DB2 Data Management Software

IBM

e business software

クラスターテクノロジー
到達点とその未来

@IT ﾃﾞｰﾀﾍﾞｰｽﾃｸﾉﾛｼﾞｰﾐｰﾃｨﾝｸﾞ with DB2
2002年10月11日
日本アイ・ビー・エム㈱ﾃｸﾆｶﾙ・ｻﾎﾟｰﾄ ｼｽﾃﾑ&ｳｪﾌﾞｿﾘｭｰｼｮﾝｾﾝﾀｰ
ICP-ｺﾝｻﾙﾃｨﾝｸﾞITｽﾍﾟｼｬﾘｽﾄ 出羽 奏太郎

IBM Software Group

お断り

- 当資料に含まれる情報は正式なIBMのテストを受けていません。また、明記にも暗黙にも、何らの保証もなく配布されるものです。この情報の使用、評価、実施は使用者の責任で使用者の環境に合わせて行ってください
- 当資料の他社情報は一般公開されている資料を参照したものであり、IBMは内容および実際の稼動を保証しません
- ｱｰｷﾃｸﾁｬｰ比較と今後の動向については資料作成者の私見であり、IBMの見解ではありません
- 当資料で取り上げた新製品の内容は出荷時点で変更になる場合があります
- 記載の会社名と製品名はそれぞれの会社の登録商標および商標です

DB2 Data Management Software

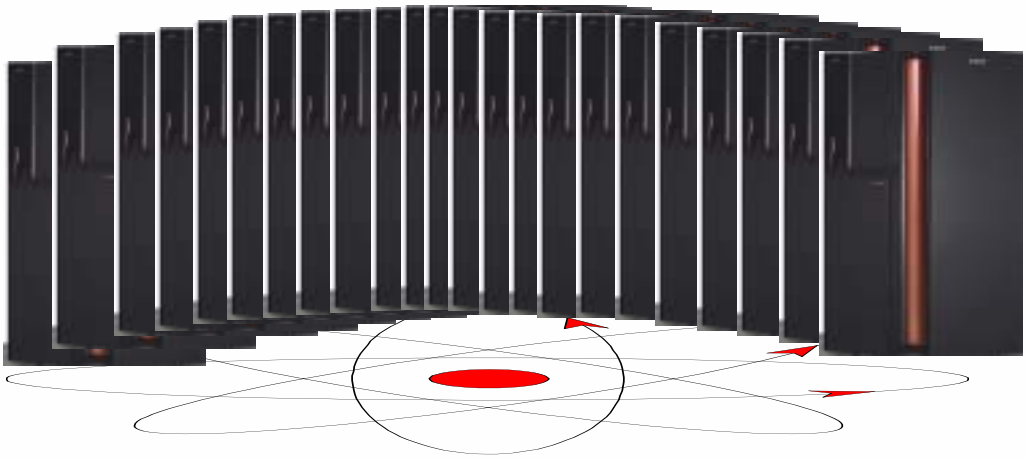
IBM

目次

- 区分データベース
- スケーラビリティ
- OLTPパフォーマンス
- 高可用性
- ｸﾗｽﾀｰ・ｱｰｷﾃｸﾁｬｰ比較
- DB2 ESE V8.1と今後

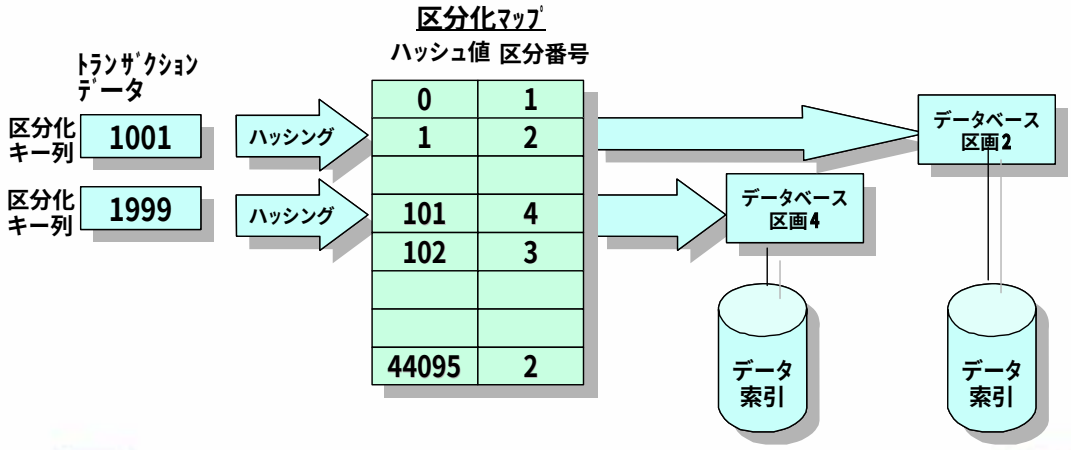
区分データベース

- データベースの区分したデータをもつサーバー群をパラレルオプティマイザーと高速の通信経路で接続したものです
  - ►全体で1つのﾃﾞｰﾀﾍﾞｰｽです
  - ►ひとつひとつの区分が独立したﾃﾞｰﾀﾍﾞｰｽｻｰﾊﾞｰに属します



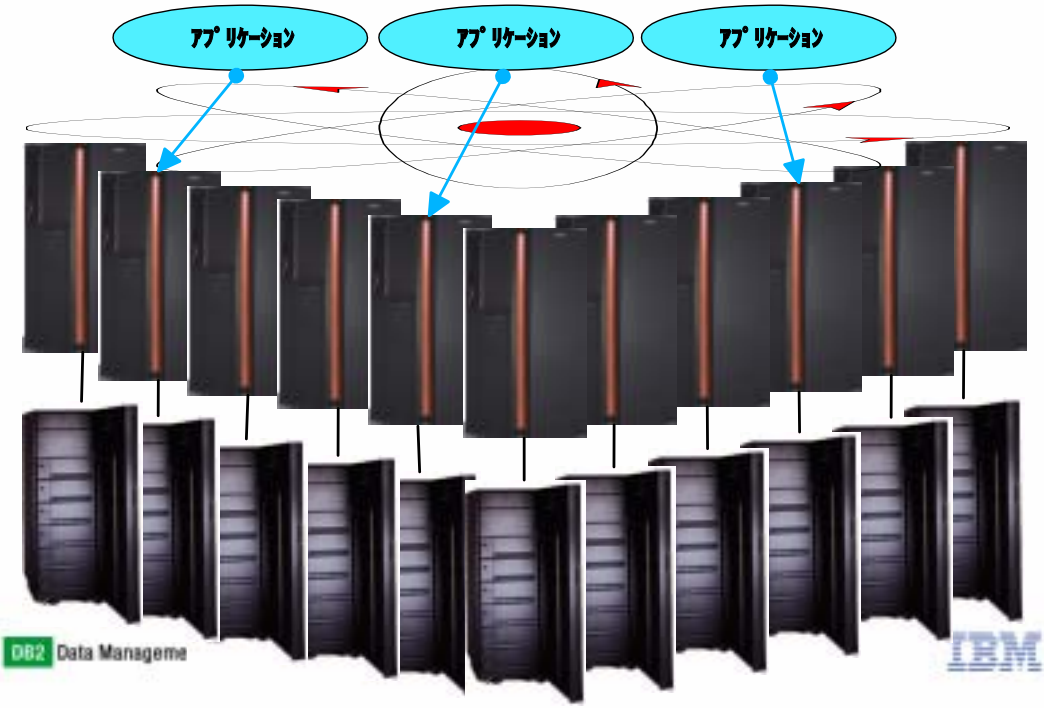
区分化のしくみ

- データをどの区分に置くのか？
  - ►各表で区分化のための列（複数可）を定めます（区分化キー列）
    - • CREATE TABLE文に指定します
    - –区分化キーの値でハッシングします
    - –お客様のﾃﾞｰﾀにあわせて均等化するためにｶｽﾀﾏｲｽﾞした区分化ﾏｯﾌﾟの間接参照で区分を決めます
  - ►各行の属する区分は軽い計算によって高速に決まります



