THE DOCUMENT COMPANY
FUJI XEROX

共有型高速ストレージ“GigaExpress”を使ったデータベースの高速化

富士ゼロックス株式会社
光システム事業開発部

アジェンダ

THE DOCUMENT COMPANY
FUJI XEROX

- ディスクI/Oの課題
- GigaExpress(半導体ディスク装置)の紹介
- データベースでのGigaExpress検証事例
- まとめ

ディスクI/Oの課題

THE DOCUMENT COMPANY
FUJI XEROX

- 企業が扱うデータが急増 (年率30%～70%)
- ＣＰＵの性能はムーアの法則に従い向上
  ⇔一方HDDは速度性能の向上スピードが遅い
  (パフォーマンス・ギャップ)

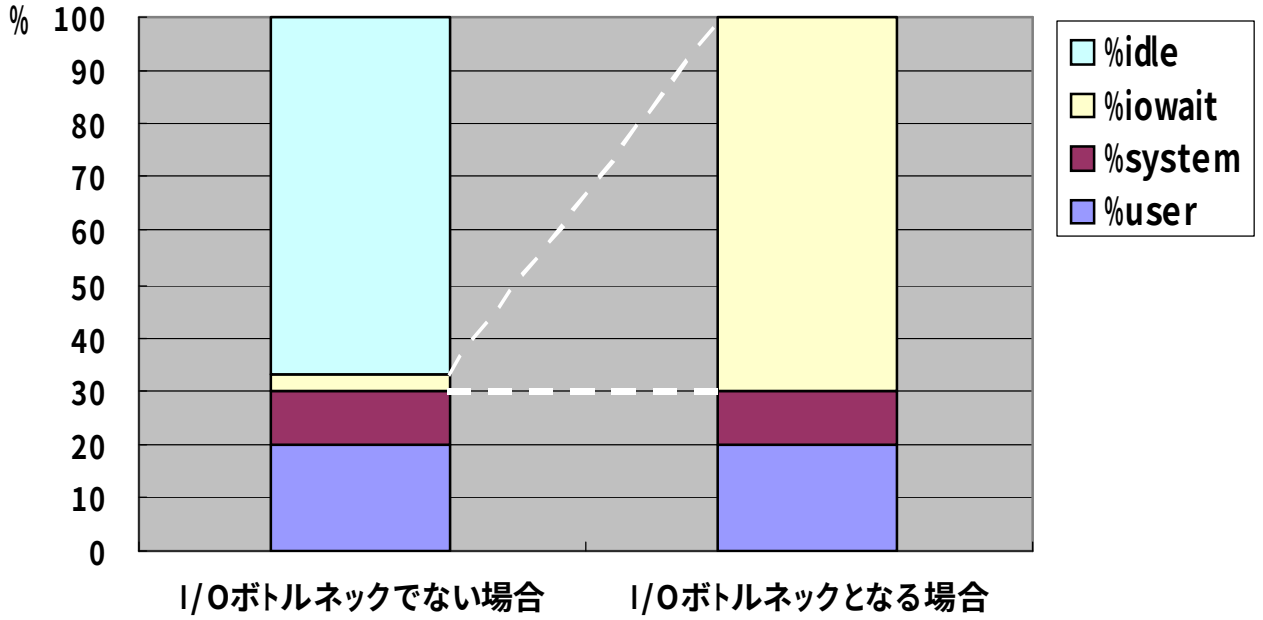
データが増えれば増えるほど、
CPUの性能が上がれば上がるほど、
ディスクI/Oがボトルネックとなる可能性が浮上

