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

今すぐできるNAS環境の高速化！
Oracle DatabaseのI/Oに最適化されたNFSクライアントの使用方法とSSDを組み合わせた活用方法

日本オラクル株式会社　製品事業統括技術本部　基盤技術部　エンジニア　岩本知博

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。

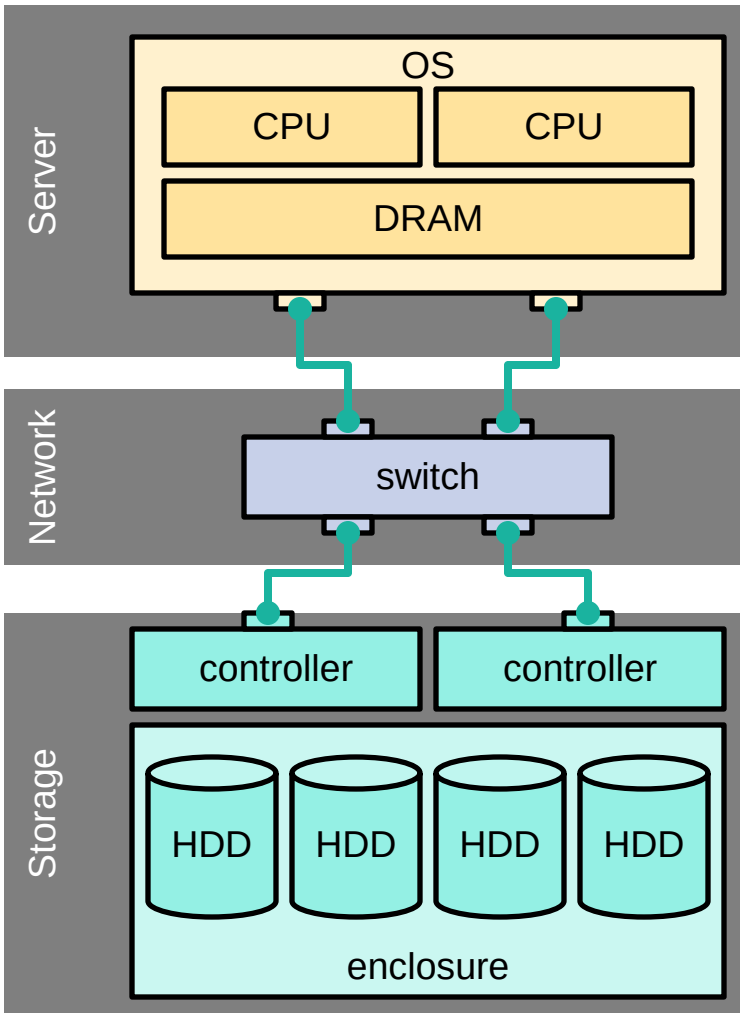
Agenda

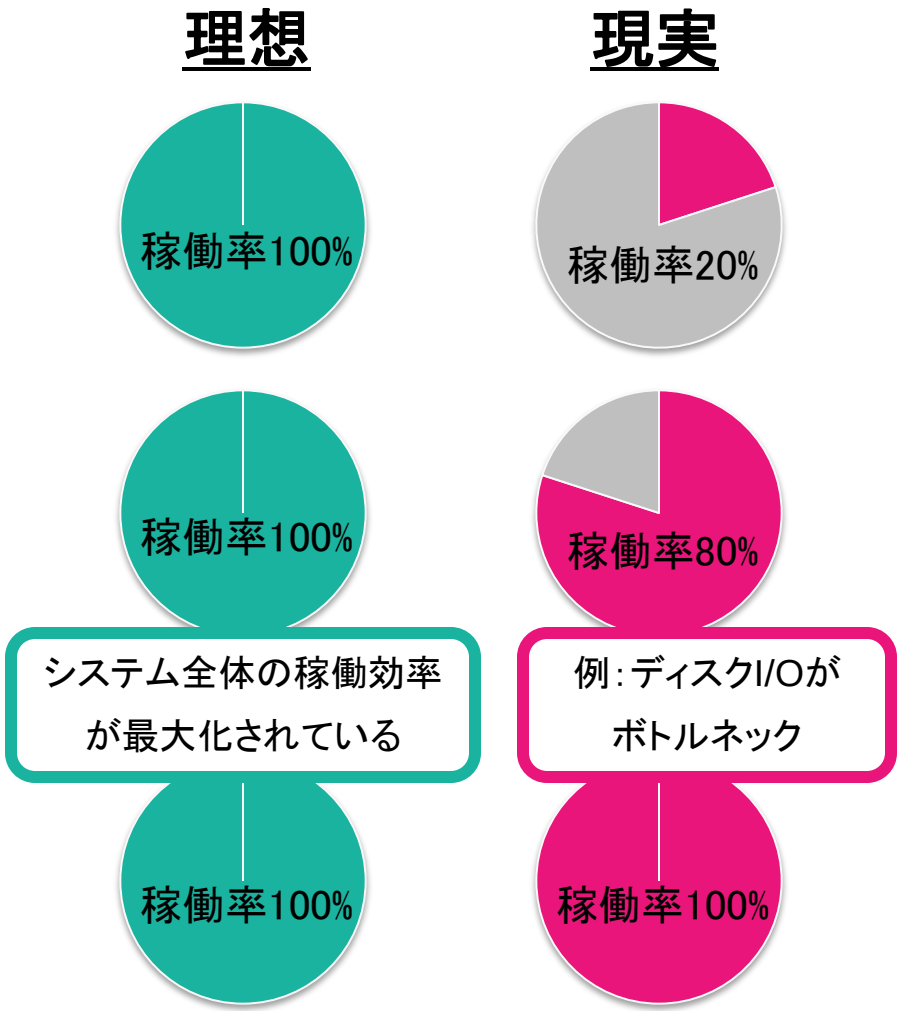
- Introduction
  - システム性能の最大化とディスクI/Oのボトルネック
  - Storage Area Network（SAN）とNetwork Attached Storage（NAS）
  - データベースの配置先としてのNetwork Attached Storage（NAS）
- Direct NFSのご紹介
  - Direct NFSの適用ケースの分析
  - 設定方法
  - Direct NFSによるネットワーク帯域のスケーラビリティ
- Direct NFSのまとめ
- Direct NFS活用例
  - DB Smart Flash CacheによるSSDの効果的な活用法
  - dNFSとの組み合わせによるDB統合の集約密度向上

Introduction

はじめに

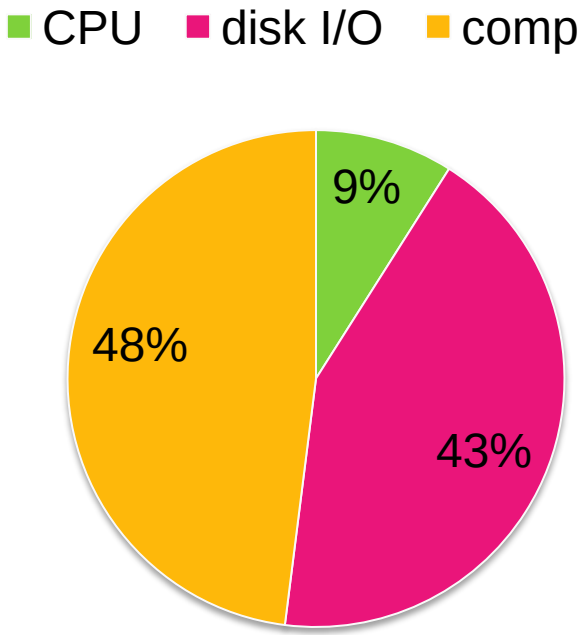
限られた予算でシステム性能を最大化するには





ディスクI/Oがボトルネックになりがち

- ストレージ設計は、容量以上に「性能」を考慮することが重要



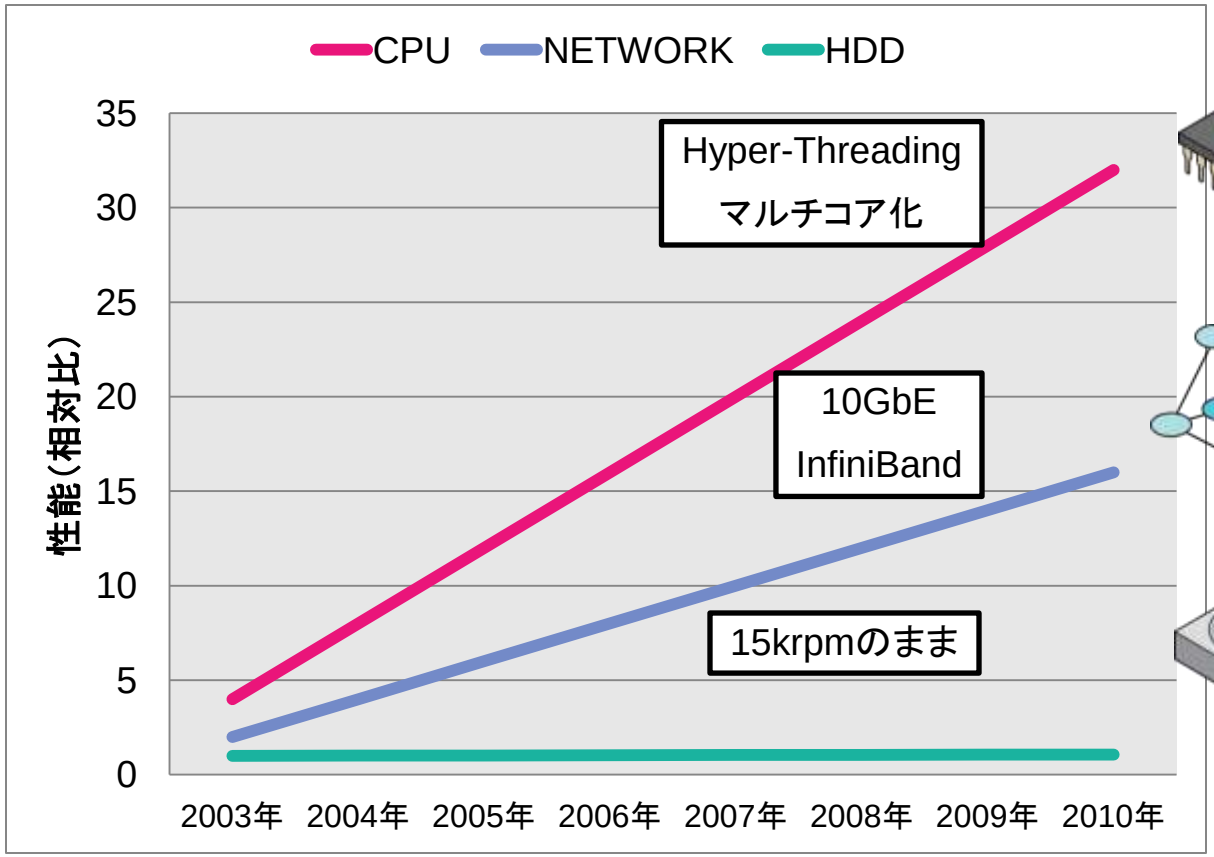
DBシステム性能のボトルネックの要因

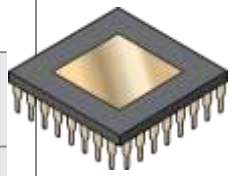
- CPU: 9%
- ディスクI/O: 43%
- 非効率なSQL文、索引の設計等: 48%

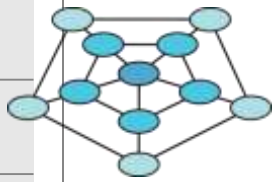
*Oracle Direct パフォーマンス・クリニック・サービス

【参考】http://www.oracle.com/lang/jp/direct/service/pc.html

なぜディスクI/Oがボトルネックになるのか







2003 vs 2010

処理性能

x 32

トランジスタ数

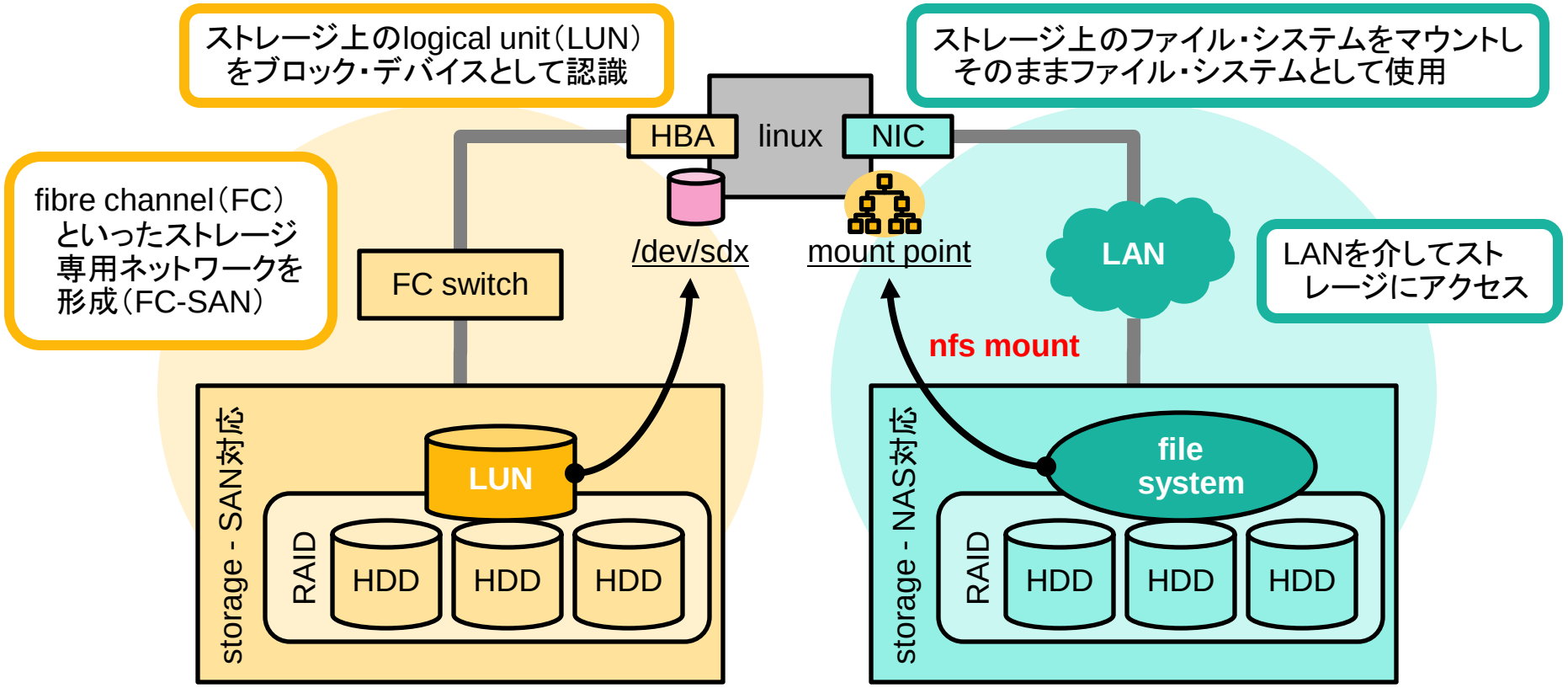
拡大傾向

x 1.X

回転数（rpm）

外部ストレージ接続形態（SANとNAS）

HDDの扱い方と接続方法



SAN（Storage Area Network） NAS（Network Attached Storage）

外部ストレージ接続形態（SANとNAS）

特徴と一般的な用途

- SAN（FC-SAN）の特徴
  - 直接ブロック・デバイスにアクセスすることで、I/Oオーバーヘッドを最小化
  - FCで形成された広帯域なストレージ専用ネットワーク

