

# オープンソース BI スイート「Pentaho」を Sun プラットフォームでを使用した際のサイジング

Sizing Study for Pentaho Open Source Business Intelligence Suite on Sun Platforms

Krishnan Shankar

June 2009

Sun Microsystems, Inc.

This article was published by BigAdmin at:

[http://www.sun.com/bigadmin/features/articles/pentaho\\_sizing.jsp](http://www.sun.com/bigadmin/features/articles/pentaho_sizing.jsp)

To keep track of the latest content published by BigAdmin, subscribe to the BigAdmin newsletter: <http://www.sun.com/bigadmin/newsletter/>.

日本語翻訳：株式会社 KSK ソリューションズ

Copyright © 2009 Sun Microsystems, Inc., 4150 Network Circle, Santa Clara, California 95054, U.S.A. All rights reserved.

U.S. Government Rights - Commercial software. Government users are subject to the Sun Microsystems, Inc.

standard license agreement and applicable provisions of the FAR and its supplements. Use is subject to license terms.

This distribution may include materials developed by third parties.

Parts of the product may be derived from Berkeley BSD systems, licensed from the University of California. UNIX is a registered trademark in the U.S. and in other countries, exclusively licensed through X/Open Company, Ltd.

X/Open is a registered trademark of X/Open Company, Ltd.

AMD, Opteron, the AMD logo, the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation in the United States and other countries.

Sun, Sun Microsystems, the Sun logo, Java, JDBC, MySQL, Solaris, Sun BluePrints, Sun Fire, and SunSolve are trademarks or registered trademarks of Sun Microsystems, Inc. or its subsidiaries in the United States and other countries.

This product is covered and controlled by U.S. Export Control laws and may be subject to the export or import laws in other countries. Nuclear, missile, chemical biological weapons or nuclear maritime end uses or end users, whether direct or indirect, are strictly prohibited. Export or reexport to countries subject to U.S. embargo or to entities identified on U.S. export exclusion lists, including, but not limited to, the denied persons and specially designated nationals lists is strictly prohibited.

DOCUMENTATION IS PROVIDED "AS IS" AND ALL EXPRESS OR IMPLIED CONDITIONS, REPRESENTATIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE HELD TO BE LEGALLY INVALID.

Sizing Study for Pentaho Open Source Business Intelligence Suite on Sun Platforms 2

1. イントロダクション .....	5
1.1 概要 .....	5
1.2 この仕様書の構成 .....	5
1.3 用語 .....	5
1.4 キーパフォーマンスメトリクス .....	5
1.5 プロジェクト評価と要件 .....	6
1.6 参照リンク .....	6
2. テスト .....	7
2.1 テストに関する詳細 .....	7
2.1.1 Pentaho ETL .....	7
2.1.2 Pentaho レポート .....	7
2.1.3 Pentaho Mondrian OLAP サーバー .....	7
2.2 デプロイメントトポロジー .....	8
2.3 ハードウェアとソフトウェア仕様 .....	9
3. パフォーマンス結果 .....	10
3.1 Pentaho ETL .....	10
3.2 Pentaho Report Designer .....	10
3.2.1 クエリー 1: 1-GB の DWH で 2 つのテーブルの簡単な Join を行う .....	11
3.2.2 クエリー 2: 30-GB の DWH での WHERE 節を使った単一テーブル .....	12
3.3 Pentaho Mondrian OLAP サーバー .....	13
3.4 テスト結果の考察 .....	13
4. 結果の概要 .....	15
4.1 パフォーマンスデータの概要 .....	15
4.2 分析と結論 .....	15
4.3 推奨ハードウェア .....	16
5. 追加情報 .....	17
5.1 パラメータ調整の基準 .....	17
5.2 ワークロード詳細 .....	17
6. 補足 .....	19
6.1 アプリケーションソフトウェアのデプロイメント詳細 .....	19
6.1.1 VNC ソフトウェア設定 (任意) .....	19
6.1.2 Apache Tomcat Server と MySQL 接続ドライバの設定 .....	19
6.1.3 データウェアハウス Workload Kit と DBGEN データ生成ツールの設定 .....	20
6.1.4 MySQL データベースの設定 .....	20
6.1.5 Pentaho ETL 設定 .....	20

