

1 はじめに

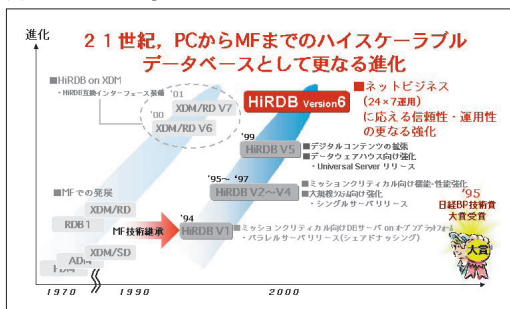
HiRDBは、永年にわたり基幹系メインフレームで培った高い信頼性、柔軟な運用性を継承し、オープンプラットフォームにおけるデータベースの基本要件を徹底的に追求した純国産RDB製品である。本冊子では、これからのネットビジネスを支えるためにサポートした「HiRDB Version 6」の新機能を中心に、HiRDBの特長や機能、サービスについて解説する。

1.1 | HiRDBの歩み

データベースの重要な要素である「信頼性」「運用性」「スケーラビリティ」を重視したHiRDBは1994年に登場し、以来ミッションクリティカルな分野で豊富な実績を積み重ね進化しつづけてきた(図1-1)。1990年代はPCやサーバの高性能化により、オープンプラットフォームへのダウンサイジングが進展し、クライアント/サーバシステムやWebベースのシステムが広まった。そのなかで、ミッションクリティカルなシステムをオープンプラットフォームで構築するニーズも高まり、安価なサーバを複数台並べてスケーラビリティを確保する並列DBMSが登場してきた。HiRDBはVersion 1からその並列技術への取り組みを行い、シェアドナッシング方式の平行サーバをリリースした。1995年には、それらの並列技術の1つである並列一括更新技術と並列リカバリ技術が評価され、日経BP技術賞大賞を受賞した。

SQL仕様がSQL99へと改訂されつつあるなかで、DBMSの世界でもオブジェクト指向の概念がクローズアップされるようになり、HiRDBは1999年にVersion 5でSQL99、オブジェクト指向技術に対応した。それにより数値・文字といったリレーショナルデータに加え

図1-1：HiRDBの歩み



て、テキスト・イメージ・映像・地図情報といったデジタルコンテンツをデータベースで一元管理でき、アプリケーションプログラムからのシームレスな操作を可能にした。また、同時期にそれらデジタルコンテンツを管理する上位アプリケーション群を用意し、「HiRDB Universal Server」として発売した。

そして、最新の「HiRDB Version 6」では、24時間365日稼働するネットビジネスで求められる高い信頼性と柔軟な運用性を更に追求するとともに、ネットビジネス基盤に求められる先進性を強化した。

1.2 | HiRDB Version 6のねらい

昨今、eマーケットプレイスといったネットビジネスに代表されるIT技術の急速な社会への浸透は、ビジネスを24時間365日の環境に変えつつある。このような環境においては、ノンストップビジネスを支える高い信頼性・運用性、ビジネスチャンスを逃さないスピードはもちろん、新しいサービスや価値を創造する変革が必要であり、先進のWeb技術への対応とともに、スケーラビリティと堅牢性を持ったシステムの追求が重要になってくる(図1-2)。

1.2.1 24時間365日ノンストップビジネスを支える信頼性と運用性

サービスの停止がビジネスチャンスの重大な喪失に直結するネットビジネスの世界において、システムダウンをいかに最小限にとどめるかという信頼性(耐障害性)と、サービスをストップせずデータベースのメンテナンスやバッチ処理などの定期的業務を行える運用性が重要となる。

HiRDB Version 6では、ノンストップサービスに耐えられるよう、これまでのミッションクリティカルな分野

図1-2：HiRDB Version 6のねらい

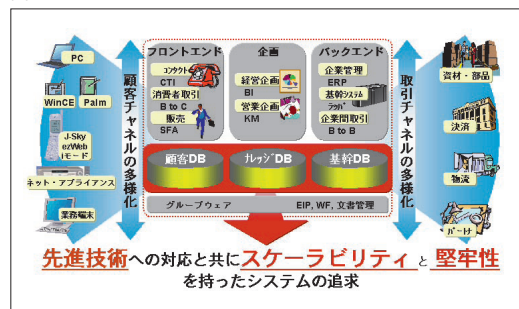
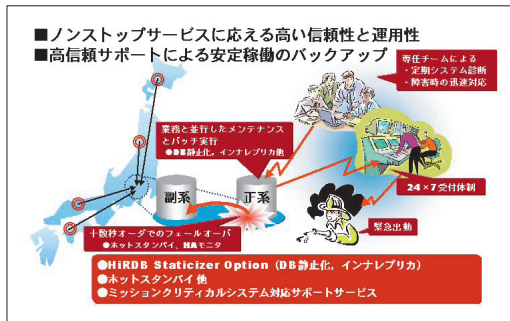


