



# ビッグデータ時代を乗り切るための OSS + サーバー技術のご紹介

2012年3月16日

日本アイ・ビー・エム株式会社

織 学 (orim@jp.ibm.com)

石川 公基 (kixs@jp.ibm.com)



## はじめに

- この資料の内容に関しては正式なIBM のテストを受けておりません。この資料は、資料作成時における最新情報をご参考のために提供することを目的として記載されており、IBMは、情報の正確性、完全性または有用性について何ら保証するものではありません。また、内容は予告なしに変更または更新されることがあります。
- この資料の内容は、限られた検証環境における結果に基づくものであり、全ての環境で同一の結果を保証するものではありません。お客様固有の環境に対し、適切であるかどうか、また、正確であるかどうかは十分検証されていません。この資料の情報に基づき導入・設定を実施される場合には、十分な検証テストを行ってください。また、予め製品のマニュアルおよびディストリビューターが提供する情報をご覧ください。
- この資料の情報に基づいて導入・設定・運用した結果について、IBMはいかなる保証も責任も負いかねますので予めご了承ください。

当資料をコピー等で複製することは、日本アイ・ビー・エム株式会社および執筆者の承認なしではできません。

## ビッグデータ時代の到来

1.8 ZB = 1,800,000,000 GB

2011年に全世界で生成されたデータ量

データ量 50倍, ファイル数 75倍

2020年のデータセンターについての2011年時点での予測

90%以上が非構造化データ

テキスト、画像、音楽、映像、センサー情報…etc.

出典 : IDC, 2011 Digital Universe Study : Extracting Value from Chaos, Jun 2011  
IDC, 2010 Digital Universe Study : A Digital Universe Decade ? Are You Ready?, May 2010

## Watsonをご存知ですか？

2011年2月16日（米国時間）、IBMリサーチの4年間にわたる研究成果である質問応答システム「Watson」が、米国の人気クイズ番組「Jeopardy!」（以下「ジヨパディ！」）に挑戦しました。

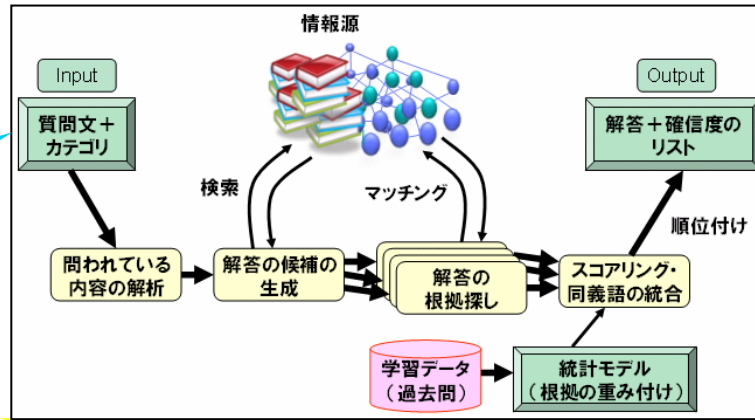


**Watsonとは、IBMの研究部門が開発した質問応答システムです。**

問題(文)の内容を分析し、事前に収集された大量のテキスト情報から問題の解答候補とその根拠・確信度を計算し、高い確信度の候補が得られた場合に解答する、という一連の知的処理を高速に実行するコンピューター・システム