6.1.6 Pentaho BI-CE 2.0.0 サーバー設定 .....	21
6.1.7 Pentaho Report Designer 設定 .....	21
6.1.8 Mondrian OLAP サーバー設定 .....	21
7. 謝辞 .....	22
8. 製品情報 .....	22
8.1 Sun 製品 .....	22
8.2 Pentaho 提供 .....	23
9. 参照 .....	24
10. ライセンス情報 .....	26

# 1. イントロダクション

## 1.1 概要

このレポートは、Sun プラットフォーム上にあるオープンソースビジネスインテリジェンス (OSBI) スイート Pentaho に関する、Pentaho のアプリケーションツールで行ったテストも含めてのサイジングについて記述しています。このガイドは、ユーザーがビジネスインテリジェンス (BI) アプリケーション用に適切な Sun プラットフォームを選択するときの参考にするために作られています。

このテストでは Solaris 10™ 5/08 OS、MySQL™ データベース、Sun Fire™ X4540 サーバーを使用しています。この仕様書の Sun と Pentaho の技術について詳しくは、このドキュメントの 8 章「製品情報」、9 章の「参照」を参照してください。

## 1.2 この仕様書の構成

ハードウェア・ソフトウェア要件に従ってプロジェクトの要件を記載しています。次に Pentaho アプリケーションツールのデプロイメントとテストのセットアップに関して述べています。パフォーマンス結果は表形式で示しています。レポートの最後では結果の分析とテストしたデータ量に対して推奨されるハードウェアについて書かれています。

## 1.3 用語

表 1: 用語

用語	定義
仮想 CPU (Virtual CPU)	物理 CPU でコアのデータの流れを共有する処理粒度 (例えば、軽量処理 [LWP])
データ変換 (Transformation)	Pentaho の Kettle アプリケーション (ETL ツール) で行われる処理。ソースの中からデータを抽出して変更します。
キューブ、メジャー、ディメンジョン、レベル (Cube, Measures, Dimension, Levels)	Pentaho Mondrian サーバーのマルチディメンション表現 (MDX) クエリー言語の構造物。

## 1.4 キーパフォーマンスメトリクス

表 2: キーパフォーマンスメトリクス

パフォーマンスメトリクス	定義
スループット	Pentaho ETL でロードされるデータ (GB/時間) Pentaho レポートで取得されるデータ (GB/時間)
応答時間	クエリーを処理してアウトプットを取得するのにかかる時間
CPU 使用率	Solaris 10 OS の vmstat ツールから取得される CPU 使用率の 情報
メモリ使用率	vmstat アウトプットが示す計算をしていない状態の“free”値
平均インプット/アウトプット使用率	Solaris 10 OS iostat ツールからの kr/s カラムと kw/s カラムか らの情報

## 1.5 プロジェクト評価と要件

ここではプロジェクトの評価について説明します。

- ・ スケーラビリティはクラスターでないシステム上で 2-4 プロセッサで行っています。高可用性 (HA) やフェイルオーバーは必要ありません。設定はプロセッサ使用率 20-80 パーセント、予期していないロードのために 10 パーセントのヘッドルームで測られています。
- ・ 予期されるデータベースの増大やピークロードの基準はありません。これらはお使いの環境によって変化します。
- ・ ユーザーミックスは、オンライントランザクション処理 (OLTP) の特徴というよりは、オンライン分析処理 (OLAP) の特徴といえます。想定されるのは、数百人のユーザーが中～大規模サイズのデータを頻繁に操作し、大規模なデータへ同時にクエリーすることはまれであるといった状況です。メトリクスでは ETL、レポートクエリー、Mondrian OLAP クエリー (G byte/時間と応答時間) について定義されています。
- ・ システム設定要件では、中～大規模 ETL ロードまたはレポートや OLAP クエリーで受け入れる応答時間等を設定しており、この両方で CPU 使用率が高いものと低いものを設定しています。