データベースシステムでディスクI/Oの課題が顕在化

ディスクI/Oボトルネック

THE DOCUMENT COMPANY
FUJI XEROX

- ディスクI/O量の確認方法
  - ① パフォーマンス監視ツールを利用する利用
  - ② OSに付属のツール (iostat) を利用する方法

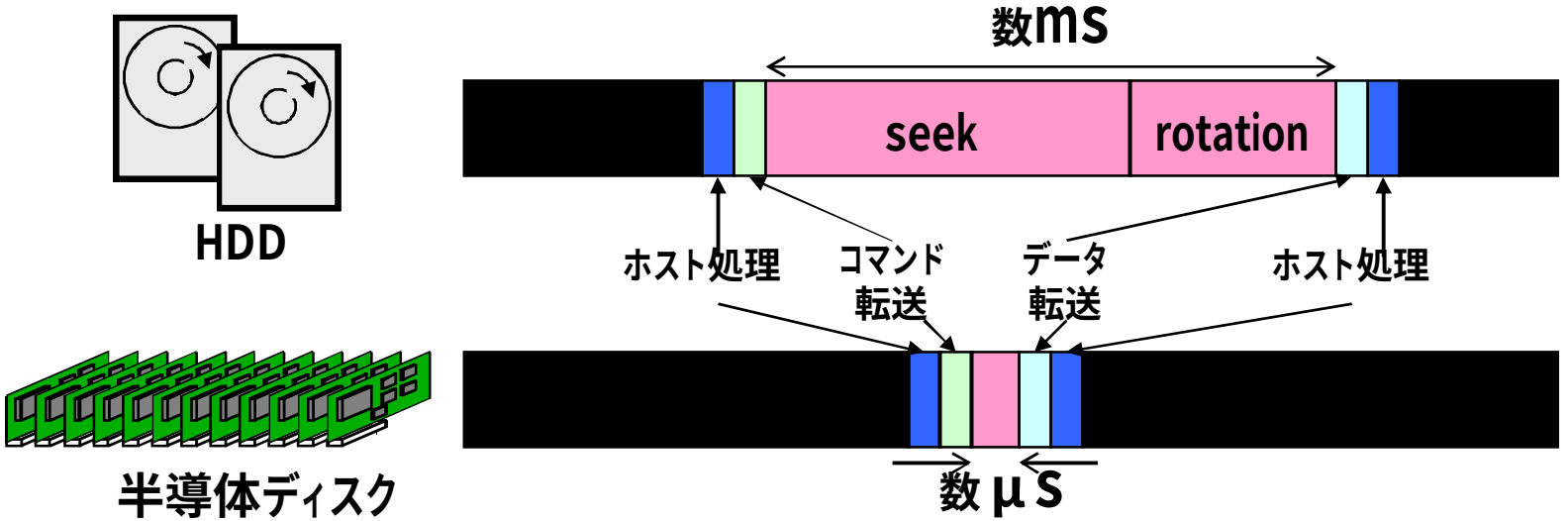


GigaExpress (半導体ディスク) が有効

GigaExpressとは①

THE DOCUMENT COMPANY
FUJI XEROX

- GigaExpress = 半導体ディスク装置
  - ✓HDD(磁気ディスク)はディスクアクセスにシーク待ち時間と回転待ち時間を要するため数msのディレイがかかり、特にランダムなデータアクセスにおいて遅延が長いというデメリットがある
  - ✓半導体ディスクは記憶媒体にメモリーを使用するため、アクセス遅延をHDDに比べて二桁向上できる



GigaExpressとは②

THE DOCUMENT COMPANY
FUJI XEROX

- コストパフォーマンスに優れる半導体ディスク装置が登場
  - ✓メモリー価格の低下により、現実的な価格で装置をご提供可能、2～3年前と比較して装置価格は1/3～1/5
  - ✓メモリーの高速・高密度化や高速なデータ伝送技術により装置性能もアップ
- データの高速処理に対する要求にマッチするソリューション
  - ✓HDDと使い分けることによって、コストパフォーマンスに優れたシステム構築が可能

GigaExpressの主な特長①

THE DOCUMENT COMPANY
FUJI XEROX

➢DRAMを使うことで高速なデータアクセスを実現

- ✓ GigaExpressはDRAMベースの半導体ディスク
- ✓ メモリーバスに当社独自の光伝送技術を採用しており、**1ポートでIOPS値80,000***という高速なデータアクセスを実現

* ブロックサイズ512Byte、ランダムリード100%での実効値（実際に使用した場合の値）。
ただし、サーバー性能に依存します。

➢PCI Express ×4による高速転送

- ✓ 10GbpsのスピードをもつPCI Express ×4を外部I/Fに採用し、**1ポートで650MB/s***の高速転送を可能
- ✓ 付属のホストカードをホストサーバーのPCI Expressスロットに挿入し、本装置との間を付属のケーブル接続するだけで設置可能

* ブロックサイズ32KB、ランダムリード100%での実効値（実際に使用した場合の値）。
ただし、サーバー性能に依存します。

GigaExpressの主な特長②

THE DOCUMENT COMPANY
FUJI XEROX

➢ラックマウントタイプ

- ✓1ポートタイプ(xxG-1P)は8/16/32GBのメモリー容量をEIA規格の1Uラックマウントサイズ*で実現
- ✓2ポートタイプ(xxG-2P)は2Uラックマウントサイズで冗長電源を備え、より信頼性の高いモデル