図1-3：高度な信頼性と運用性を提供



でのメインフレームの実績、高い品質の継承に加え、万一障害がおきた場合に短時間でのサービス再開を実現する高速フェールオーバー機能などの高度な技術で対応する。運用面では、オンライントランザクションに影響を与えることなく、バックアップや再編成といったデータベースのメンテナンス業務やバッチ業務の並行処理を実現。加えて、開発者による迅速かつ確かなサポートを提供しており、製品とサポートの両面からシステムの堅牢性を高めている (図1-3)。

1.2.2 ネットビジネス対応の

Webアプリケーションとの親和性

eコマースやeマーケットプレイスといったネットビジネスへスピーディーに対応するためには、柔軟なシステム間連携を実現するXML (eXtensible Markup Language) やWebアプリケーションの開発言語として広まりつつあるJava™といった先進かつコモディティな技術とのインターフェイスがDBMSでは必要となる。HiRDB Version 6では、これらXMLやJava™といったインターネット技術を使ってWebアプリケーションからデータベースを利用できる。これによりネットビジネスアプリケーションのスピーディーかつ柔軟な開発を支援し、イン

図1-4：先進技術への対応とスケラビリティ

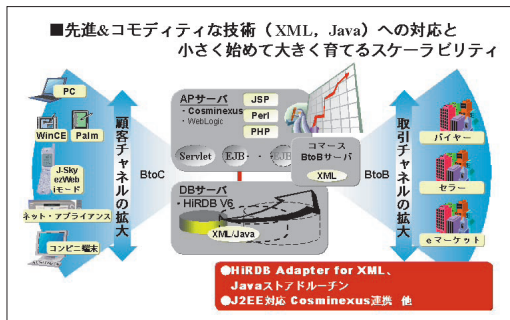
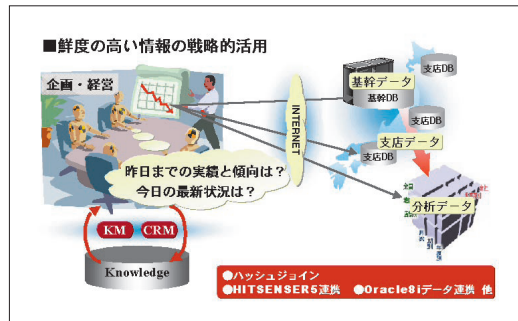


図1-5：強力なデータウェアハウス基盤を提供



ターネット環境との親和性を高める (図1-4)。

1.2.3 データ分析をビジネスに活かす データウェアハウス基盤の強化

「昨日までの実績と傾向は?」、「今日の最新状況は?」。こうした過去の履歴データの多面的把握と蓄積されたノウハウがあってこそ、鮮度の高い情報を戦略的に活用できる。HiRDB Version 6では、情報活用の局面で行われるアドホックな結合検索や、既存の異なるDBMSで管理しているデータとの連携を強化することで、散在した情報の共有とタイムリーな活用を促進する。

2 高信頼ノンストップDBサーバ

24時間365日のノンストップサービスが当然となっている現在のネットビジネスにおいては、オンラインサービスと並行したデータベースメンテナンスやバッチ業務の実行、あるいは耐障害性、さらには運用の容易性などが重要なポイントとなる。HiRDB Version 6は数多くの機能によってこれらのポイントに対応し、ネットビジネスにおけるノンストップサービスの実現を強力にバックアップしている。

2.1 | オンラインでの運用を支えるコア技術

■データベース静止化

データベース静止化とは、メモリ内 (DBバッファ) の更新情報をディスクに強制出力し、ディスクの整合性 (トランザクションの完結) を保証した状態にすることである。このデータベース静止化は、日立メインフレームデータベース製品の「TMS-4V」や「XDM/RD」で豊富な実績を持つ技術であり、HiRDB Version 6にも継承