## 1.6 参照リンク

このドキュメントの最後にある 9 章の「参照」をご覧ください。

## 2. テスト

### 2.1 テストに関する詳細

#### 2.1.1 Pentaho ETL

1. データウェアハウスワークロードキット (data warehouse workload kit: 9 章の「参照」を見てください) をダウンロードして、スキーマを作成します。
2. TPC から DBGEN データジェネレータツールを構築します。
3. Pentaho データ統合ツールバージョン 3.1.0.をダウンロード・インストールします。
4. Pentaho Spoon データ変換ツールを通して Pentaho Kettle データ変換を作成し、データベースにデータをロードするために pan ツールを実行します。(9 章参照)
5. 毎回システムリソースを変えつつ、1-100 Gbyte のフラットファイルインプットロードを繰り返します。
6. スループットを想定し、システムリソース使用率をモニタします。

#### 2.1.2 Pentaho レポート

1. Pentaho Report Designer (RD) 1.7.1.をダウンロード・インストールします。
2. Pentaho BI Server Community Edition (BI-CE) v2.0.0. ダウンロード・インストールします。
3. Pentaho RD GUI を使用したレポートを作成して、BI-CE サーバーにレポートをパブリッシュします。
4. システムリソースを毎回変更しつつ、一般的なユーザークエリーを繰り返し実行します。
5. システムリソースを毎回変更しつつ、スループットを測定します。

#### 2.1.3 Pentaho Mondrian OLAP サーバー

1. Apache Tomcat 5.5.27.をダウンロード・インストールします。
2. Pentaho Mondrian Server 3.0.4.をダウンロード・インストールします。
3. Apache Tomcat で Mondrian サーバーを設定します。
4. Mondrian サーバーのデモデータベーススキーマとテーブルをインストールします。
5. 毎回システムリソースを変えてデモクエリーを実行します。
6. 応答時間を測定して、システムリソース使用率をモニタします。

## 2.2 デプロイメントトポロジー

ここでは、使用したハードウェアとソフトウェアについて説明します。

Pentaho アプリケーションと MySQL™ データベースは同じ物理サーバーで実行します。そして Microsoft Windows クライアントは Pentaho ETL の Spoon (GUI) でデータ変換を作成するのに使用されます。ネットワーク接続は 1Gbps、10BaseT Cisco switch です。個人用インターコネクトとロードバランサはありません。ストレージは内部の SATA-II ディクスで構成されます。



Pentaho ETL の Spoon (GUI) 用の Microsoft Windows クライアント



アプリケーションとデータベース用の Sun Fire™ X4540 server

図 1: デプロイメントトポロジー

## 2.3 ハードウェアとソフトウェア仕様

表 3: ハードウェアとソフトウェア仕様

アプリケーション コンポー ネント	ソフトウェア (バージョン)	ハードウェア設定			
		プロセッサ	メモリ	ネットワーク	ストレージ
アプリケーションとデータベースは Sun Fire X4540 サーバーでホストされています。	Pentaho Data Integration 3.1.0 Pentaho Report Designer 1.7.1 Penatho Mondrian Server 3.0.4 Pentaho BI Server Community Edition 2.0.0 MySQL Community Edition v5.1.30, data warehouse workload kit Apache Tomcat 5.5.27、Solaris 10 5/08 OS	Sun Fire X4540 Server	64 GB	1 GB NIC	48 pluggable、SATA-II、各 930 GB の内部ディスク

### 3. パフォーマンス結果

#### 3.1 Pentaho ETL

PentahoETL Kettle データ変換を使ってフラットファイルを MySQL 5.1.30 データベースのデータウェアハウススキーマにロードするときアーカイブされるスループットです。

- ・ ディスク: データベースはアプリケーションやインプットファイルから異なるディスク上にあります。
- ・ ネットワーク: 1-Gbyte NIC でプライベートインターコネクトは使用されていません。
- ・ データベース: デフォルトの MySQL 5.1.30 64-bit binary を使用しています。
- ・ データ変換: 簡単な join と WHERE 節を使用。
- ・ 処理ヒープサイズ: ヒープサイズ 2,048 m