* 1Uラックマウントサイズとは、ラックに取り付けるユニットの単位で、高さ1.75インチ。この規格の最小単位になります。

➢メモリー容量の拡張が容易

- ✓複数のPCI Expressのポートをもつサーバーに装置単位でメモリー容量を拡張可能
- ✓1つのボリュームにすることも、パーティションに分けての利用も可能

➢サーバーシステムへの導入が容易

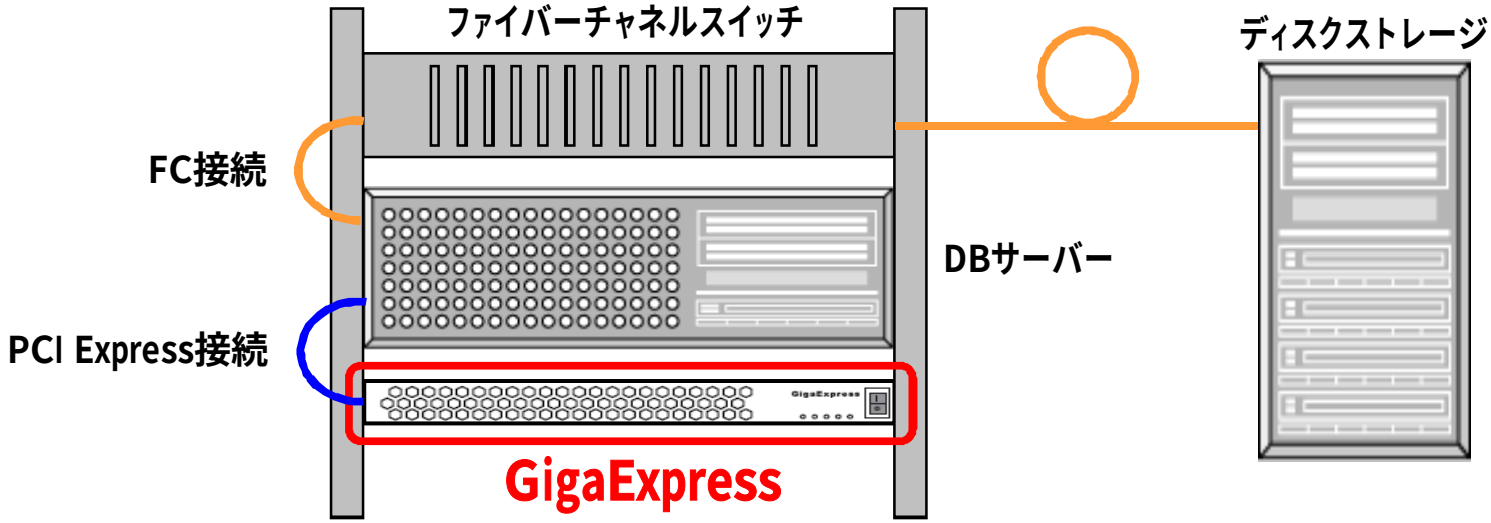
システム構成例①

THE DOCUMENT COMPANY
FUJI XEROX

➢ 例1：1ポートタイプを利用したホットファイル格納構成

✓アクセス頻度の高いデータをGigaExpressに配置し、それ以外のデータを従来のディスクストレージに配置するハイブリッド構成により最適なシステムを構築可能

例1:1ポートタイプ



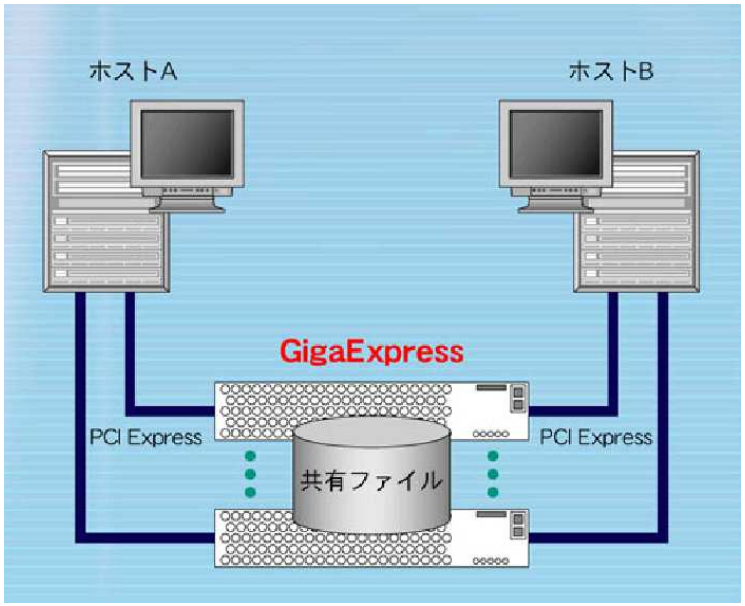
システム構成例②

THE DOCUMENT COMPANY
FUJI XEROX

➢例2：2ポートタイプを利用した共有ストレージ構成

- ✓2台のホスト(サーバー) で共有することが可能で、より信頼性の高い冗長構成を実現
- ✓実績のあるクラスターソフトウェアを利用することでフェールオーバー構成や、OracleのRAC構成も構築可能

例2:2ポートタイプ



PostgreSQLベンチマーク①