区分データベース：透過的データアクセス

- アプリケーションはどの区分に接続してもよい
  - ▸パラレルオプティマイザーが透過的なデータアクセスを提供



DB2 Data Manageme

IBM

区分データベースの発展

1995 - DB2 PE 1.1
- DB2 PE on RS/6000 SP (SP2)
- 完全なMPP対応
- 業界初のTPC-Dベンチマーク結果の出版

1996 - DB2 PE 1.2
- Outer Join
- Case表式
- クエリーガバナー
- No Log表

1997 DB2 UDB EEE 5.0
- オブジェクトリレーショナル
  - LOB
  - ユーザー定義型
  - ユーザー定義関数
  - 参照整合性
  - チェック制約
  - default
  - トリガー
  - 表関数
  - 再帰SQL
- OLAP
  - スターバーストオプティマイザー
  - 共通表式
  - クエリーの自動書き換え
  - スタージョイン
  - 動的ビットマップインデックスANDing
- パラレリズムとスケーラビリティ
  - パラレル I/O
  - 非同期書き出し
  - ブロック読み取り
  - シークエンシャルプリフェッチ
  - インデックスプリフェッチ
  - 複数バッファープール
- ラージデータベース、高可用性
  - 表スペース
  - 動的スペース割り振り
  - 表スペースポイントインタイムリカバリー
- データロードの拡張
  - オートローダー
  - 各種データ型のサポート
  - SMPパラレリズム
- GUIコントロールセンター
  - パフォーマンスモニター
  - ヴィジュアルエクスプレイン

1998 - DB2 UDB EEE 5.2
- 区画内と区画外クエリーパラレリズムの融合
- 自動サマリー表
- レプリケーテッドサマリー表
- ハッシュジョイン
- オートローダーの改善
- データ再配分のステップ化

1999 - DB2 UDB EEE 6.1
- Javaコントロールセンター
- OLAP SQL
- スタースキーマの拡張
- サマリー表の拡張
- 管理機能の拡張
- インデックスアドバイザー
- オンライン索引再編成
- パフォーマンスモニター
- 双方向索引
- 索引列長の拡大
- ページサイズの拡張
- ストアードプロシージャビルダー
- Javaサポート
- クエリーパトローラー

2000 - DB2 UDB EEE Version 7.1.
- 統合ウエアハウス
- 統合OLAP
- XMLエクステンダー
- 地図情報エクステンダー
- 抽象データタイプ
- テンポラリーテーブル
- 生成列
- SQLプロシージャー
- OLEDB
- 32GBログファイルサイズ
- ロードユーティリティ拡張

2001 - DB2 UDB EEE Version 7.2
- 高可用性の拡張
  - 増分バックアップ
  - スプリットイメージからのバックアップ
  - クラスターソフトウエアサポートの拡張
- Linuxサポートの拡張
- ペリーラージメモリーの活用
- WebSphere統合
- MQシリーズ統合
- リレーショナルコネクトの機能強化
- Oracle, SQL Server, Sybase, Informixからの移行容易性の強化
- ウエアハウスの強化
- コンテンツ管理の強化

2002 - DB2 UDB ESE Version 8.1
- 革新的管理ツール
- WebServiceコンシューマサポート
- XML data操作の向上
- フル64ビットサポート
- 256GB活動ログ、無限ログスペース
- コネクションコンセントレーター
- 動的構成変更
- In-placeオンライン再編成
- オンラインロード
- オンライン・ストレージ管理
- null defaultの圧縮
- レプリケーションの拡張
- 新クライアントアーキテクチャー
- マルチディメンジョナルデータクラスタリング
- デベロップメントセンター
- WebSphere開発環境統合
- Microsoft開発環境統合

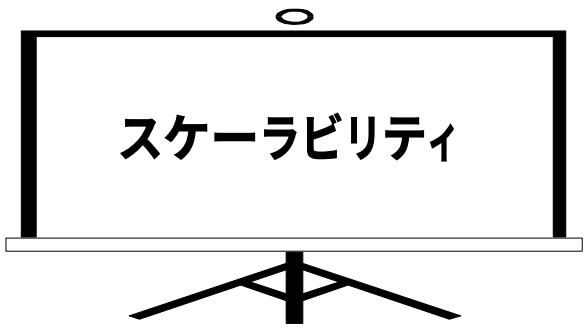
DB2 Data Management Software

IBM

多くのお客様事例-金融業界のお客様

多くのお客様事例-流通業界のお客様

多くのお客様事例-通信ﾒﾃﾞｨｱ・公共事業・官庁・製造業界のお客様



スケーラビリティ

- 最大1000区分までの単一データベース
- お客様での大規模な本番システムの実例多数
  - ►400ノード以上
    - –ライフサイエンス業界のお客様
    - –プロセッサー　Linuxクラスター
    - –http://www-3.ibm.com/solutions/lifesciences/success/
      - •2001年1月時点で216ノード
  - ►130ノード
    - –金融業界のお客様
    - –プロセッサー　RS/6000 4ウェイ　POWER3-II SMP
    - –ディスク容量　37テラバイト(RAID-5を含む)
  - ►66ノード
    - –通信業界のお客様
    - –プロセッサー　RS/6000 8ウェイ　POWER3-II ﾊｲﾉｰﾄﾞSMP
    - –ディスク容量　170テラバイト(ミラーリングを含む)

市場調査レポートから見る大規模BI/DWシステム

- 出典
  - ►survey.com　High End BI/DW competitive Analysis Report
  - ►北米のハイエンドデータウエアハウス･データマートのユーザー147社にインタビュー
  - ►IBM DB2 UDB EEE, NCR Teradata, Oralce 8i OPSまたは9i RACについて調査、比較
- 調査サマリー
  - ►ハイエンドの平均的データウエアハウス
    - –30CPU、350同時ユーザー
      - •単純な照会から複雑な照会まで
  - ►お客様の求めるBI/DWシステムの重要性トップ3項目
    - •可用性
    - •SQLの照会機能の充実
    - •スケーラビリティ