Virtual Processors	DB にロードされるデータ (GB)	CSV ファイルからのロード時間 (分)	ロードスループット (GB/時間)	プロセッサ使用率 (%)	メモリ使用率 (%)	インプット/アウトプット使用率の最大値 (kw/s)
1	1	23	2.6	100	18	2323
2	1	19	3.2	63	17	2435
	10	188	3.2	66	20	2531
4	1	18	3.3	37	20	2442
	10	163	3.7	42	20	2573
	30	485	3.7	38	20	2558
	100	1637	3.7	33	23	2596
8	1	18.5	3.2	25	23	2108
	10	174	3.5	30	20	2515

#### 3.2 Pentaho Report Designer

表 5 と表 6 は、Pentaho Report Designer ツールからパブリッシュした大きいクエリーレポートを Pentaho BI-CE サーバーが実行するときの応答時間を示しています。

- ・ ディスク: データベースはアプリケーションやインプットファイルから異なるディス

ク上にあります。RAID は設定されていません。

- ・ ネットワーク: 1-Gbyte NIC でプライベートインターコネクトは使用されていません。
- ・ データベース: デフォルトの MySQL 5.1.30 64-bit binary が使用しています。

### 3.2.1 クエリー 1: 1-GB の DWH で 2 つのテーブルの簡単な Join を行う

a.) レポートのパフォーマンス (SQL コンポーネントとレポートコンポーネントでの外部アクションシーケンス[xaction]) :

- ・ アウトプット: 1 行当たり 256B で 2,000,000 行
- ・ 取得されるデータ量: 488.28MB

表 5: レポートのパフォーマンス (SQL コンポーネントとレポートコンポーネントでの Xaction)

Virtual Processors	DB からの FETCH 時間 (分)	FETCH スループット (GB/時間)	プロセッサ使用率 (%)	メモリ使用率 (%)
1	18	1.59	100	16.4
2	14.5	1.97	67	16.3
4	10	2.86	45	16.4
8	12.2	2.35	27	16.5

b.) クエリーを実行するプラットフォームのパフォーマンス (SQL コンポーネントでの xaction) :

- ・ アウトプット: 200,000 行
- ・ 1、2、4CPU では、FETCH 時間は 3.5 分で取得されるデータは 0.82 GB/時間です。
- ・ 8CPU では、FETCH 時間は 4.3 分で、取得されるデータは 0.67 GB/時間です。
- ・ プロセッサ使用率は、各 1、2、4、8CPU で、それぞれ 100、52、26、13%です。
- ・ メモリ使用率はほんの少しです。
- ・ 300,000 行を超えるクエリーでは長い時間がかかります。

c.) RAW クエリーのパフォーマンス（シェルコマンドラインでクエリーを選択）：

- ・ アウトプット: 200,000 行

CPU	FETCH 時間 (mm:ss)	CPU 使用率
1	6:52	100%
2	6:42	52%
4	6:58	26%
8	7:14	13%

### 3.2.2 クエリー 2: 30-GB の DWH での WHERE 節を使った単一テーブル

a.) レポートのパフォーマンス（SQL コンポーネントとレポートコンポーネントでの xaction）：

- ・ アウトプット: 24,000,000 行のテーブルから、一行当たり 146.85B で 2,000,000 行
- ・ 取得するデータ量: 280.09MB

表 6: レポートのパフォーマンス（SQL コンポーネントとレポートコンポーネントでの xaction）

Virtual Processors	DB からの FETCH 時間 (分)	FETCH スループット (GB/時間)	プロセッサ使用率 (%)	メモリ使用率 (%)
1	19.5	0.84	100	18.5
2	19.3	0.85	60	18.7
4	18.3	0.9	34	17.8
8	18.6	0.88	29	17.5

