

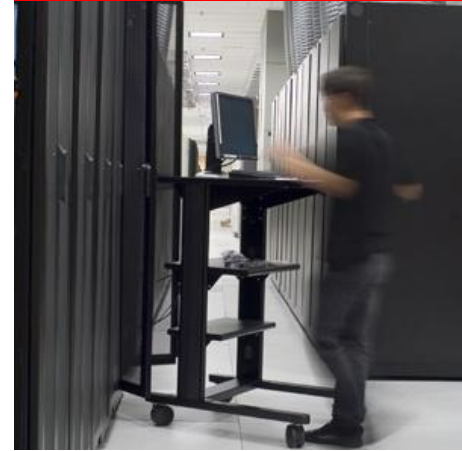


ORACLE®

Oracle Automatic Storage Management

日本オラクル株式会社

ORACLE®

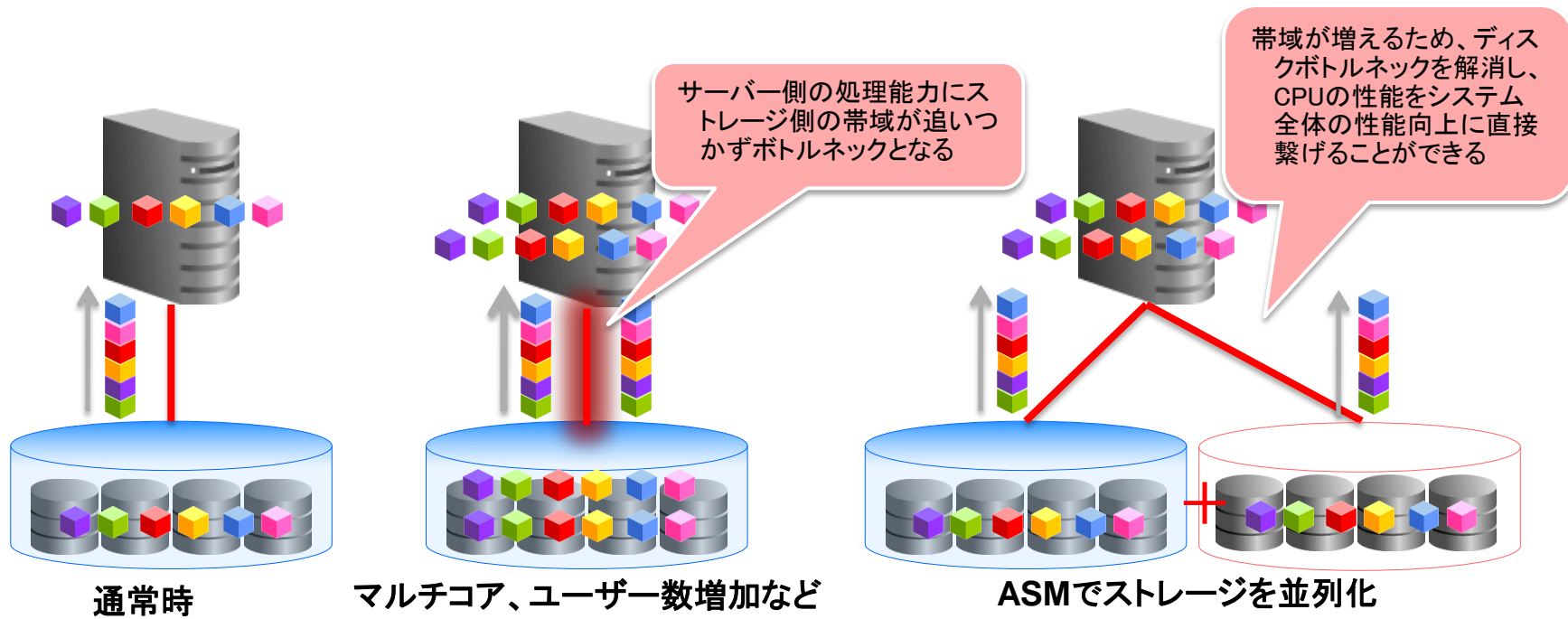


1. ASMの性能向上のためのアーキテクチャ

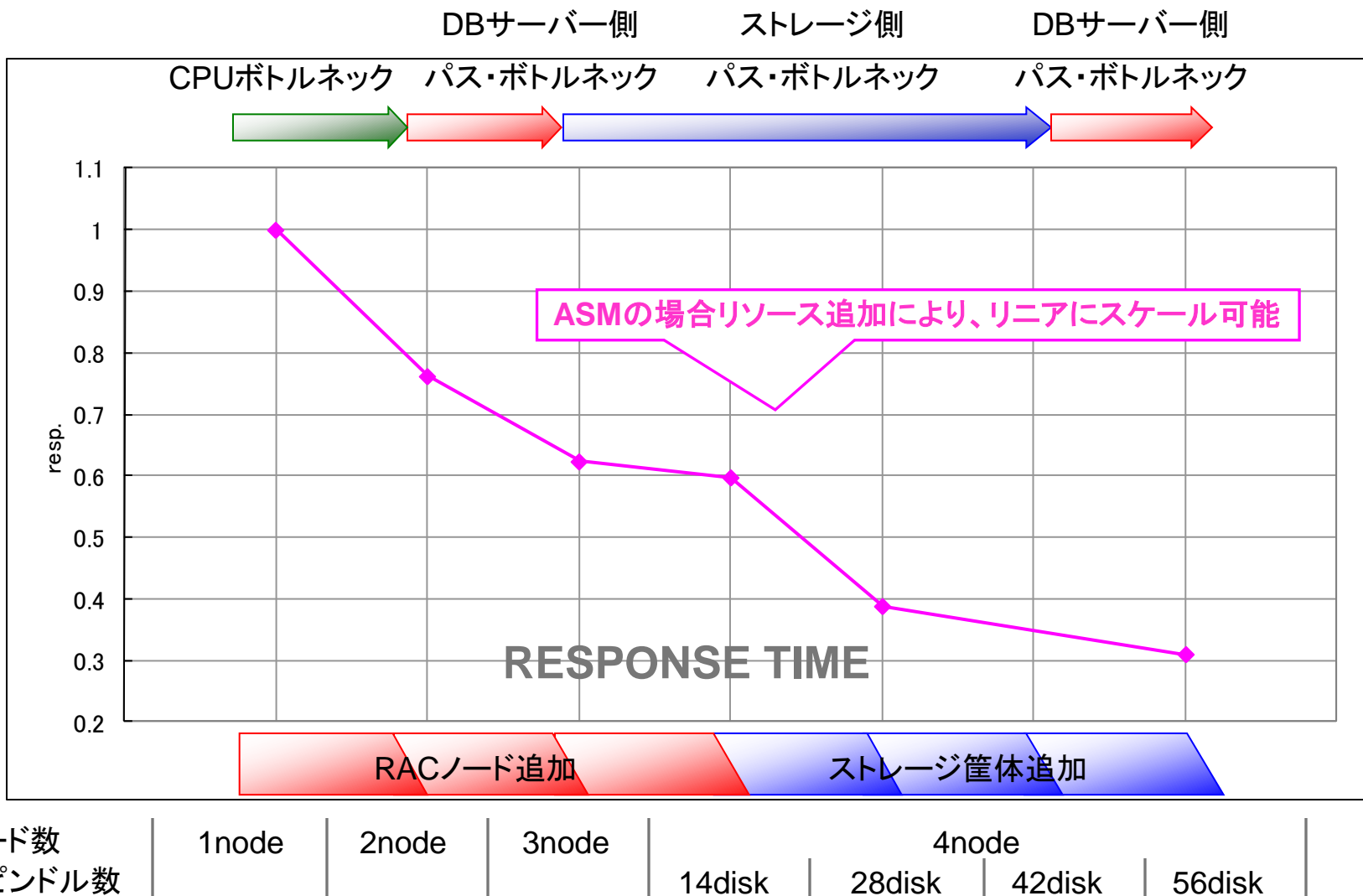
ASMのスケールアウト型アーキテクチャ

並列化による性能向上

- OracleのASMはマルチコア化により向上するCPU性能を最大限活用し、ストレージ側のボトルネックを解消
 - Disk I/O性能がCPU性能に対して不足し、データベースサーバー全体性能のボトルネックになりがち
 - ASMは安価なストレージを並列化することで帯域を確保し、性能向上を図る

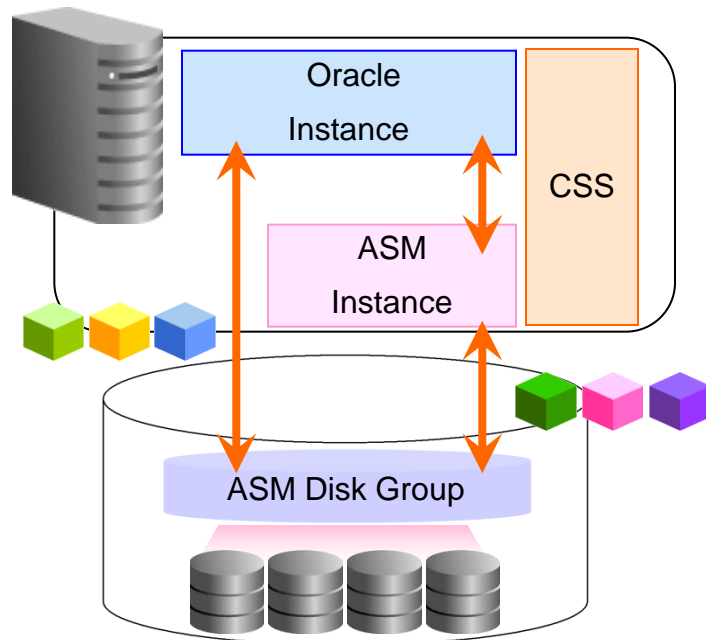


リソース追加によるASMのスケラビリティ 検証結果



Oracleデータベースのストレージ仮想化

- Oracle Automatic Storage Management(以下、ASM)
 - Oracle 10g～実装
 - Oracleデータベースに対してボリューム・マネージャ兼ファイルシステムとして機能し、ディスク構成を仮想化