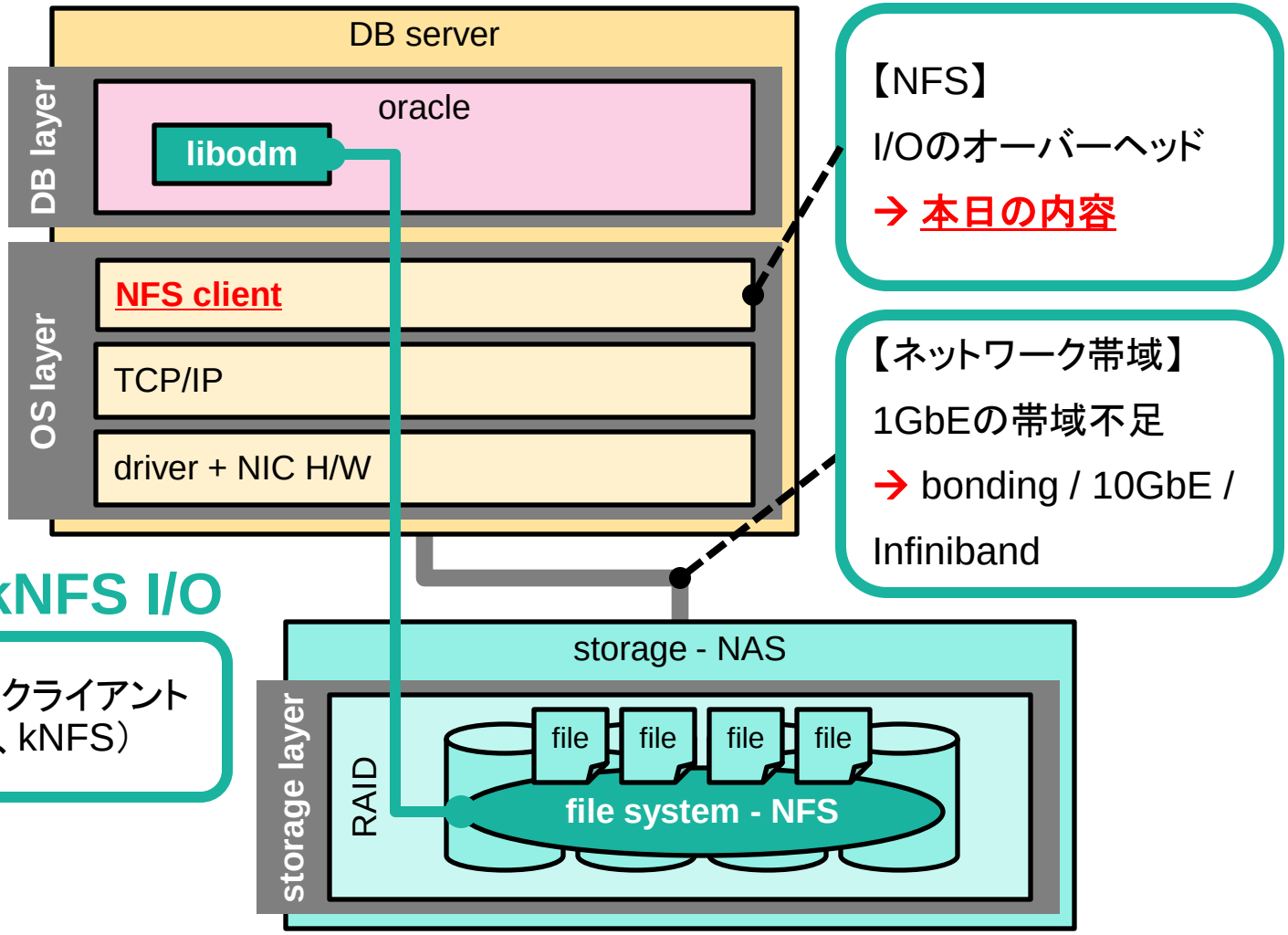
→ DBシステムのストレージとして一般的

- NASの特徴
  - SANと比較して、導入が容易かつ低コスト
  - ファイル・システムのインターフェイスを介した使い勝手の良さ

→ ファイル・サーバ用途として広く普及

データベースの配置先としてのNAS

性能への影響



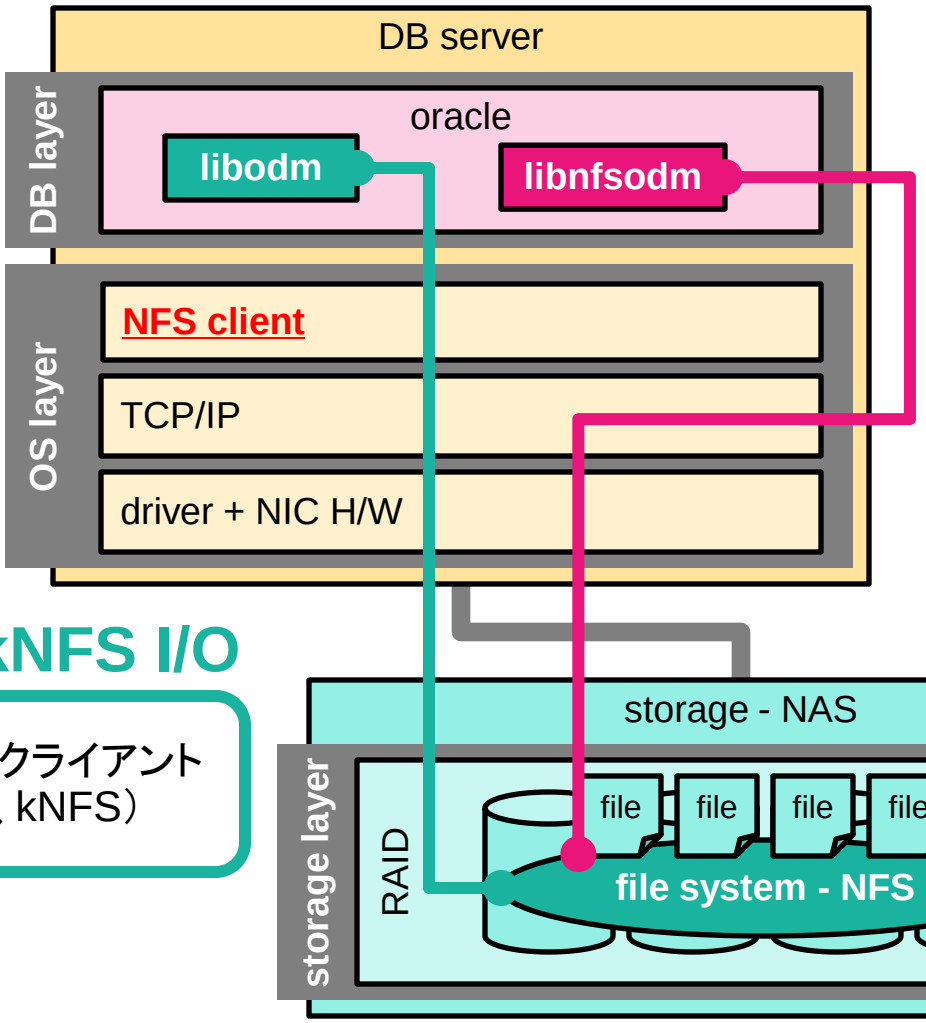
kNFS I/O

oracleはOSのNFSクライアントを介してI/O（以降、kNFS）

Direct NFS

Direct NFS（dNFS）

OracleはNASへのI/Oオーバーヘッドを減らす機能を実装



dNFS I/O

OS層をバイパスすることでオーバーヘッドを減らし、I/Oを高速化

kNFS I/O

oracleはOSのNFSクライアントを介してI/O（以降、kNFS）

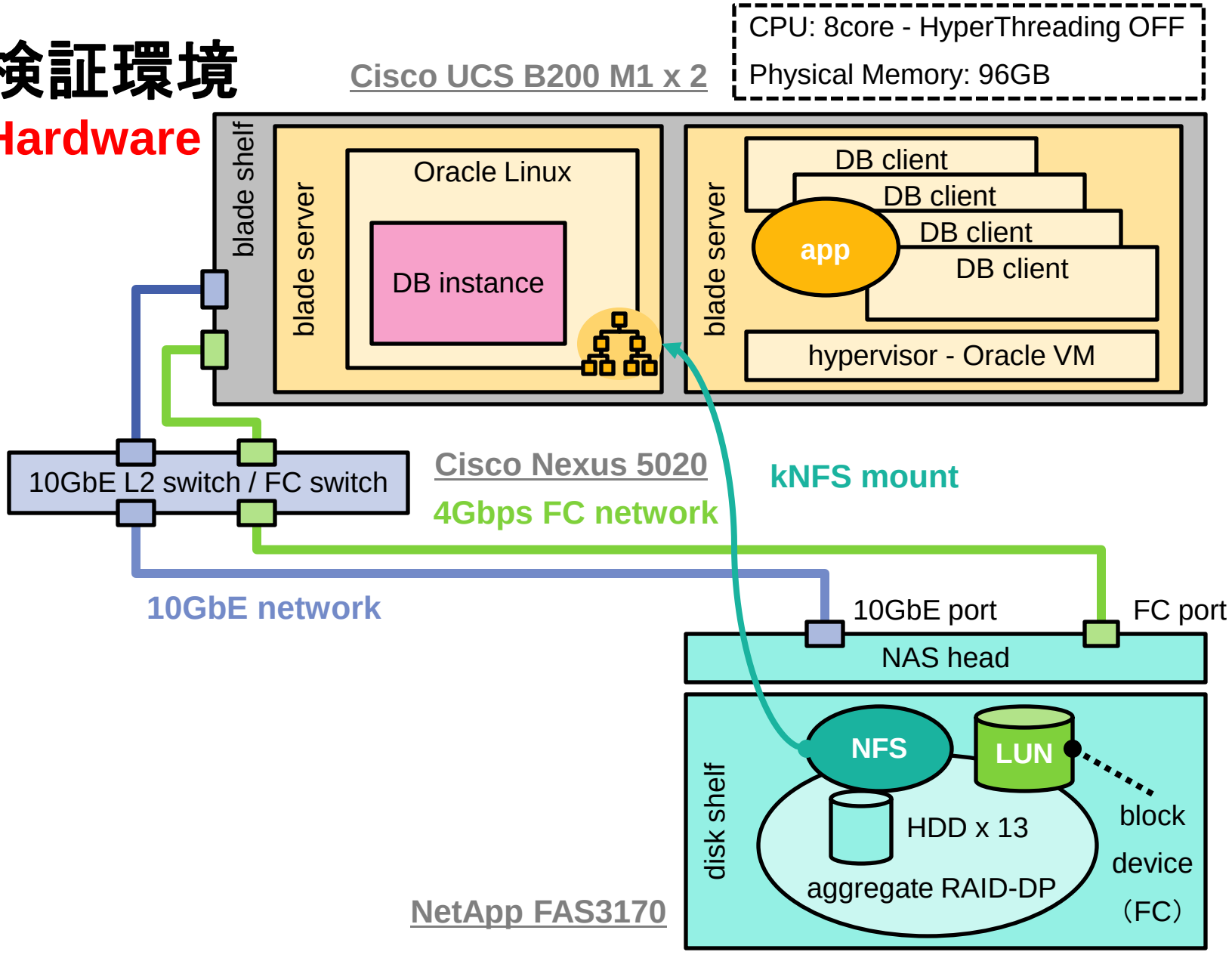
Direct NFS（dNFS）機能概要

- dNFSとは
  - Oracle Database内部に実装されたNFS（v3）クライアント機能
  - Oracle Database 11g Release 1から使用可能

- dNFSの特徴
  1. OSカーネルのNFSクライアント（kNFS）より高いディスクI/O性能
  2. 簡単な手順で機能を有効化
     - アプリケーションの書き換えは必要ない
     - ストレージの構成や運用に影響はない
     - 複数イーサネット・ポートを使用したネットワーク帯域のスケーラビリティの設定が簡単