THE DOCUMENT COMPANY
FUJI XEROX

➢pgbench試験環境

✓pgbenchのスケーリングファクターは300
データサイズは約4.5GB (3,000万件)

✓データをサーバーのHDDに配置した場合とGigaExpressに配置した場合で1秒間に処理するトランザクションの数（tps）を比較

✓システム環境は以下の通り

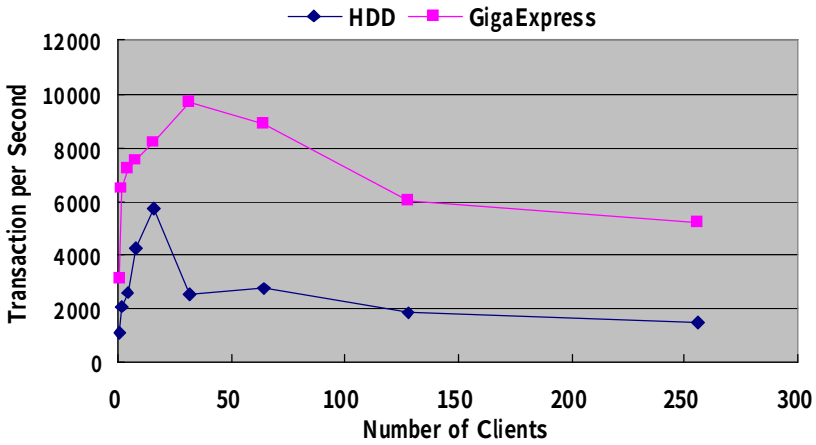
CPU	Xeon 5160 (3GHz×2)デュアルコア　× 2
メモリー	FB-DIMM 8GB
HDD	2.5" SAS 73GB　× 3 (RAID5)
OS	MIRACLE Linux V4.0 (32bit)
PostgreSQL	バージョン8.1

PostgreSQLベンチマーク②

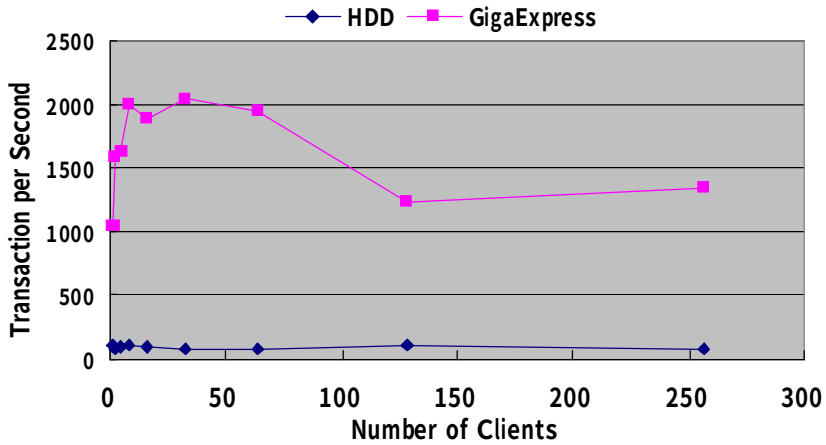
THE DOCUMENT COMPANY
FUJI XEROX

➢pgbench試験結果

✓SELECTの結果、TPC-Bの結果は以下の通り



selectのtps比較



TPC-Bのtps比較

✓データをGigaExpressに配置した場合の方がtpsははるかに大きい値となり、SELECTだけではなくUPDATEも含まれる複合トランザクションの場合においてGigaExpressの効果は特に大きい

