

Oracle DBA & Developer Days 2011

日本オラクル、今年最大の技術トレーニングイベント

2011年11月9日(水)～11月11日(金) シェラトン都ホテル東京



ORACLE®

Oracle Exadataはなぜ速い？ 圧倒的なパフォーマンスの秘密を徹底解剖

日本オラクル株式会社 テクノロジー製品事業統括本部
プリンシパルセールスコンサルタント 秋山真一

以下の事項は、弊社の一般的な製品の方向性に関する概要を説明するものです。また、情報提供を唯一の目的とするものであり、いかなる契約にも組み込むことはできません。以下の事項は、マテリアルやコード、機能を提供することをコミットメント(確約)するものではないため、購買決定を行う際の判断材料になさらないで下さい。オラクル製品に関して記載されている機能の開発、リリースおよび時期については、弊社の裁量により決定されます。

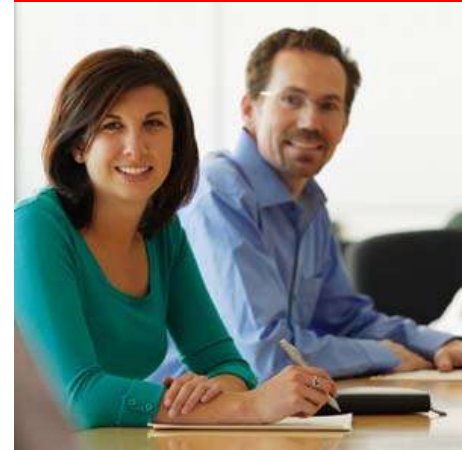
OracleとJavaは、Oracle Corporation 及びその子会社、関連会社の米国及びその他の国における登録商標です。文中の社名、商品名等は各社の商標または登録商標である場合があります。

Agenda

- Oracle Exadata 概要
- Oracle Exadataが速い7つの理由



Oracle Exadata 概要



「工業製品化」された次世代プラットフォーム

Engineered Systems

Fabric for the 21st Century Datacenter



ORACLE

Oracle Exadata Database Machine

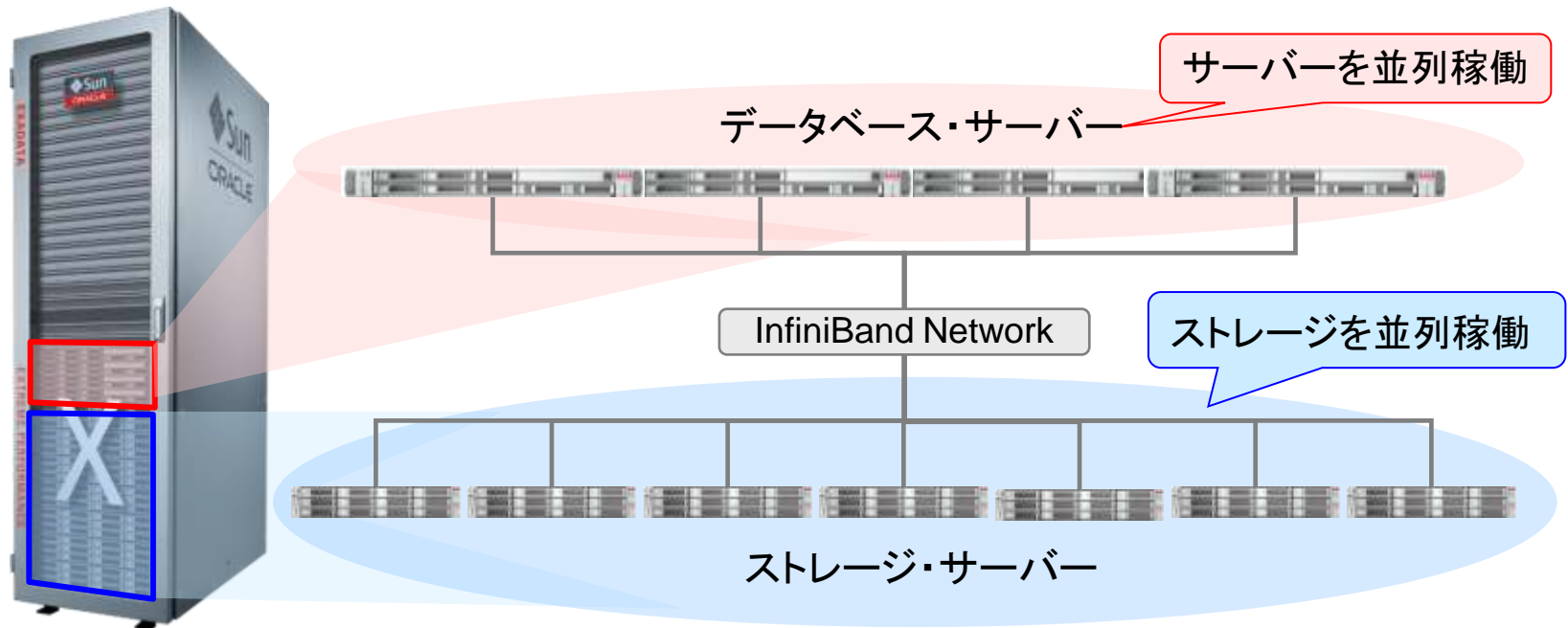
Oracle Databaseに最適化されたEngineered System



- 実績あるOracle Databaseを搭載
 - 既存資産の有効活用
- 大量データを高速に処理
 - 性能のボトルネックになりやすいI/Oを効率化
- 標準化された構成
 - 迅速な構築とサポートの一元化
- 統合DB基盤の実現
 - あらゆるワークロードに対して最適なパフォーマンスを提供