# WatsonはPower Systems上のLinuxで稼動しています

前処理の  
アノテーション付与に  
UIMAとHadoopを  
活用しています



- 下流プロセスになるほど多くの計算が必要
- 各解答候補の信頼度を並列に計算

Watsonは  
Power Systems上の  
Linuxで稼動しています

- 高い並列性とCPU高負荷の計算実行
- 高速のテキスト分析処理を支えるメモリ機構
- 対戦時の安定性
- 商用システムをそのまま利用



- 各ノードは4個のPOWER7チップから構成され、各チップは8個のCPUコアを含む → 合計2,880コア
- 2~3秒での応答を実現



# ビッグデータ/テキストデータの活用をもっと簡単に

## ■ IBM® InfoSphere BigInsights

Apache Hadoop を使いやすくパッケージングしたソフトウェア製品

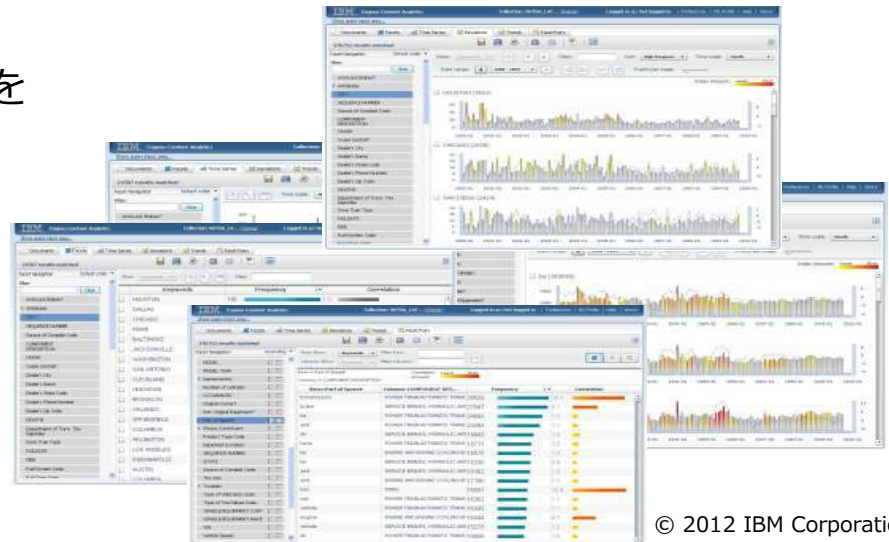
- 管理GUI 対応
- スプレッドシート型分析インターフェース
- アプリケーション開発用言語追加
- …etc.



## ■ IBM Content Analytics

非構造化情報からの収集・分類・分析を一貫して提供するソフトウェア製品

- UIMA対応
- 日本語構文解析対応
- 大容量データ対応
- レポートツールとの連携
- ..etc.