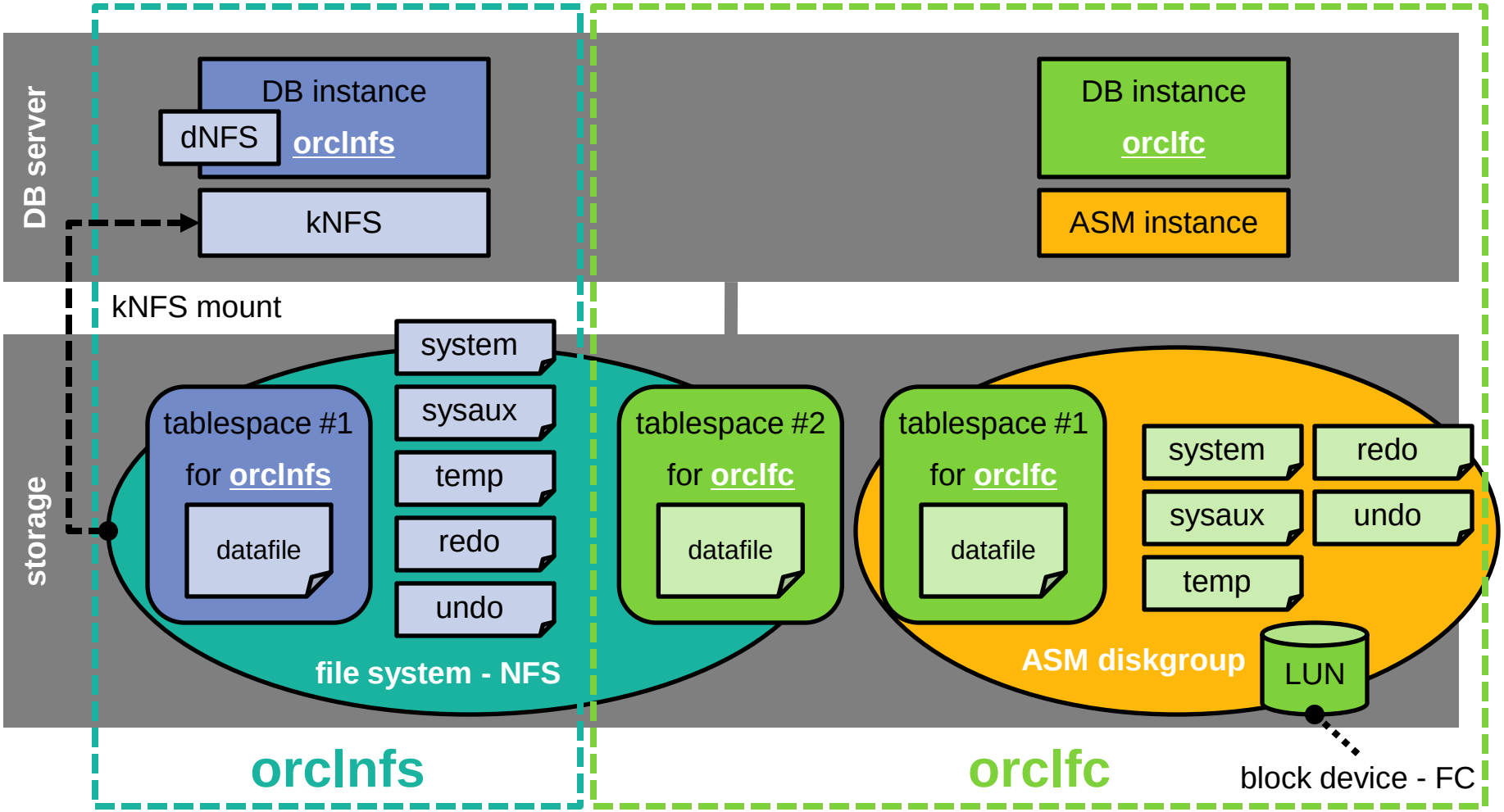
検証環境

Hardware



検証環境

Oracle Configuration



kNFSより高いI/O性能

検証内容

- dNFSの適用ケースを判断するため、以下の内容でkNFSとdNFSの性能を確認する

1. Webショッピング・サイトを想定したOLTP

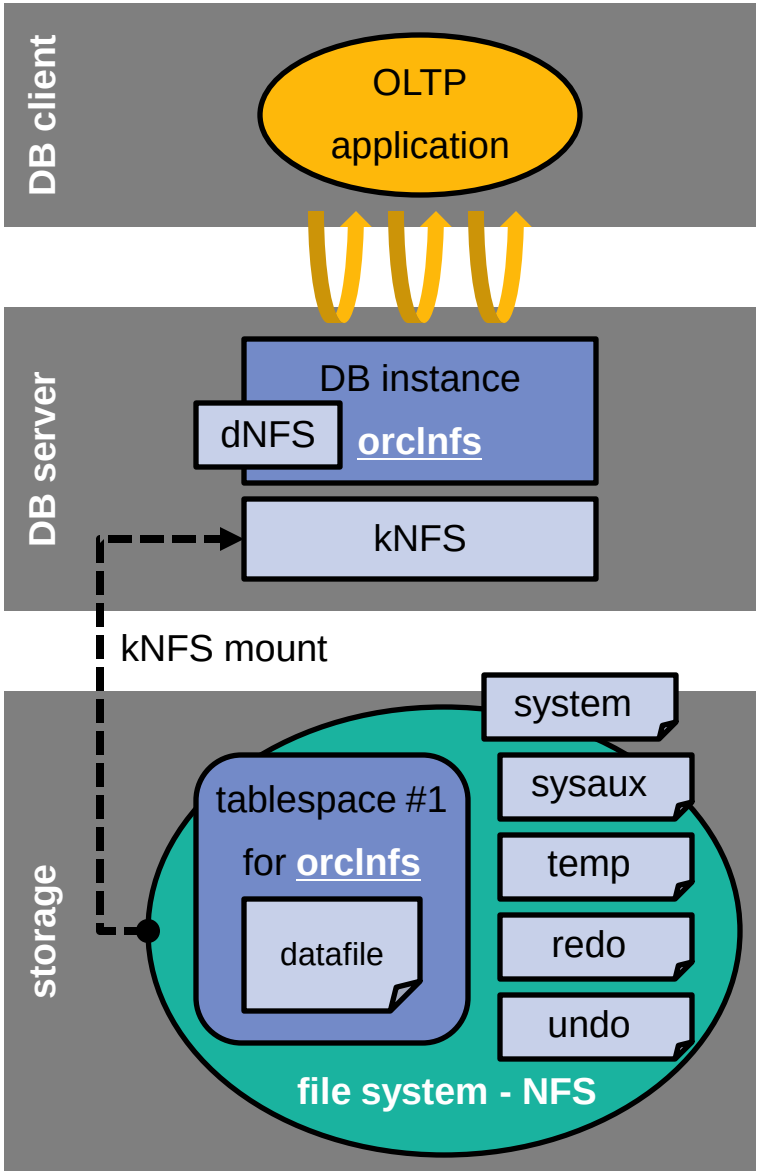
2. DWH / BATCH
   - 2-1. SELECT
   - 2-2. INSERT
   - 2-3. UPDATE

kNFSより高いI/O性能

OLTP

- Webショッピング・サイトを想定したOLTP

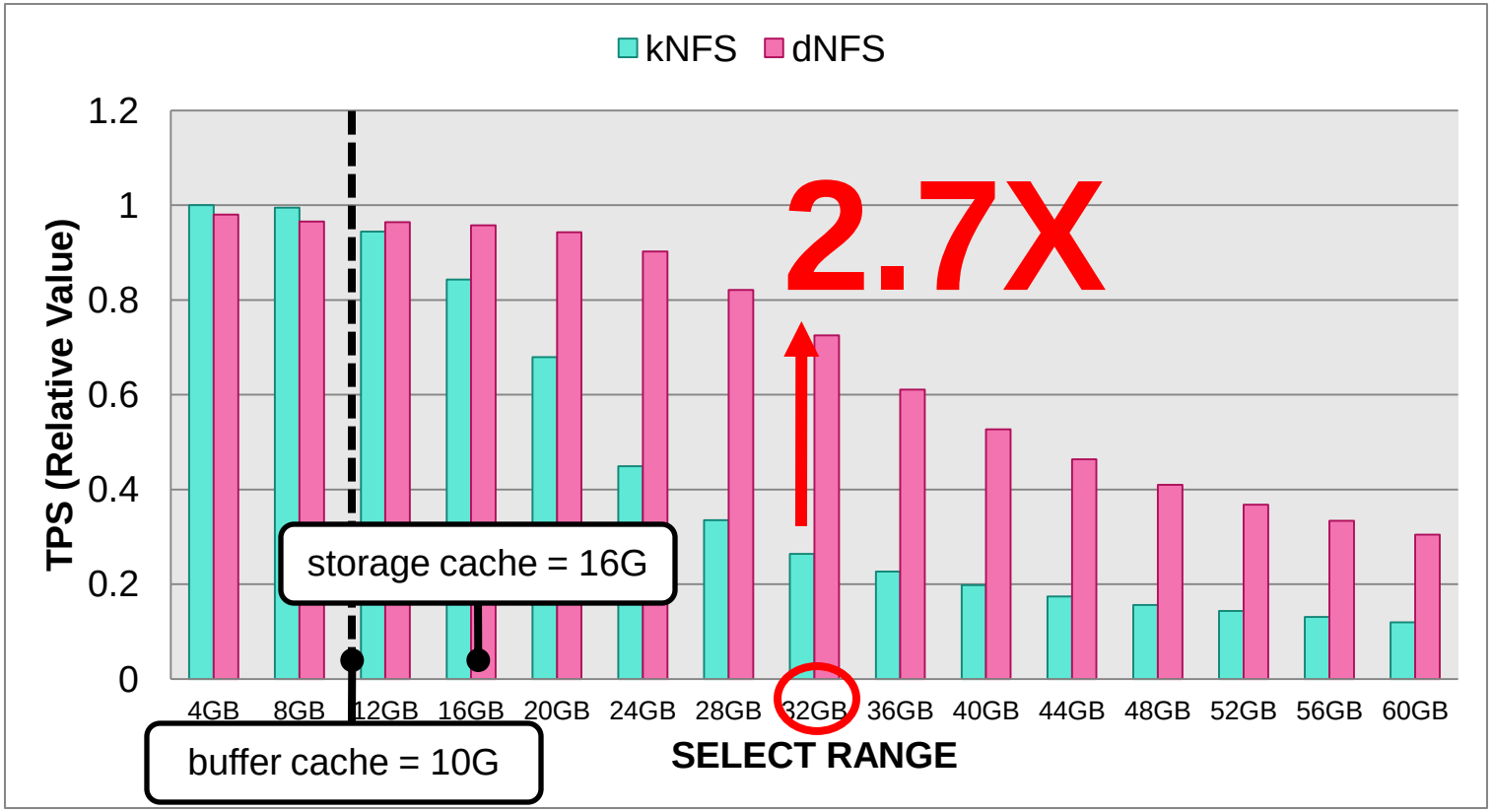
- トランザクションの割合
  - 商品検索のみ（SELECT）のトランザクション = 80%（SELECT）
  - 商品購入を含む（SELECT & INSERT & UPDATE）トランザクション = 20%



kNFSより高いI/O性能

OLTP

- OLTPで発生するI/Oの性能が向上することを確認
  - キャッシュ・ヒット率が低く、I/O量が多い環境ほど有効



kNFSより高いI/O性能

OLTP: AWRレポート

 # kNFS

Event	Waits	Time(s)	Avg wait (ms)	% DB time	Wait Class
db file sequential read	2,132,908	268,121	126	87.8	User I/O
log file sync	170,284	35,137	206	11.5	Commit
DB CPU		1,636		.5	
Disk file operations I/O	1,620	358	221	.1	User I/O
enq: TX - index contention	1,059	218	206	.1	Concurrenc

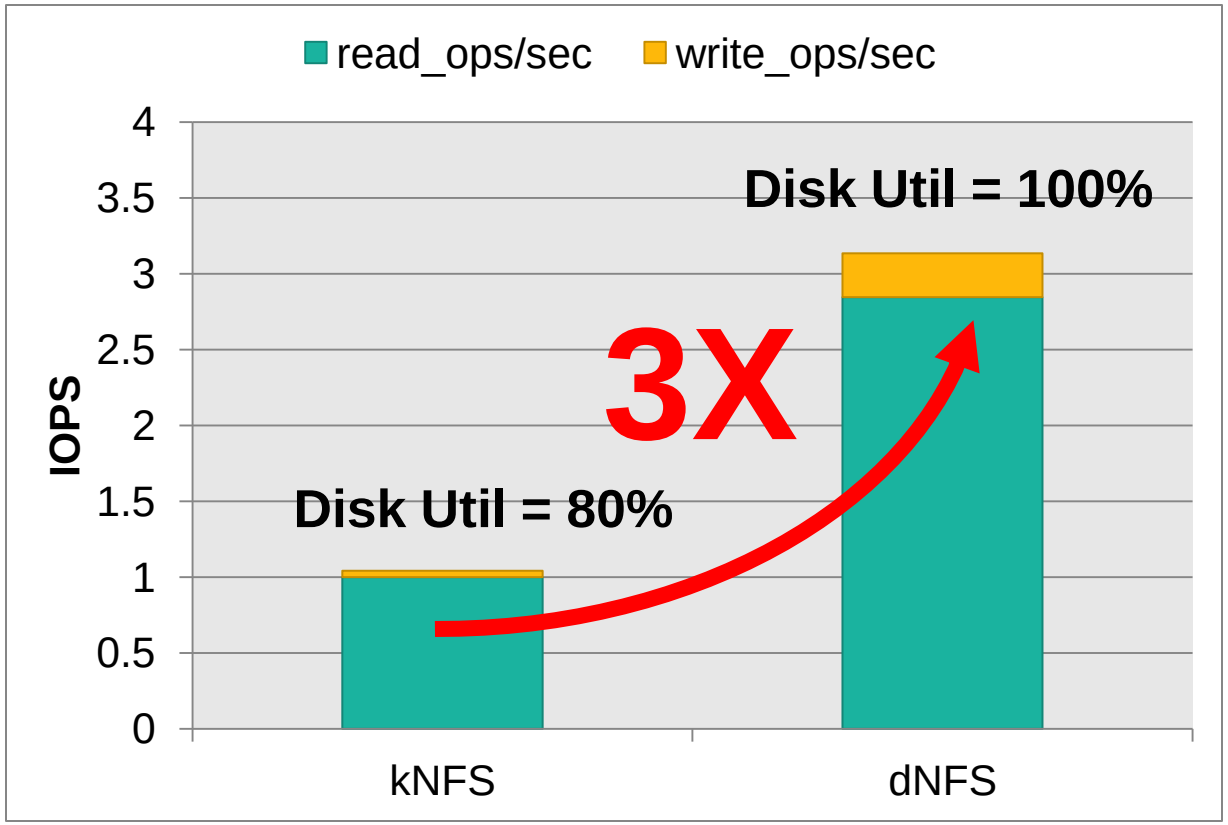
 # dNFS

Event	Waits	Time(s)	Avg wait (ms)	% DB time	Wait Class
db file sequential read	6,074,413	133,800	22	94.2	User I/O
log file sync	468,167	5,427	12	3.8	Commit
DB CPU		3,003		2.1	
read by other session	1,479	89	60	.1	User I/O
cursor: pin S	13,004	51	4	.0	Concurrenc

kNFSより高いI/O性能

OLTP: ストレージ統計

- dNFSであれば、HDDの限界性能まで引き出せている
  - IOPSが約3倍になり、disk util = 100%

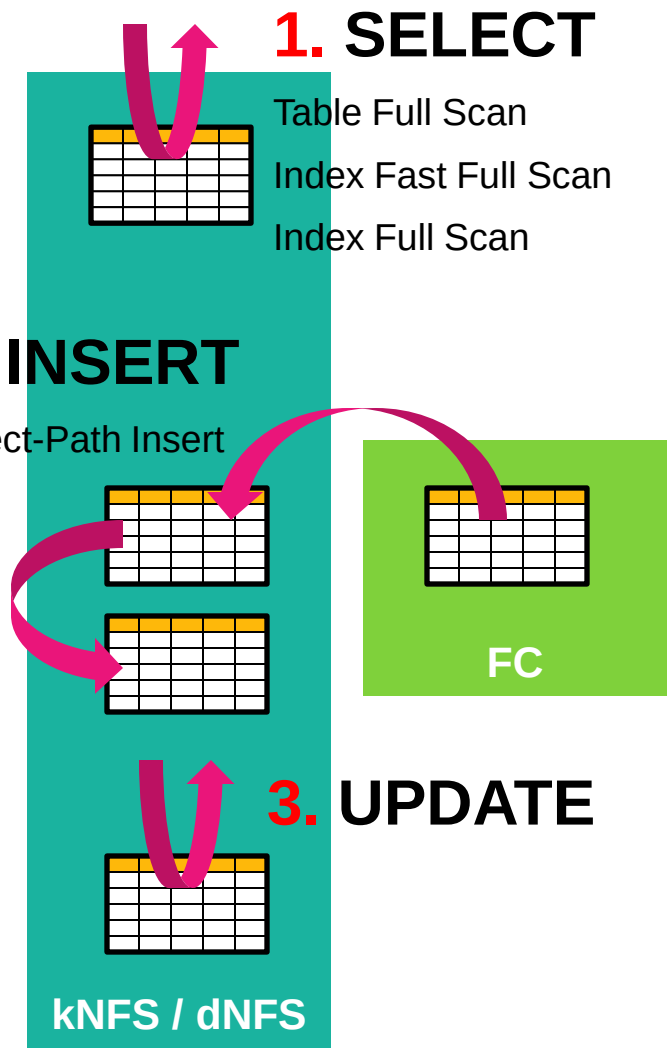


kNFSより高いI/O性能

DWH / BATCH

1. SELECT
   - 表の行数をカウント（960,000,000行）
     - 同じSQLに対して、ヒント句を用いてアクセス・パスを制御
2. INSERT
   - 960,000,000行のINSERT処理
     - NFS上の表からNFS上の表へDirect-Path Insert
     - FCのLUN上の表からNFS上の表へDirect-Path Insert
3. UPDATE
   - 960,000,000行の表を全行UPDATE