    重要性トップ3項目については3社への満足度は僅差

ディスク容量の分布

- 1－5TBが最多、20TB以上も10%程度
  - ▶ディスク容量の2/3が使用域

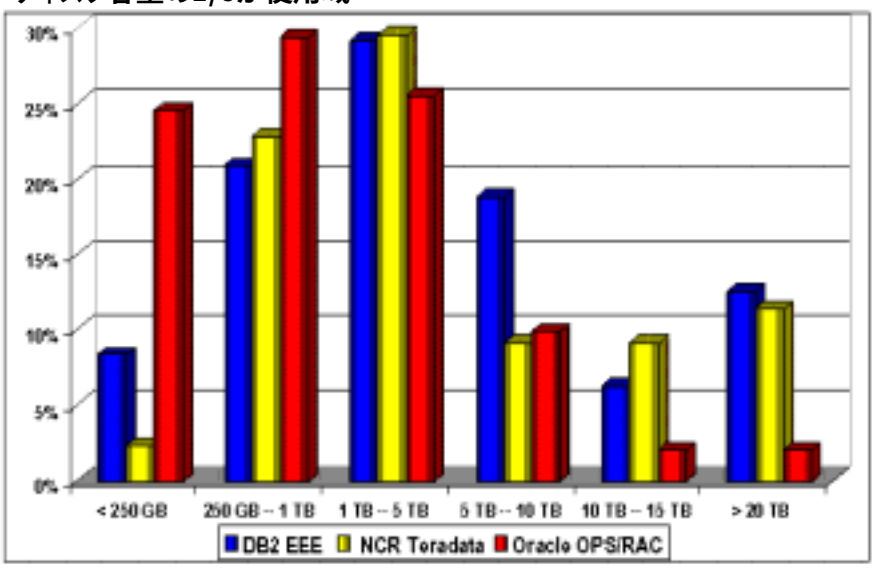


Figure 2.2. Total disk attached to server in the warehouse/mart.

CPU数の分布

- 8－9CPUが最多、平均約30CPU、100CPU以上も1割程度

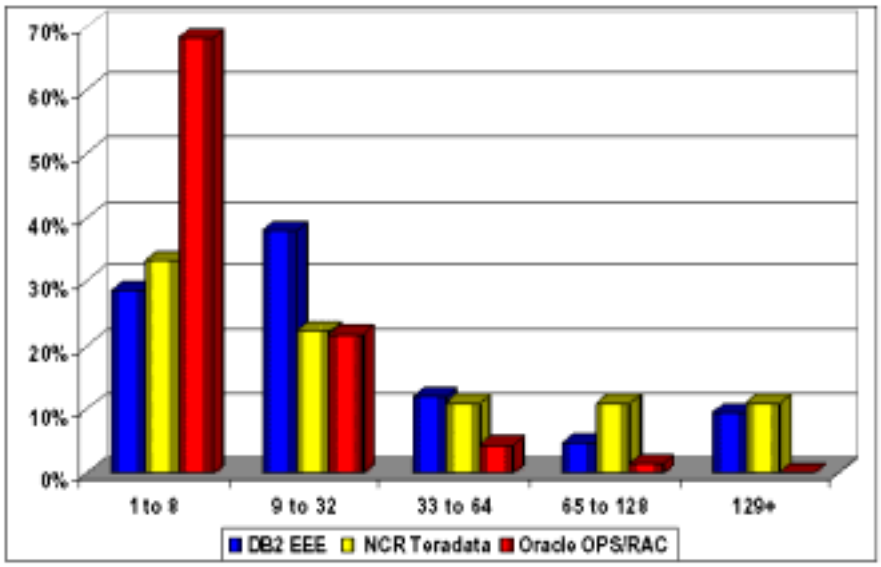


Figure 2.4. Number of processors in the warehouse/mart server.

データベース管理者数の分布

- DBAは1－2人が最多、平均は3人程度、大規模では10人以上も

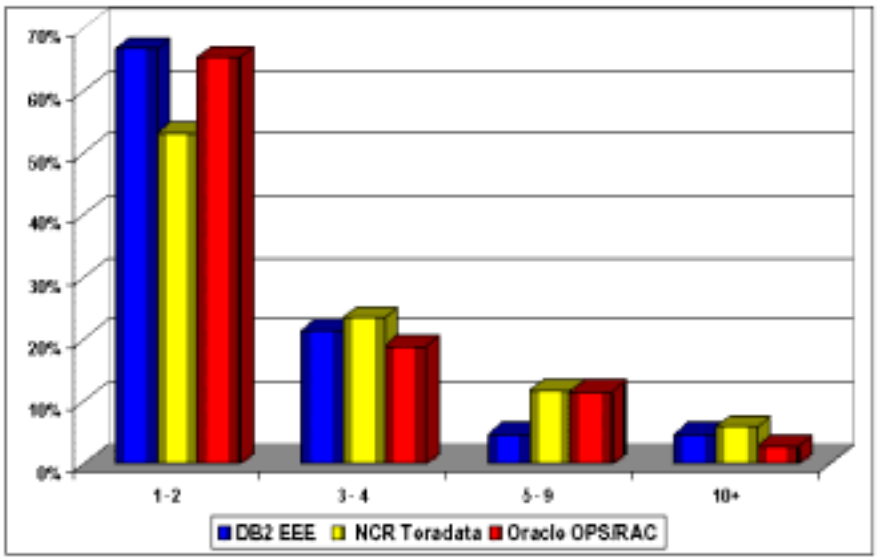


Figure 2.6 Number of Database Administrators in the warehouse/mart

ITスタッフがサポートするデータ容量の分布

- DBA一人が管理するディスク容量は1.4-3TB
- SQLプログラマー当たりのディスク容量は0.7-1.7TB

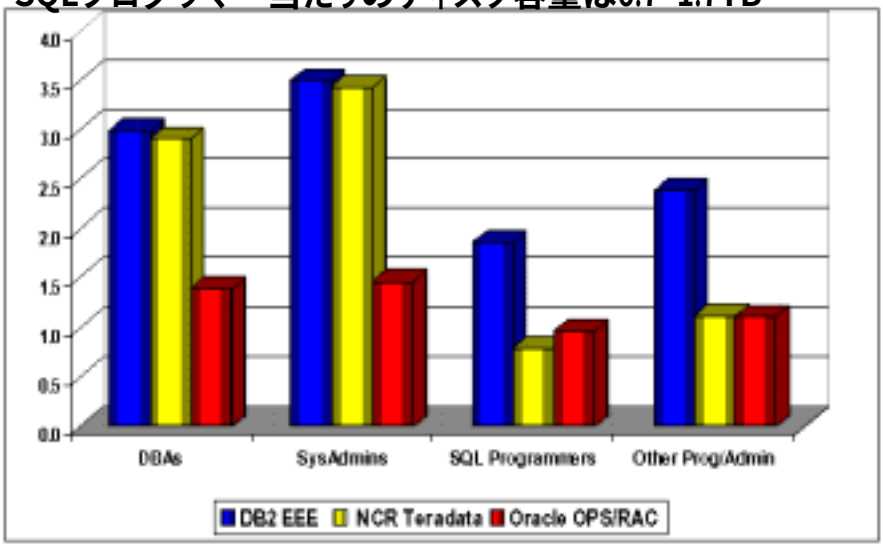
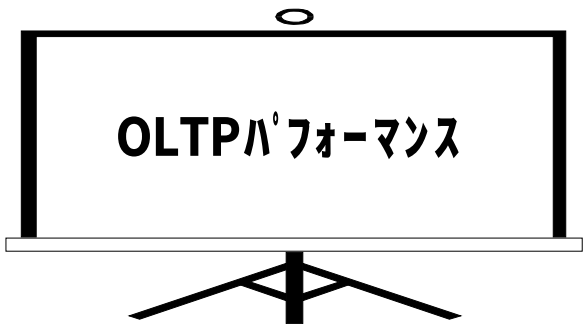


Figure 2.10. Terabytes of disk supported per IT staff member (higher is better).