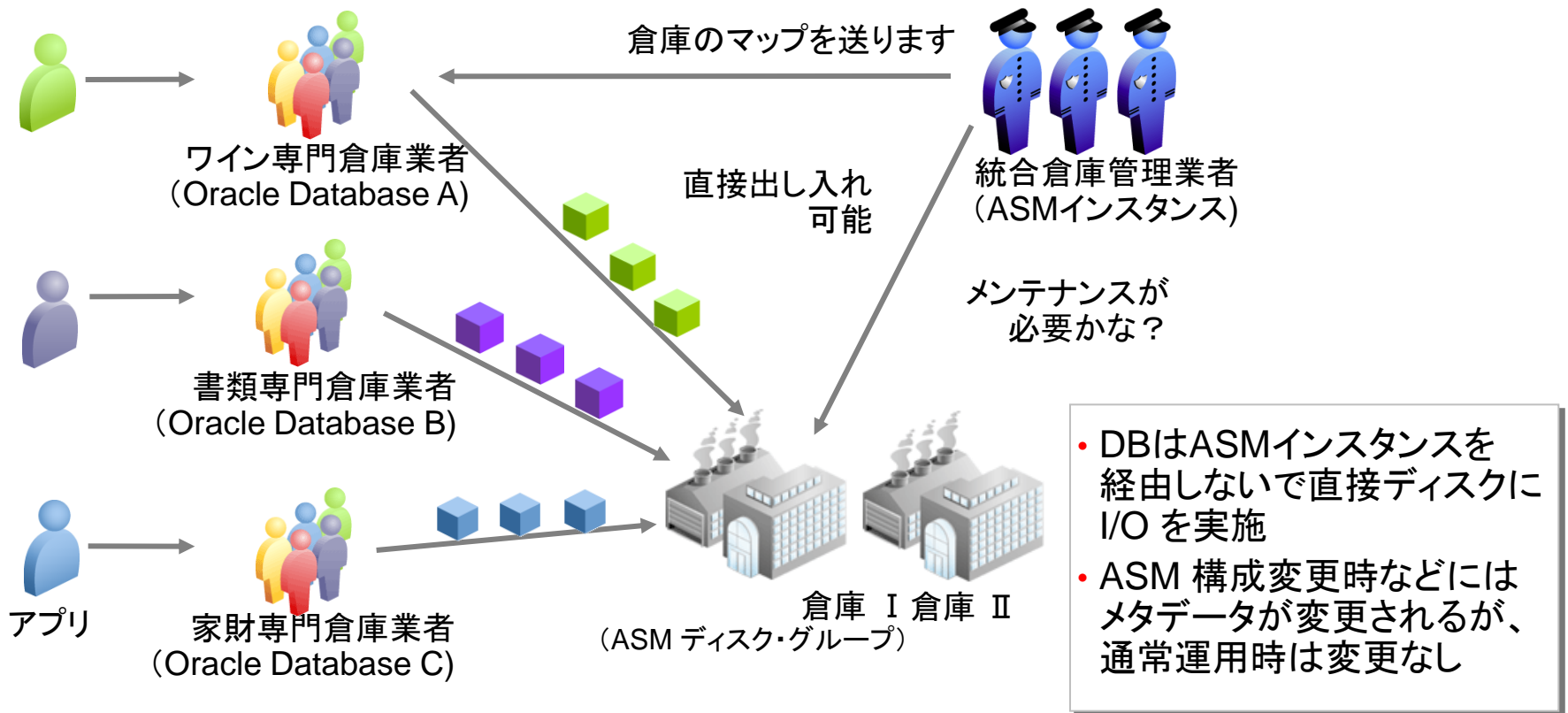


- Oracle Databaseにフラットなディスク・プールを提供 + ディスク管理工数を大幅削減
- 複数のディスク・アレイにまたがってディスクを仮想化し、ディスク追加 / 削除でもデータを透過的に再配分
- GRIDCenter検証では、標準使用

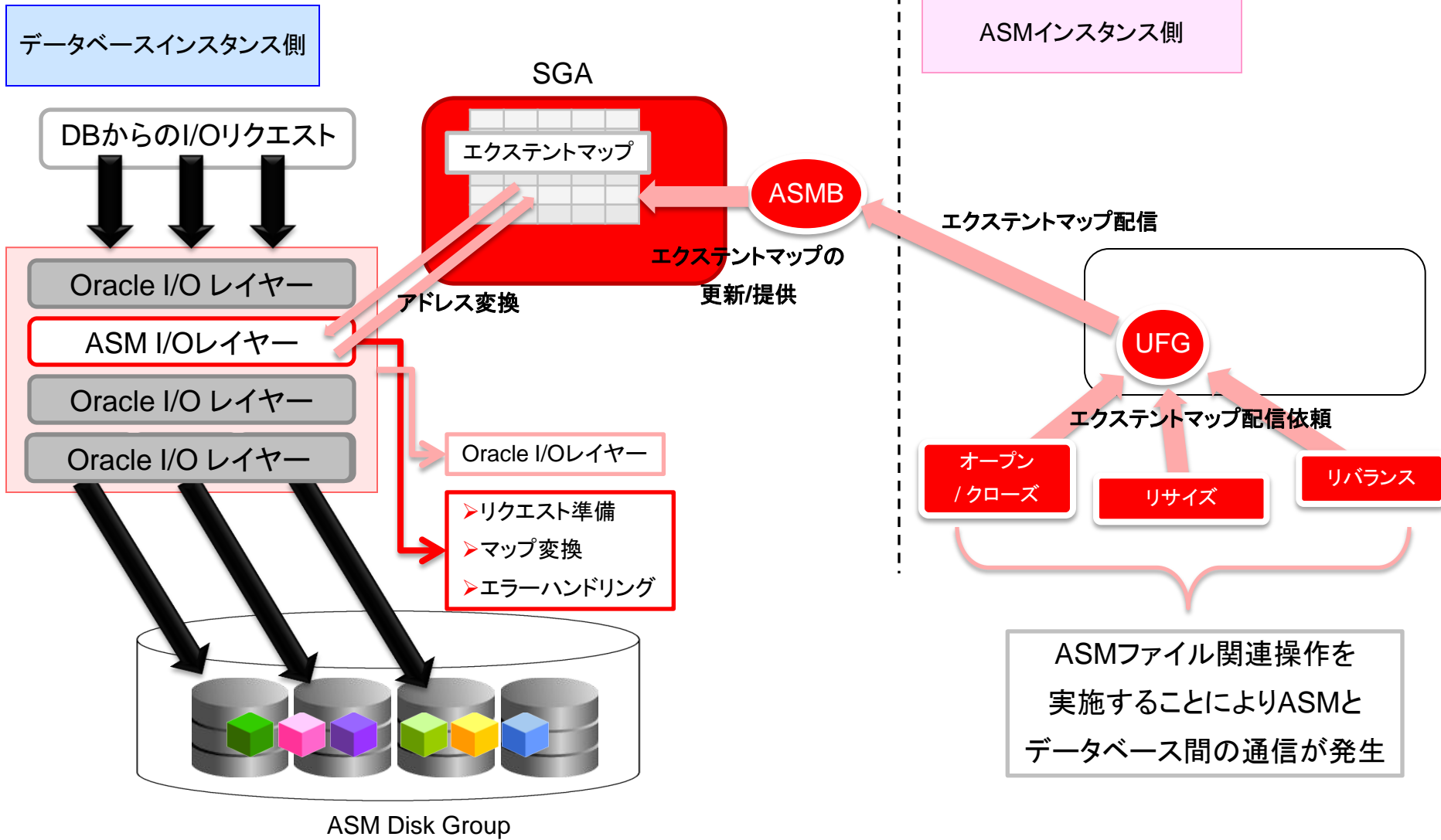
RAC環境でのASM 使用率は約 **6 割**
11g だと約 **9 割**で採用

ASM のアーキテクチャ・イメージ

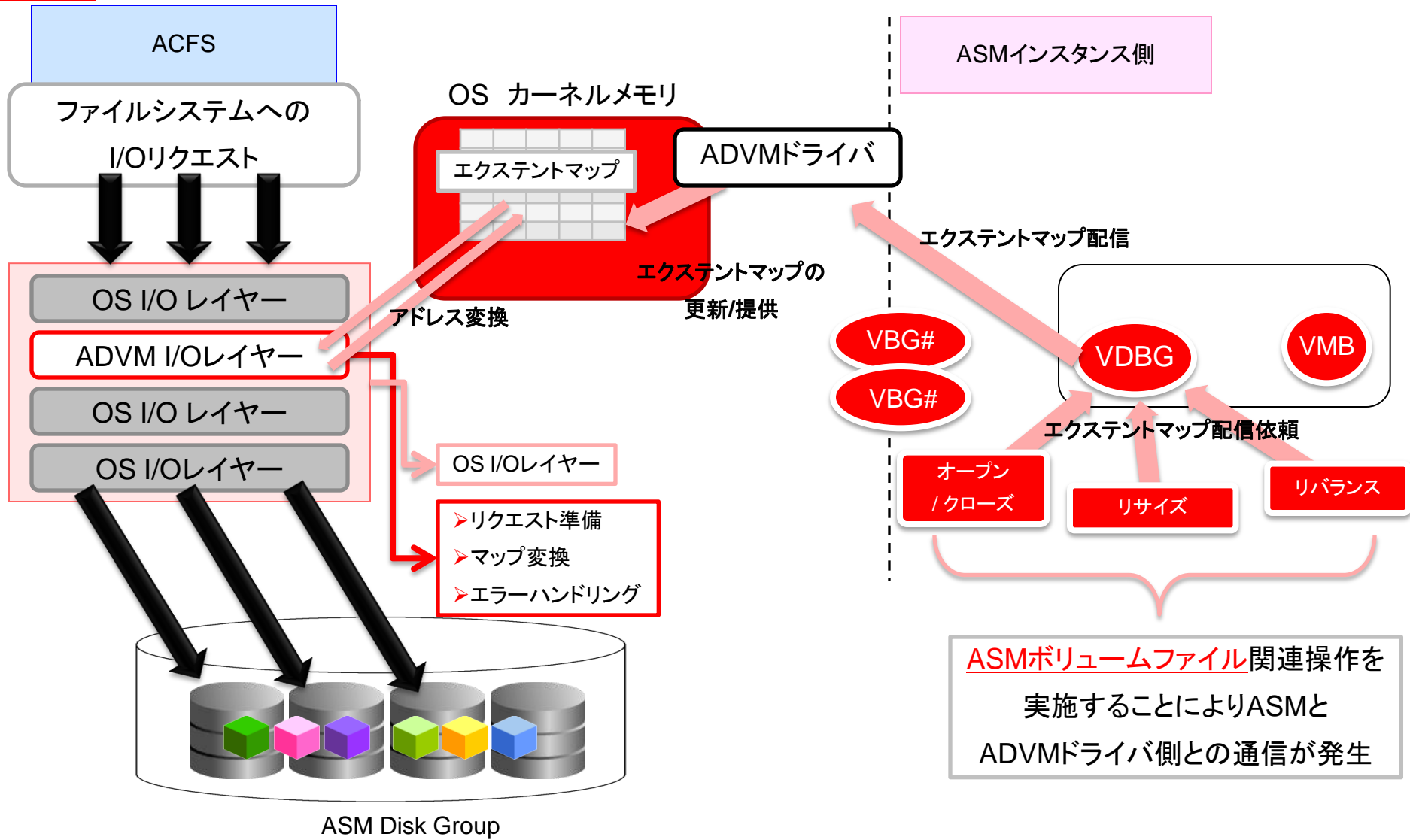
- ASM は複数のデータベースから利用可能
- メタ・データは各データベースにキャッシュ