Oracle Exadata アーキテクチャ

- Oracle Real Application Clustersにより、データベース・サーバーを並列稼働させ、高可用性と高拡張性を実現
- Oracle Automatic Storage Managementにより、ストレージ・サーバーを並列稼働させ、高いI/O性能と高可用性・高拡張性を実現
- さらに、Oracle Exadata Storage Serverが処理の一部をオフロードし、大量データの高速処理を実現



Oracle Exadata Database Machine ファミリー

データウェアハウス、OLTP、DB統合とあらゆるワークロードに対応

Oracle Exadata X2-2



- Quarter, Half, Full and Multi-Racks
 - QuarterからHalf、HalfからFull、Fullから複数ラックへと拡張可能

Oracle Exadata X2-8



- Full and Multi-Racks
 - Fullから複数ラックへと拡張可能

Oracle Exadata 各モデルのH/W構成

	X2-8 Full	X2-2 Full	X2-2 Half	X2-2 Quarter
Database Servers	2	8	4	2
Database Grid Cores	128	96	48	24
Database Grid Memory (GB)	2048	768 (max 1152)	384 (max 576)	192 (max 288)
InfiniBand switches	3	3	3	2
Ethernet switch	1	1	1	1
KVM	No	Yes	Yes	Yes
Exadata Storage Servers	14	14	7	3
Storage Grid CPU Cores	168	168	84	36
Storage Grid Disks	168	168	84	36

Oracle Exadata 各モデルのキャパシティ(未圧縮時)

		X2-8 or X2-2 Full Rack	X2-2 Half Rack	X2-2 Quarter Rack
Raw Disk Capacity	High Perf Disk	100 TB	50 TB	21.6 TB
	High Cap Disk	336 TB	168 TB	72 TB
Raw Flash Capacity ¹		5.3 TB	2.6 TB	1.1 TB
Usable Mirrored Capacity ^{1,2}	High Perf Disk	45 TB	22.5 TB	9.5 TB
	High Cap Disk	150 TB	75 TB	31.5 TB
Usable Triple Mirrored Capacity ^{1,3}	High Perf Disk	30 TB	15 TB	6.5 TB
	High Cap Disk	100 TB	50 TB	21.5 TB

1- Capacity calculated using normal space terminology of 1 TB = 1024 * 1024 * 1024 * 1024 bytes.

2 - Actual space available for a database after mirroring (ASM normal redundancy) and allowing one disk (Quarter and Half) or two disks (Full Rack) of free space to automatically remirror after disk failures.

3 - Actual space available for the database computed after triple mirroring (ASM high redundancy).

Oracle Exadata 各モデルのI/O性能

		X2-2 or X2-8 Full Rack	X2-2 Half Rack	X2-2 Quarter
Disk Data Bandwidth ^{1,3}	High Perf Disk	25 GB/s	12.5 GB/s	5.4 GB/s
	High Cap Disk	14 GB/s	7 GB/s	3 GB/s
Flash Cache Data Bandwidth ^{1,3}	High Perf Disk	75 GB/s	37.5 GB/s	16 GB/s
	High Cap Disk	64 GB/s	32 GB/s	13.5 GB/s
Disk IOPS	High Perf Disk	50,000	25,000	10,800
	High Cap Disk	25,000	12,500	5,400
Flash IOPS ^{2,3}		1,500,000	750,000	375,000
Data Load Rate ⁴		12 TB/hr	6 TB/hr	3 TB/hr

1 - Bandwidth is peak physical disk scan bandwidth achieved running SQL, assuming no compression. Effective data bandwidth will be much higher when compression is factored in.

2 - IOPS – Based on read IO requests of size 8K running SQL. Note that the IO size greatly effects flash IOPS. Others quote IOPS based on 2K, 4K or smaller IOs that are not relevant for databases. Exadata Flash read IOPS are so high they are typically limited by database server CPU, not IO.

3- Actual Performance varies by application.

4 – Exadata load rates are typically limited by database server CPU, not IO. Rates vary based on load method, indexes, data types, compression, and partitioning

Oracle Exadataが速い 7つの理由





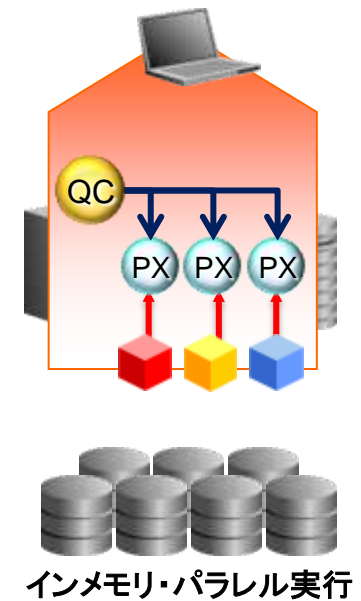
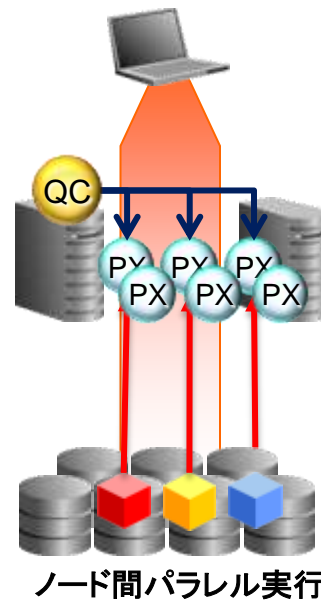
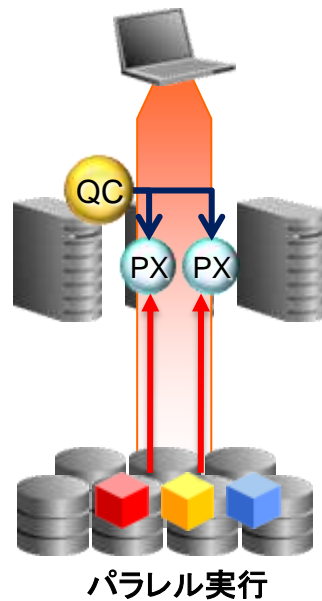
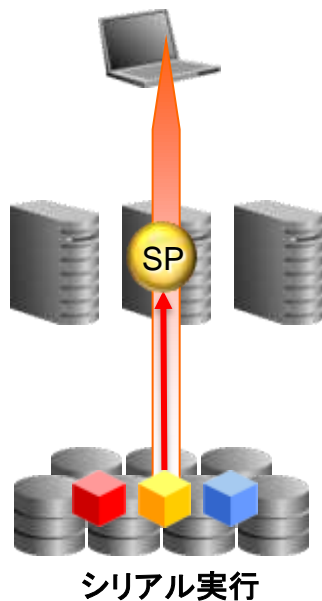
1 パラレル実行