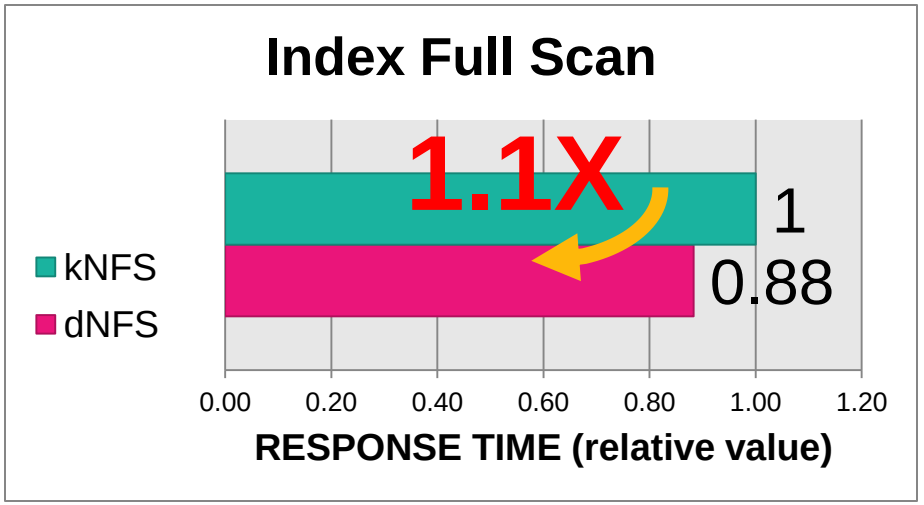
- Parallel Query / DMLを使用

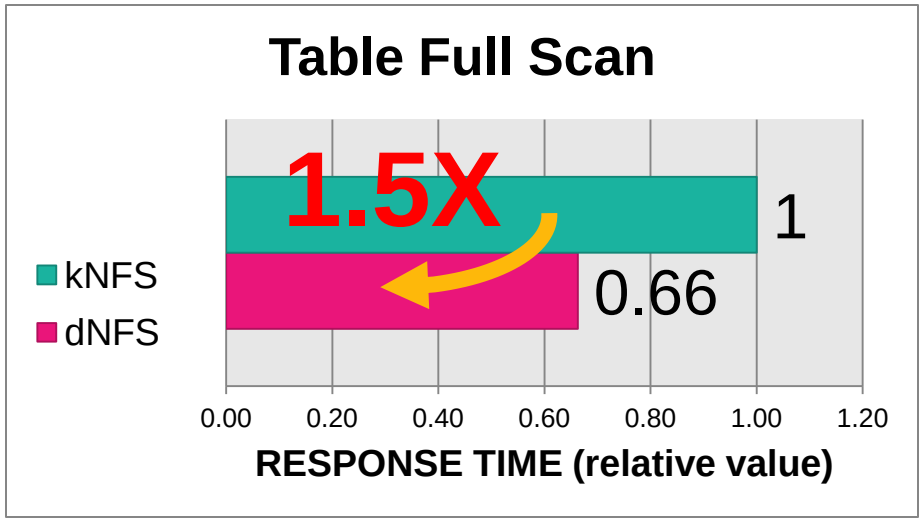


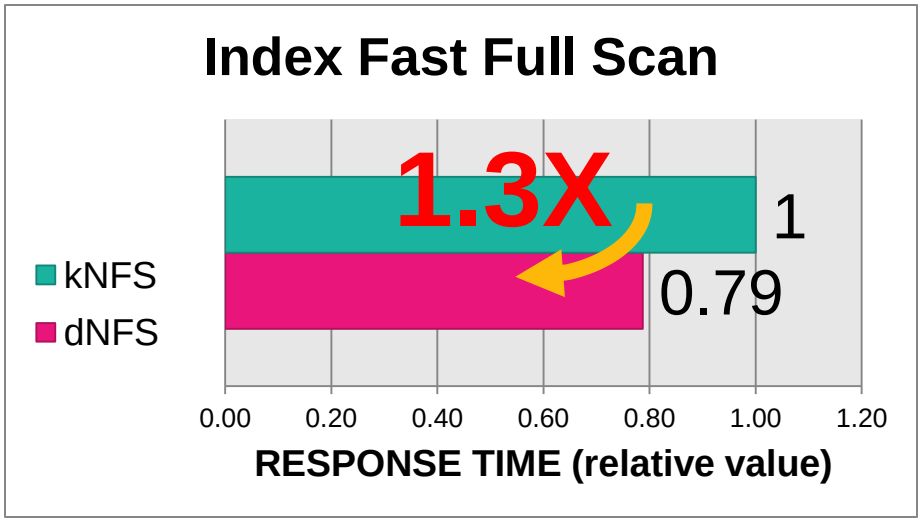
kNFSより高いI/O性能

DWH / BATCH: SELECT（パラレル処理）

- dNFSを使用することで、いずれのアクセス・パスでも高速化されることを確認
  - 特にTable Full Scanで効果が高い（約**1.5倍**）



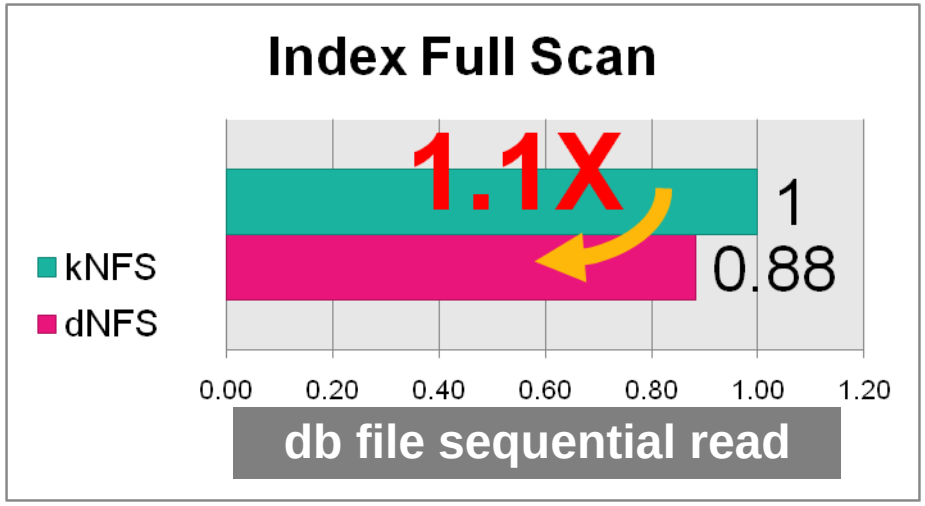


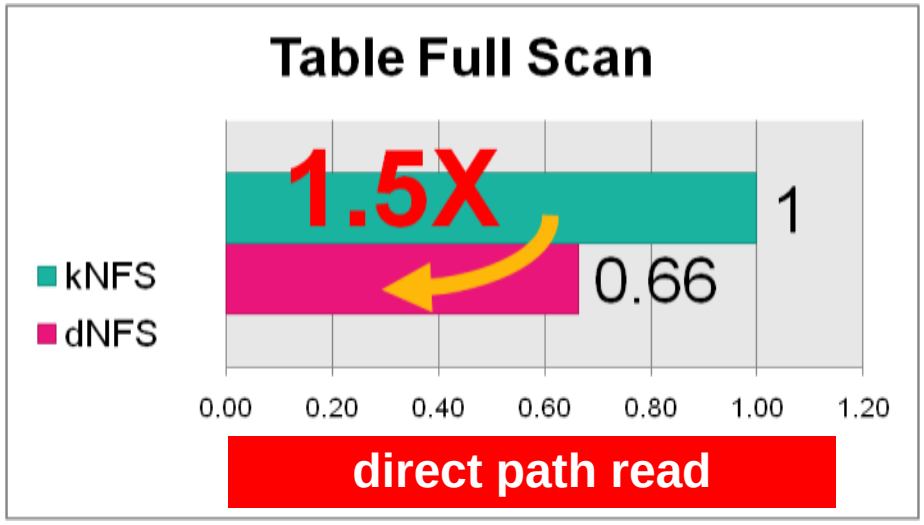


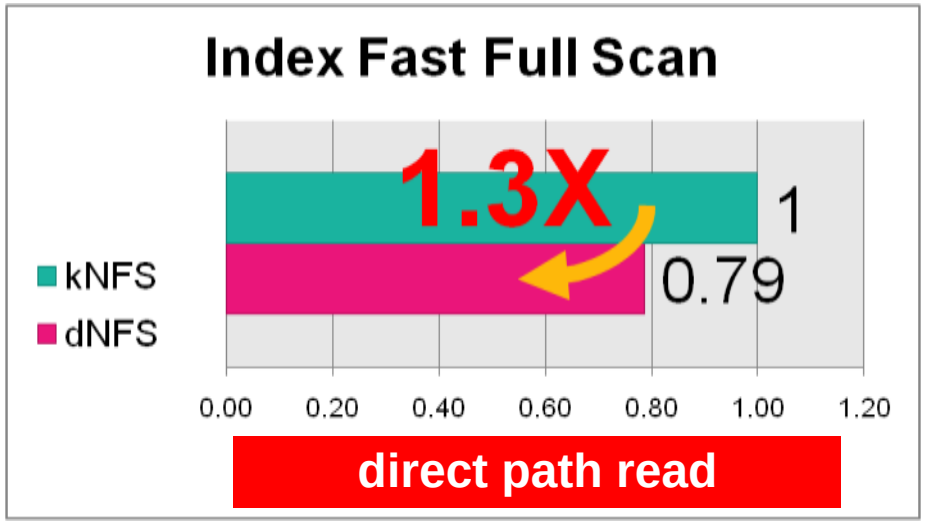
kNFSより高いI/O性能

DWH / BATCH: SELECT（パラレル処理）

- 効果が高いケースは、データ・ブロックの読み込み処理が、direct path readで行われている



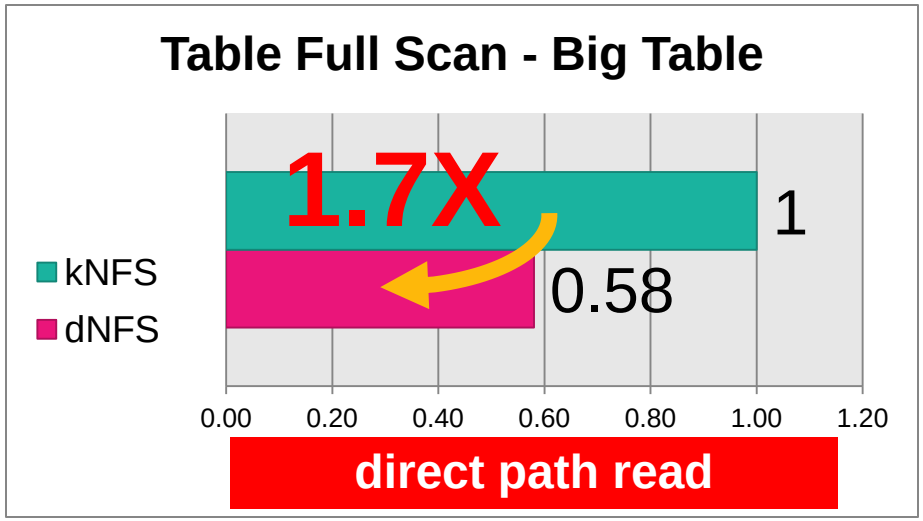


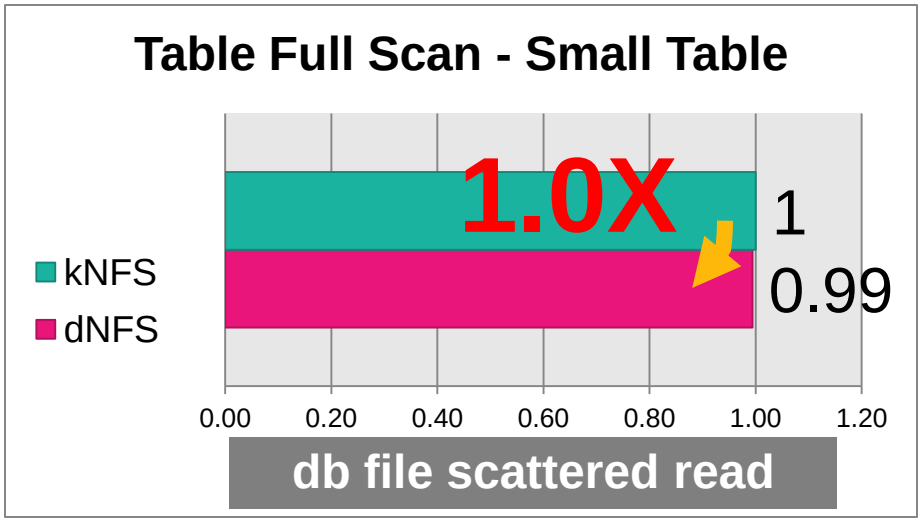


【補足】kNFSより高いI/O性能

DWH / BATCH: SELECT（シリアル処理）

- データ・ブロックの読み込み処理がdb file scattered readで行われている場合、効果が小さい傾向
  - ※ パラレルでのTable Full Scanはdirect path readで行われる
  - ※ シリアルでのTable Full Scanは、DB buffer cacheと表のサイズに次第でdirect path readを選択（11g Release 1以降）