ASMインスタンスとDBインスタンスの関係



ASMインスタンスとADVMDライバの関係

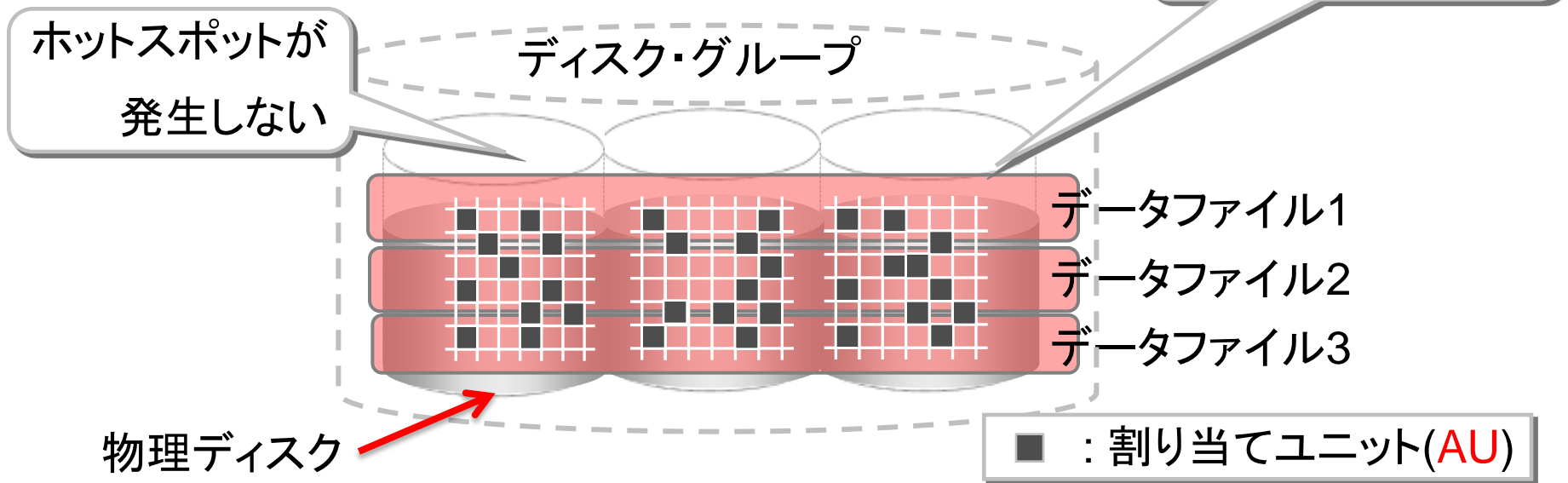


ASMボリュームファイル関連操作を実施することによりASMとADVMDドライバ側との通信が発生

性能の最適化機能

ASMのストライピング

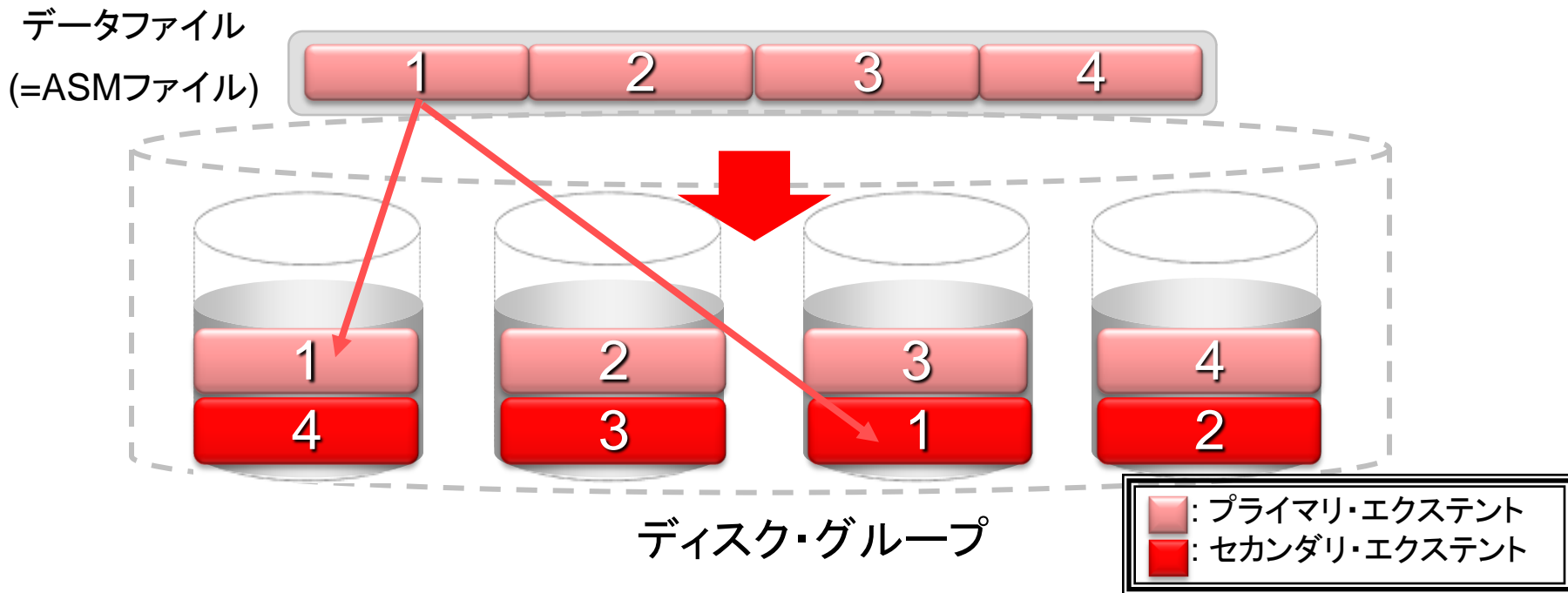
- ディスク・グループ内の、全てのディスクでストライピング
- 割当てユニット(AU)単位で領域を割当て
 - 1, 2, 4, 8, 16, 32, 64MB の可変サイズから選択、デフォルトは 1MB
 - Oracle Exadata では 4MB を採用しシーケンシャル I/O に最適化
- 全てのディスクの使用率が同じになるように割当て



可用性担保機能

ASMのミラーリング

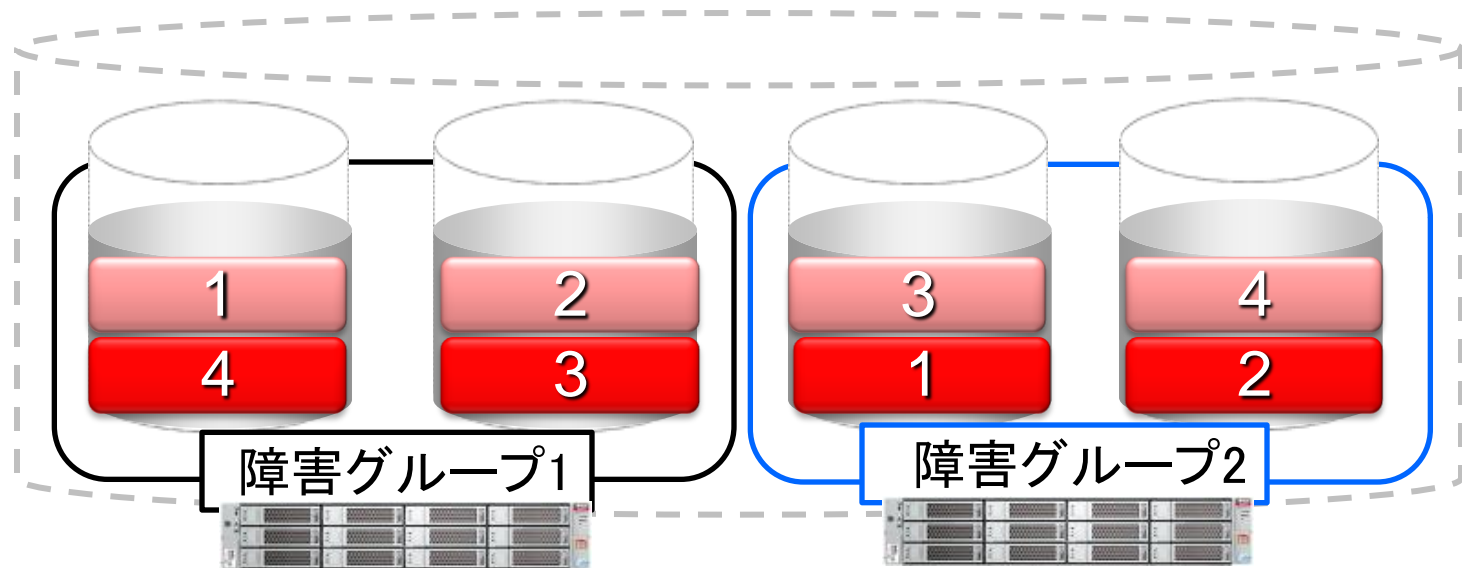
- ファイルの重要度に合わせてミラーリング可能
 - ミラー無し・二重化・三重化から選択
 - ディスク・グループ単位で指定



可用性担保機能

障害グループ

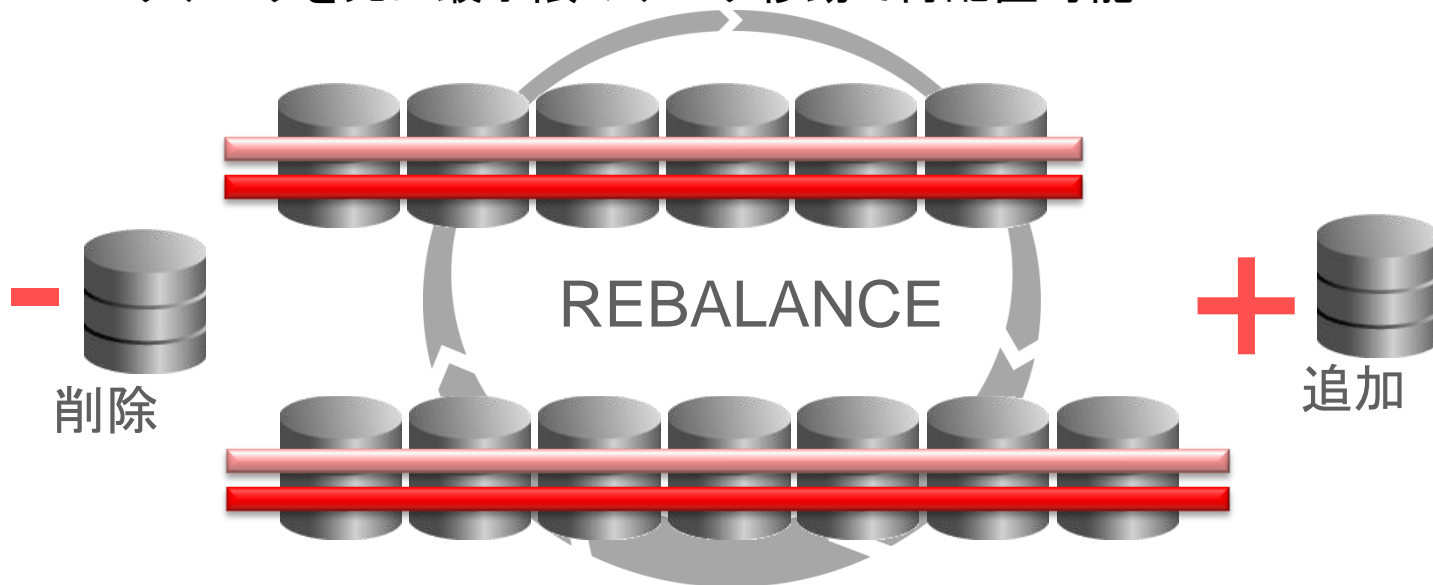
- 障害グループを設定し、オリジナルとミラーの同時損失を回避可能
- 障害グループとは
 - リソース(電源など)を共有しているディスクのグループ(筐体・コントローラー)
 - ミラーリングは、異なる障害グループに属しているディスク間で行われる
- ストレージ筐体を意識しないASMだから実現可能



データ再配置機能

リバランシング

- S.A.M.E 構成を維持したまま、ストレージ構成の変更が可能
 - ASM ディスクグループにディスクが追加されると、自動的に既存のデータファイルが再配置される
 - データベースからは透過的、データベース無停止で実施可能
 - 管理者が手作業でストライピングを行う必要無し
 - メタデータを元に最小限のデータ移動で再配置可能