1 対 192

SQLの平行実行

マルチコアCPU、複数ノードの有効活用

- 平行実行：1つのSQLを内部的に並列化
- ノード間平行実行：1つのSQLを複数ノードで並列化
- インメモリ・平行実行：
物理メモリ上にキャッシュされたデータに対する平行実行



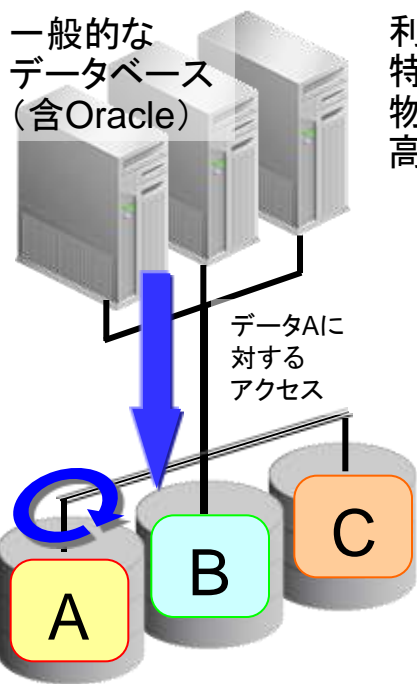
ストレージも パラレル実行 (ストライピング)

4 対 168

Oracle ASMによるデータのストライピング

Oracle ASMがストレージ管理を自動化
最高のI/Oパフォーマンスを実現するための高度なデータ・ストライピング

従来のデータベースシステム

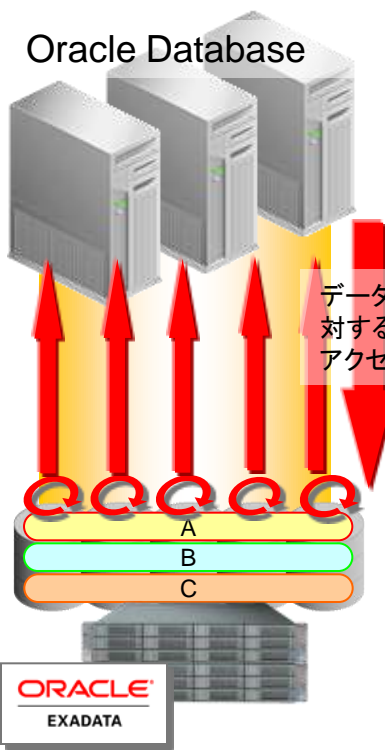


利用アプリケーションの特性に応じたデータの物理配置設計やより高価で高速なディスク装置が必要

アクセスするデータに依存し、ディスク毎の稼働率に偏りや遊びが発生

アクセス対象のデータが格納されたディスクしか稼働しない

Oracle Exadata



物理設計が不要となり、Storage Server追加の際にはディスク処理性能と転送性能がリニアに向上

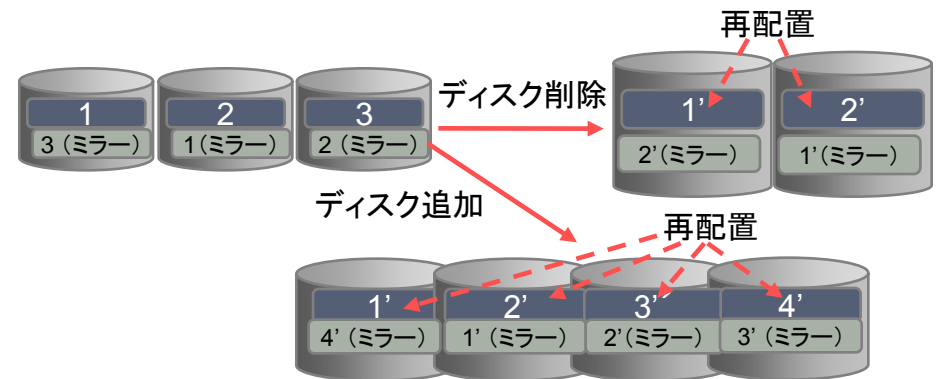
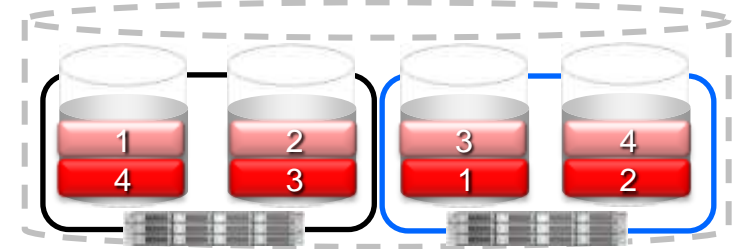
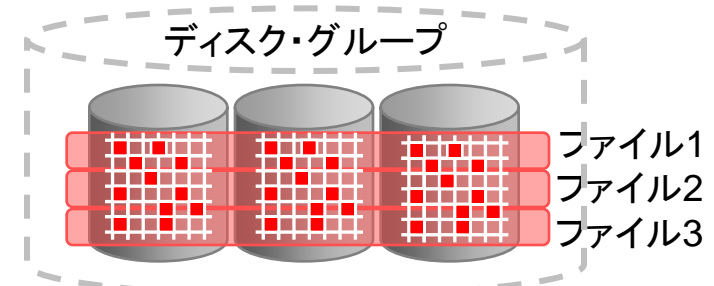
データの分散配置により、複数のStorage Serverが並列稼働し短時間でデータ処理