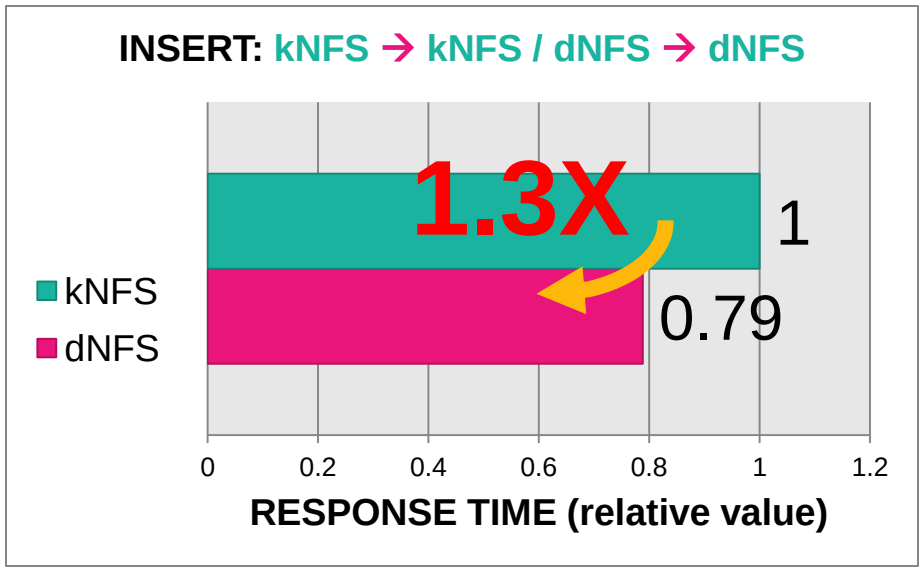


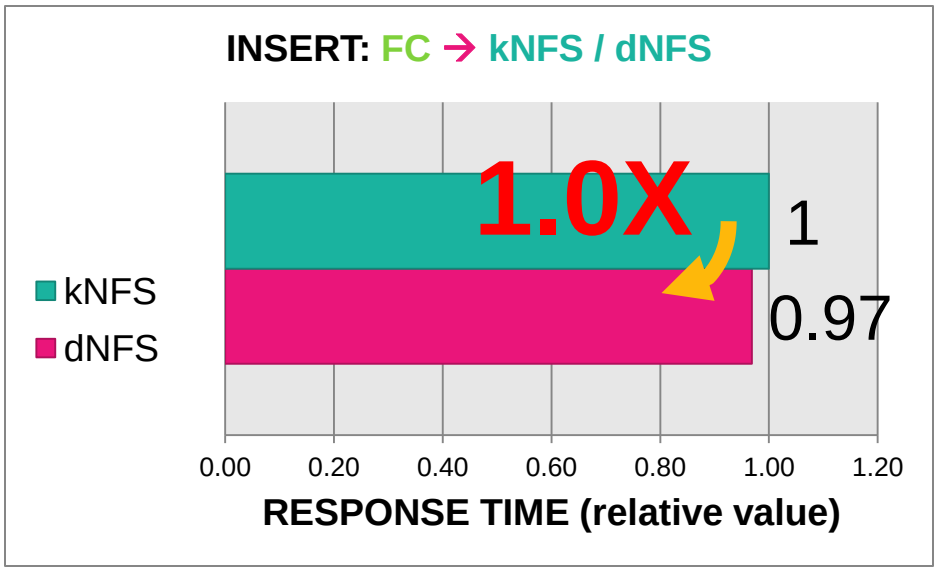


kNFSより高いI/O性能

DWH / BATCH: INSERT

- 書き込み処理においては効果が低い傾向
  - dNFS → dNFSのケースは、読み込み処理（direct path read）の高速化による性能向上

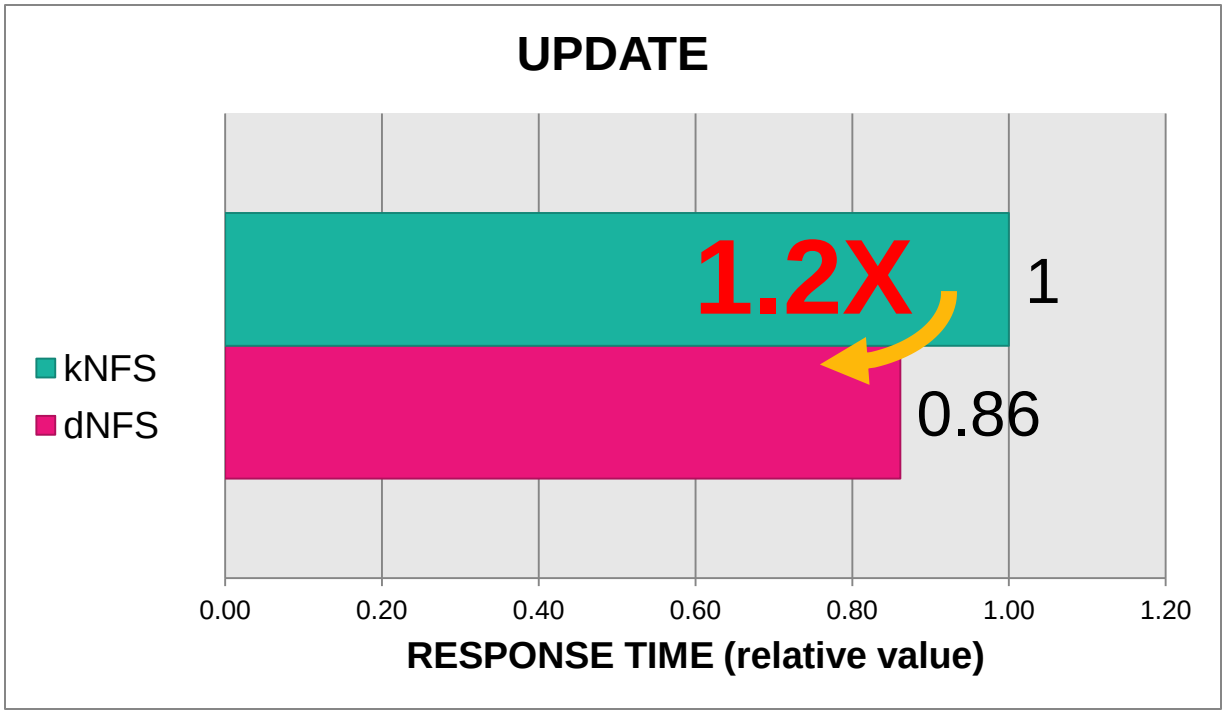




kNFSより高いI/O性能

DWH / BATCH: UPDATE

- 大量データの更新処理が高速化されることを確認



Direct NFS
設定方法

dNFSの有効化

- dNFSを有効にするには、以下の2つの手順を行う

1. 設定ファイル（oranfstab）の作成と編集

2. 使用するライブラリ・ファイルを変更

【補足】oranfstabについて

- oranfstabは、dNFSに対してoracle固有のオプションを指定するファイルという位置づけ
  - dNFSでは、現在マウントしているボリューム（/etc/mtab）の構成に基づいてマウント・ポイント設定が決定される

  → そのため、dNFSでアクセスするボリュームも、kNFSでマウントする必要がある
    - kNFSのマウント・オプションは、従来通り、ボリュームの用途（配置するファイルの種類）に適したマウント・オプションを指定

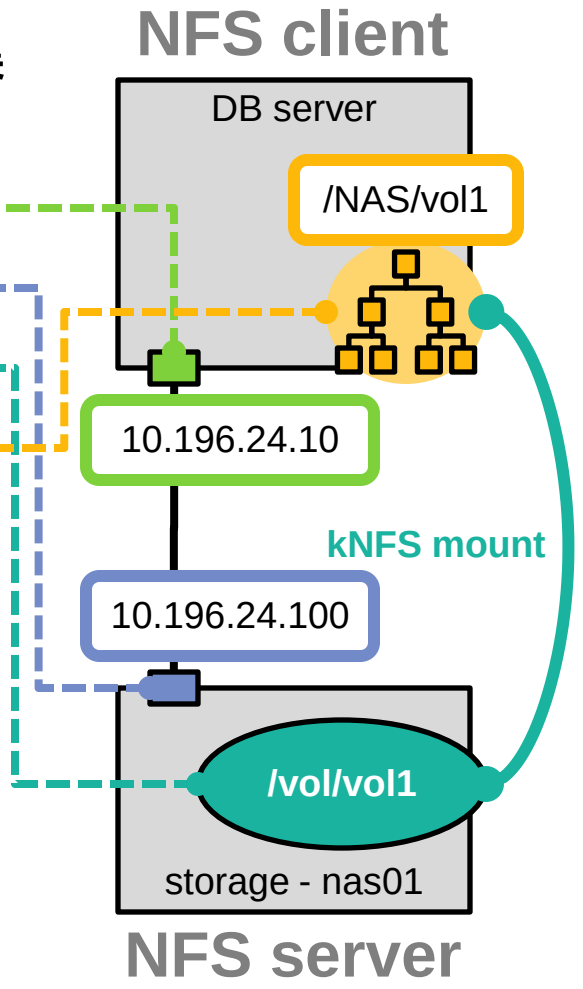
    ※ 具体的なマウント・オプションは、各NAS製品ベンダー様にご確認ください

dNFSの有効化（1/2）

oranfstabの編集

1. $ORACLE_HOME/dbs/oranfstabを編集
   - server: NFSサーバ名
   - local: NFSクライアントのIPアドレス（最大4つ）
   - path: NFSサーバのIPアドレス（最大4つ）
   - export: NFSサーバからexportされたパス
   - mount: NFSクライアントのmountポイント

```
server: nas01
local: 10.196.24.10
path: 10.196.24.100
export: /vol/vol1 mount: /NAS/vol1
```



dNFSの有効化(2/2)

ライブラリ・ファイルの入れ替え

2. DBインスタンスを停止
3. 標準のOracle Disk Mananger(ODM)ライブラリのかわりに、NFSクライアント機能が実装されているODM NFSライブラリを使用する

```
cd $ORACLE_HOME/rdbms/lib

make -f ins_rdbms.mk dnfs_on
```

4. DBインスタンスを起動
   - DBインスタンス起動時に、機能が有効化される
   - 以下のメッセージがalert logに出力されることを確認

```
Oracle instance running with ODM: Oracle
Direct NFS ODM Library Version 3.0
```

dNFSの有効化(2/2)

ライブラリ・ファイルの入れ替え

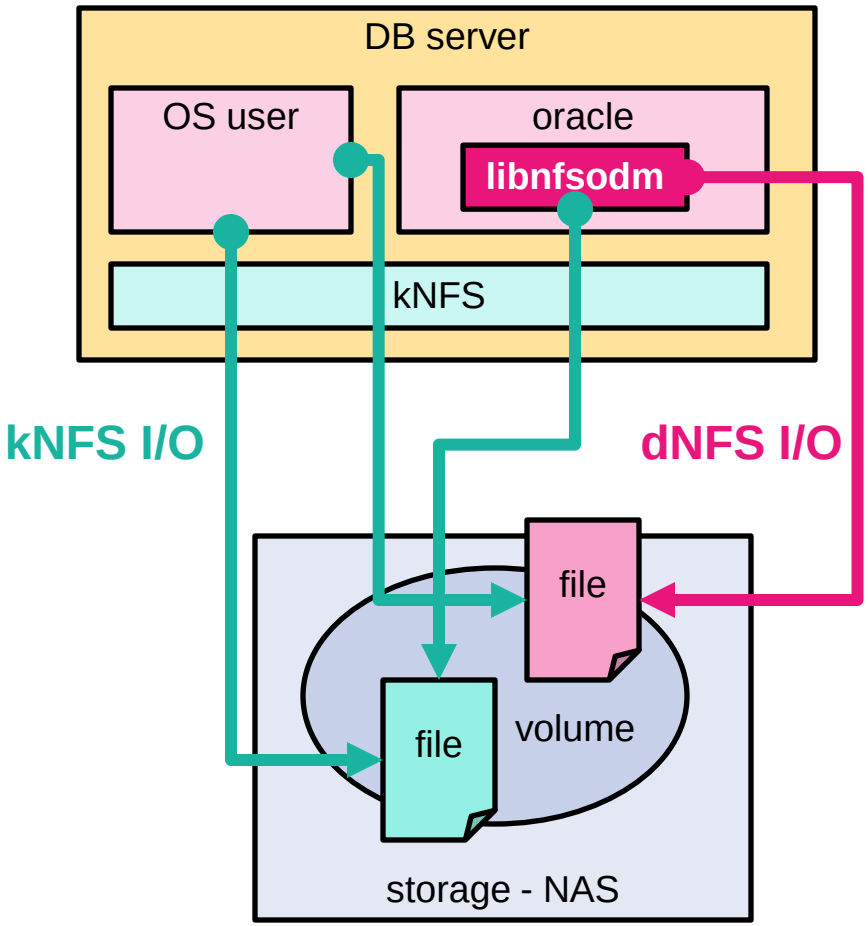
- dNFSの対象はボリューム単位ではなく、ファイル単位

→ サポート対象のファイルのみ、dNFS I/Oが有効化

  - それ以外のファイルは従来通りkNFS I/O対象

- dNFSでアクセスするボリュームは、kNFSマウント済

→ 従来通りのファイル・システムのオペレーションが可能

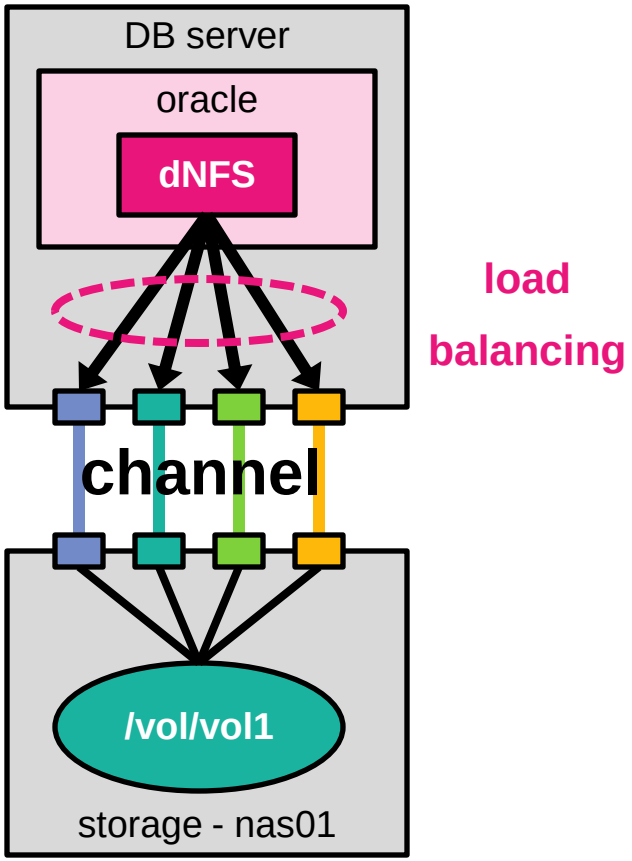


ネットワーク帯域のスケーラビリティ

- oranfstabにチャネル（localとpathのペア）を追加するだけ（最大4チャネル）

```
server: nas01
local: 10.196.24.10
path: 10.196.24.100
local: 10.196.25.11
path: 10.196.25.101
local: 10.196.26.12
path: 10.196.26.102
local: 10.196.27.13
path: 10.196.27.103
export: /vol/vol1 mount: /NAS/vol1
```