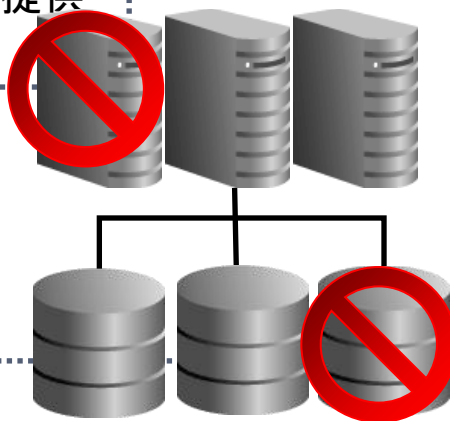
2. ASMの障害時動作



ASM 障害時の動作

インスタンス障害

- ◆ ディスク・グループ整合性は自動担保
- ◆ 正常インスタンスが ASM メタデータ情報を元に処理を再開あるいはロールバック
- ◆ 11gR2より自動再起動の仕組みを提供 (RAC, Single問わず)

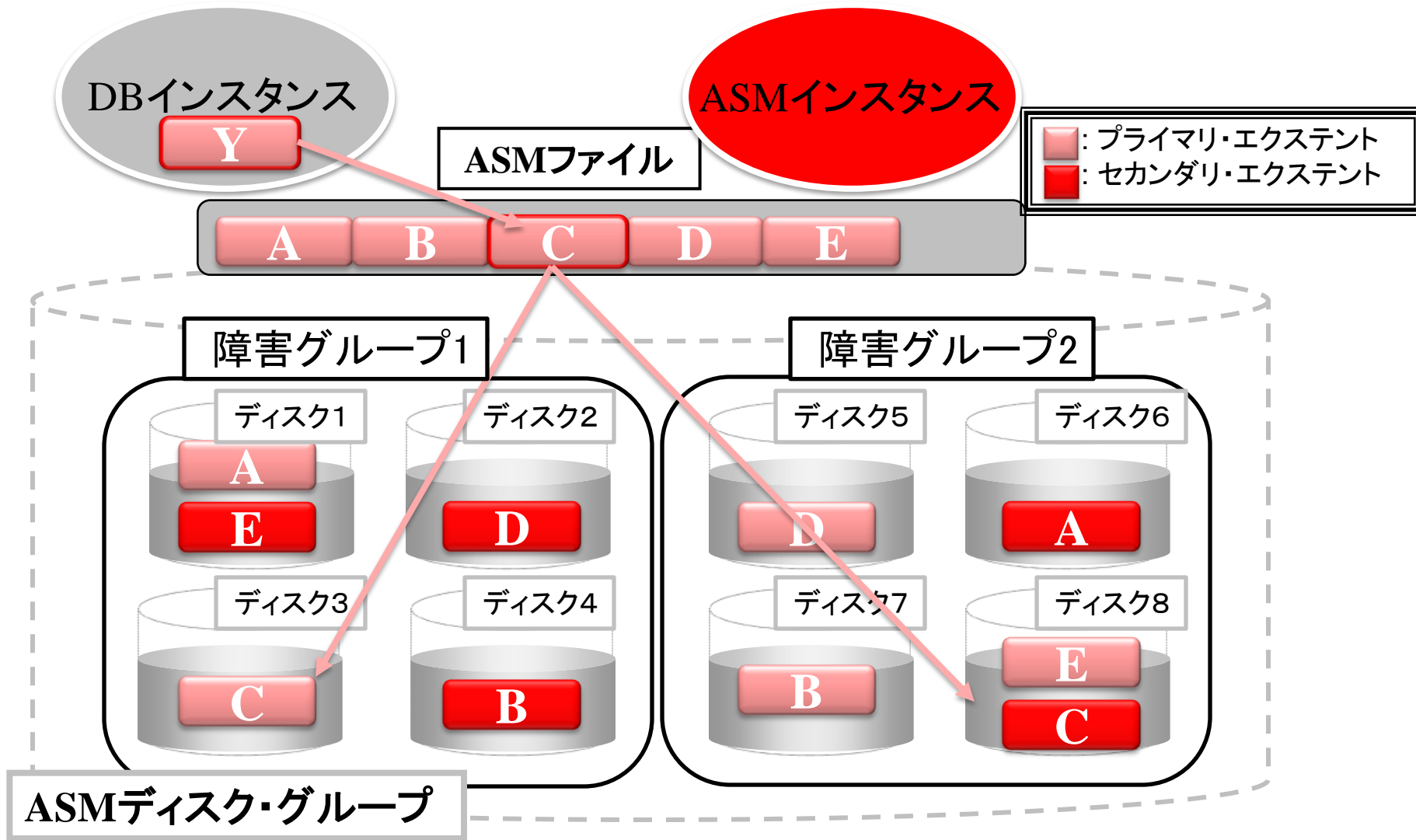


ディスク障害

- ◆ ミラーリング構成時、ASM により自動障害対応
- ◆ 書き込み処理時に I/O エラーを検知すると障害ディスクを自動でオフライン処理
- ◆ 読み込み処理時に I/O エラーを検知するとセカンダリから読み取り、不良ブロックは自動修復