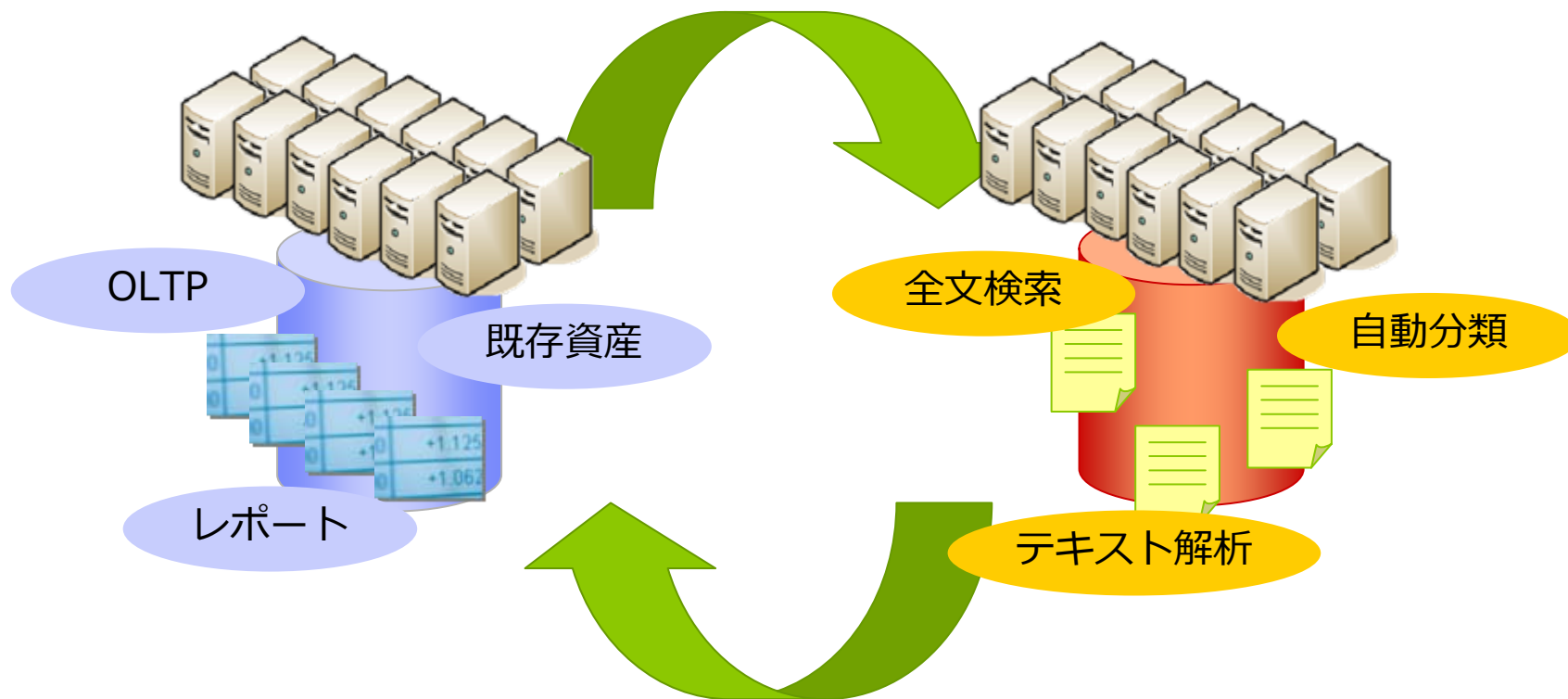
# とはいえ使い分けが重要です

## ■ リレーショナル・データベース

- ✓ 構造化データの保持
- ✓ 豊富なトランザクション保護機能
- ✓ 既存のソフトウェア資産も多数あります

## ■ New Technology Trends

- ✓ 非構造化データの保持
- ✓ データの解析・分析処理
- ✓ 新規開発これまでになかったようなアプリケーション



## OSS-DB をハードウェアでもっと使いやすくする

- 最近 IBM ではPostgreSQLに関連したこんな活動をしています
  - ✓ 高速高可用性PostgreSQL折紙付構成の共同発表
  - ✓ PostgreSQL 9.2のPPC64プラットフォームでのスケーラビリティ向上