channel 1
channel 2
channel 3
channel 4



ネットワーク帯域のスケーラビリティ

v$dnfs_channels

```
SQL> select a.pnum, b.program, a.svrname, a.path, a.local, a.ch_id
from v$dnfs_channels a, v$process b where a.pnum = b.pid;

PNUM PROGRAM                        SVRNAME    PATH           LOCAL          CH_ID
----- ---------------------------- ---------- -------------- -------------- ------
   10 oracle@dbsrv01 (DBW0)         nas01      10.196.24.100 10.196.24.10     0
   10 oracle@dbsrv01 (DBW0)         nas01      10.196.24.100 10.196.24.11     1
   10 oracle@dbsrv01 (DBW0)         nas01      10.196.24.100 10.196.24.12     2
   10 oracle@dbsrv01 (DBW0)         nas01      10.196.24.100 10.196.24.13     3
   11 oracle@dbsrv01 (LGWR)         nas01      10.196.24.100 10.196.24.10     0
   11 oracle@dbsrv01 (LGWR)         nas01      10.196.24.100 10.196.24.11     1
   11 oracle@dbsrv01 (LGWR)         nas01      10.196.24.100 10.196.24.12     2
   11 oracle@dbsrv01 (LGWR)         nas01      10.196.24.100 10.196.24.13     3
   ....
   ...
   ..
   .
```

dNFSのまとめ

- kNFSより高いディスクI/O性能
  - 【OLTP】キャッシュ・ヒット率が低く、ディスクI/Oが頻発している環境で、大きな効果が期待できる
  - 【DWH / BATCH】多くのケースで効果が期待できる
    - 特にdirect path readが行われるSQLで効果大
- 簡単な手順で機能を有効化
  - アプリケーションの書き換えは必要ない
  - ストレージの構成や運用に影響はない
  - 複数イーサネット・ポートを使用したネットワーク帯域のスケーラビリティの設定が簡単

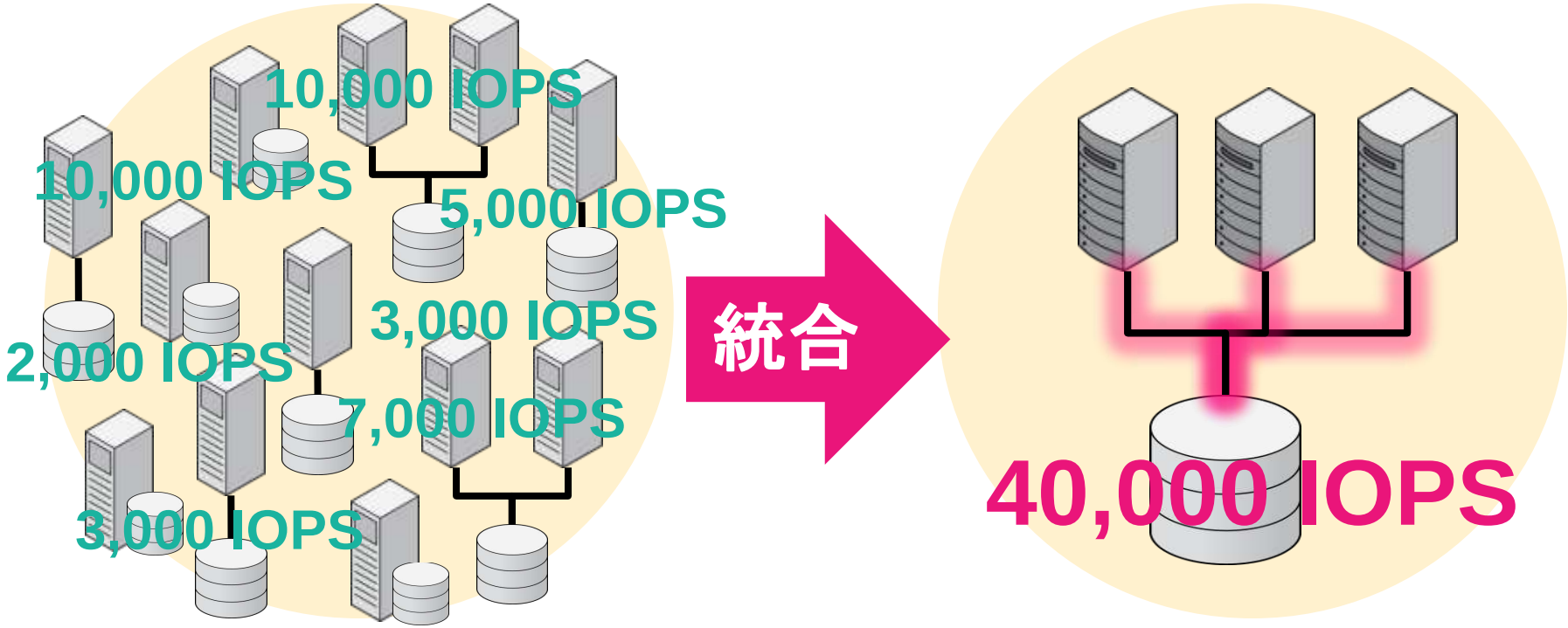
Direct NFS 活用例

dNFS活用例

DB統合とストレージ要件

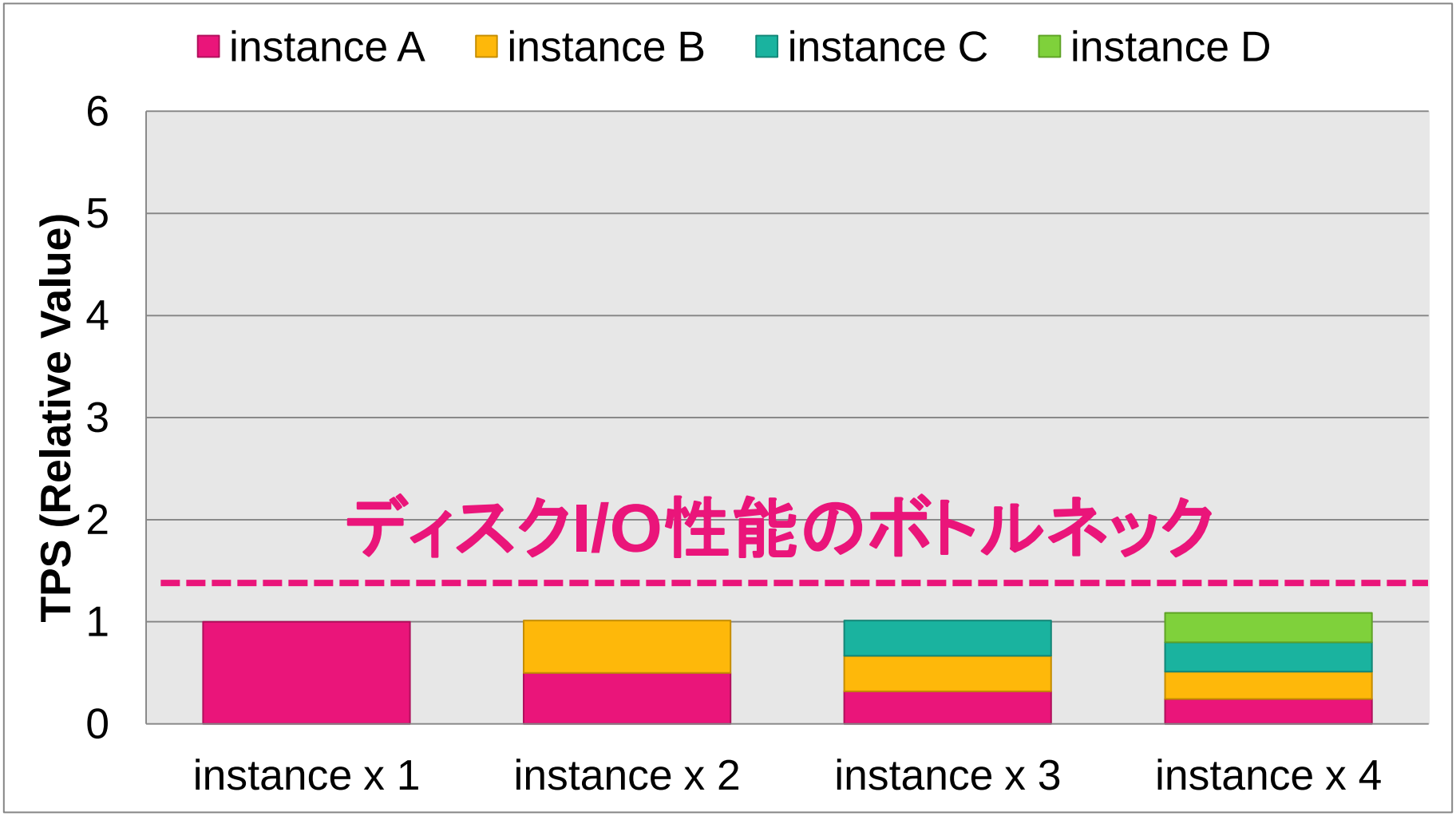
- CPUのマルチコア化によって集約密度が向上するか

→ 高密度集約を実現するカギはディスクI/O性能



※ 複数のOLTPシステムを統合した例

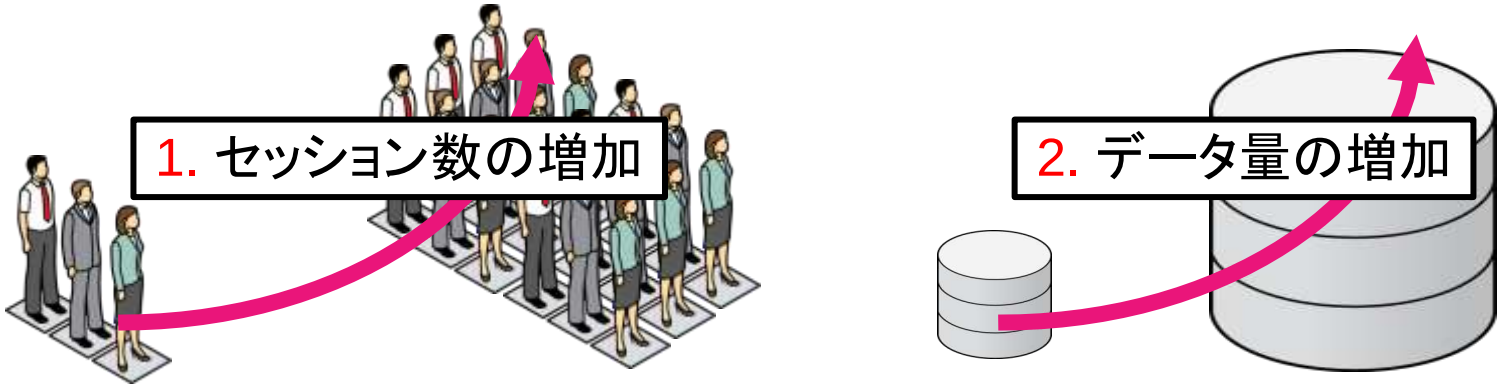
ディスクI/O性能を考慮せずにDB統合すると



ディスクI/O性能ボトルネックを改善するには

OLTPでは、ディスクI/Oを発生させないのが理想だが・・・

- 近年、同時に処理するデータ量が飛躍的に増加している



- メモリの追加（Database Buffer Cacheサイズを増加）
  - 高密度なDRAMは高価
  - サーバのスロット数には限りがある
  - 統合環境の場合、割当可能なサイズが減少

→ メモリ上だけで処理するのが困難

ディスクI/O性能ボトルネックを改善するには
ディスクI/Oそのものの性能向上させる

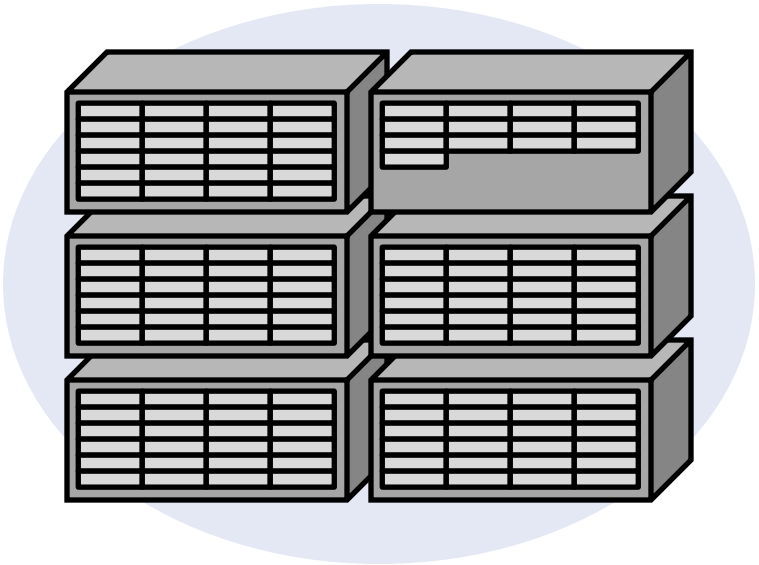
- HDDの追加
  - 設置スペース、重量、消費電力の問題
- SSDという選択
  - SSDはその特性を正しく理解し、賢く使いこなすことが重要
  - small random readのI/Oワークロードにおいて、