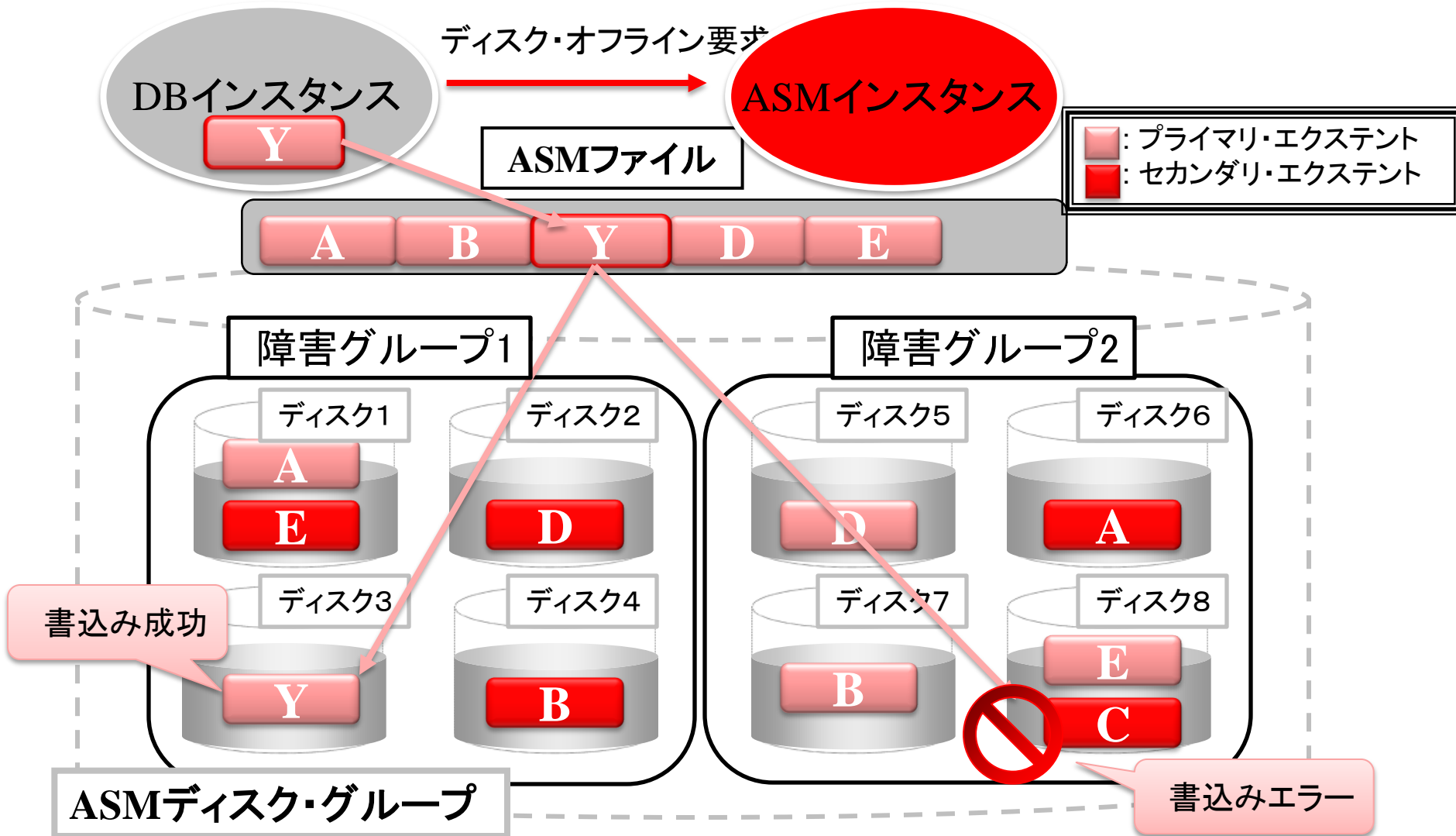
書き込みI/Oエラー時の動作

データ「C」への更新要求



書き込みI/Oエラー時の動作

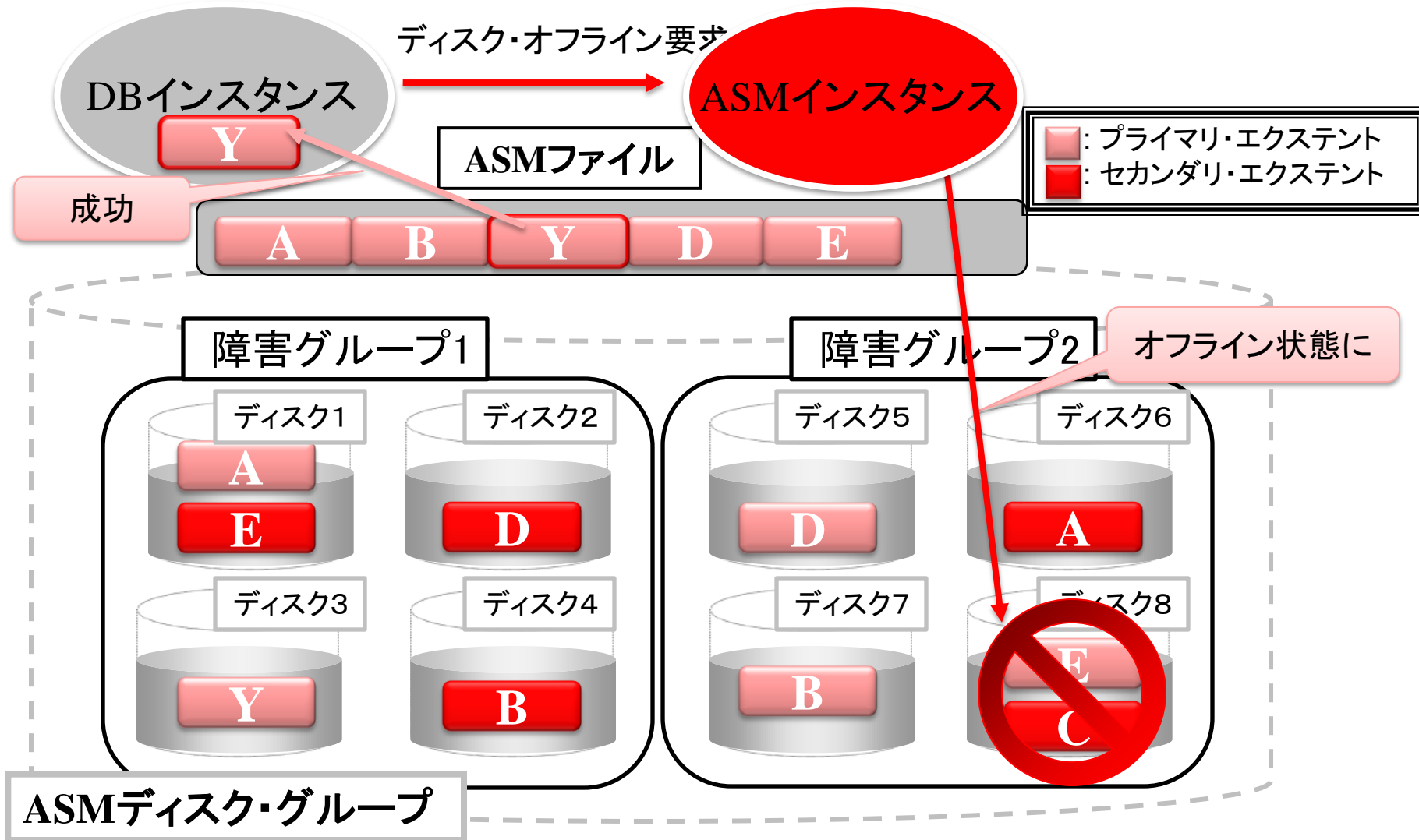
ディスク8で書き込みエラー発生



ORACLE

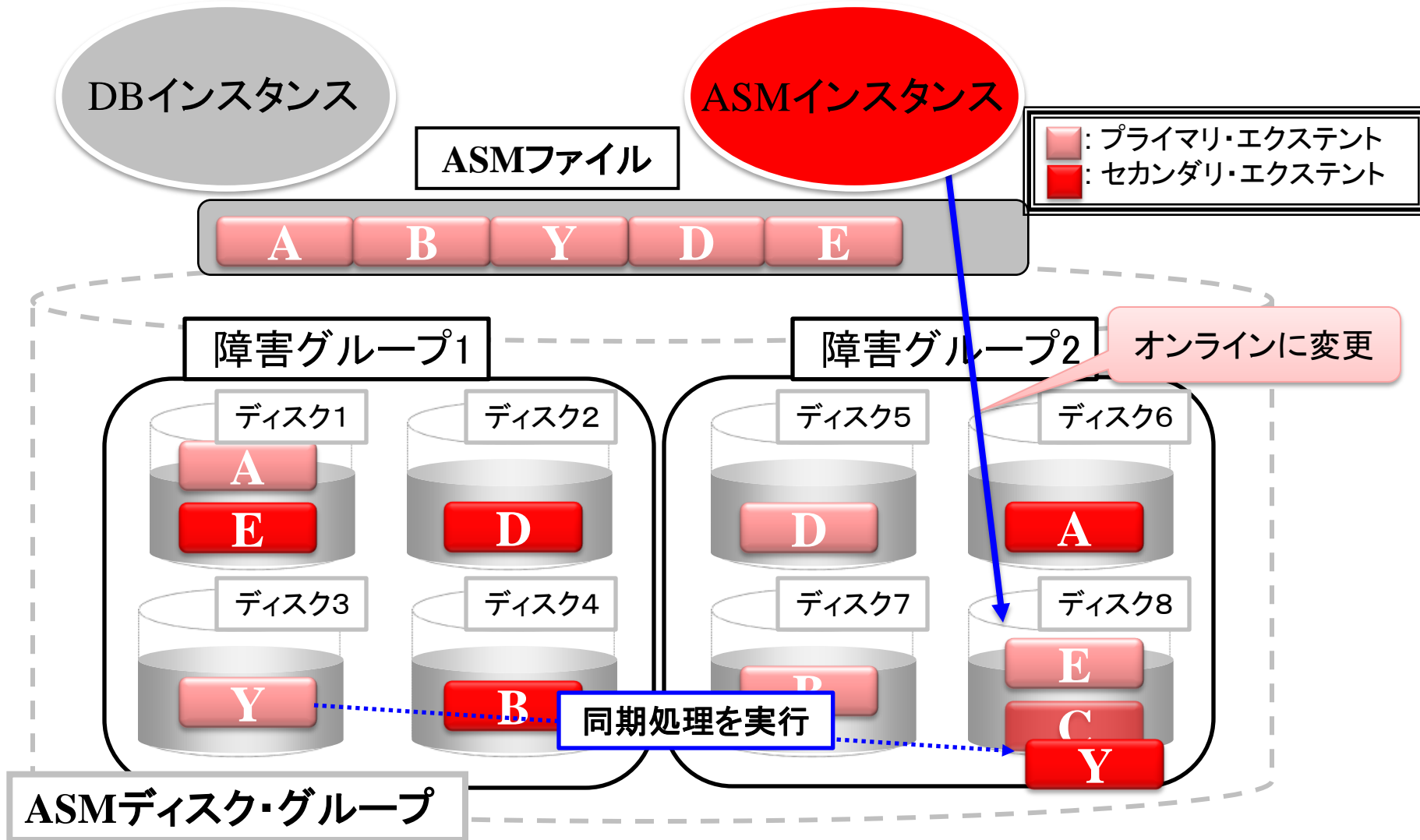
書き込みI/Oエラー時の動作

ディスク8をオフライン状態に変更



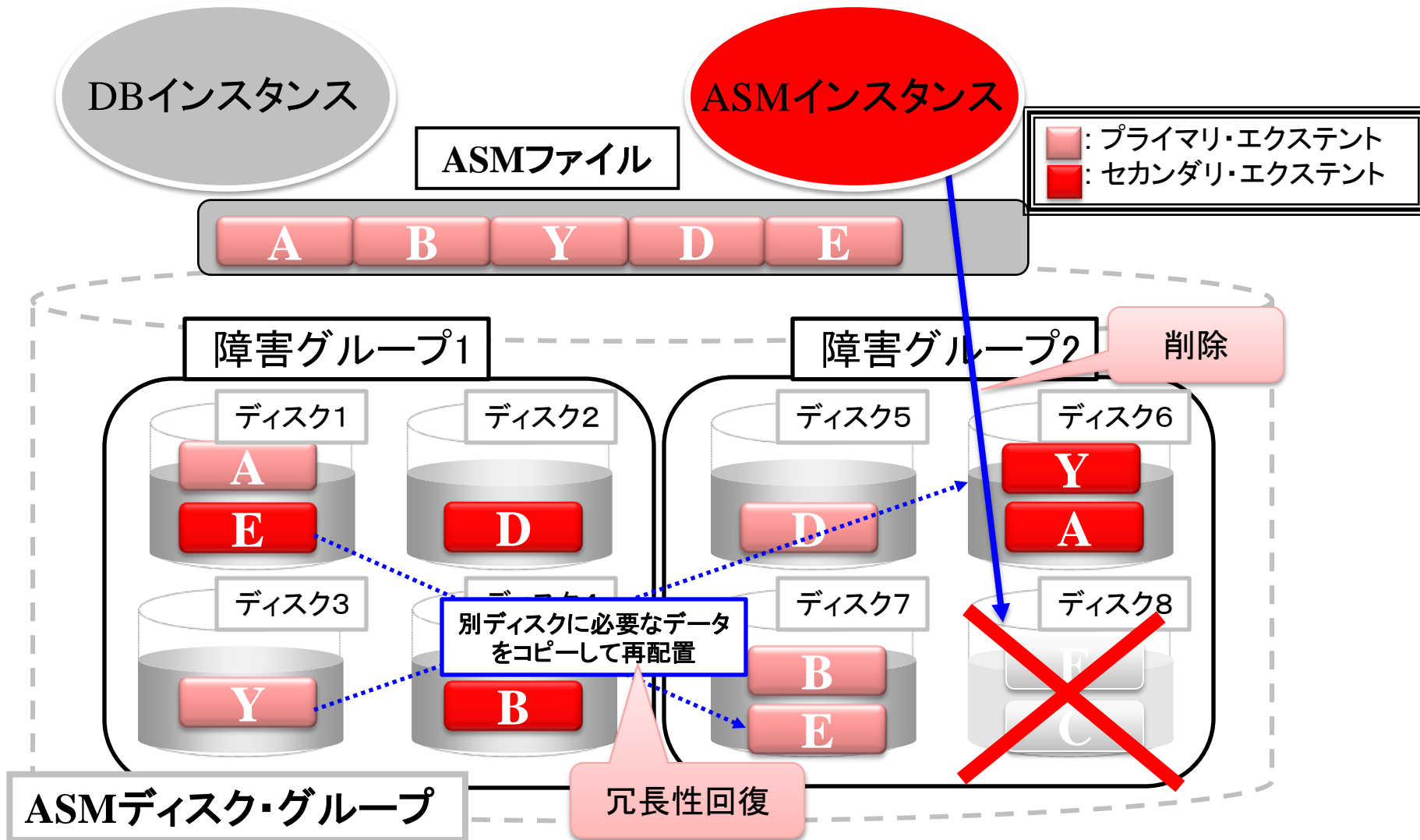
書き込みI/Oエラー時の動作

高速ミラー再同期による冗長構成の回復



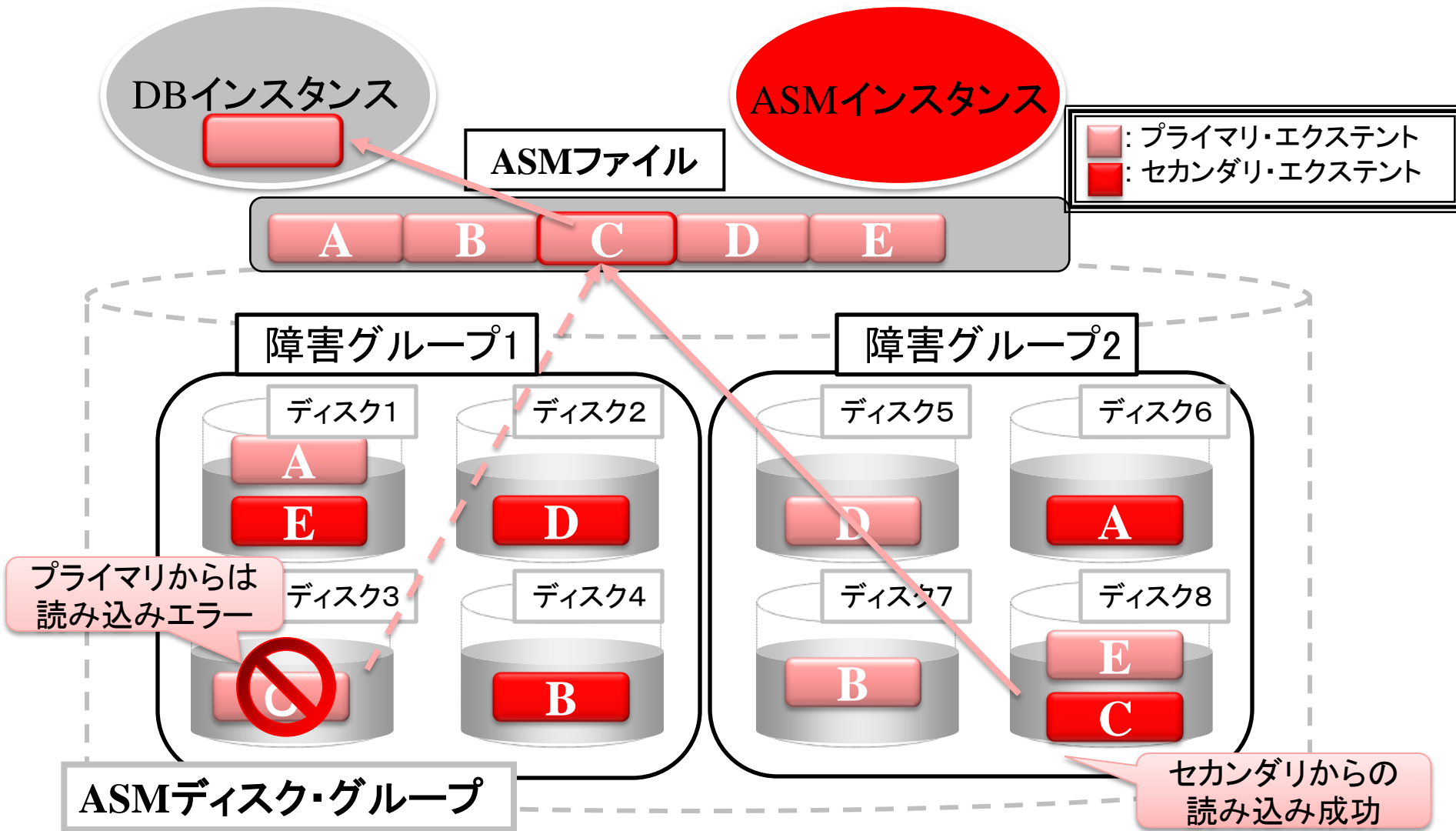
書き込みI/Oエラー時の動作

リバランスによる冗長構成の回復



読み込みI/Oエラー時の動作

データ「C」への読取り要求



ORACLE

読み込みI/Oエラー時の動作

ログへの出力例

- OracleがI/Oエラーを検出すると、自動でミラーへI/Oを行う