PostgreSQLベンチマーク③

THE DOCUMENT COMPANY
FUJI XEROX

➢DBT-1試験環境

- ✓Webベースのトランザクション性能をベンチマーク
- ✓データベースのデータ量は約10GB程度
- ✓エミュレートするユーザ数 (以降eu) を800、1000、1200および1400と変化させ、それぞれにおいてデータベースのデータをローカルディスクに配置した場合とGigaExpressに配置した場合で各インタラクションの平均応答時間を測定
- ✓システム環境は以下の通り

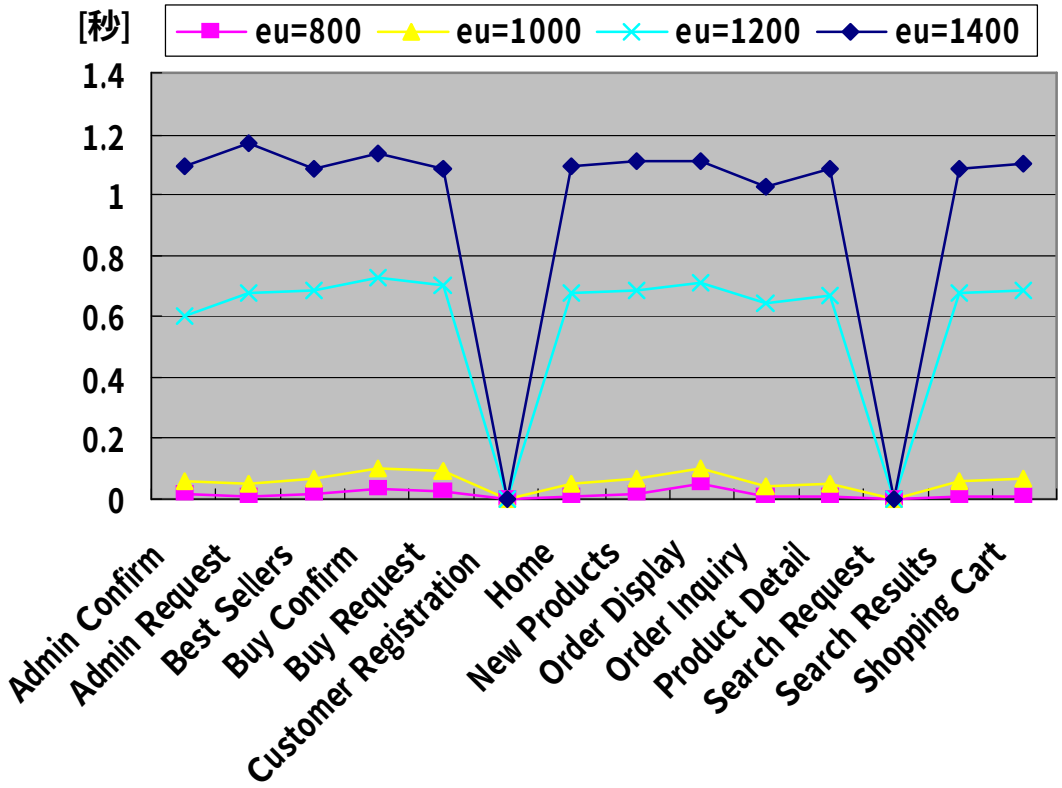
CPU	Xeon 5355(2.66GHz×2)クワッドコア　× 2
メモリー	FB-DIMM 4GB
HDD	2.5" SAS 73GB　× 4 (RAID5)
OS	MIRACLE Linux V4.0 (64bit)
PostgreSQL	バージョン8.0.6

PostgreSQLベンチマーク④

THE DOCUMENT COMPANY
FUJI XEROX

➢DBT-1試験環境

- ✓いずれのeuにおいてもGigaExpressにデータベースのデータを配置した場合の平均応答時間はローカルディスクに配置した場合より短縮
- ✓GigaExpress使用時とローカルディスク使用時の平均応答時間の差を各インタラクションごとに示す