    HDDの**20～30倍**高速

【例】40,000 IOPSを達成するには

small random readの場合

- HDDの場合
  → HDD x 133

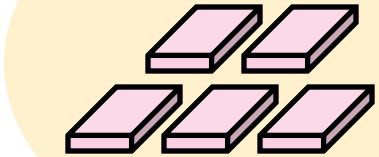


- SSDであれば
  → SSD x 5

【価格性能比】

H/Wコストは約**1/10**

消費電力は約**1/80**

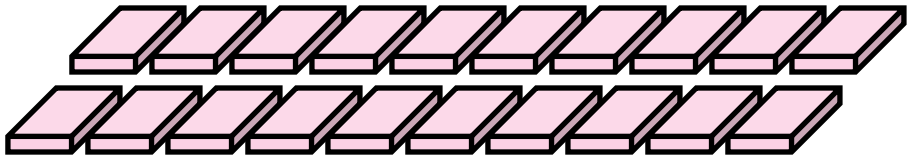


【例】高速なSSD上に、1TBのデータを配置するには

- SSD上に全てのデータを配置するのは高コスト

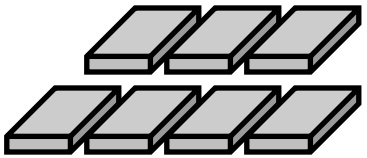
→【価格容量比】H/WコストはHDDの**10倍**

SSD x 25: 1TB



SSD（100GB） x 25（RAID1+0）

HDD x 7: 1TB



HDD（300GB） x 7（RAID1+0）

- アクセス頻度の高いデータのみをSSDに配置する

→ 運用に手間がかかる（アクセス頻度の分析、データの移動）

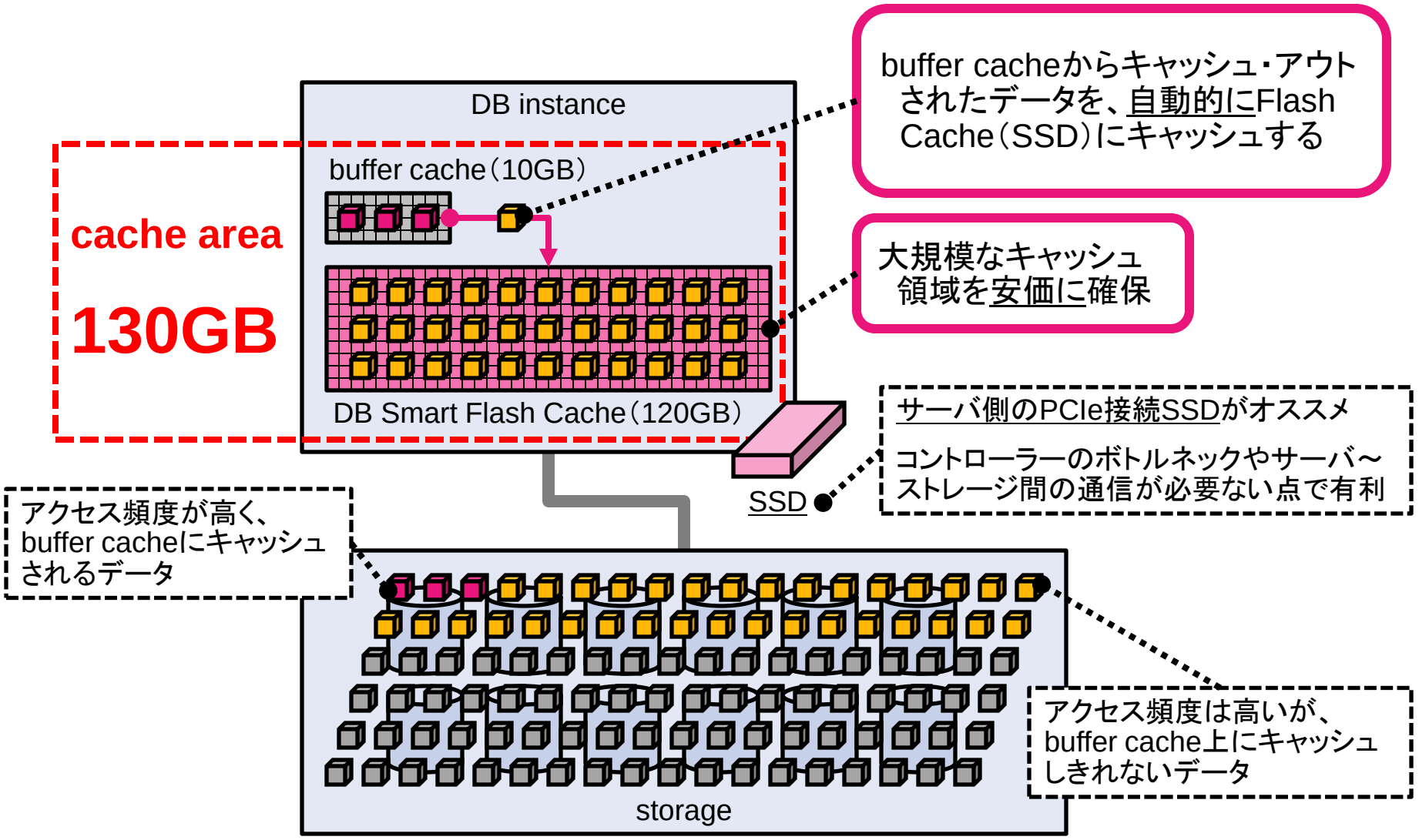
Oracle Databaseのアプローチ

- SSDをキャッシュとして活用する機能を実装

→ Database Smart Flash Cache（DB Smart Flash Cache）

  - Oracle Database 11g Release 2 Enterprise Editionの標準機能
  - LinuxとSolarisで使用可能

DB Smart Flash Cache

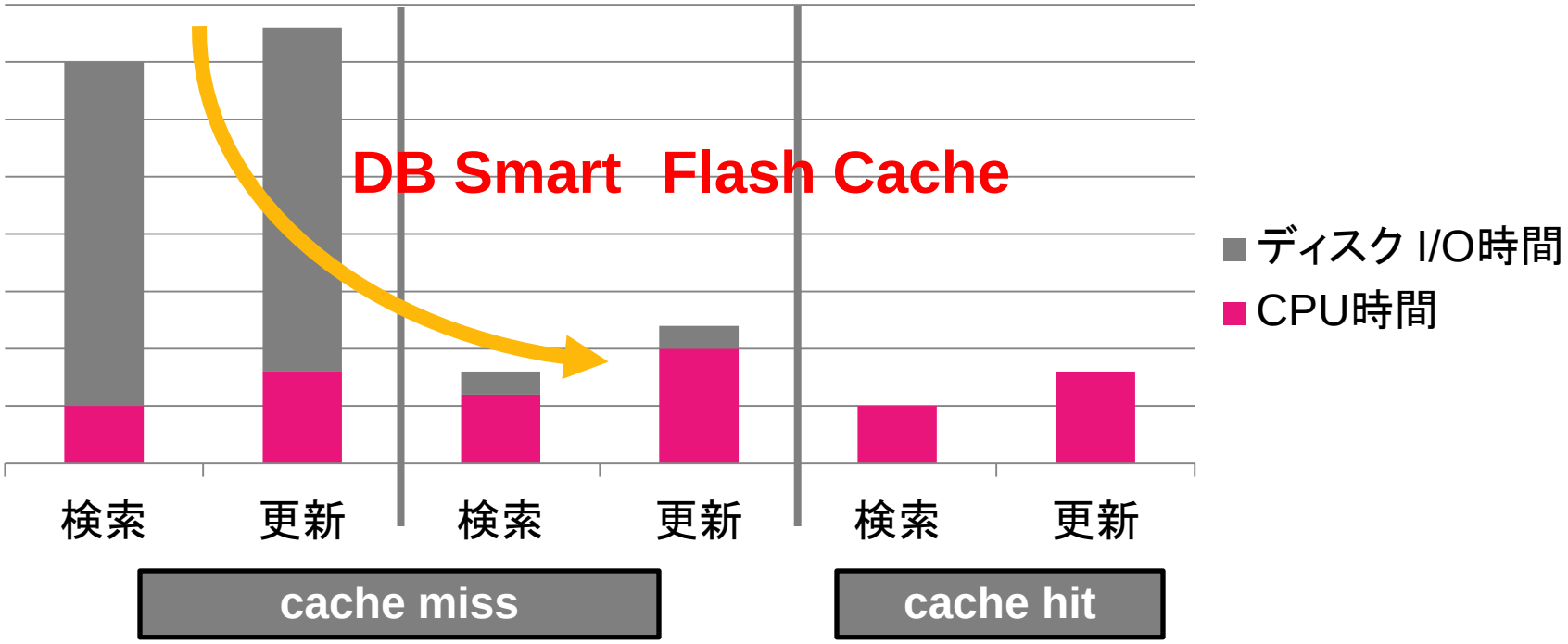


DB Smart Flash Cacheの効果

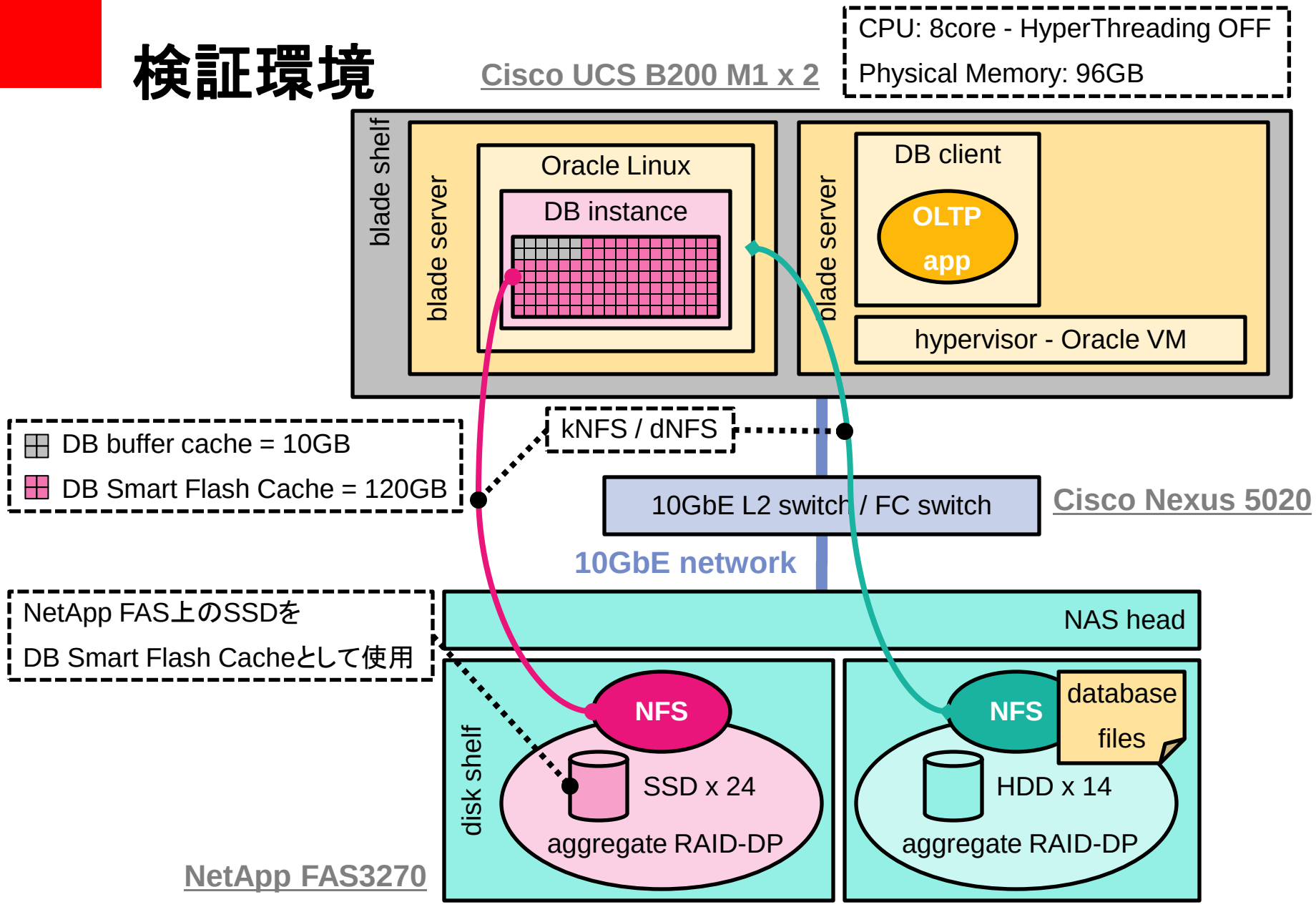
SQL処理時間の内訳イメージ

- DB buffer cacheでキャッシュ・ミスした場合でも、
  I/O待ち時間を大幅に削減

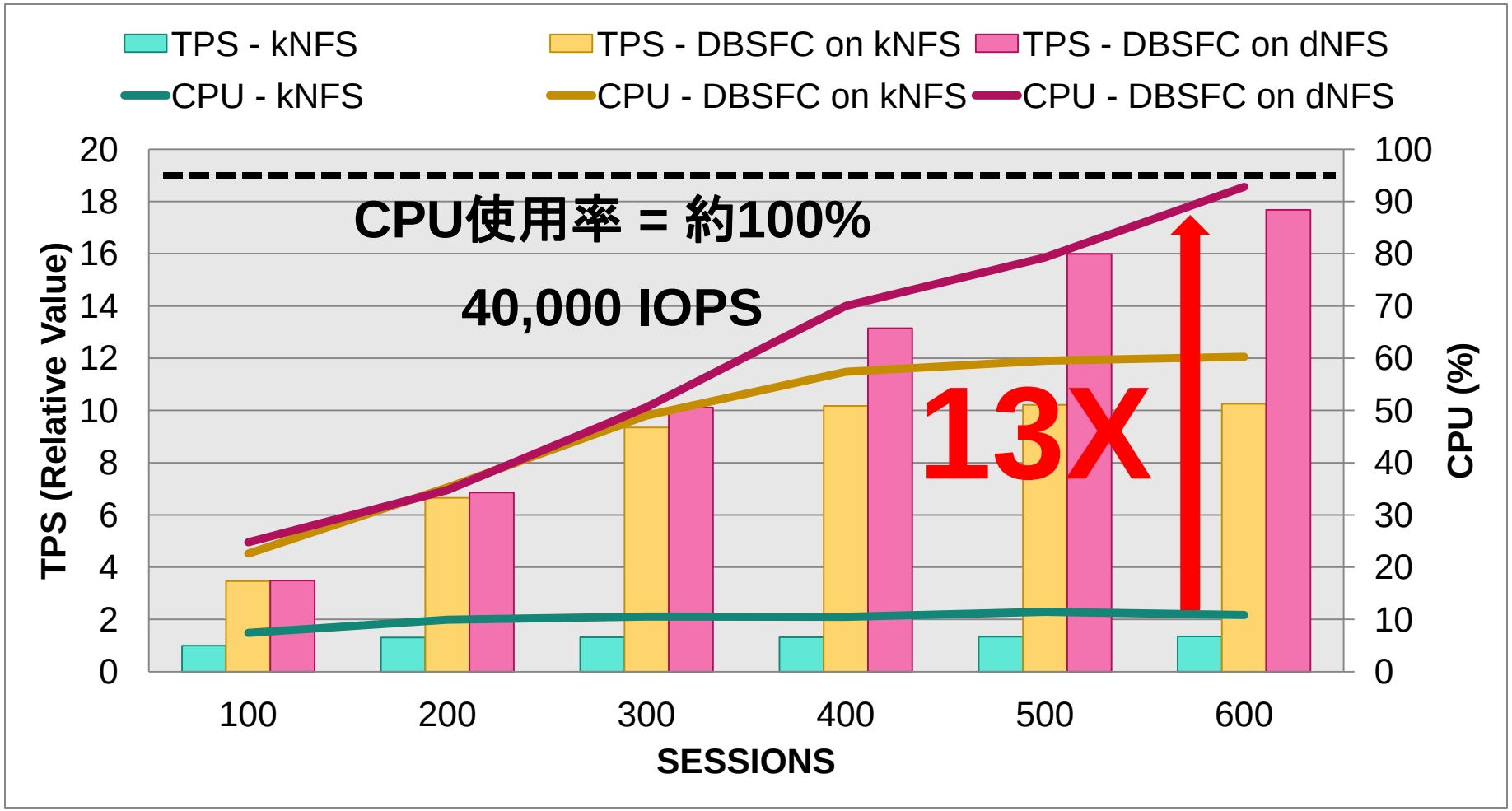
→ キャッシュ・ヒットした場合と同等のレスポンス・タイムを実現



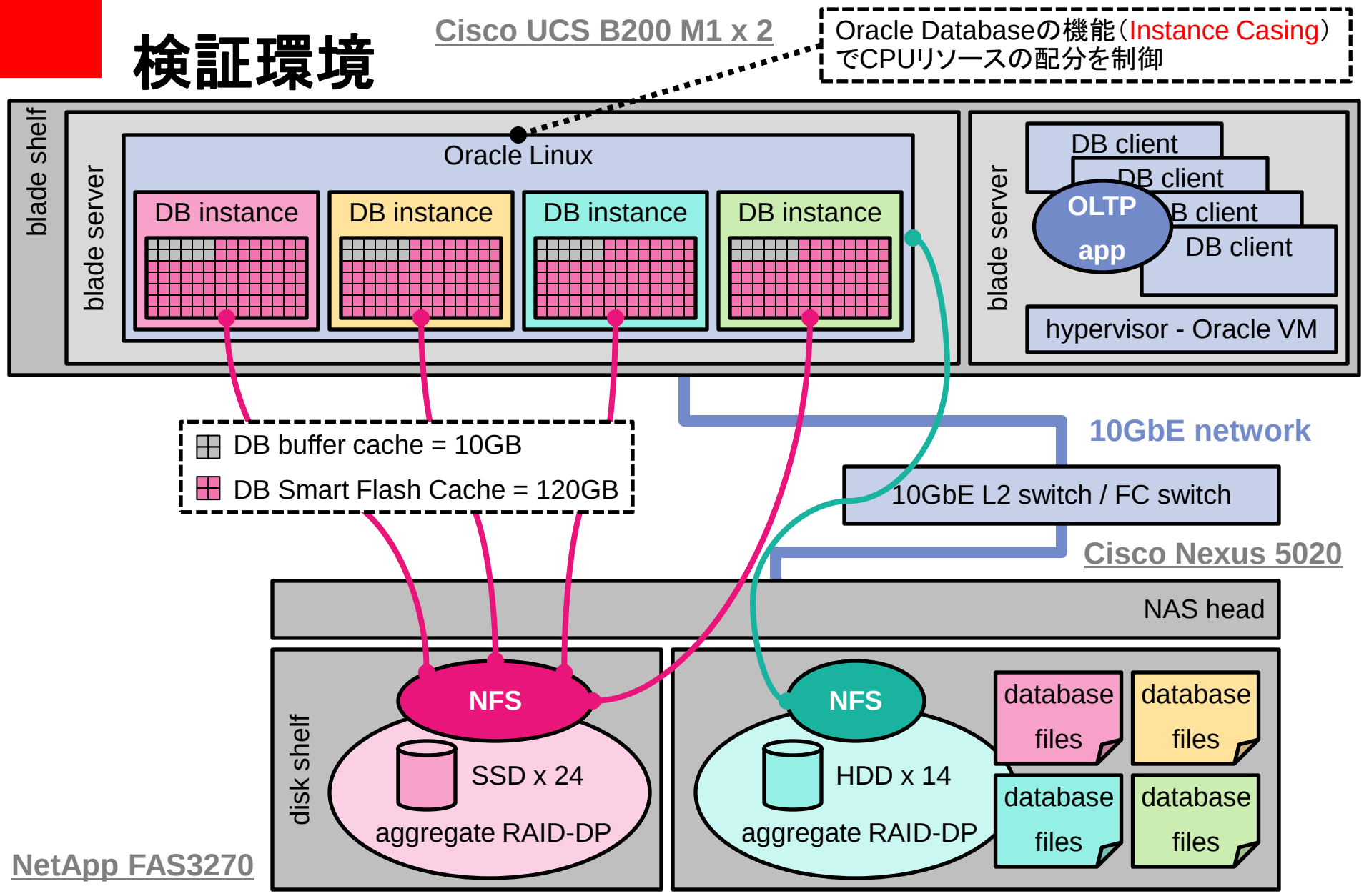
検証環境



DB Smart Flash Cacheの効果とdNFSとの組み合わせ

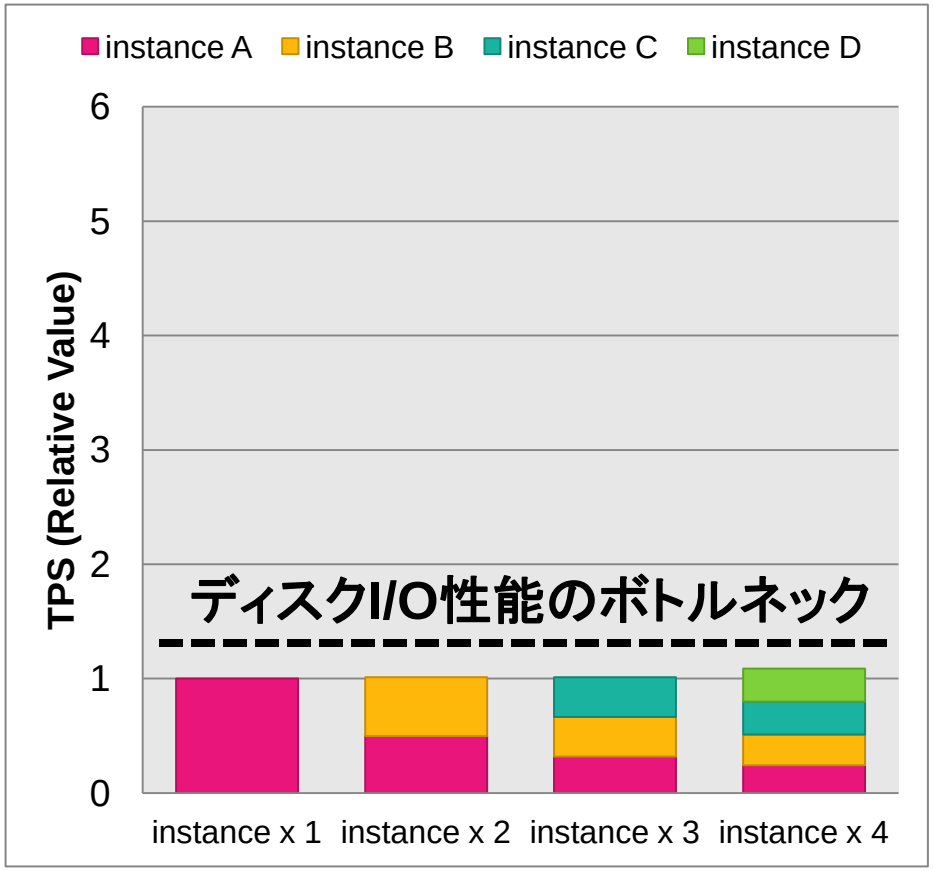


検証環境

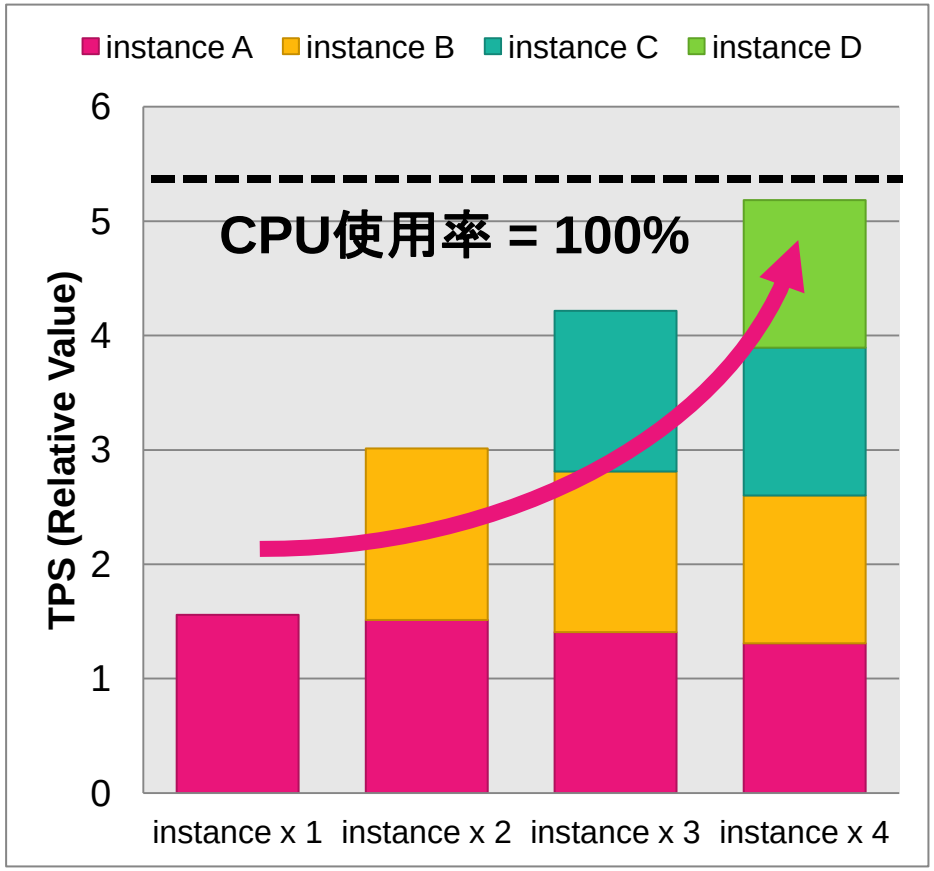


DB統合の集約密度を最大化

DB Smart Flash Cache with dNFS: OFF