 - 以降はミラーへI/Oを行う
 - アプリケーションに対しては透過的(エラーはなし)
- DBインスタンスのアラート・ログ(一部抜粋)

...

ORA-27061: 非同期I/Oの待機に失敗しました。

...

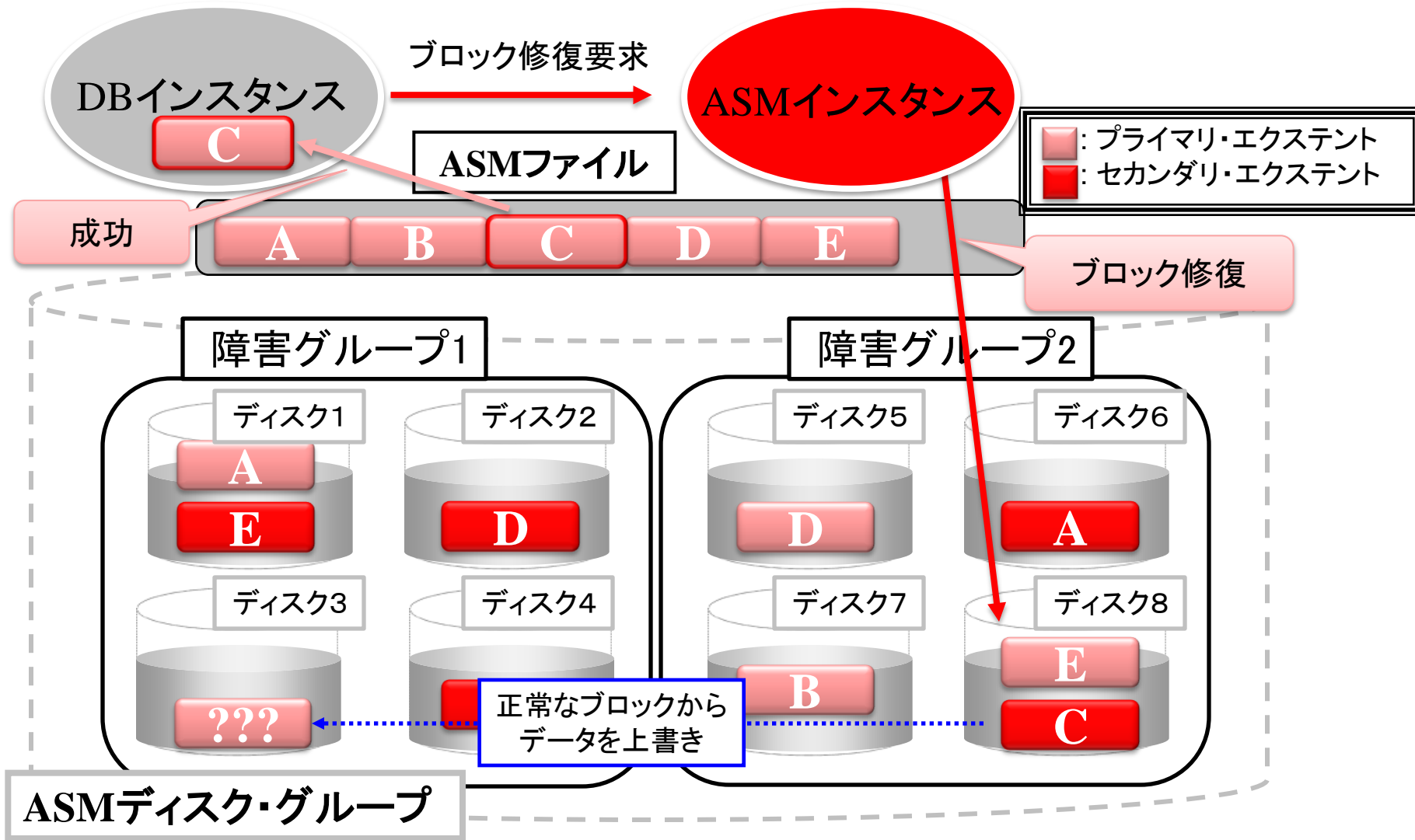
WARNING: failed to read mirror side 1 of virtual extent 0 logical extent 0 of file 263 in group 3 from disk 2 allocation unit 50908; if possible, will try another mirror side

NOTE: **successfully read mirror** side 2 of virtual extent 0 logical extent 1 of file 263 in group 3 from disk 1 allocation unit 316