半導体ディスクの効果的な使い方

THE DOCUMENT COMPANY
FUJI XEROX

半導体ディスクの使い方		メリット/デメリット
①	Webサーバー/APサーバーに接続し、キャッシュとして機能させる	◇DBにアクセスがいかないため、DBサーバーの負荷を軽減できる ◆Webサーバー/APサーバーの台数が多い場合はコスト増
②	DBデータ全てを入れる	◇キャッシュヒット率が低い場合でもDBのレスポンスが高速化できる ◆データ量によっては大容量の半導体ディスクが必要でコスト増
③	DBデータの一部 (ホットスポット) だけを入れる	◇コストを抑えながらDBのレスポンスを向上できる ◆ホットスポットをダイナミックにディスクに入れる仕組みが必要
④	Oracleであれば、REDOログ・ファイルだけを入れる	◇小容量の半導体ディスクの利用で性能の効果が見込める ◇複雑な運用の仕組みが不要
⑤	Oracleであれば、ASM (Auto Storage Management) に管理させる	◇全データを半導体ディスクに入れなくても性能の効果が見込める ◇複雑な運用の仕組みが不要

Oracleベンチマーク①

THE DOCUMENT COMPANY
FUJI XEROX

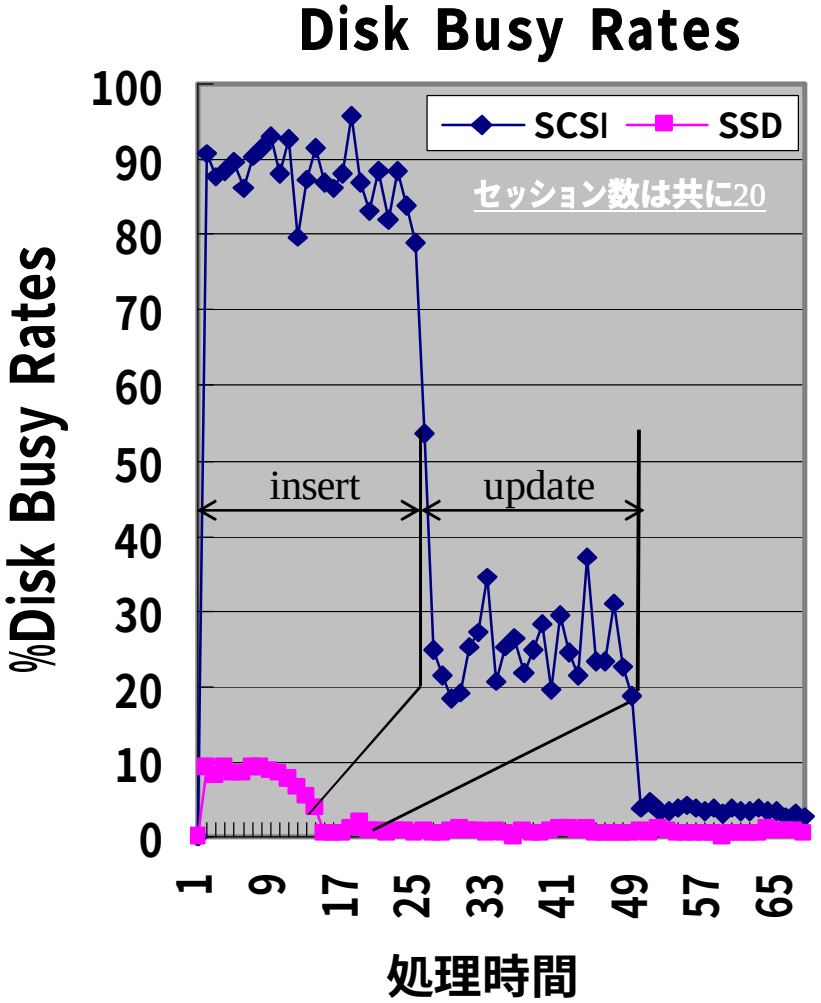
➢REDOログファイルの格納先別性能比較

◆評価条件

- ✓ SCSIディスクのDisk Busy Ratesが100%近くになるような大量バッチ処理を想定したスクリプトを実行し比較
- ✓ 環境
  Xeon 3.6GHz x 2 (with HT)
  メモリー　8GB
  Adaptec 39320A SCSIカード
  SCSI 15,000rpm x 2 (RAID0)