大規模BI/DWシステムのDB2 UDB EEEの傾向

- DB2 UDB EEEはハイエンドシステムでも多く使用されている
  - ►CPU数、ディスク数：ハイエンドにも多く分布
- DB2 UDB EEEではデータベース管理者やSQLﾌﾟﾛｸﾞﾗﾏｰの生産性が高い
  - ►DBA数：少ない方に分布
  - ►DBA一人当たりのディスク容量：多い方に分布
  - ►SQLﾌﾟﾛｸﾞﾗﾏｰ一人当たりのディスク容量：多い方に分布

- 参考）DB2 UDB EEEでは要らないﾃﾞｰﾀﾍﾞｰｽ管理項目
  - ►ロールバックセグメント
  - ►FreeList/PCTUSED
  - ►ITL slotやロールバックセグメント空き
  - ►照会ヒント



OLTP性能の優位性

- http://www.tpc.org/ より
  - 2002/5

Top TPC-C results by vendor

	Database	System Availability	tpmC	Price/tpmC	System	Clustered
DB2による最高結果	IBM DB2 V7.1	12/07/2000	440,879	19.35 US $	IBM eServer xSeries x370	Yes
Oracle9i非ｸﾗｽﾀｰによる最高結果	Oracle 9i	11/22/2002	403,255	19.51 US $	IBM eServer pSeries p690	No
Oracle9iｸﾗｽﾀｰによる最高結果	Oracle 9i	06/04/2002	137,260	19.25 US $	HP Parallel Database Cluster	Yes

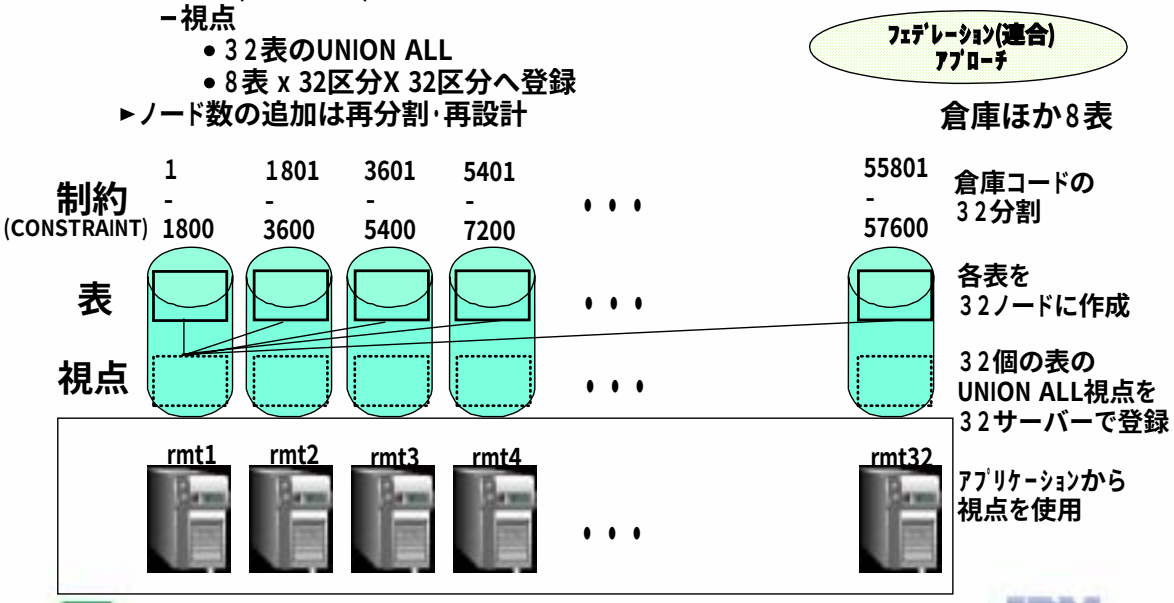
OLTP性能の優位性

- 比較的安価なインテルサーバーを束ねて超高性能クラスターを作るDB2クラスター技術
  - ►DB2 UDB EEE for Windows 32ノード構成が、サーバーの最高峰IBM p690のOracle 9i を今だに上回る
  - ►今年発行のOracle 9i クラスター構成のトランザクション性能をも上回る
- TPCﾍﾞﾝﾁﾏｰｸは馬とびゲーム
  - ►TPCﾍﾞﾝﾁﾏｰｸは後攻のﾍﾞﾝﾀﾞｰが最新のハードウエアを使って先行の結果を追い越すのが常
  - ►DB2 UDB EEE for Windows 32ノード構成の結果は2年前の2000/7の公表

SQL Server2000のtpccﾍﾞﾝﾁﾏｰｸでのﾚﾝｼﾞ分割

- SQL Server2000 TPCC構成
  - ►8表、32区分について
    - －制約(constraint) 8表x32区分
    - －視点
      - 32表のUNION ALL
      - 8表 x 32区分X 32区分へ登録
  - ►ノード数の追加は再分割･再設計

出典：TPCﾌﾙﾃﾞｨｽｸﾛｰｼﾞｬｰﾚﾎﾟｰﾄ
http://www.tpc.org/results/FDR/TPCC/compaq.256p.091901.fdr.pdf

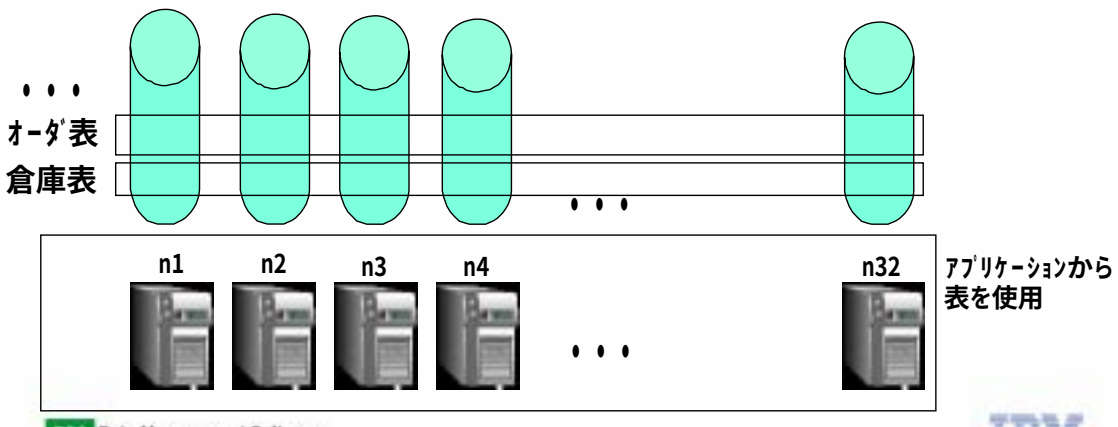


IBM

DB2 UDB EEEのtpccﾍﾞﾝﾁﾏｰｸでのﾊｯｼｭ分割

- DB2のTPCC構成
  - ►8表32区分について
    - －表を倉庫コードでハッシュパーティショニング
      - CREATE TABLE文で区分化キーに倉庫ｺｰﾄﾞ列を指定
  - ►ﾉｰﾄﾞ数に依存しない拡張性
    - －ハッシュパーティショニングによるデータの自動的な均等配置