b.) クエリーを実行するプラットフォームのパフォーマンス（SQL コンポーネントでの xaction）：

- ・ アウトプット: 200,000 行
- ・ FETCH 時間は 8CPU までは 2 秒以下でした。

c.) RAW クエリーのパフォーマンス（シェルコマンドラインでクエリーを選択）：

- ・ アウトプット: 2,000,000 行

CPU	FETCH 時間	CPU 使用率
1	9:15	8%
2	8:54	4%
4	9:00	2%
8	8:58	1%

### 3.3 Pentaho Mondrian OLAP サーバー

表 7 は、Pentaho Mondrian サーバーが小さなデータベースに対して一連のクエリーを実行するときの応答時間を示しています。

- ・ ディスク: アプリケーションやインプットファイルから異なるディスク上にあります。RAID は設定されていません。
- ・ ネットワーク: 1-Gbyte NIC でプライベートインターコネクトは使用されていません。
- ・ データベース: デフォルトの MySQL 5.1.30 64-bit binary を使用します。
- ・ クエリー: FoodMart Demo クエリーと 135MB のデータベースを使用します。
- ・ キャッシュ: キャッシュは無効です。データベースアプリケーションサーバーは実行するごとに再起動されます。

表 7: Pentaho Mondrian サーバーが小さなデータベースに対してクエリーを実行するときの応答時間

Virtual Processors	データベースからの FETCH 時間 (秒)	CPU 使用率 (最大%)	メモリ使用率 (%)
1	588	100	19
2	522	54	19
4	502	27	19
8	484	13	19

### 3.4 テスト結果の考察

Pentaho ETL と Pentaho レポートクエリーにおいて、スループット測定は 1 時間当たりロードされるデータを GB で出しています。ロード時間は最初にデータウェアハウスからデー

タのロードを始めた時点から、最後のデータウェアハウステーブルのロードが完了した時点までを測定しています。ETL データロードはフラットファイルからです。

Pentaho レポートでは、応答時間を測定しました。ロー（JOIN クエリーで）サイズの合計と取得したローの数を掛けて、時間を推定するという計算をしています。

Pentaho Mondrian サーバーMDX クエリーでは、応答時間は 1 時間に取得される GB 数であらわされます。これには最初の FoodMart デモクエリーがブラウザでクリックされてから最後のクエリー結果が取得されてスクリーンに表示されるまでの時間が含まれています。

## 4. 結果の概要

### 4.1 パフォーマンスデータの概要

Pentaho ETL では、さまざまなバーチャルプロセッサでの異なるサイズのデータロードでは、違いが出てきます。中から大規模のソフトウェアベンダでよくあるデータ量において、プロセッサ、メモリ、そしてインプット/アウトプット使用率が調べられています。

Pentaho レポートでは、あるクエリーシナリオは2つのテーブルの JOIN をし、別のものでは単一テーブルでのクエリーをしています。これらは巨大なデータベースを使用している中～大規模ベンダのデータ量です。テーブルは大きなデータベースから選ばれます。スループットはGB/時間で示されます。テーブルはインデックスを付けられていません。プロセッサとメモリ使用率は示されます。

Pentaho Mondrian では、シナリオは実行中の FoodMart デモだけです。これは小～中規模のベンダでの現実的なデータ量を想定しています。応答時間はさまざまなプロセッサ数で示され、データ量のコンテキストの中で判断されます。

メモリとインプット/アウトプットはさほど重要ではありません。表ではバーチャルプロセッサと応答時間を記載しています。またプロセッサとメモリ使用率を記載しています。

### 4.2 分析と結論

この章では、Sun ハードウェアプラットフォームで一般的なデータ量のもと、システムがどのように動くかを検証しています。これは BI ユーザーが行うあらゆるデータ量を想定したものではありません。

Pentaho ETL:

- ・ 4 プロセッサはインプットロードサイズに関係なくスループットが最大になります。
- ・ 10GB ロードでもっともスループットの値が高くなります。
- ・ 8CPU セットアップでは CPU あたりの使用率が最も低く、3 パーセントとなります。  
An 8-CPU setup yielded the lowest “per cpu” utilization at 3%.
- ・ 1GB のデータベースでは、平均スループットは 3.1GB。
- ・ 10GB のデータベースでは、平均スループットは 3.5GB。
- ・ 30GB のデータベースでは、平均スループットは 3.7GB。
- ・ 100 GB のデータベースでは、平均スループットは 3.3GB。
- ・ 1 プロセッサの平均スループットは 2.6GB/時間。
- ・ 2 プロセッサの平均スループットは 3.2GB/時間。