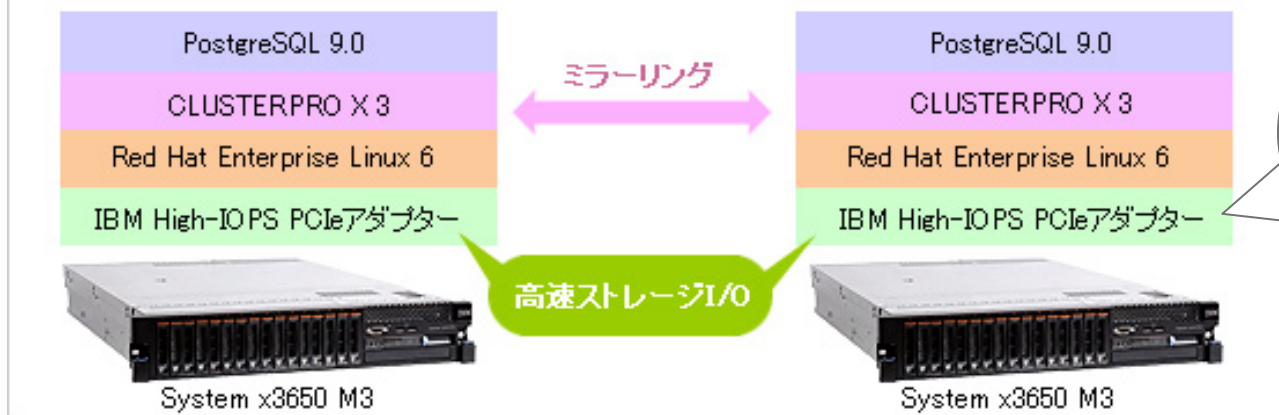


## 高速高可用性PostgreSQL折紙付構成 とは

- パフォーマンスと可用性を両立させたデータベースソリューション
- SRAOSS Inc.、NEC、Red Hat、Fusion-io、IBM 共同検証



ソリューション構成 概略図

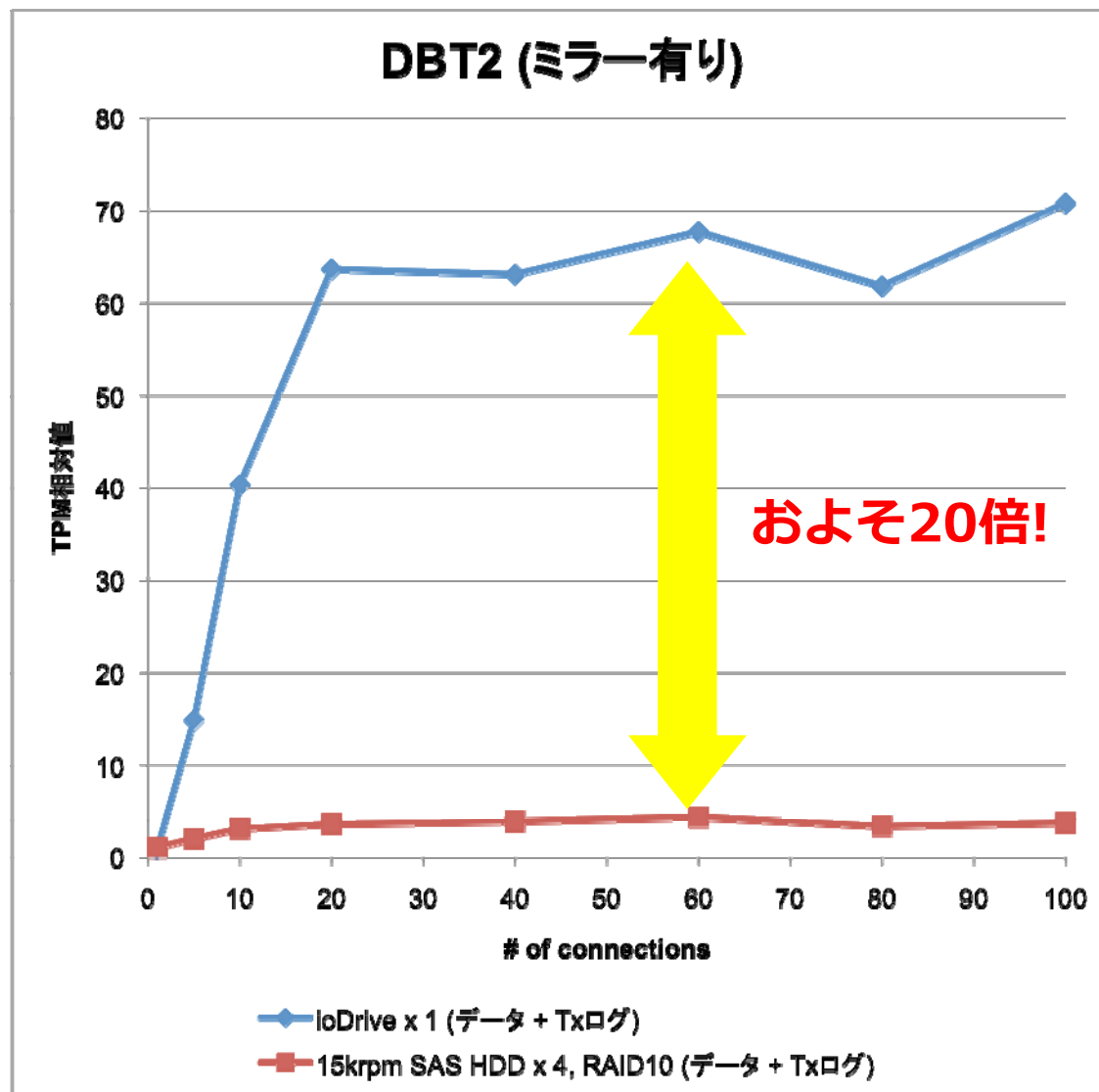


フラッシュメモリ  
ストレージ