された。HiRDBでは、定期的なシンクポイントのタイミングでDBバッファの更新情報をディスクに書き込んでいるが、このデータベース静止化では任意の時点でディスクの整合性を確保できるのが特長である。

データベースの静止化は、HiRDBデータベース格納領域であるRDエリアの閉塞コマンド（pdholdコマンド）にて行い、コマンド実行時の仕掛り中トランザクションの扱いについては、次のオプションから選択できる。

- ・仕掛り中の全トランザクションの完了を待って、DBバッファの更新情報をディスクに書き込む。
- ・仕掛り中のトランザクションの完了を待たずに、DBバッファの更新情報をディスクに書き込む。その後の更新情報がディスクに出力される場合は、その物理ログを取得する。

また、データベース静止化中に入ってくる新たなトランザクションの扱いについては、次のオプションから選択できる。

- ・参照は可能にして、更新をエラーにする。
- ・参照は可能にして、更新は静止化が完了するまでの間（数秒程度）待ち状態にする。
- ・参照も更新も可能にする。
- ・更新された場合はその物理ログを取得する。

上記オプションのうち、データベース静止化中に更新も可能な場合は、物理ログと組み合わせた運用でディスクの整合性確保が可能である。例えば静止化したデータベースをバックアップする場合、静止化が完了した時点でバックアップを作成しても更新トランザクションが途中の状態で存在している場合があるので、これをそのままリストアすることができない。そこで、バックアップに静止化中の物理ログを被せることで静止化開始後の任意の時点で回復できる。

データベース静止化は、あらかじめ二重化して運用していたミラーディスクのうち片方を、整合性を保証した状態で切り離すための有効な手段となる。日立のディスクアレイサブシステム「SANRISEシリーズ」のペアボリューム作成機能MRCF (Multiple RAID Coupling Feature) や、その他のミラーリングソフトウェアとの組み合わせで、整合性の取れた多重化データベースを実現できる。例えばSANRISEとの組み合わせの場合、RDエ

リア（HiRDBデータベース格納領域）に対応した論理ボリュームをMRCFで多重化し、静止化後にペアボリュームを切り離すことで、整合性を確保した複製データベースを作成できる。

■ データベースを多重化するインナレプリカ

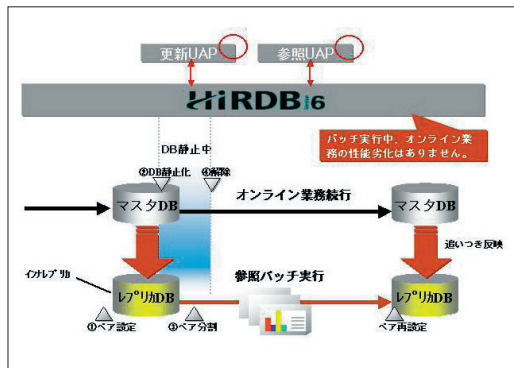
インナレプリカとは、SANRISEのMRCFやミラーリングソフトウェアを使って多重化（n重化）したRDエリアを、多重化した状態で管理する機能をいう。1つのHiRDBシステム内で多重化したデータベースのどれにアクセスするのか切り替えることも可能である。これまでのデータベース二重化とは、一般的に信頼性向上を目的としたストレージのレベルでの二重化がほとんどであり、DBMS自身が二重化を管理していたわけではない。したがって、多重化したデータベースの任意面に対してアクセスすることができなかったが、HiRDB Version 6はこれを可能にした。

インナレプリカを利用するには、あらかじめデータベース定義を多重化しておく。この多重化は、データベース格納領域であるRDエリアを単位として行う。レプリカRDエリアの定義はデータベース構成変更（pdmod）ユーティリティを用い、replicate rdarea制御文で行う。正側のデータベースをアクセスするか副側（レプリカ側）のデータベースをアクセスするかは、クライアント単位や運用コマンド（pddbchg）で選択でき、クライアント主体あるいはサーバ主体での運用を選択できる。例えば、二重化したRDエリア（データベース定義）に対応する論理ボリューム（実体）の片側を切り離す（ペア解除）ことによって、副側ボリュームを同一の表定義を持った物理実体の異なるデータベースとしてアクセスできるようになる。