- ・ 4 プロセッサの平均スループットは 3.6GB/時間。
- ・ 8 プロセッサの平均スループットは 3.3 GB/時間。
- ・ メモリ使用率は全てのロードタイプとプロセッサで 17-23%。
- ・ インプット/アウトプット使用率は全てのロードタイプとプロセッサで 16-19%。
- ・ インプットのスタートアップコピーの実行、または「並行で実行」でも、パフォーマンス結果は同等です。
- ・ アプリケーションとデータベースは、I/O 制限を防ぐために別々のディスクに置いています。

Pentaho レポートではレポートのパフォーマンス（SQL コンポーネントとレポートコンポーネントでの xaction）は 1GB のデータベースで 4 プロセッサを使用した時 2.86GB/時間というスループットの最大値を出します。スループット、CPU、そしてメモリ使用率はすべて、両方のクエリーにおいて 4 プロセッサで測定されます。

Pentaho Mondrian OLAP サーバーでは、8 プロセッサで最も短い FETCH 時間の 484 秒となります。プロセッサ使用率は、プロセッサの数によって測定されます。メモリ使用率は定数です。

### 4.3 推奨ハードウェア

規模の小さな ETL ロード、あまり複雑でないレポートのクエリー、レポートで応答時間がある程度許容される場合、OLAP クエリーで CPU 使用率が 70%を超えると、ハードウェアは **1 quad-core Intel または、2.5 GHz でメモリが 16GB の AMD Opteron プロセッサを搭載したサーバー**を推奨します。

中から大規模の ETL ロード、中から高程度の複雑なレポートのクエリー、レポートの応答時間が短い必要がある、OLAP クエリーで CPU 使用率が 70%以下の必要がある場合、ハードウェアは **2 quad-core Intel または、2.5 GHz でメモリが 16GB の AMD Opteron プロセッサを搭載したサーバー**を推奨します。

## 5. 追加情報

Pentaho ETL では、Spoon GUI は Solaris 10 OS for x86 プラットフォームでは提供されていない gtk ライブラリを必要とします。Microsoft Windows クライアントはベーシックなデータ変換を作成する Spoon GUI の起動に用いられます。

Pentaho レポートでは、レポートバーストテストに Pentaho BI-EE サーバー (Enterprise Edition v3.0.0 RC2) が必要になります。パートナー評価ライセンスを取得しています。ですが、BI-EE をインストールすると、エンタープライズコンソールツールは Solaris OS できちんと動作しなくなり、そのためライセンスもインストールされませんでした。結果、バーストテストは省略されました。Tsung シュミレーターツールを使用してロードシュミレーションを試みています。(9 章「参照」をご覧ください) しかし、ログインシーケンスのシュミレーションで Tsung が識別しなかった特別文字を、Pentaho BI-CE 2.0.0 はアウトプットしています。また、多くのコンカレントユーザーは Pentaho BI の一般的なシナリオでは想定されていません。その代わりに、規模の大きい OLTP クエリーでユーザー数が少ない場合相殺された形になります。(Instead, a small number of users with large OLTP queries is the likely mix.)

そのため、このテストは追及されていません。

Pentaho Mondrian OLAP サーバーでは、デモファイルの FoodMart.xml が Mondrian Schema Workbench から Pentaho サーバーにパブリッシュされますが、エントリはアナリシス画面には表示されません。Apache Tomcat Server と Pentaho BI-CE サーバーは、別のポート番号を設定されていますが同時に起動しないこともあります。これを解決するには、BI-CE のインストールスクリプトの CATALINA\_HOME をきちんと設定することを推奨します。

### 5.1 パラメータ調整の基準

Pentaho ETL では、Filter Rows 変換のスタートアップコピーが 10 個適用されています。最大ヒープスペースは 2048m に設定されています。スレッドプライオリティ管理はデータ変換で有効です。

Pentaho レポートでは、最大ヒープサイズは 1800M に設定しました。Pentaho Mondrian サーバーでは、ヒープスペースは 1800M には設定していません。キャッシュは Mondrian サーバーとデータベースを各実行の前に再起動することによって事実上無効になります。

