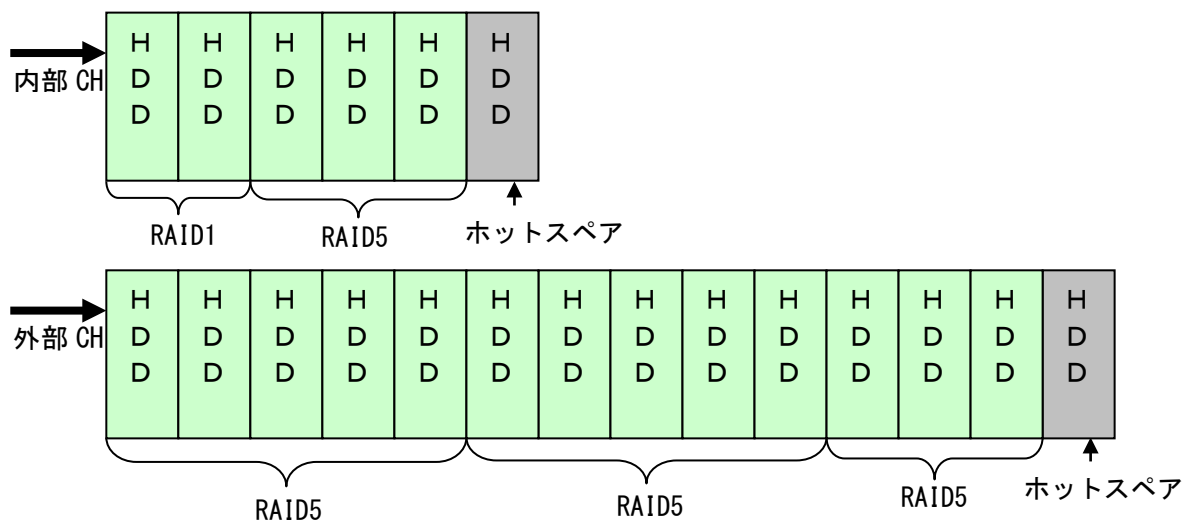
①小規模システム例（物理デバイス3台）



②中規模システム例（物理デバイス6台）



③大規模システム例（物理デバイス6台・内部ケージ物理デバイス 14 台・外部 DISK 増設筐体）



5.4 安定運用のために

RAID コントローラは RAID システムを構築し制御することで高速化、大容量化および高信頼性を提供します。ただし、RAID システムを構築するには複数の物理デバイスを必要とします。物理デバイスは技術の粋を結集した非常に高度な精密機械であり、デリケートな要素を持っています。RAID システムにおいて高信頼性を確保するには下記に示す操作を行う必要があります。

5.4.1 パトロールリード

パトロールリードは、RAID 構成済みドライブのシステム専用領域を含む、コントローラに接続されているドライブのすべてのセクタを定期的に確認します。すべての RAID レベルとすべてのスペアドライブに対して、パトロールリードを実行できます。コントローラが定義された期間中にアイドル状態で、他にバックグラウンドアクティビティがない場合にのみ、パトロールリードを開始できます。パトロールリードは工場出荷時に有効になっています。

5.4.2 整合性チェック

整合性チェック操作は、RAID レベル 1、5、6、10、50、60 を使用している論理ドライブのデータの正確性を検証します。たとえば、パリティ付きシステムの整合性チェックでは、1 つのドライブ上のデータを計算し、その結果をパリティドライブの内容と比較します。

フォールトトレラントな論理ドライブの整合性チェックは定期的に行われます。RAID0 はデータ冗長性を提供しないため、RAID 0 論理ドライブの整合性チェックは実行できません。

整合性チェックを実行するには、最初に整合性チェックのプロパティを設定してから、次のいずれかを実行します。

- ・ 定義した間隔で整合性チェックを実行するようにスケジュールを設定する。
- ・ 整合性チェック操作をただちに開始する。

整合性チェックの優先度は、1～100 の範囲です。

整合性チェックでは、以下のモードが使用できます。

- ・ Concurrent (同時) - すべての論理ドライブで整合性チェックを同時に実行します。
- ・ Sequential (順次) - 論理ドライブの整合性チェックを 1 つずつ順番に実行します。
- ・ 無効 - 整合性チェックを無効にします。

5.4.3 RAID システム管理ユーティリティの使用について

- ・ RAIDコントローラを使用する場合は、必ずRAIDシステム管理ユーティリティをシステムにインストールしてください。RAIDシステム管理ユーティリティをインストールしていない場合、RAIDシステムの障害検出ができません。
- ・ RAIDシステム管理ユーティリティを使用する場合は、管理者権限のあるユーザ(administrator等)でログオンしてください。管理者権限を持たないユーザでログオンした場合は、RAIDシステム管理ユーティリティが動作しない、または操作できない場合があります。

5.4.4 RAID コントローラ用ドライバ、RAID システム管理ユーティリティのアップデート

既知の問題や機能強化を実施した RAID コントローラ用ドライバ、RAID システム管理ユーティリティの最新バージョンは、Web サイトに随時公開しています。定期的に確認し適時アップデートしてください。

NEC サポートポータルサイト (URL <https://www.support.nec.co.jp/>)

5.4.5 RAID 構成物理デバイス台数の設定による保守運用性の向上

障害事例:

1 台の RAID コントローラに物理デバイスを 16 台接続し、16 台の物理デバイス全容量を使用して 1 ディスクアレイ (RAID5) を構築した。整合性チェックを実施すると、約 18 時間を要し夜間作業にて整合性チェックが終了しなかった。そのため、定期的な整合性チェックが実施されることなく運用が継続され、障害発生時のリビルドにてリードエラーが起り、復旧に失敗した。

改善:

論理ドライブを構成する物理デバイスの台数を少なめに設定します。

(1) 耐障害性・冗長性の改善

1 つのディスクアレイを構成する物理デバイス台数を少なくする(ディスクアレイを細分化させる)ことで、冗長性が向上(障害発生率が低下)します。

(2) 保守運用性の改善

ディスクアレイを構成する物理デバイス台数を少なくすることで、リビルド時間も減少します。また、整合性チェックに要する時間も、RAID システムの構成を細分化することで、大幅に減少します。

[注意事項]

- ・ 物理デバイスのマルチデッドによるシステム障害の発生を低減させる観点から、各アレイの物理デバイス搭載数は 8 台以下を目安とした RAID 構成を推奨します。
- ・ 大容量物理デバイスにて RAID を構築する場合、障害復旧時に長時間のリビルドが必要です。その間冗長性が失われますので、より信頼性を高めるためにも物理デバイス 2 台の障害に対応する RAID6 あるいは RAID60 でのご利用を推奨します。