出典：TPCﾌﾙﾃﾞｨｽｸﾛｰｼﾞｬｰﾚﾎﾟｰﾄ
http://www.tpc.org/results/FDR/tpcc/ibm.x370.c5.fdr.01111401.pdf



IBM

SQL Server2000 とDB2 UDB EEE TPCC比較

- DB2 UDB EEEはMS SQL Serverより1年前のサーバー機種でありながら、サーバーCPU1個あたりのtpmCは依然としてSQL Server 2000を上回っています
  - ►TPCﾍﾞﾝﾁﾏｰｸは馬とびゲーム

TPCCｸﾗｽﾀｰ順位	Database	System Availability	tpmC	Server CPU	Server CPU 1個当たりの tpmC
4	IBM DB2 V7.1 EEE	12/07/2000	440,879	Intel PentimIII Xeon 700MHz x 128	3444
1	Microsoft SQL Server2000	10/15/2001	709,220	Intel PentimIII Xeon 900MHz x 272	2607
5	Microsoft SQL Server2000	10/15/2001	410,769	Intel PentiumIII Xeon 900MHz x 136	3020

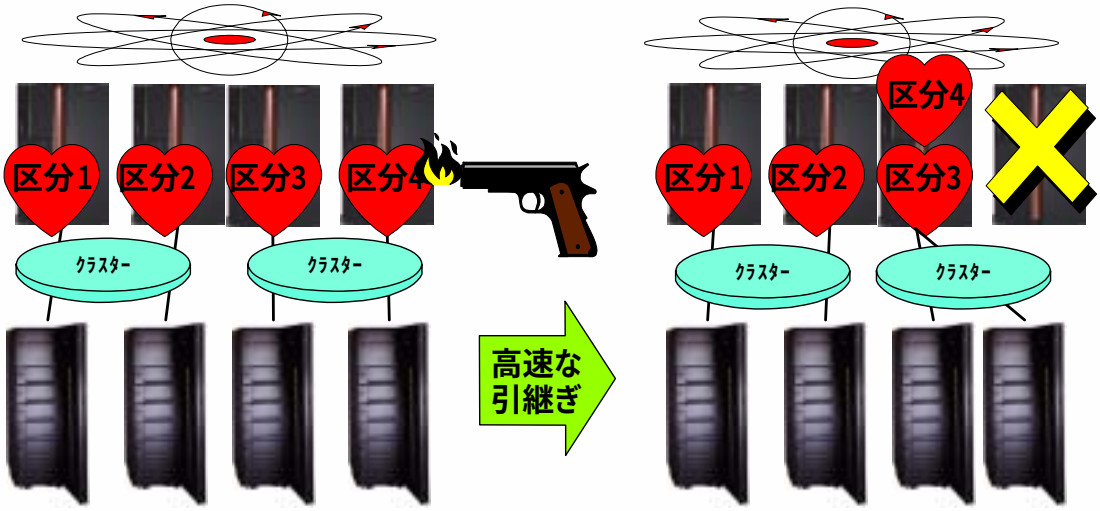
出典:
http://www.tpc.org/

DB2 Data Management Software

IBM

高可用性

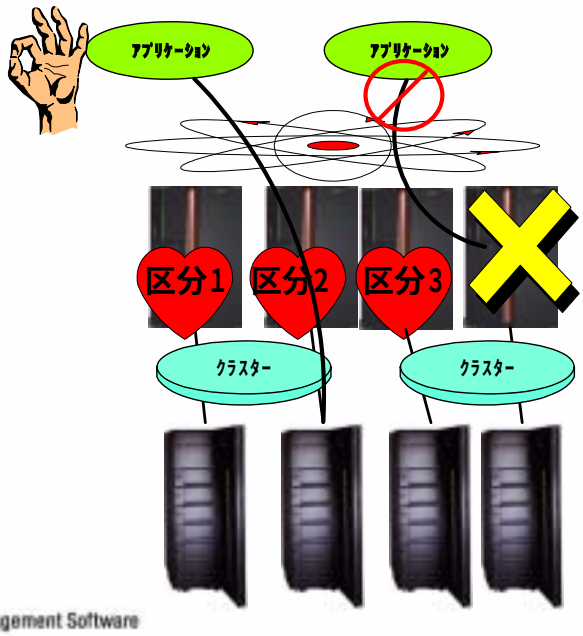
- 高可用性ｸﾗｽﾀｰと連動して高速な引継ぎが可能
  - ローデバイス表スペース、ログ、とコンカレントアクセスモードの使用
  - チェックポイント間隔、ロールバック量の適正な使用



- 適切な構成･デザインにより1.5分程度の引継ぎが可能 (構成やｱﾌﾟﾘｹｰｼｮﾝに依存します)
- クラスター構成(N:M)はクラスターソフトウエアにより自由に選択可能