自動的にミラーから読み込む

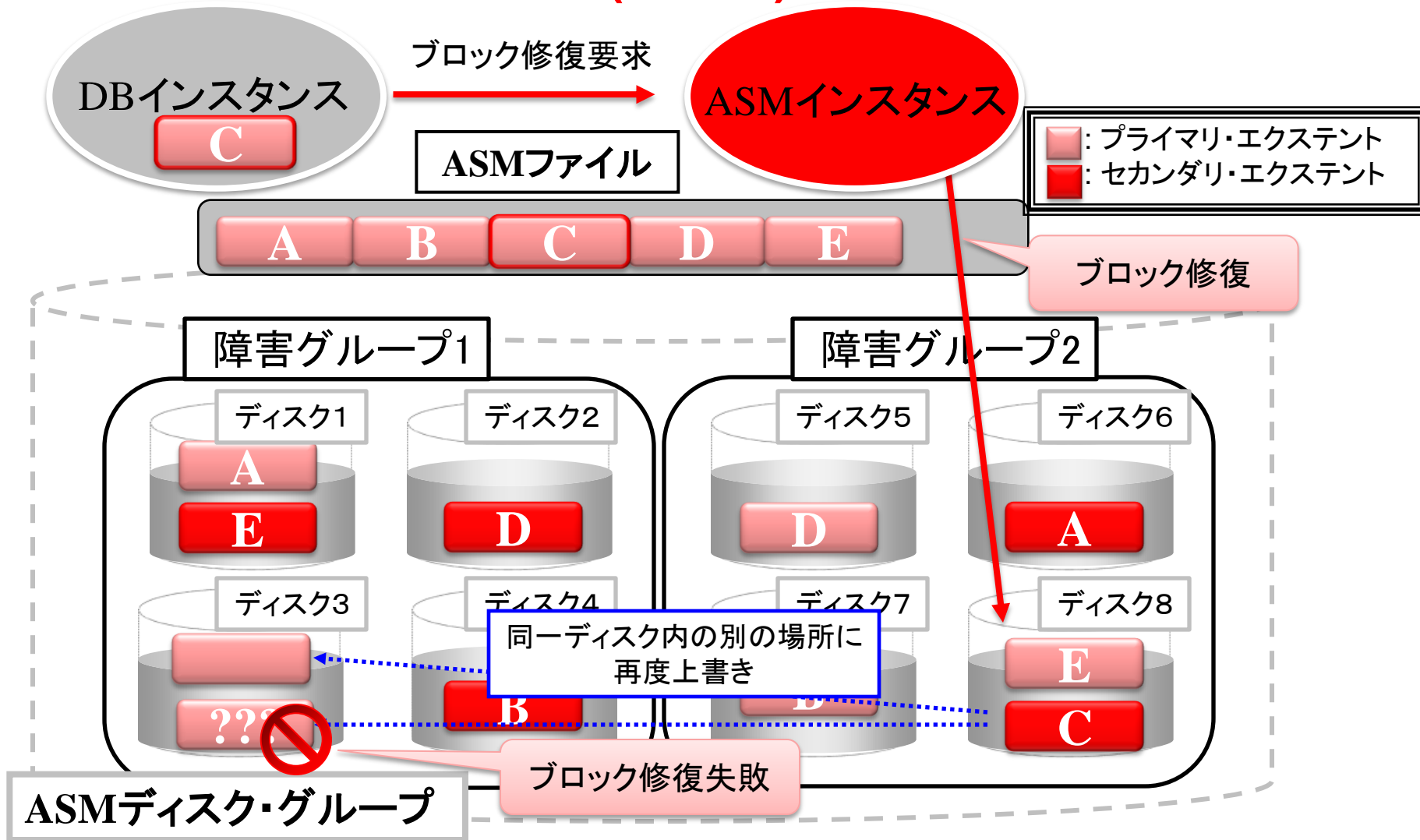
読み込みI/Oエラー時の動作

データ「C」のブロック修復



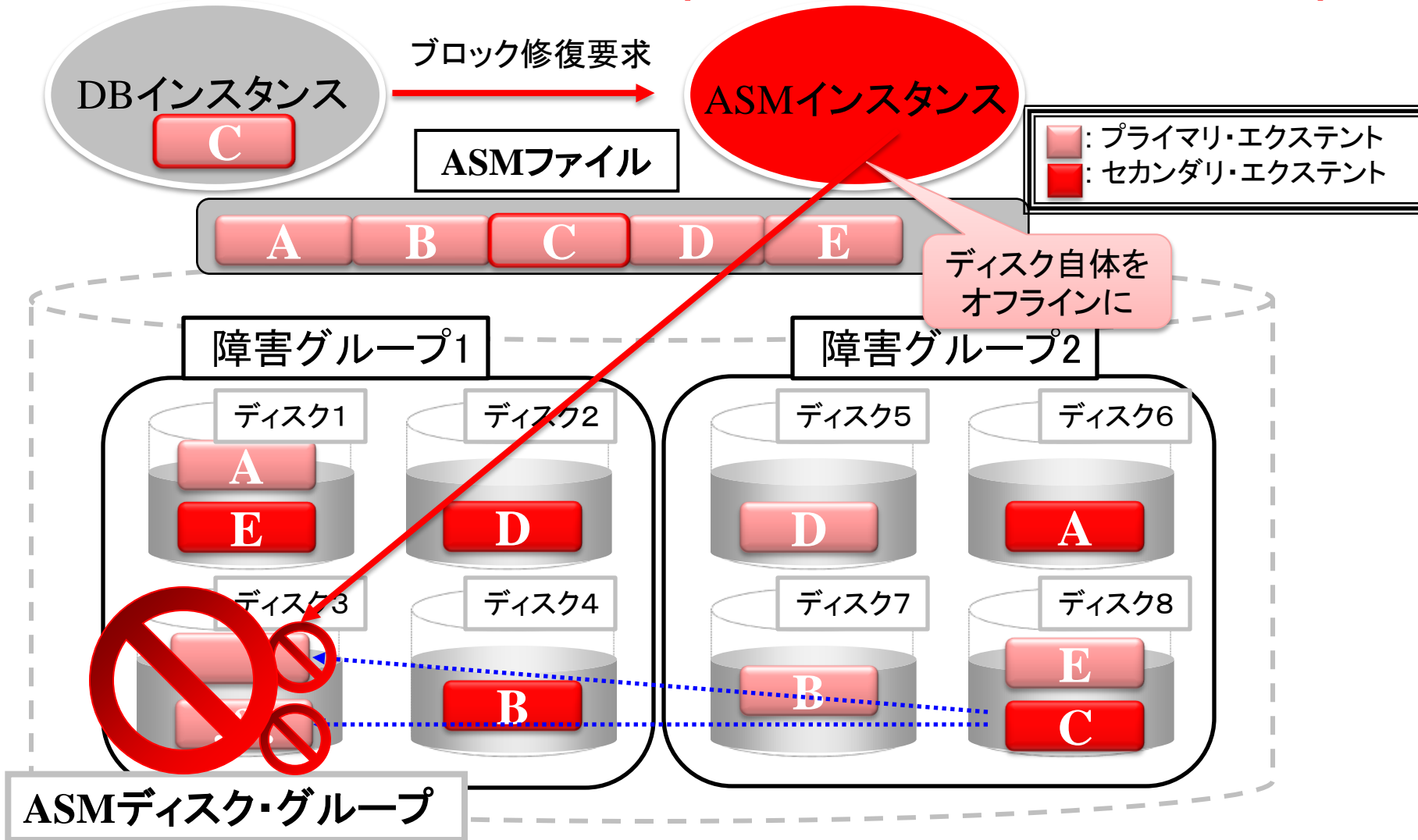
読み込みI/Oエラー時の動作

データ「C」のブロック修復(再実行)



読み込みI/Oエラー時の動作

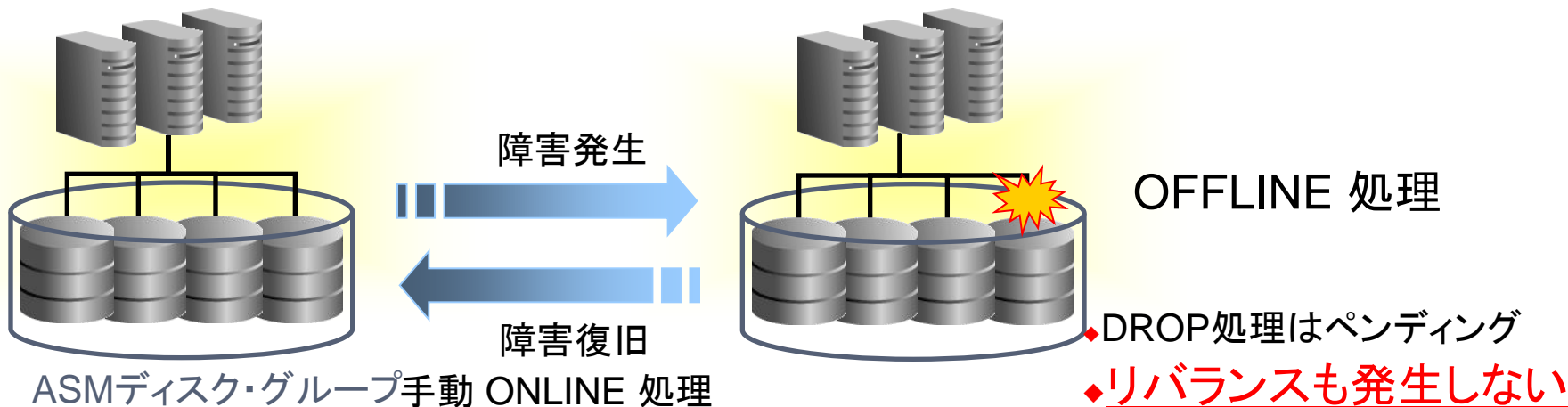
ディスク3をオフライン状態に(高速ミラー再同期機能が動作)



高速ミラー再同期機能

インテリジェントな障害復旧

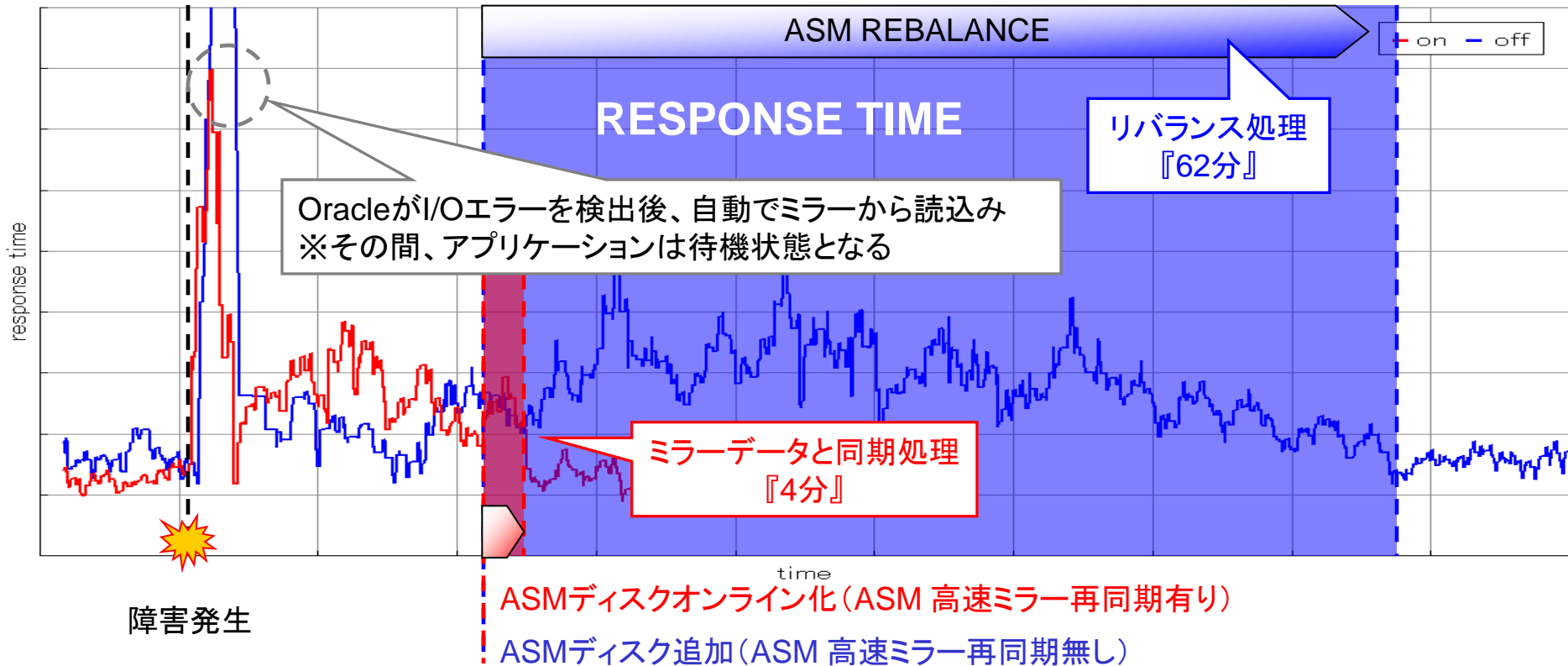
- 一時的なディスク障害の復旧を高速化
 - 11g より以前のバージョンでは、ディスク・パス障害や電源障害などの一時的な I/O 障害でディスクが使用できない場合、データ破損がない場合でもリバランスが行われていた
 - ディスクの自動削除をペンディングさせることが可能
 - ASM ディスクの手動 ONLINE/OFFLINE 処理が可能
 - ONLINE 復旧時、必要最小限のエクステントだけを同期



高速ミラー再同期の有効性

業務への影響と障害復旧までの時間

- 高速ミラー再同期により、障害発生時の作業コストが減少し、障害復旧までの時間を短縮できることを確認



自動不良ブロック修復機能

- メディア障害時、不良ブロックのデータをミラーからリストア可能
 - ASM によるミラーリング構成をとっている場合
 - 不良ブロック読み込み時にI/Oエラーが返ってくると自動で実行
 - ASMCMDユーティリティあるいはEMを使っての手動実行も可

メンバー ディスク

表示

すべて選択 | 選択解除

ディスク

NORMAL1 FG NORMAL2 FG

ディスクの不良ブロックのリカバリ: NORMAL1

ASMにより、検出された不良ブロックが自動的に修復されます。ディスク修正などその他の操作で不良ブロックを検出した場合、この操作を使用して、破損ブロックの範囲を修復できます。ストレージ・サブシステム・ハードウェアの場合と同様に物理ブロック番号を指定してください。