前述のデータベース静止化とインナレプリカを組み合わせることで、バックアップやデータベース再編成といったメンテナンス、および参照系バッチ業務を、オンライン業務の性能に影響を与えないで並列に実行することを可能にする。例えば、月末のある時点でデータベース静止化を行い、ペア設定したディスクの副側ボリューム（レプリカDB）を切り離す（ペア解除）。これにより正側ボリューム（マスタDB）を使ったオンライン業務を続行したまま、日次/月次データの一括集計といった定型バッチ処理を副側ボリュームを使って並行実行することが可能になる。

さらに、ペア解除している間に正側ボリュームのマス

図2-1：DB静止化およびインナレプリカによる運用例



タDBに対して実行されたトランザクションについては、ベア再設定により副側ボリュームのレプリカDBに反映することもできる。SANRISEのMRCFを使えば、高速にデータ内容を追いつかせることが可能となる。

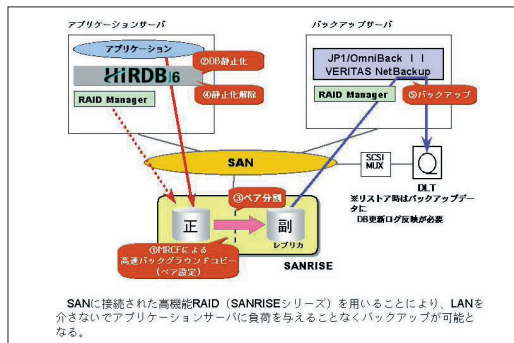
■ LANフリー高速バックアップ (SAN連携)

24時間365日システムを稼働させるためには、バックアップの取得は欠かせない。HiRDBのデータベース静止化機能とネットビジネス向けのストレージ「SAN」の機能を組み合わせた運用により、アプリケーションサーバとLANのサービスに影響を与えることなく、データベースのバックアップを行うことが可能になる。

例えばSANRISEのMRCFを用いると、以下の手順となる。

- ①あらかじめバックアップ対象となるRD エリアが格納されているボリュームをMRCFで二重化 (ベア設定) する。
- ②バックアップ対象となるRDエリアを静止化する。
- ③静止化完了後、対象ボリュームをベア分割する。

図2-2：LANフリー高速バックアップの運用例



- ④対象RDエリアの静止化を解除する。
- ⑤バックアップサーバから「JP1/OmniBack II」などのバックアップソフトウェアで、分割した副側ボリュームに格納されている複製RDエリアをDLT (Digital Linear Tape) 装置などにバックアップする。

このように、専用のバックアップサーバとの間をSANで接続することにより、バックアップサーバで直接バックアップを取得することが可能になる。つまり、アプリケーションサーバからデータを読み込み、LANを介してバックアップサーバに転送する必要がないため、アプリケーションサーバとLANへの負荷を与えることがない。

■ オンラインでのデータベース再編成

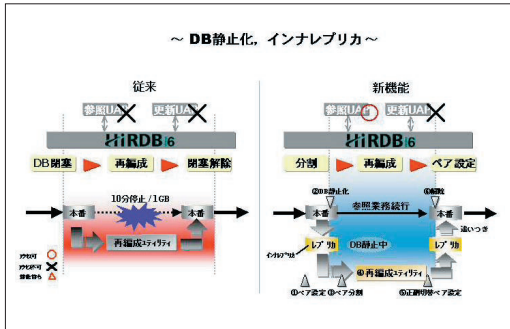
一般に、長い期間にわたってデータベースを運用していくと、レコードの追加と削除が重なった結果、DB内の格納効率が低下する (断片化が発生)。この時、ガベージが発生し、アクセス速度が遅くなることがある。通常この状態を解消するために、データベースの再編成 (データベースをアンロードしてリロード) を行う。しかし、従来のアンロード/リロードによる再編成では、どうしてもオンラインサービスを停止する必要があった。HiRDB Version 6のDB再編成 (pdorg) ユーティリティでは、オンライン業務 (参照サービスのみ) に影響を与えずにデータベースを再編成することが可能となっている。例えばSANRISEのMRCFを用いると、以下の手順となる。

- ①あらかじめ再編成の対象となるRDエリアが格納されているボリュームをSANRISEのMRCFで二重化 (ベア設定) する。
- ②対象RDエリアを更新不可モードで静止化する。
- ③静止化完了後、対象ボリュームをベア分割する。
- ④ベア分割完了後、分割した副側ボリュームに格納されている複製RDエリアをDB再編成 (pdorg) ユーティリティにて再編成する。
- ⑤再編成完了後、分割したボリュームを正と副を入れ替えて再ベア設定し、再編成した副側ボリュームの内容を正側ボリュームにバックグラウンドで反映する。
- ⑥対象RDエリアの静止化を解除する。