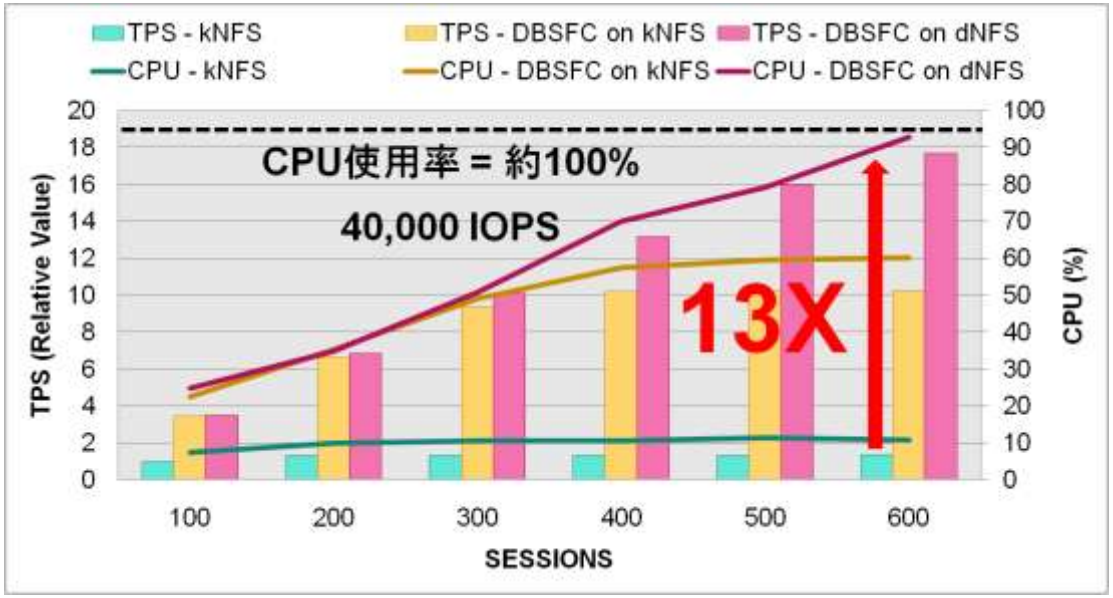
DB Smart Flash Cache with dNFS: ON

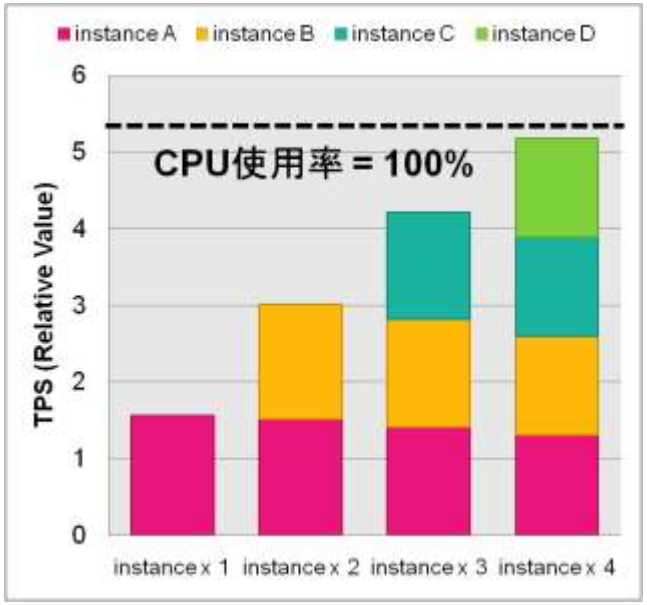


まとめ

DB Smart Flash CacheとdNFSによるDB統合

- 最小限のコストで最大限のI/O性能（IOPS）が得られる機能であり、集約密度を向上させたDB統合の実現を加速
  - DB Smart Flash CacheによりSSDを効率的に活用
  - dNFSによりH/Wリソース（マルチコア、SSD）を最大限活用
- これらを実現できるのはOracle Databaseだけ





Oracle GRID Center / oracletech.jp サイトにて公開中

http://www.oracle.co.jp/solutions/grid_center/netapp

http://oracletech.jp/products/pickup/000250.html

http://oracletech.jp/products/pickup/000262.html

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

ORACLE TECHNOLOGY NETWORK
OTNセミナー オンデマンド

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

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

ORACLE

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

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

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

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

技術コラム — SQLスクリプト、索引メンテナンスetc. 当たり前の運用/機能が見違える!?

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

オラクルエンジニア通信

oracletech.jp

ITエンジニアの皆様に向けて旬な情報を楽しくお届け

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。
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®