◆評価結果

- ✓ 上記検証において、GigaExpressを使用した場合、Disk Busy Ratesが低く、処理時間も大幅に短縮された



Oracleベンチマーク②

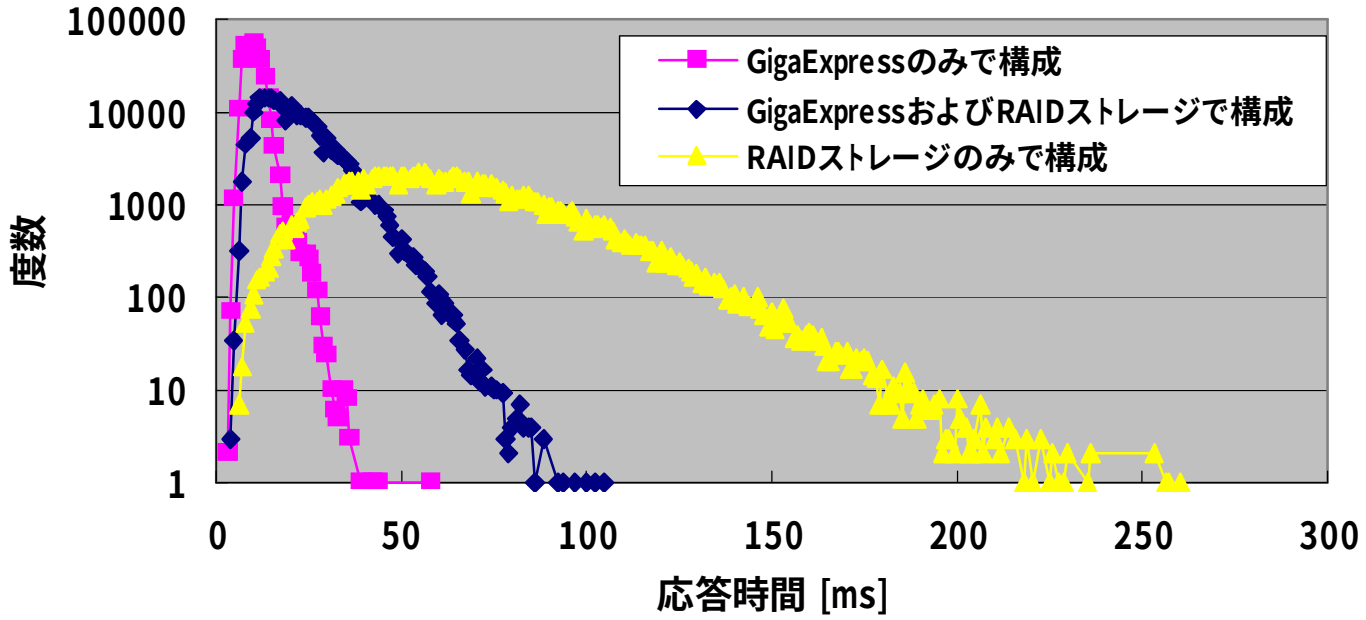
THE DOCUMENT COMPANY
FUJI XEROX

➢ASM (Auto Storage Management)

✓ASMを利用しGigaExpressと通常のRAIDストレージを併用した場合の効果を検証

使用するディスクの構成をGigaExpress (16GB×2) のみで構成した場合、GigaExpress (16GB×2) とRAIDストレージ (30GB×2) で構成した場合およびRAIDストレージ (30GB×2) のみで構成した場合の、Query※応答時間分布の比較 (データサイズは約30GB)