<http://www.ibm.com/systems/jp/x/solution/postgres/>

## 高速化の結果

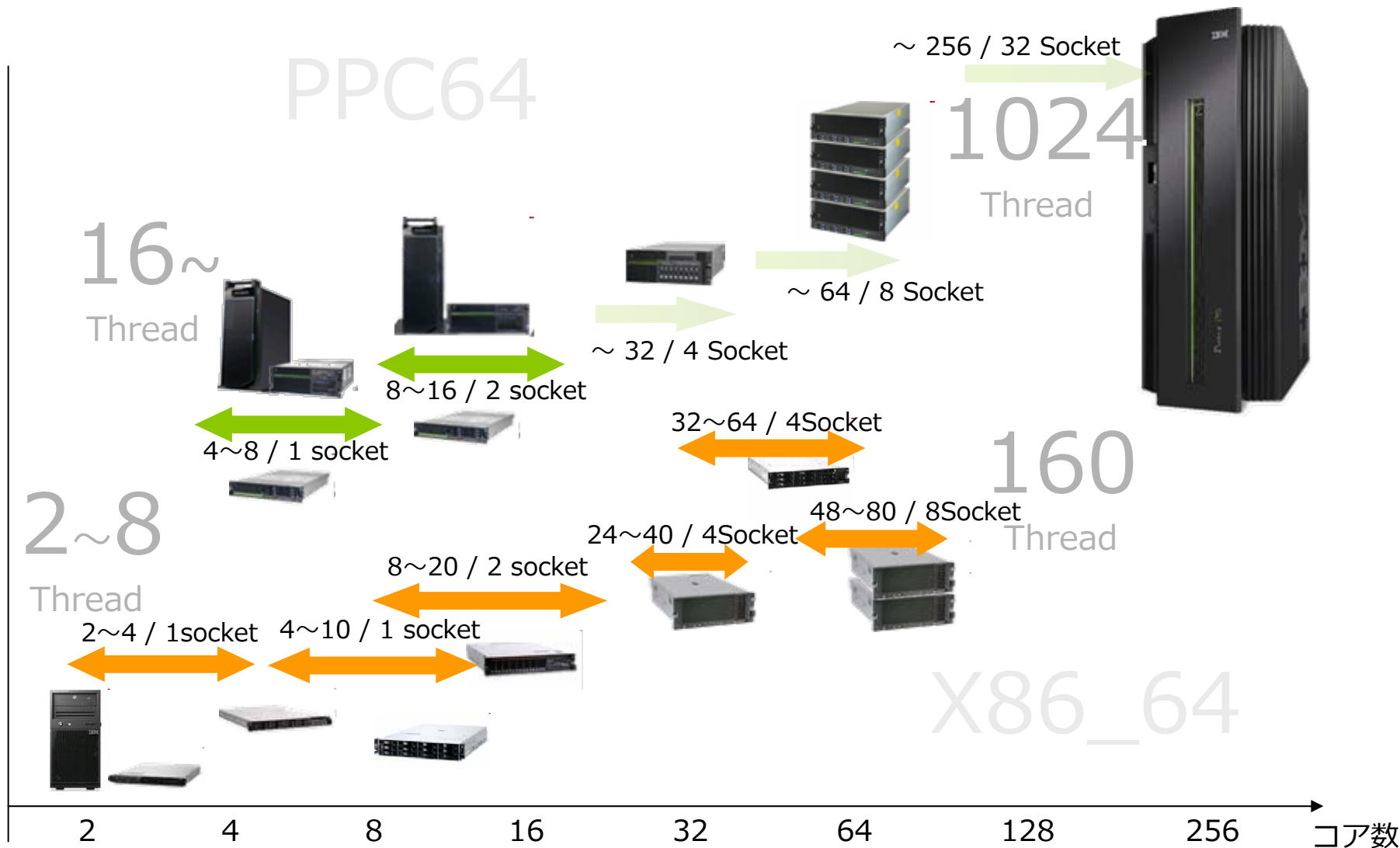
- RAID10 のおよそ 20 倍！
- 縦軸は トランザクション数/分
- 数値は 相対値  
RAID10・ミラーリング有り x 1  
接続時の TPM を 1
- ディスク・ミラーリング有り  
CLUSTERPROによる  
ネットワーク経由のミラー