■ オンラインでのデータ再配置

サービスの開始後にデータ量が増加すると、それに応

図2-3：オンラインでのデータベース再編成

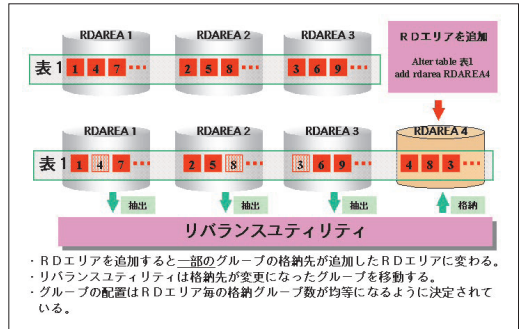


じてデータエリアを追加する必要が出てくる。特にインターネット関連のビジネスでは、サイズの大きなデジタルコンテンツなどのデータもあるため、その規模が急激に増加することがあり、頻繁にデータエリアを追加しなければならない。ところが、従来のフレキシブルハッシュ分割した表に対する単純なRDエリアの追加方法では、特定のディスクにI/Oが集中してしまう傾向にあった。

これを解消したのがHiRDB Version 6のリバランス (pdrbal) ユーティリティである。このユーティリティは、サービスを続行したまま表にRDエリアを新規追加しても、ディスクへのI/Oが偏らないようにデータのバランスを保つことができる。また、DB容量が膨大な場合、1日では全データの再配置が完了しないケースも考えられる。このような場合では、本ユーティリティを時間指定で複数回に分けて実行する事ができる。例えば夜間のようなトランザクションが比較的少ない時間帯にのみ再配置を実施し、昼間のトランザクションが多い時間帯では再配置処理を一時中断する運用ができるのだ。これを何日か繰り返すことにより業務への影響を最小限にしながら、次第にバランスのとれたDBの状態にすることができ、どこかにI/Oが集中してしまうといったボトルネックが解消できる。以下に、このリバランス (pdrbal) ユーティリティの具体的な利用方法を示す。

- ①DB 構成変更 (pdmod) ユーティリティによってRD エリアを追加する。
- ②追加後、対象となる表に対してRDエリアを追加登録する (ALTER TABLE文)。
- ③リバランス (pdrbal) ユーティリティをバックグラウンドで実行する。

図2-4：リバランス機能の概要 (RDエリアの追加とデータの移動)



このようにリバランス (pdrbal) ユーティリティは、サービスを続行したままバックグラウンドで追加したRDエリアと既存のRD エリアとのデータ量を均一化する。DB管理者がデータを移動させる必要もまったくなく運用も容易だ。

さらに、このデータ均一化処理を高速化する工夫も施されており、高性能のDB再配置を実現している。以下に、どのように高速化しているのかを説明する。

ハッシュ分割した表には、行データをあらかじめハッシュキー値によって1,024個のグループに分割しておき、グループごとに各RDエリアに格納している。このため、RDエリアには複数のグループが割り当てられることになる。表にRDエリアを追加してリバランス (pdrbal) ユーティリティを実行すると、既存のRDエリア中から特定のグループだけを対象にデータを移動し、各RDエリアのデータ量を均一化する。これにより、より高速にデータの再配置を行うことができるのだ。

また、リバランス (pdrbal) ユーティリティは表を排他モードで占有することで高速に実行するオプションもサポートしている。これを利用すれば、オンライン中に

図2-5：リバランス機能の概要 (グループへのマッピング)

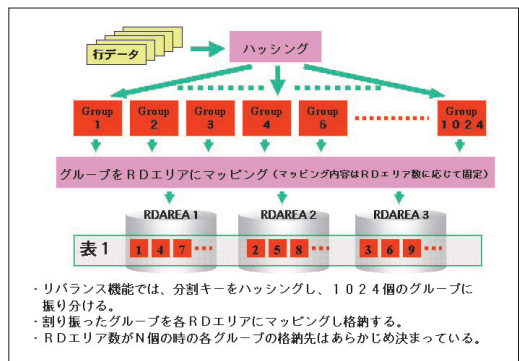
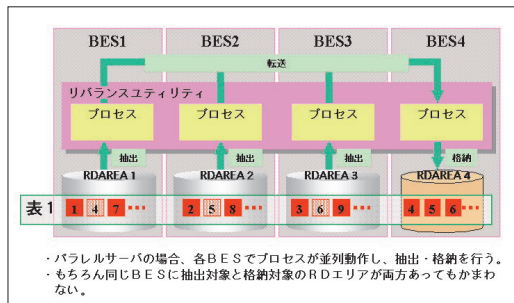