データは各ディスクに自動で分散配置され、アクセス時には各ディスクが並列で稼働

Oracle Automatic Storage Management

Oracle Databaseのストレージ仮想化技術

- 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
(ホットスポットが発生しない)
→ **性能の維持**
- ミラーリング
 - ファイルタイプに応じて、ミラーリング
(2重化 / 3重化)
→ **可用性の担保**
- 動的リバランシング
 - ディスクの追加 / 削除時に自動的にデータを再配置
→ **拡張性**



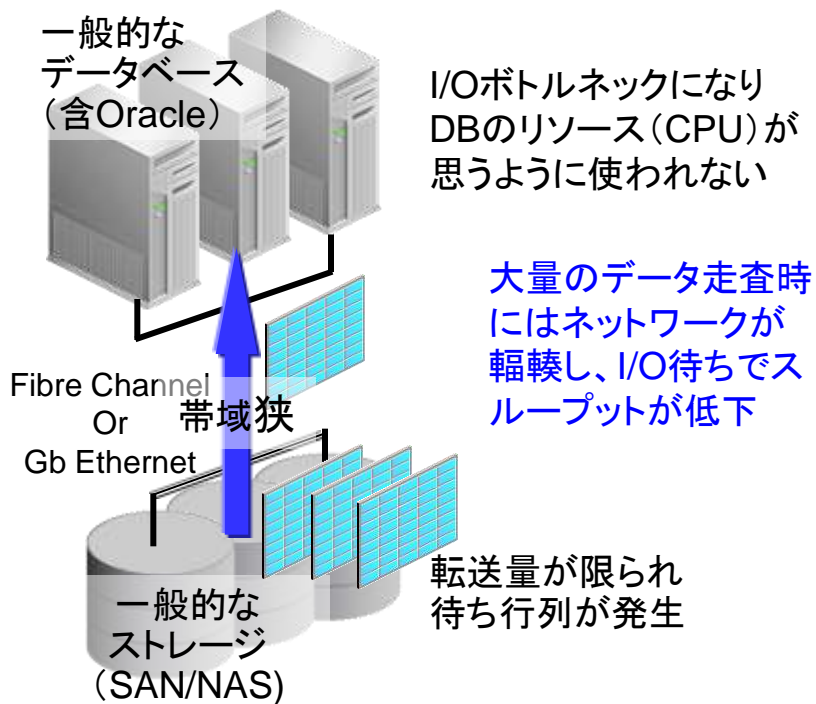
広帯域、低遅延 ネットワーク

8 对 40

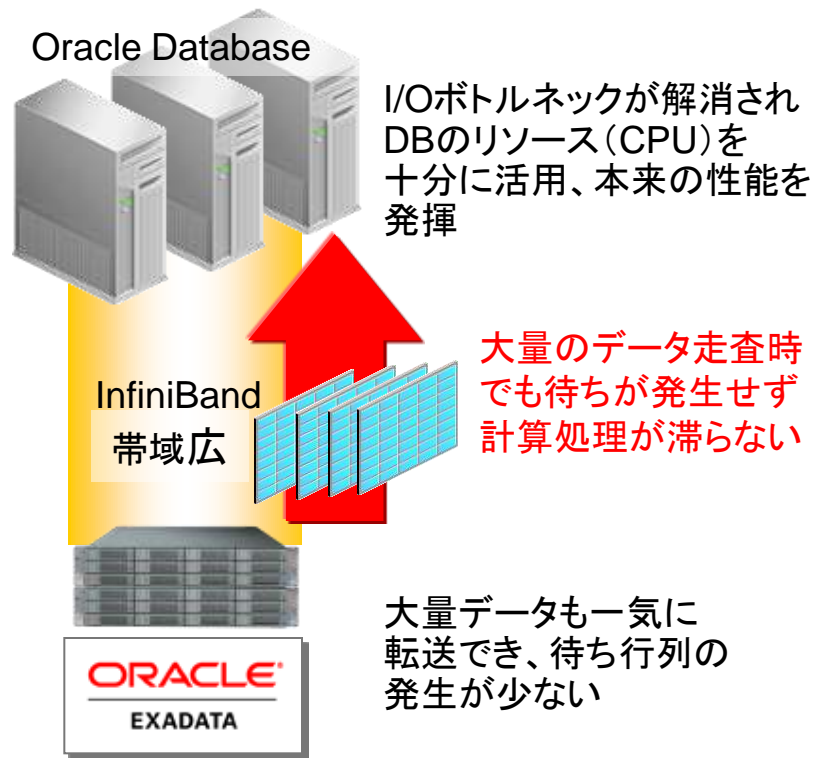
InfiniBandによるデータ転送帯域

40Gbpsもの帯域を持つInfiniBandにより、
大量のデータ転送にも輻輳せず、効率的なデータアクセスを実現

従来のデータベースシステム



Oracle Exadata



InfiniBand Network

- 広帯域、低遅延のInfiniBandを採用
 - 40Gbpsのネットワーク(片方向)
 - SANのような機能(Zero copy, buffer reservation)
 - IPネットワークのような簡易な管理
- ネットワークの統合
 - ストレージネットワーク
 - RACインターコネクト
 - 外部接続 (optional)
- プロトコル
 - ZDP RDS v3
 - Oracleが実装し、Open Sourceとして公開
 - 非常に低いCPUオーバーヘッド
(3 GB/sec の転送に 2% しかCPUを使用しない)
 - Internet Protocol over InfiniBand (IPoIB)
 - 通常のEthernet ソフトウェアのように使用できる
(tcp/ip, udp, http, ssh,...)