# PostgreSQL 9.2 ではスケーラビリティがHot Topic です

\* この資料は配布しません

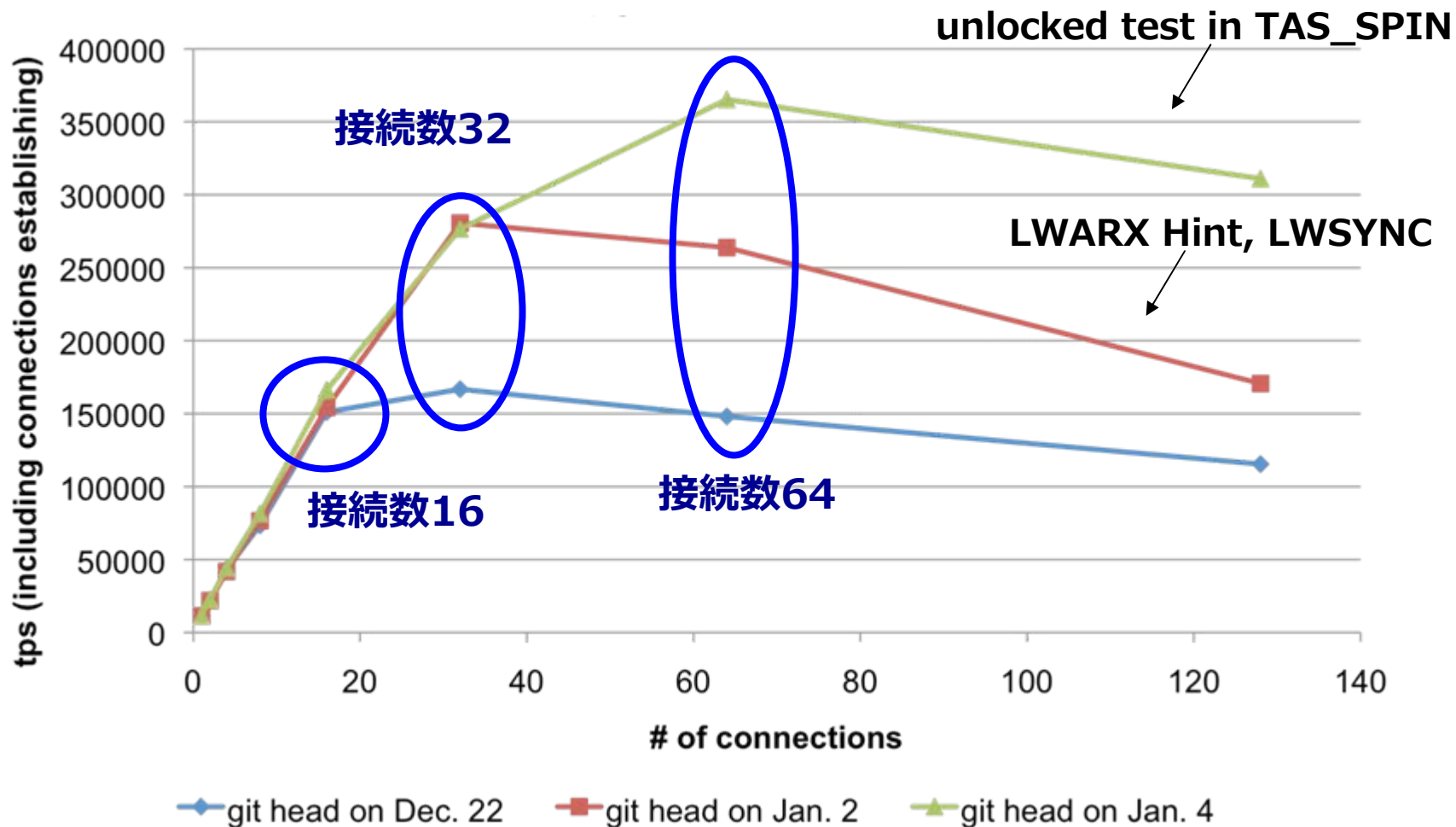
# 最近のハードウェアのSMPスケールビリティの例



出典 : <http://www.ibm.com/systems/jp/x/product/> & <http://www.ibm.com/systems/jp/power/hardware/>

# PPC64でもちゃんとスケールします (するように改良しました)

pgbench -S , scale factor 100, 500 sec, 32-core POWER7 on Power 750 (128 threads)  
max\_connections = 128, shared\_buffer = 8GB