※10のSELECTをまとめて1トランザクションとした

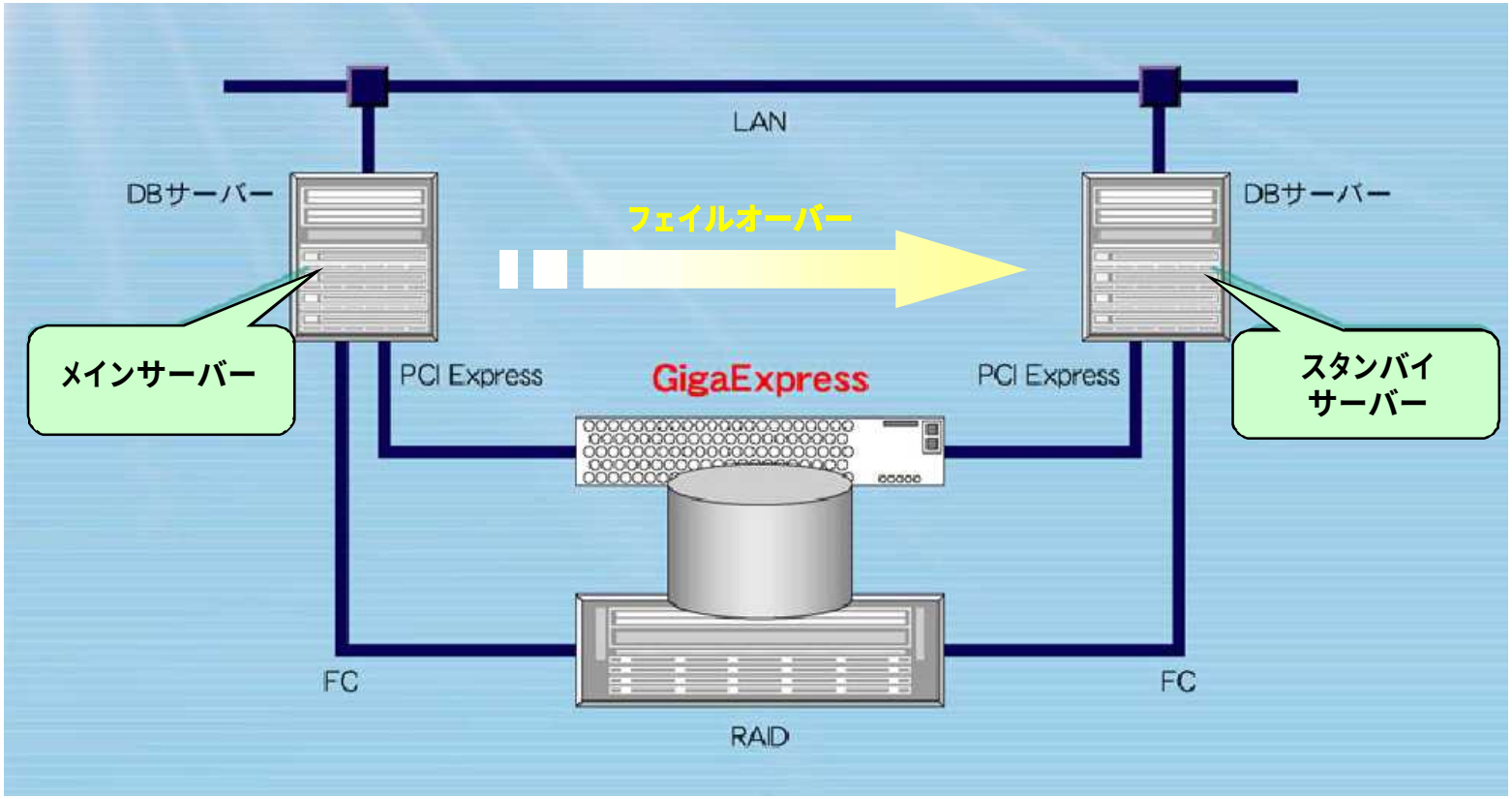


データベースシステムの可用性検証

THE DOCUMENT COMPANY
FUJI XEROX

➢商用ソフトを利用した高可用構成

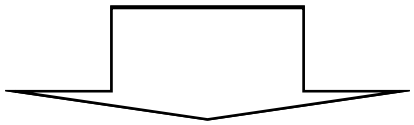
✓実績のあるストレージ管理ソフトウェアやクラスタソフトウェアを利用してシステム構築できることを随時検証中



まとめ

THE DOCUMENT COMPANY
FUJI XEROX

- 高速なディスクI/Oを実現するGigaExpress (半導体ディスク装置) をデータベースシステムへ適用した場合の効果検証例を示しました
- ディスクI/Oがボトルネックとなる状況であれば性能面で大幅な効果が期待できることを示しました
- 実績のあるソフトウェアと組み合わせることで性能面に加え信頼性面でも高いシステムを構築することができると考えています



HDDとGigaExpressをうまく使い分けることで
コストパフォーマンスの高いシステム構築が可能

お問合せ先

THE DOCUMENT COMPANY
FUJI XEROX

装置の無償評価貸出を行っています。
ご要望、ご質問がございましたら遠慮なくご連絡ください。

富士ゼロックス株式会社
光システム事業開発部
お電話　：　０４６５－８０－２１２１
ＵＲＬ　　：
http://www.fujixerox.co.jp/product/gigaexpress/

E-mailでのお問い合わせは上記URLのメニュー「ご相談窓口」からお願いします