Flashの活用

1 対 30

Exadata Smart Flash Cacheによる ホットデータのキャッシュ

- 利用頻度の高いデータをフラッシュ・ストレージに自動的にキャッシュ
- 秒間150万I/Oを処理可能(ディスクの30倍)
- 非圧縮データにおいて50GB/秒の検索スループット(ディスクの2倍)

従来のデータベースシステム

磁気ディスクをメインとした
データストレージでは、
シークの発生により
ランダムI/Oの性能に限界

300 IOPS



- ディスクドライブは大量のデータを保持できるが、性能はディスクあたり約300 IOPS程度が限度

Oracle Exadata



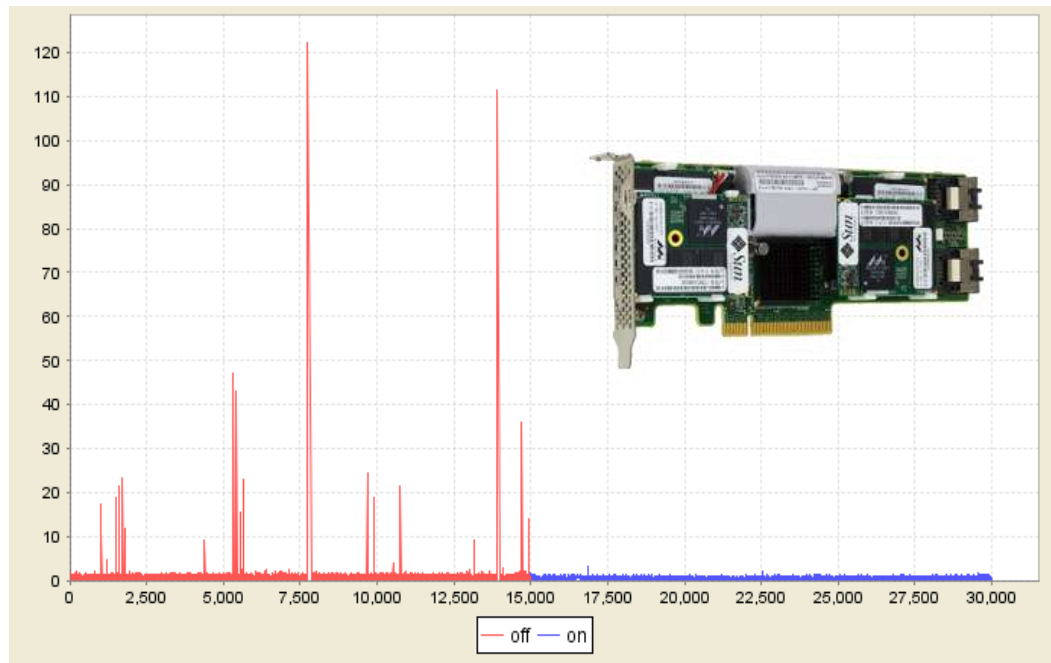
筐体あたり最大5.3TBの
Flash Storageを搭載し、
アクセス頻度が高いホット
データをキャッシュする
ことで、安定したスルー
プットを実現

数万 IOPS



Exadata Smart Flash Logによる ログ書き込みの高速化

- Flash MemoryとDisk Controller Cacheの両方に同時にRedo書き込みを行い、どちらかが完了したら書き込みを完了とする
- レスポンスタイムの向上と、待機の異常値を削減し、データベース全体のスループットの向上を実現