物理ブロック番号

* 開始

* 終了

ホスト 資格証明

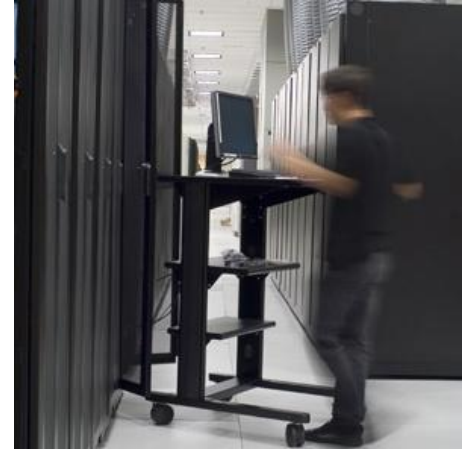
優先資格証明の使用

優先資格証明の上書き

* ユーザー名

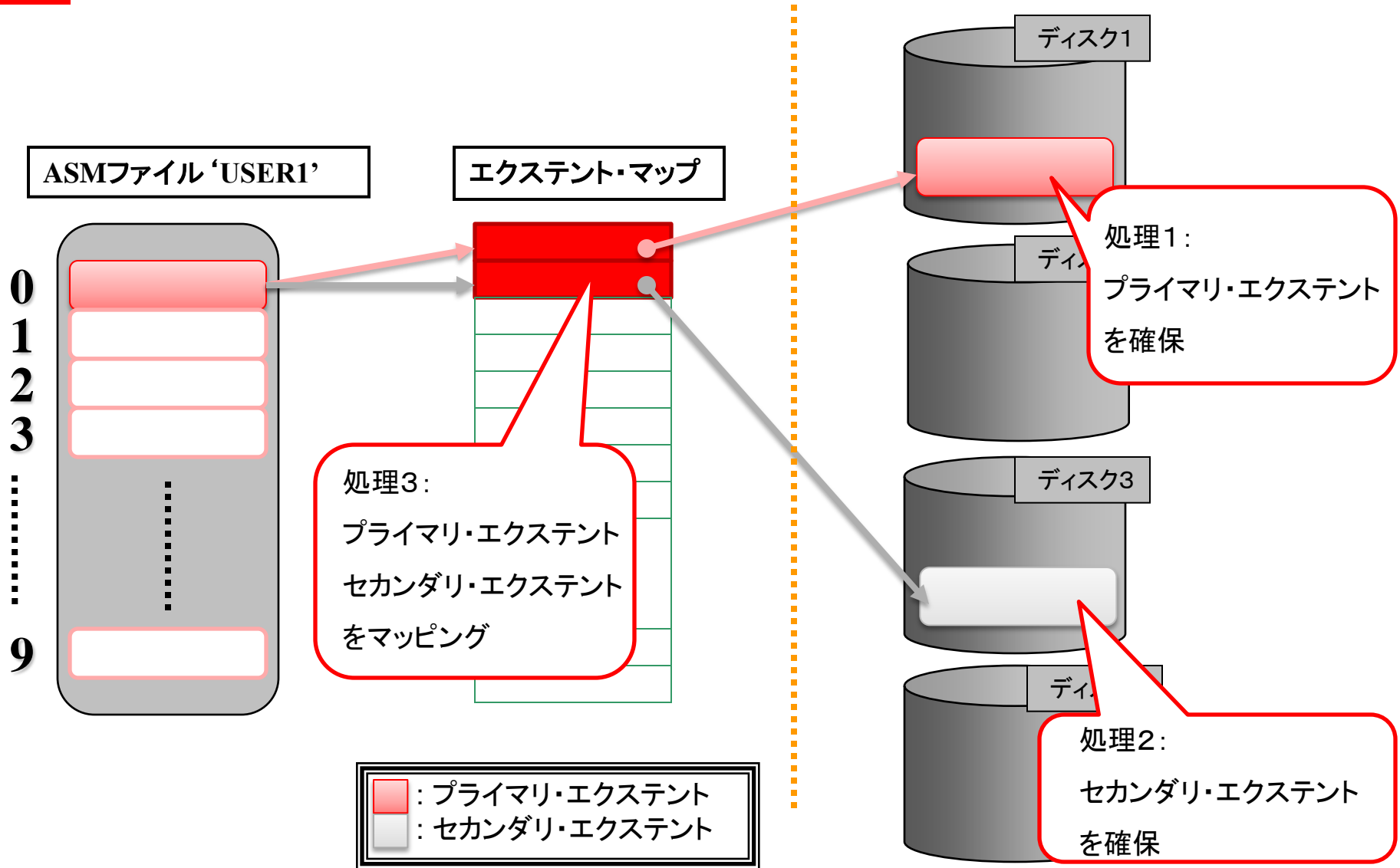
* パスワード

優先資格証明として保存



3. ASMの整合性担保のための仕組み

ASMファイルの領域確保の仕組み



ASMファイル処理のREDO情報

Active Change Directory

- ASMファイルの構成変更に関連する処理(resize / create / drop / rebalanceなど)は全てACDに記録される
- ASMファイル領域確保中に万が一ASMインスタンスが停止しても、最低限の整合性を確保できるようにする仕組み

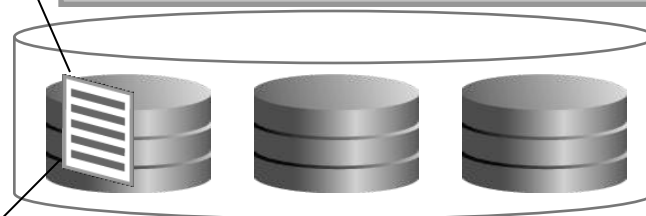
ACD

処理1:プライマリ・エクステント確保

処理2:セカンダリ・エクステント確保

処理3:エクステント・マップ更新

ASMディスク・グループ

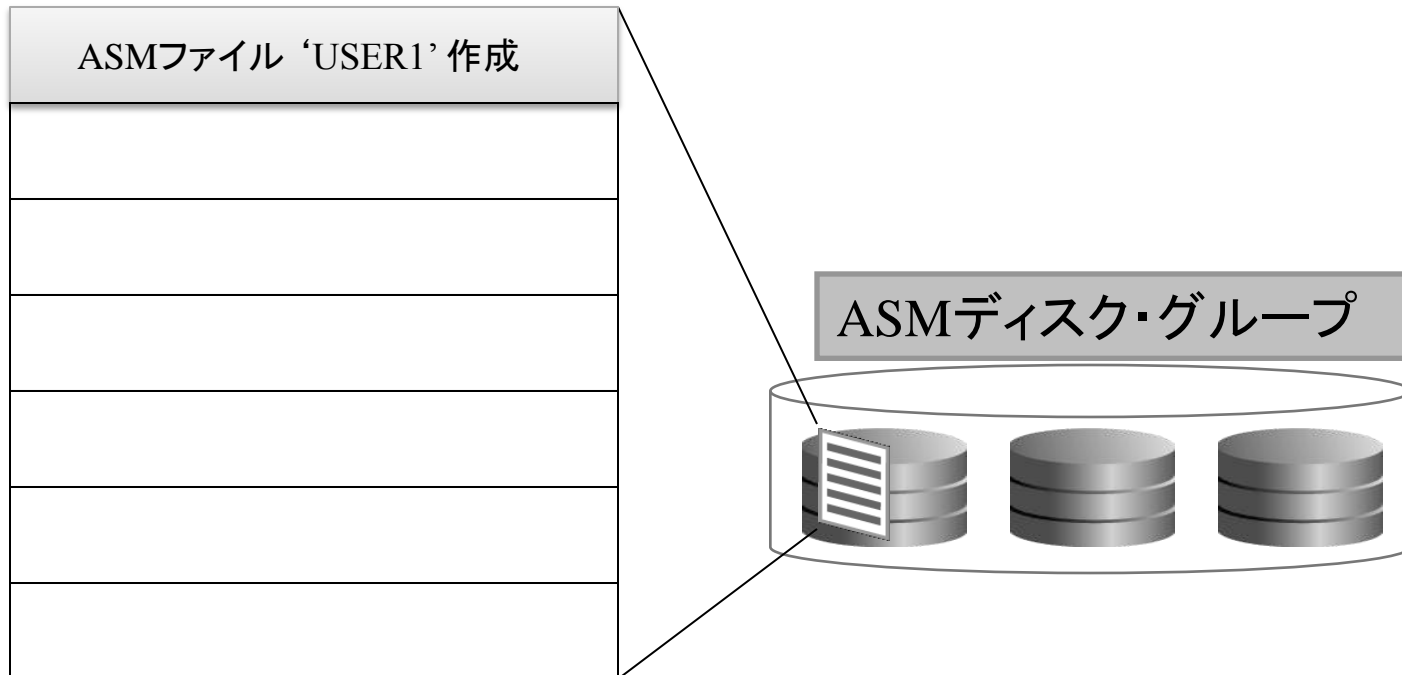


ASMファイル処理のUNDO情報

Continuing Operation Directory

- ASMファイルに対する実行中の処理を記録
- ASMファイル作成途中でASMインスタンスが停止すると、まずACDを適用して整合性を保った状態でCODをフラッシュ
- ユーザがファイル作成操作を再実行可能な状態に戻すための仕組み

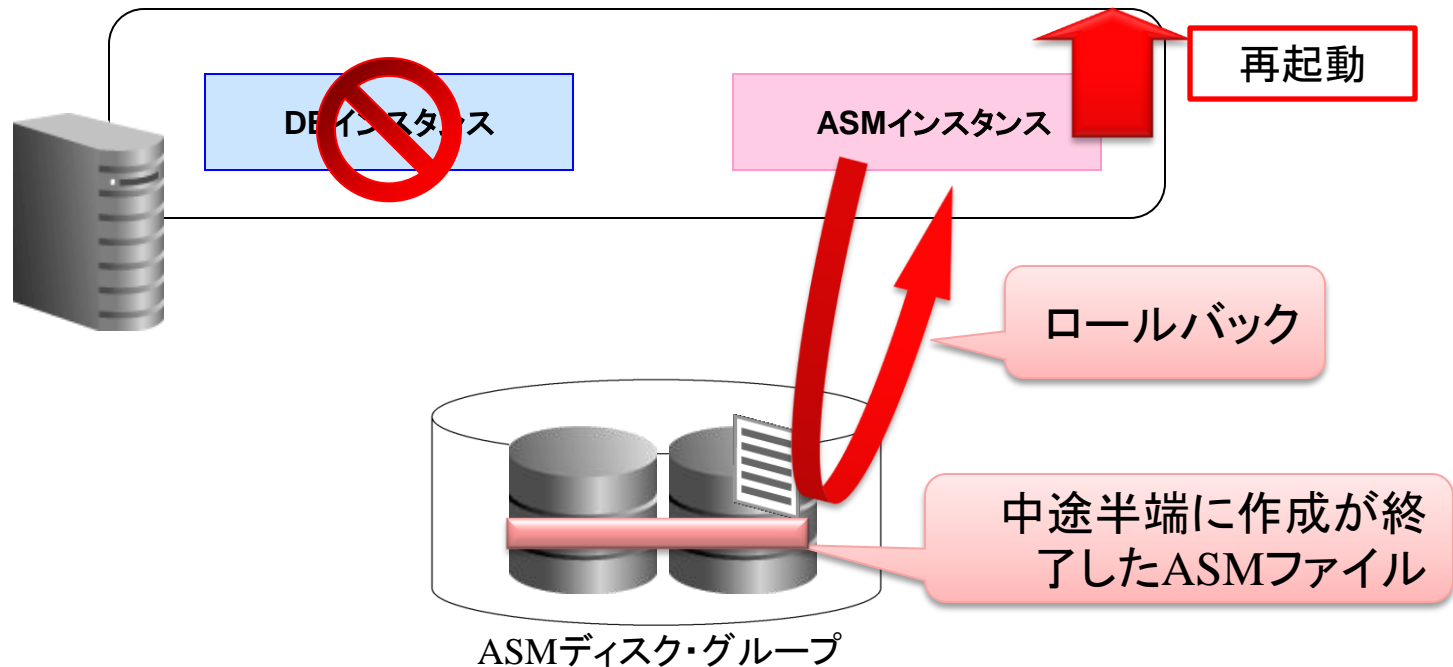
COD



ASMファイル処理のロールバック

シングル・インスタンス環境の場合

- ACDとCODの仕組みにより、作成途中でASMインスタンスが停止した場合はそのインスタンスの再起動時に整合性の取れた状態で処理がロールバックされる
- シングル環境の場合、ASMインスタンスに接続しているDBインスタンスはASMインスタンス停止の影響を受ける



ASMファイル処理のロールバック

クラスタ環境の場合

- クラスタ環境の場合、停止したノード以外の生き残っているノードでASMファイル作成のロールバックが引き継がれる
- クラスタ環境の場合は停止したASMインスタンスに接続していたDBインスタンスだけが停止の影響を受ける