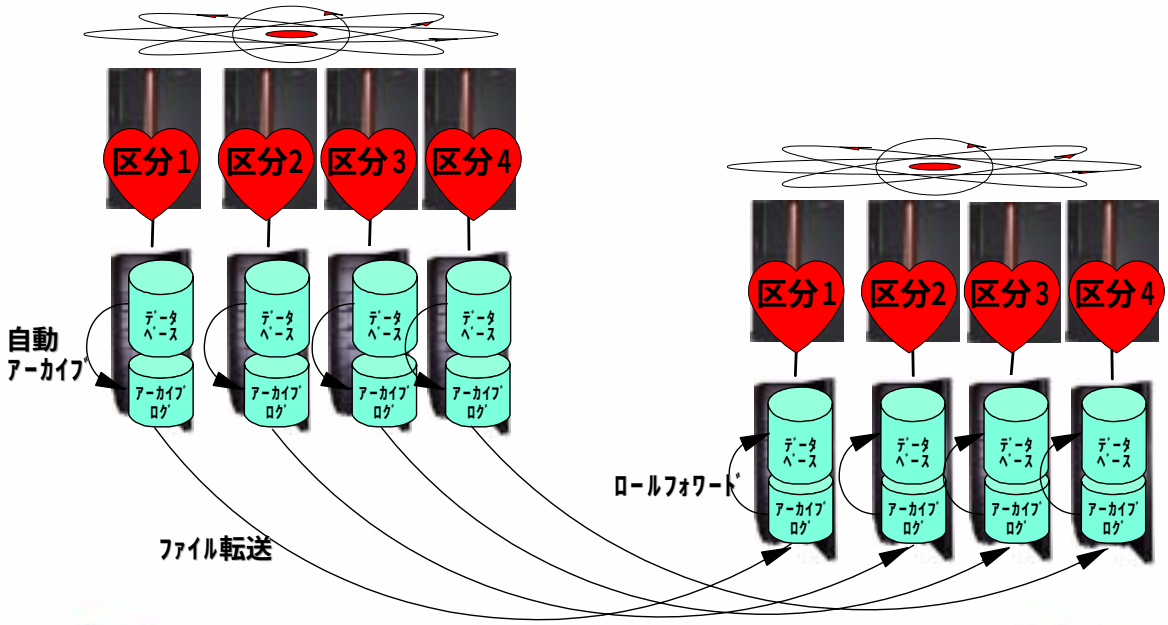
障害範囲の局所化

- 復旧中のデータベース区分を使わないｱﾌﾟﾘｹｰｼｮﾝは稼動し続けます



ログ転送による災害対策システム

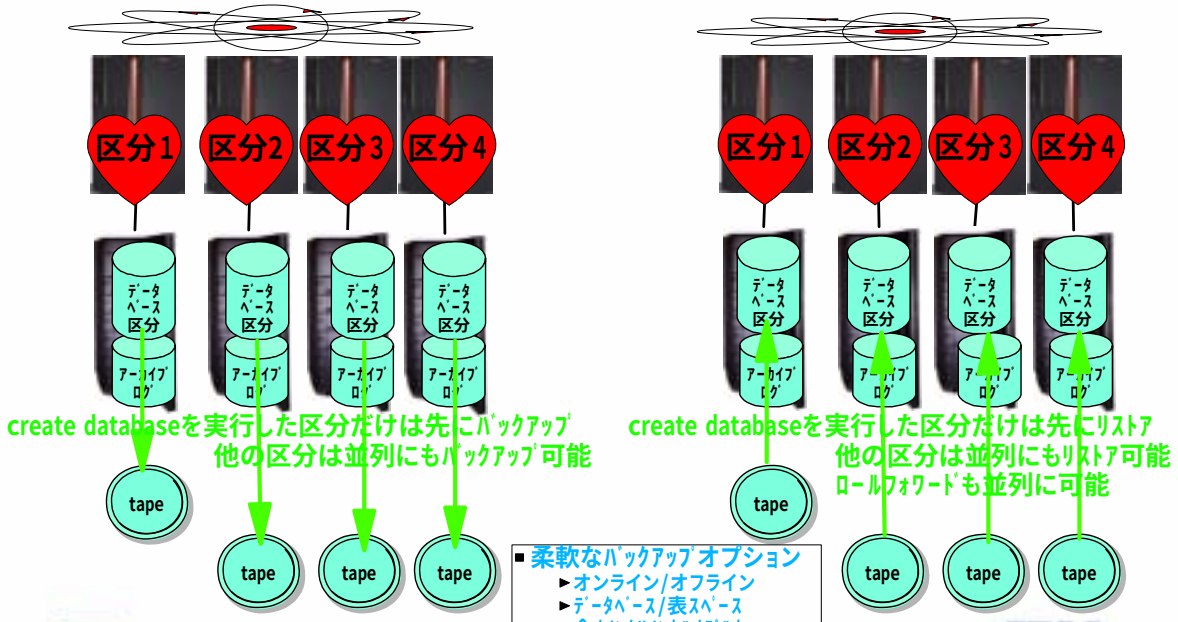
- アーカイブログを他システムへ転送してロールフォワード



IBM

ﾊﾟﾗﾚﾙ・ﾊﾞｯｸｱｯﾌﾟ、ﾊﾟﾗﾚﾙﾘｽﾄｱ、ﾊﾟﾗﾚﾙﾛｰﾙﾌｫﾜｰﾄﾞ

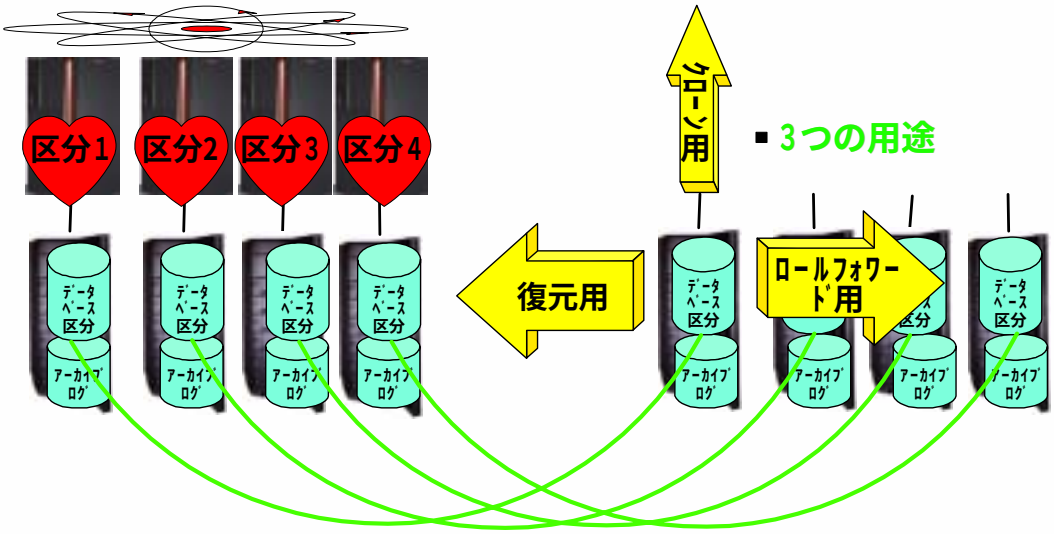
- 1区分ﾃﾞｰﾀﾍﾞｰｽのﾊﾞｯｸｱｯﾌﾟ回復操作を複数区分へ自然な拡張
  - ►並列操作による時間短縮、分割された1ﾊﾞｯｸｱｯﾌﾟ対象容量



IBM

分割ミラー

- ディスクサブシステムのミラー機能と連動した分割ミラーイメージの取得
  - ►短時間の大容量バックアップ取得：バックアップ用、ロールフォワード用、データベースクローン用



- オンライン中にディスクサブシステムのコピー機能を利用してディスクコピー
  - ►短い時間だけデータベースへの書き込み中断、再開
    - –db2コマンドで指示：：アプリケーションは短い待ち状態後自動的に再開

IBM

クラスター・アーキテクチャーの比較

シェアドナシングアーキテクチャーとシェアドディスクアーキテクチャーの方向性の違い

■ シェアドナシング・アーキテクチャー

►高いスケーラビリティ
- −もともと区分間の資源の競合が最小
- −一部の共有資源の競合を改善

►高可用性
- −ハイアベイラビリティクラスターソフトウェアや最新ディスクサブシステムと連携した高可用性ソリューションを利用

►クラスターデータベース固有機能の例
- • パラレルオプティマイザー
- • 高速ノード間通信
- • グローバルなデッドロック検知デーモン
- • パラレルユーティリティ