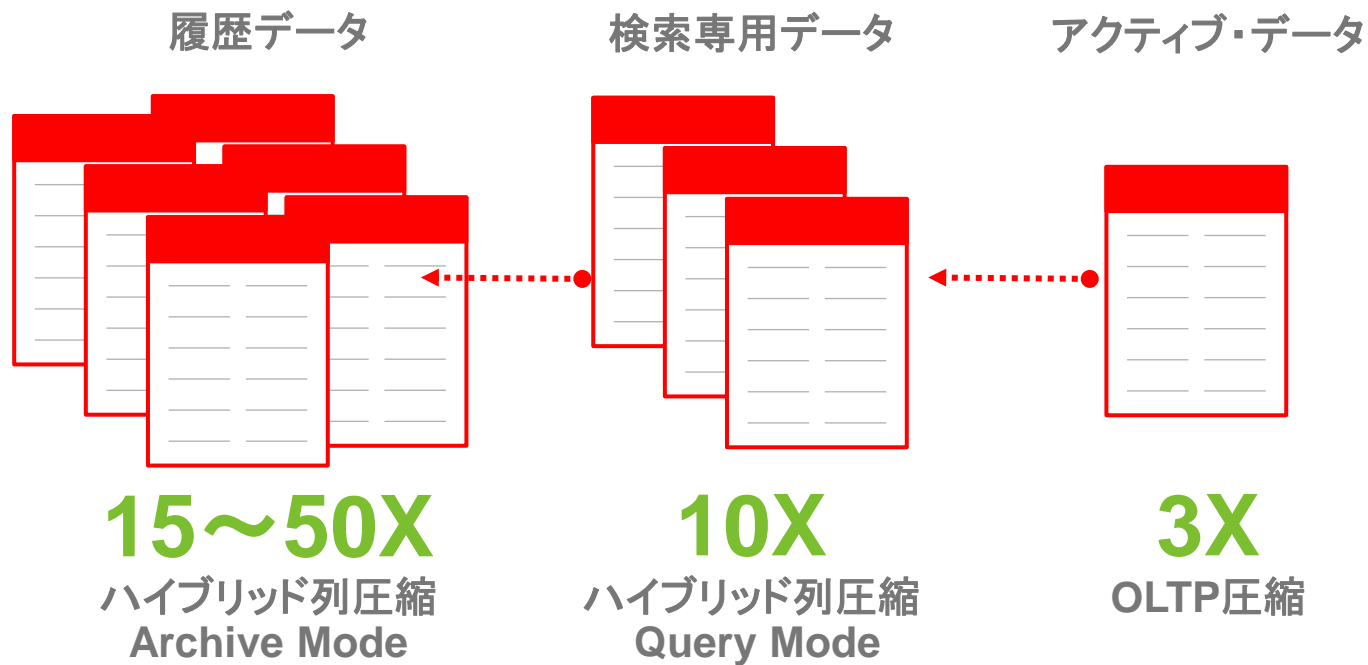


圧縮による I/O削減

10 対 1

ハイブリッド列圧縮によるI/O削減

- 列ベースでデータを格納し、データ圧縮
- Query Mode : データウェアハウス・テーブル用、平均で10倍の圧縮率
- Archive Mode : 履歴データ用、平均で15倍の圧縮率



不要なI/Oの削減

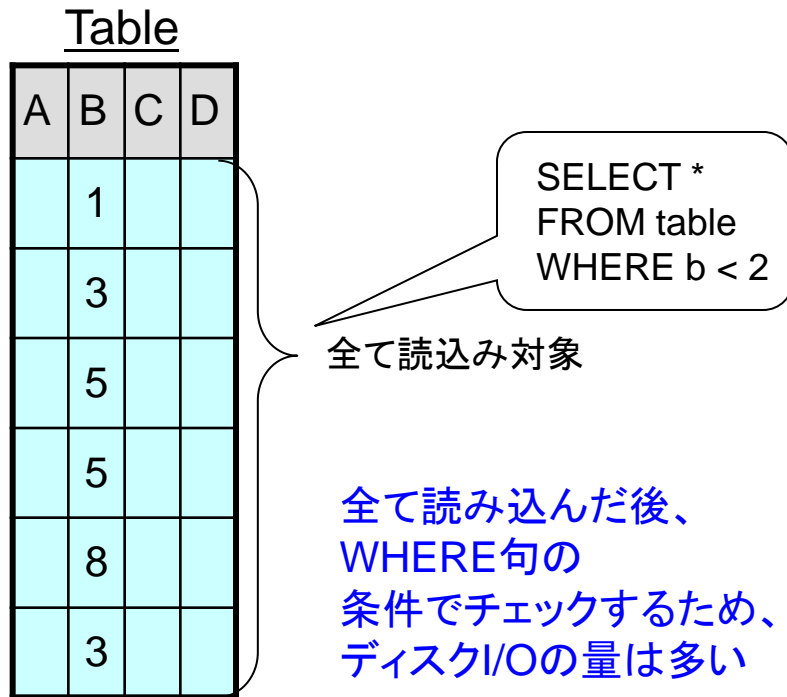
10 対 1

Storage Indexによる不要なI/Oの削減

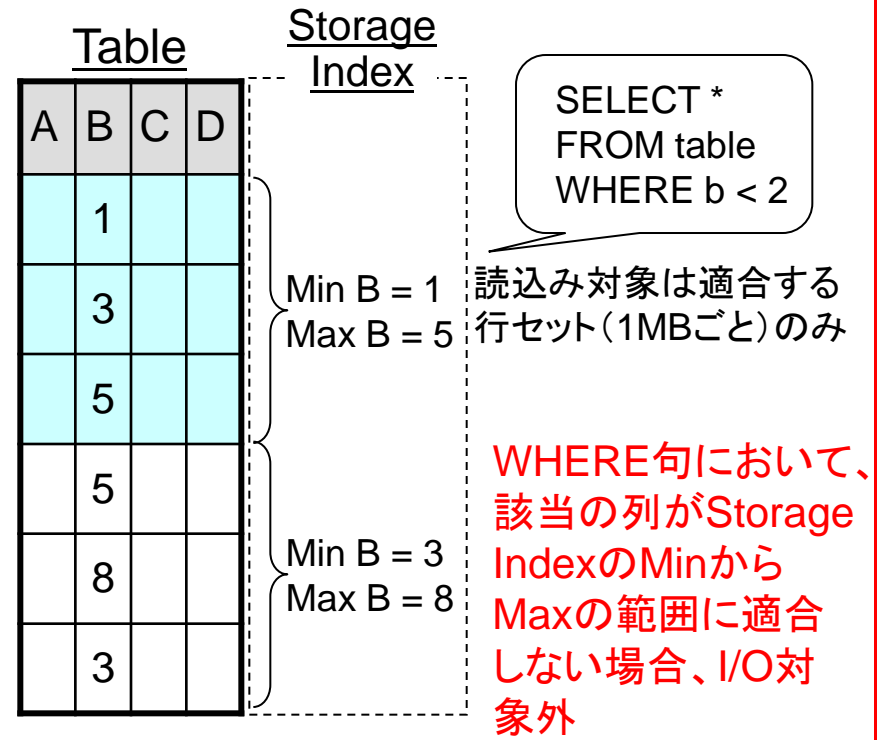
表データのサマリ情報をメモリ上で管理することで、
アクセス対象データセットを選別し不要な I/O を削減