図2-6：リバランス機能の概要（パラレルサーバでのリバランス）



データ再配置をバックグラウンドで実行するよりも、より高速に実施することも可能だ。

■ オンラインでのデータ・インデクス一括登録

HiRDB Version 6では、データの一括登録やそれに伴う全文検索インデクスの再作成など、負荷の高いバッチ業務を、オンラインの参照サービスを続行したまま性能への影響を与えないで実行できる。

具体的な手順はオンラインでのデータベース再編成と同様、ペア設定したディスクの副側ボリューム（レプリカDB）に対してデータロード、インデクスの一括作成、全文検索インデクスの作成などのバッチ業務をバックグラウンドで実行する。この機能は日次/月次データの一括登録など、負荷の高いバッチ業務においてもきわめて有効である。

2.2 | 高度な耐障害性と容易な運用性

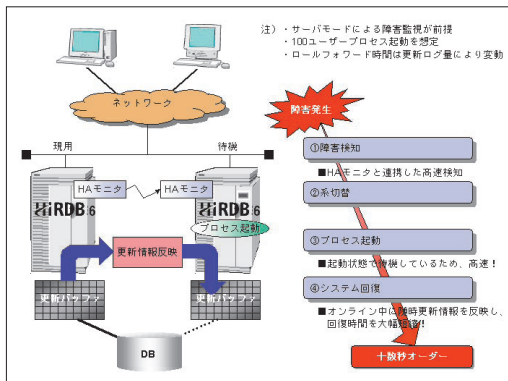
■ 高速系切り替えとホットスタンバイ

HiRDB Version 6では、障害発生時に予備のサーバへの切り替え時間を短縮する目的で、ホットスタンバイ機能をサポートしている。

従来のクラスタ構成では、障害が発生した場合にクラスタソフトウェアと連携して予備のサーバに切り替えるコールドスタンバイであった。ホットスタンバイ機能では、予備のサーバでもあらかじめHiRDBプロセスを立ち上げた状態で待機することにより、切り替え時間を短縮する。

また、障害を高速に検知するHAモニタ機能との連携により、従来数分かかっていた切り替え時間を1分オーダーに短縮可能とした。さらに、更新情報などをホットスタンバイの予備サーバへ逐次反映することによって予備のサーバでの回復時間の短縮化を実現し、2002年に

図2-7：ホットスタンバイ機能



は十数秒のオーダーまで切り替え時間を短縮できる予定だ。

■ 柔軟な障害時対応を可能にする系切り替え

業務実行用のサーバマシンとは別に予備のサーバマシンを準備しておくことで、業務処理中に障害が発生した場合でも、クラスタソフトウェアと連携して予備のサーバマシンに自動的に切り替え、回復処理の後に業務処理を続行できる。この機能をHiRDBでは「系切り替え機能」と呼ぶ。これによりシステムの停止時間を短縮することが可能となる。系が切り替わるのは、次に示すような障害が発生した場合である。ただし、HiRDBが単独で異常終了した場合はHiRDBを再開する。

- ・再開できないハードウェアの障害
- ・OSまたはリポートを伴うような障害

業務処理をしている系を実行系、待機している系を待機系と呼び、系切り替えが発生すると、実行系と待機系が入れ替わる。その場合、システム構築時や環境設定時に二つの系を区別するため、最初に実行系として起動する系を現用系、待機系として起動する系を予備系とも呼ぶ。システム起動後に実行系と待機系の交替を繰り返しても、現用系と予備系の関係は変わらない。HiRDB Single Serverの場合はシステム単位で系を切り替え、HiRDB Parallel Serverの場合はユニット単位で系を切り替える。

系切り替え機能を使用する場合、現用系と予備系とで共有する外付けのハードディスク（キャラクタ型スペシアルファイル）が必要である。このハードディスクを「共有ディスク装置」という。共有ディスク装置は、系