- −単一データベースとの高い共通性
- −1区分データベースの新機能の多くをそのまま享受可能

■ シェアドディスク・アーキテクチャー

►高いスケーラビリティ
- −資源の競合を効率よく処理する機能改善の積み重ねで解決

►高可用性
- −全資源にアクセス可能であることを利用して主にDBMS自身でソリューションを提供

►クラスターデータベース固有機能の例
- • 資源共用機能
- • リカバリー機能
- • 縮退運転時の他インスタンスの資源サポート
- • 複数redoログ共用
- • 縮退時の他インスタンス資源のを1インスタンスだけでサポート

- −クラスター専用機能の整備・機能改善の継続

DB2 Data Management Software

IBM

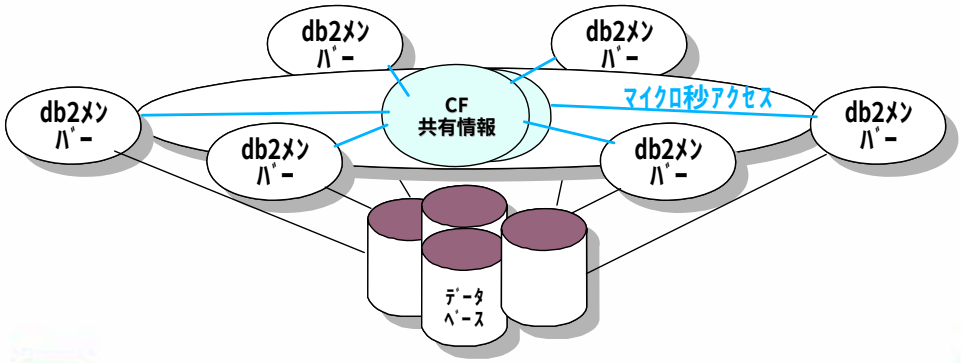
DB2 390に見る機能改善の積み重ね

■ 高いスケーラビリティのためにメンバー間の競合を最小化

- −資源共用ハードウエアCoupling Facilityを活用
  - • CFの特徴
    - ◆ マイクロ秒単位でDB2メンバーからアクセス
    - ◆ CFからCPUへのバッファー無効化シグナル等はCPU割り込みを起こさない
    - ◆ 2重化