自動でメンテナンスされ、データベースやアプリケーションからは透過的

従来のデータベースシステム



Oracle Exadata



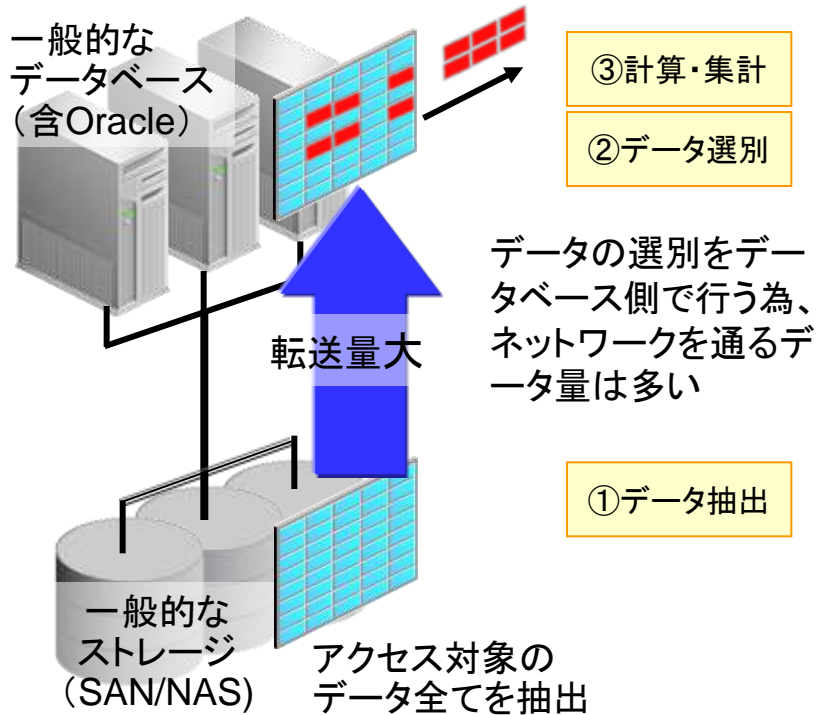
ストレージへの 処理のオフロード

10 対 1

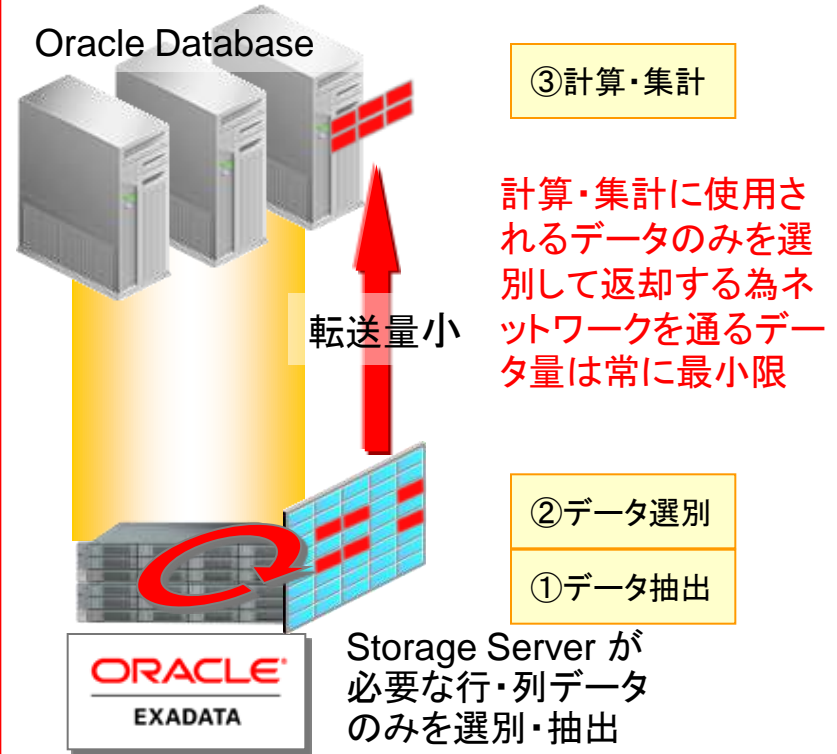
Smart Scanによるストレージへの処理のオフロード

ストレージが問い合わせを解釈し、必要なデータだけをDBサーバーへ返す
サーバーとストレージ間のI/O量を最小限に留め、安定した性能を実現

従来のデータベースシステム

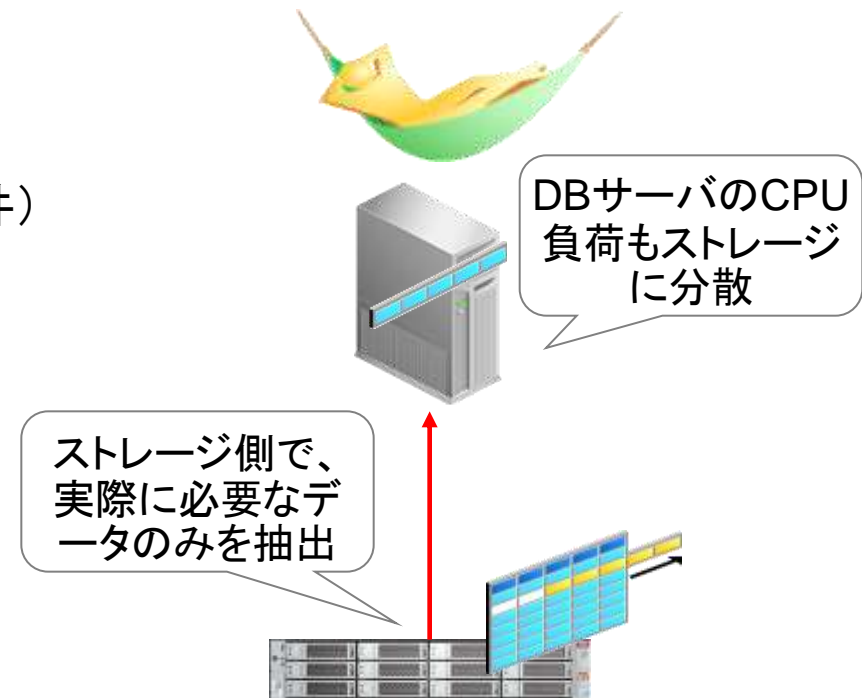


Oracle Exadata

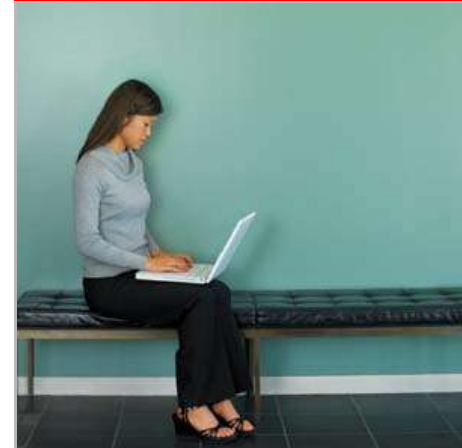


Exadata Smart Scan

- データベース処理の一部をOracle Exadata Storage Serverへオフロード
 - 結果となる行や列のみをデータベースに返す
 - 大幅なデータ削減率
- スマートスキャンの種類
 - 行のフィルタリング (WHERE句の条件)
 - 列のフィルタリング
 - ジョイン・フィルタリング
 - 増分バックアップ・フィルタリング
 - 暗号化データのスキャン
 - Data Miningモデル・スコアリング
- 完全に透過的に作動
 - 既存アプリの修正が必要ない
 - クエリー実行中に障害が発生しても透過的に作動



まとめ



まとめ

- Oracle Exadataとは
 - 驚異的なパフォーマンスを発揮する
Oracle Databaseに最適化されたEngineered System
- Oracle Exadataが速い7つの理由
 1. SQLの平行実行
 2. Oracle ASMによるデータのストライピング
 3. InfiniBandによる広帯域、低遅延ネットワーク
 4. Flashの活用によるスループットとレスポンスの向上
 5. ハイブリッド列圧縮によるI/O削減
 6. Storage Indexによる不要なI/Oの削減
 7. Smart Scanによるストレージへの処理のオフロード



OTNセミナーオンデマンド

コンテンツに対する
ご意見・ご感想を是非お寄せください。

OTNオンデマンド 感想



http://blogs.oracle.com/oracle4engineer/entry/otn_ondemand_questionnaire

上記に簡単なアンケート入力フォームをご用意しております。

セミナー講師/資料作成者にフィードバックし、
コンテンツのより一層の改善に役立てさせていただきます。

是非ご協力をよろしくお願いいたします。

OTNセミナーオンデマンド

日本オラクルのエンジニアが作成したセミナー資料・動画ダウンロードサイト

掲載コンテンツカテゴリ(一部抜粋)

Database 基礎

Database 現場テクニック

Database スペシャリストが語る

Java

WebLogic Server/アプリケーション・グリッド

EPM/BI 技術情報

サーバー

ストレージ



超入門! Oracle データベースって何
再生時間: 60分

100以上のコンテンツをログイン不要でダウンロードし放題

データベースからハードウェアまで充実のラインナップ

毎月、旬なトピックの新作コンテンツが続々登場

例えばこんな使い方

- 製品概要を効率的につかむ
- 基礎を体系的に学ぶ/学ばせる
- 時間や場所を選ばず(オンデマンド)に受講
- スマートフォンで通勤中にも受講可能



毎月チェック!



コンテンツ一覧 はこちら