# PPC64/PowerPC/POWERとは

- **Apple、IBM、Motorola が開発した RISC プロセッサー**
  - ✓ IBM製のサーバー向けは POWER と呼んでいます
- **有名なところでは、一時期 Mac に搭載されていました**
  - ✓ Power Mac G3, G4, G5
- **利用形態**
  - ✓ スパコン (Top500、Green500)
  - ✓ 組み込み系
  - ✓ ゲーム機にも採用されています
    - Pippin@, Wii, Xbox, PS3
  - ✓ 火星探査機のコンピュータ



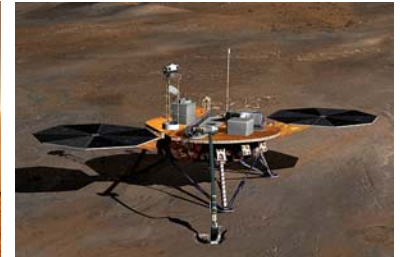
Blue Gene



Pathfinder



Spirit



Phoenix



## まとめ

- 大量データの処理にもOSSが活躍
  - ✓ 商用製品とのハイブリッドという選択も
  - ✓ サポートサービスもご活用ください
  
- データや処理の特性によって技術の使い分けは必要
  - ✓ そのアプリケーション、書き直しますか?
  
- ハードウェアもこんなに進化しています
  - ✓ フラッシュメモリを活用した高速OSS-DBソリューションの例
  - ✓ 大規模SMP環境でもスケールするOSS-DB の例

© IBM Corporation 2012. All Rights Reserved.

ワークショップ、セッション、および資料は、IBMまたはセッション発表者によって準備され、それぞれ独自の見解を反映したものです。それらは情報提供の目的のみで提供されており、いかなる参加者に対しても法律的またはその他の指導や助言を意図したものではありません。またそのような結果を生むものでもありません。本プレゼンテーションに含まれている情報については、完全性と正確性を帰するよう努力しましたが、「現状のまま」提供され、明示または暗示にかかわらずいかなる保証も伴わないものとします。本プレゼンテーションまたはその他の資料の使用によって、あるいはその他の関連によって、いかなる損害が生じた場合も、IBMは責任を負わないものとします。本プレゼンテーションに含まれている内容は、IBMまたはそのサプライヤーやライセンサーからいかなる保証または表明を引き出すことを意図したもので、IBMソフトウェアの使用を規定する適用ライセンス契約の条項を変更することを意図したものでなく、またそのような結果を生むものでもありません。

本プレゼンテーションでIBM製品、プログラム、またはサービスに言及していても、IBMが営業活動を行っているすべての国でそれらが使用可能であることを暗示するものではありません。本プレゼンテーションで言及している製品リリース日付や製品機能は、市場機会またはその他の要因に基づいてIBM独自の決定権をもっているいつでも変更できるものとし、いかなる方法においても将来の製品または機能が使用可能になると確約することを意図したものではありません。本資料に含まれている内容は、参加者が開始する活動によって特定の販売、売上高の向上、またはその他の結果が生かると述べる、または暗示することを意図したもので、またそのような結果を生むものでもありません。

パフォーマンスは、管理された環境において標準的なIBMベンチマークを使用した測定と予測に基づいています。ユーザーが経験する実際のスループットやパフォーマンスは、ユーザーのジョブ・ストリームにおけるマルチプログラミングの量、入出力構成、ストレージ構成、および処理されるワークロードなどの考慮事項を含む、数多くの要因に応じて変化します。したがって、個々のユーザーがここで述べられているものと同様の結果を得られると確約するものではありません。

記述されているすべてのお客事例は、それらのお客様がどのようにIBM製品を使用したか、またそれらのお客様が達成した結果の実例として示されたものです。実際の環境コストおよびパフォーマンス特性は、お客様ごとに異なる場合があります。

IBM、IBM ロゴ、ibm.com、Blue Gene、InfoSphere、POWER、Power Systems、POWER7、PowerPCおよびSystem xは、世界の多くの国で登録されたInternational Business Machines Corporationの商標です。

Linuxは、Linus Torvaldsの米国およびその他の国における登録商標です。

他の製品名およびサービス名等は、それぞれIBMまたは各社の商標である場合があります。現時点でのIBMの商標リストについては[www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml)をご覧ください。



ありがとうございました