■ 高可用性とスケーラビリティ向上のために数多くの細かい機能改善を段階的に継続

- −V4.1 1995, V5.1 1997, V6.1 1999, V7.1 2001

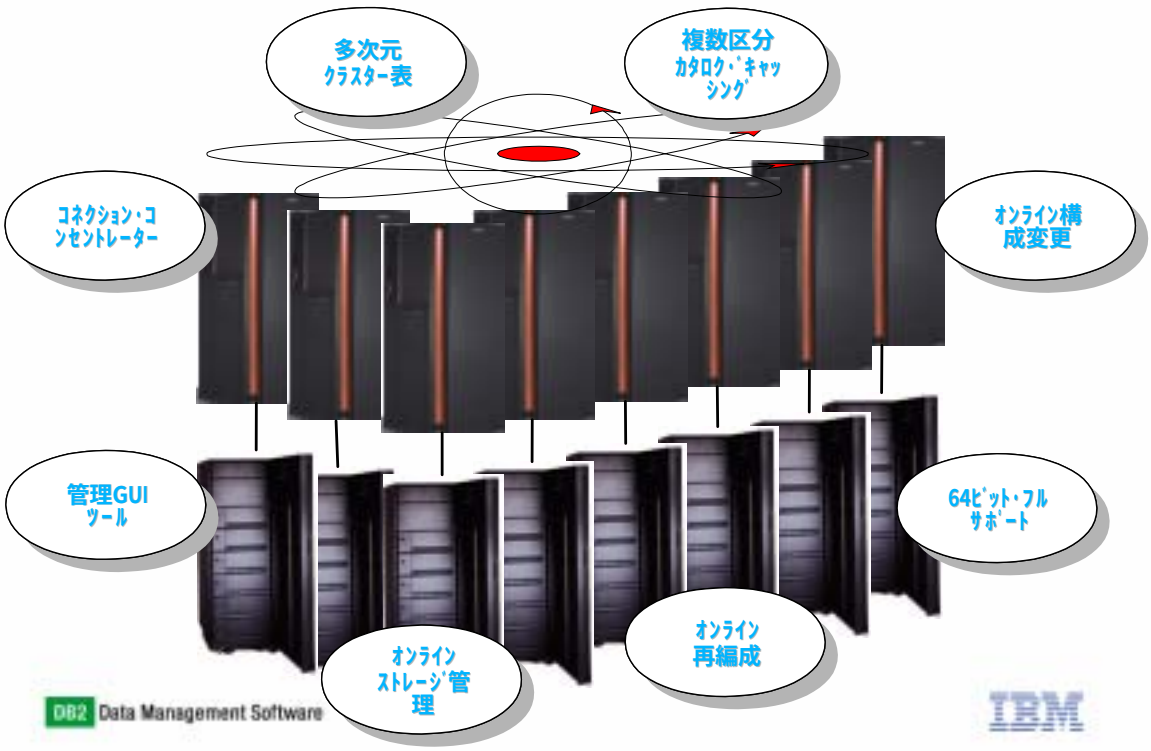


DB2 Data Management Software

IBM

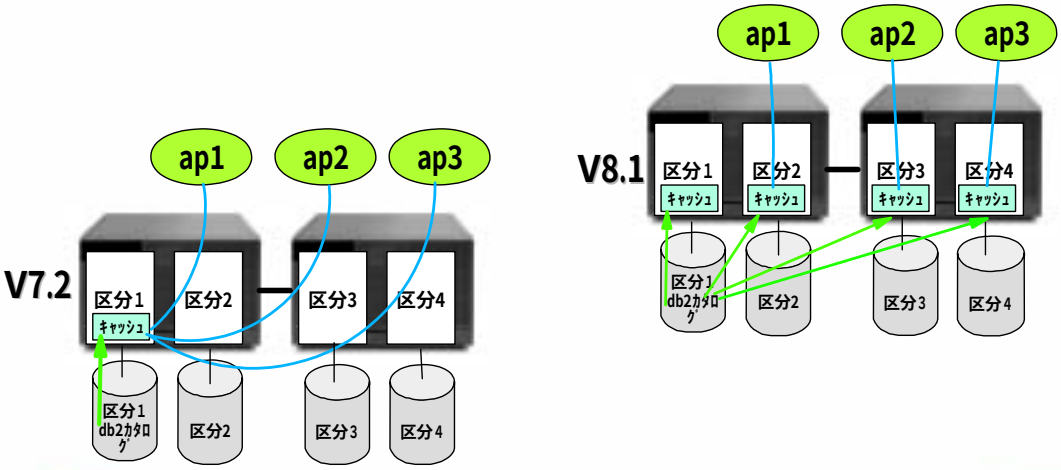
DB2 V8.1と今後

DB2 ESE V8.1 区分ﾃﾞｰﾀﾍﾞｰｽの新機能



複数区分カタログ・キャッシング

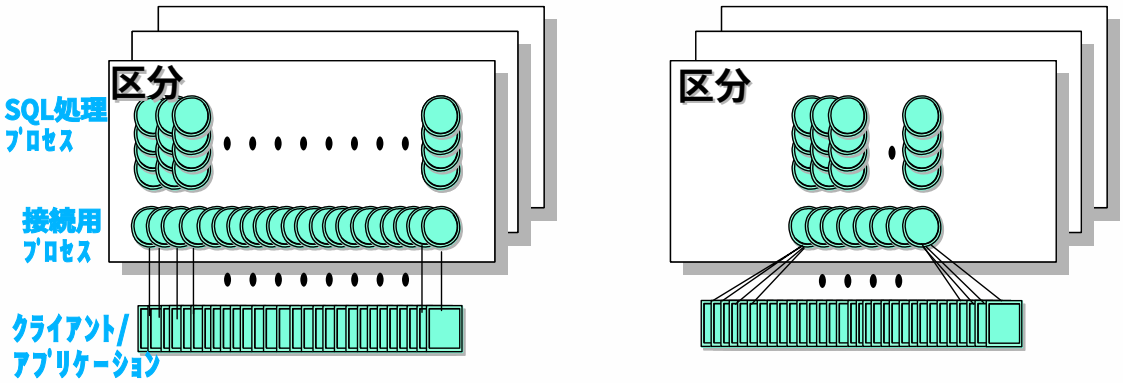
- CREATE DATABASEを行った区分へ集中していた一部の内部処理が各区分へ分散されてスループットが向上します
  - ►CREATE DATABASEを行った区分にあるDB2カタログの表や権限の情報を全区分のメモリーにキャッシングします。
  - ►SQL文の解析処理、データベース権限、ストアード・プロシージャー、ユーザー定義関数の権限チェック処理を各区分に負荷分散します



IBM

コネクション・コンセントレーター

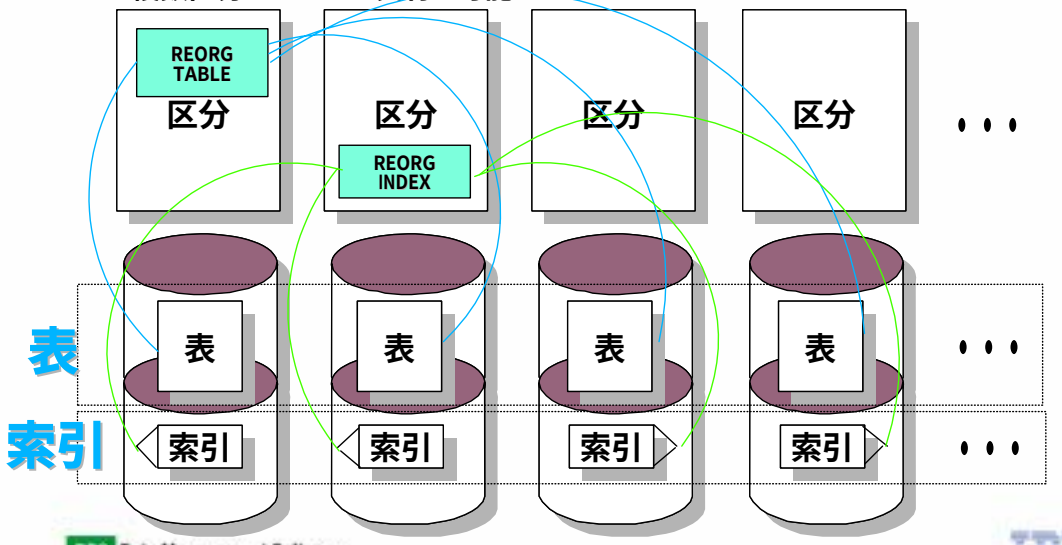
- コネクション・コンセントレーター
  - ►多くのユーザーやアプリケーションを接続・SQL処理するためのDB2のプロセス数(Windowsではスレッド数)とメモリー所要量を削減します
  - ►多数のユーザーとそのSQL処理が、DB2自身がOLTPモニターを内蔵したかのように、スリムに効率よく動きます
  - ►例えば、同時５０００接続規模の多数のユーザーの接続が可能になります



IBM

オンライン再編成

- オンライン表再編成
  - ►一時表ｽﾍﾟｰｽ不要、中断･再開可能
  - ►複数区分でのﾊﾟﾗﾚﾙ実行も可能
- オンライン索引再編成
  - ►複数区分でのﾊﾟﾗﾚﾙ実行も可能



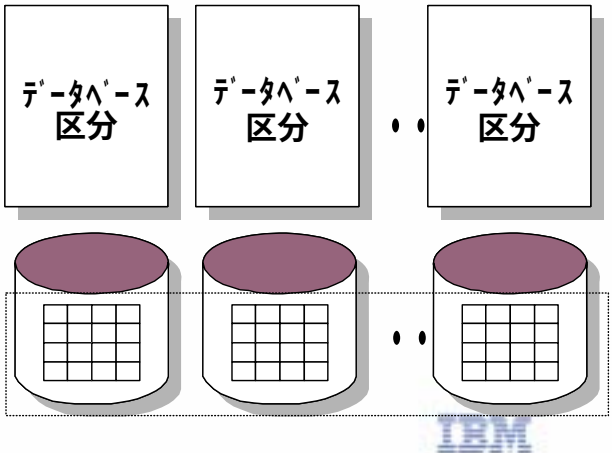
DB2 Data Management Software

IBM

多次元クラスター表

- 1つ以上の次元(列、生成列)について次元の値ごとに行が同一ブロック(エクステント) にまとめられる
  - ►INSERTやLOADが速い
  - ►次元を条件としたSELECTやGROUP BY集計が速い
  - ►次元以外の列のUPDATEで行が移動しない
  - ►次元が条件のDELETEが速い、空ブロックが再利用される
  - ►レンジパーティションの代用になり、パーティションのADD/DROPも不要
  - ►索引など従来の表の機能持つはすべて使える
  - ►複数区分でも使用可能

```
CREATE TABLE 売上表 (
製品   INT NOT NULL,
販売店       INT,
金額          INT,
販売日       DATE,
YYMM  GENERATED ALWAYS
AS (INTEGER(販売日) /100) )
ORGANIZED BY DIMENSIONS(製品,YYMM)
PARTITIONING KEY(製品)
;
-- 月単位の削除
DELETE FROM 売上表
WHERE INTEGER(販売日)/100 = 199912
;
```



多次元ｸﾗｽﾀｰ表

DB2 Data Management Software

IBM

V8.1その他の機能拡張

- 64ビットフルサポート
  - ►これまでのHP、Sun、AIXの各64ビット版のサポートに加えて、Windowsと Linuxも 64ビットモードをサポートされました
  - ►各区分で64ビットモードの大きなバッファープールやヒープサイズを利用可能
- 管理GUIツール
  - ►複数区分に対応
- オンライン構成変更
  - ►データベース構成パラメータ、データベースマネジャー構成パラメータの動的変更ができます
  - ►複数区分にも対応
- オンラインストレージ管理
  - ►従来の表ｽﾍﾟｰｽへのコンテナの追加に加え、コンテナの削除、コンテナサイズの縮小、別のデータ自動配分の範囲となるコンテナの追加が可能になりました
  - ►複数区分にも対応

今後の期待

- ソフトウエアパッケージソリューションへのバンドル
- ライフサイエンス
- Linux クラスター
- データウエアハウスの拡大
- より高い可用性と連続運転