<http://www.oracle.com/technetwork/jp/ondemand/index.html>

新作&おすすめコンテンツ情報 はこちら

<http://oracletech.jp/seminar/recommended/000073.html>

OTNオンデマンド



オラクルエンジニア通信

オラクル製品に関わるエンジニアの方のための技術情報サイト

オラクルエンジニア通信 - 技術資料、マニュアル、セミナー

Oracleエンジニアのための技術情報サイト by Oracle Japan

新着情報を知りたい

技術資料を探したい

セミナーを受けたい

About

Oracleエンジニアの方がスキルアップしていただくために、厳選した情報をお届けしています

技術資料



インストールガイド・設定チュートリアルetc. 欲しい資料への最短ルート

特集テーマ
Pick UP



性能管理やチューニングなど月間テーマを掘り下げて詳細にご説明

アクセス
ランキング



他のエンジニアは何を見ているのか？人気資料のランキングは毎月更新

技術コラム



SQLスクリプト、索引メンテナンスetc. 当たり前運用/機能が見違える!?

<http://blogs.oracle.com/oracle4engineer/>

オラクルエンジニア通信





The screenshot shows the top navigation bar of the oracletech.jp website. It features the 'oracletech.jp' logo in red and black, with the tagline '好奇心が、エンジニア人生を豊かにする。' below it. To the right is the 'ORACLE' logo, a search bar, and social media icons for Twitter, Facebook, Ustream, YouTube, and RSS. Below these are five red navigation buttons: '製品/技術情報', 'スキルアップ', 'セミナー', 'キャンペーン', and 'ちょっと一息'.

製品/技術
情報



Oracle Databaseっていくら？オプション機能も見積れる簡単ツールが大活躍

セミナー



基礎から最新技術までお勧めセミナーで自分にあった学習方法が見つかる

スキルアップ



ORACLE MASTER ! 試験頻出分野の模擬問題と解説を好評連載中

Viva!
Developer



全国で活躍しているエンジニアにスポットライト。きらりと輝くスキルと視点を盗もう

<http://oracletech.jp/>

oracletech



あなたにいちばん近いオラクル



Oracle Direct

まずはお問合せください

Oracle Direct



システムの検討・構築から運用まで、ITプロジェクト全般の相談窓口としてご支援いたします。
システム構成やライセンス/購入方法などお気軽にお問い合わせ下さい。

Web問い合わせフォーム

専用お問い合わせフォームにてご相談内容を承ります。
http://www.oracle.co.jp/inq_pl/INQUIRY/quest?rid=28

※フォームの入力にはログインが必要となります。
※こちらから詳細確認のお電話を差し上げる場合がありますので
ご登録の連絡先が最新のものになっているかご確認下さい。

フリーダイヤル

0120-155-096

※月曜～金曜
9:00～12:00、13:00～18:00
(祝日および年末年始除く)

ORACLE

Hardware and Software **Engineered to Work Together**

ORACLE®