### 5.2 ワークロード詳細

Pentaho ETL のためのデータウェアハウススキーマでインデックスまたは制約は強制され

ません。2,000,000 行が 2 つのデータウェアハウステーブルの簡単な JOIN と単一テーブルから取得されています。

クエリー 1:

```
SELECT LINEITEM.QUANTITY, SUPPLIER.NAME
FROM LINEITEM , SUPPLIER
WHERE LINEITEM.SUPPKEY=SUPPLIER.SUPPKEY
LIMIT 2000000
```

クエリー 2:

```
SELECT PARTKEY
FROM PARTSUPP
WHERE PARTKEY > 0
LIMIT 2000000
```

Pentaho Mondrian OLAP サーバーでは、大きなテストデータは現在利用できません。そのため、FoodMart データベースのデータの嵩を増す必要があります。これはまず UNIQUE キーを除去して、これらの 12 個あるデモテーブルでデータを 3 回加えることで可能になります。

12 個のテーブル:

```
customer, employee, product, salary, store, warehouse, expense_fact, inventory_fact_1997,
inventory_fact_1998, sales_fact_1997, sales_fact_1998, and sales_fact_dec_1998.
```

データを 4 倍に増やすと、バーチャルプロセッサでタイムアウトを起こします。Mondrian Server MDX クエリーはキューブ、ディメンション、メジャーから構成されています。多くのクエリーが cross-joins を含んでいるか、簡単なものから中程度の複雑さとなっています。

## 6. 補足

### 6.1 アプリケーションソフトウェアのデプロイメント詳細

このセクションでは、オープンソースビジネスインテリジェンススイートの Pentaho でテストを実行する上で必要なソフトウェア設定のための手順を説明します。

#### 6.1.1 VNC ソフトウェア設定（任意）

1. ルートユーザーとして、Solaris 10 OS に Virtual Network Computing (VNC) ソフトウェアを入手・インストールしてください。(9章「参照」をご覧ください)
2. デフォルトポート 5081 で Web ブラウザを使ってクライアントを表示してください。
3. クライアントでターミナルウィンドウを開いてください。

#### 6.1.2 Apache Tomcat Server と MySQL 接続ドライバの設定

1. ルートユーザーとして、apache-tomcat-5.5.27.tar.gz をダウンロード・解凍してください。
2. 任意で VNC ソフトウェアを通してターミナルウィンドウを使用します。
3. conf/tomcat-users.xml を編集して、admin ユーザーを追加してください。

```
<tomcat-users>
  <role rolename="admin"/>
  <user username="admin" password="admin" roles="admin,manager"/>
</tomcat-users>
```

4. MySQL 接続ファイルの mysql-connector-java-5.1.7.tar.gz をダウンロード・解凍してください。
5. mysql-connector-java-5.1.7-bin.jar を apache-tomcat-5.5.27/common/lib にコピーしてください。

### 6.1.3 データウェアハウス Workload Kit と DBGEN データ生成ツールの設定

#### 6.1.3 データウェアハウス Workload Kit と DBGEN データ生成ツールの設定

1. 標準の Data Warehouse Workload Kit をインストールして、バイナリを構築してください。
2. ロードデータ変換 (.ktr) ファイルを作成してください。(この場合、Pentaho で作成します。)
3. DBGEN ツールを使って 1GB のデータを作成してください (dbgen -vfF -s 1)。ロード中はインデックスを使用しないでください。
4. 次の章で説明されているように MySQL のインストール・設定を行ってください。そして、データウェアハウスデータベースを作成してください。
5. .ktr ファイルでインプットファイルのパス名とサーバー名を変更します。データ変換を通してデータをロードしてください。

### 6.1.4 MySQL データベースの設定

1. Root としてログインして、mysql ユーザーとグループを作成してください。
2. MySQL Community Server 5.1.30 をダウンロード・インストールしてください。
3. /usr/local/mysql/bin、/usr/local/mysql/libexec、MySQL Connector jar ファイルを環境変数のパスに設定してください。
4. /usr/local/mysql から MySQL インストールディレクトリにソフトリンクを作成してください。
5. /etc/my.cnf 設定ファイルを作成して、関連するパラメータを設定してください。
6. /mysql\_install\_db スクリプトを実行して、GRANT と特権を与えてください。

### 6.1.5 Pentaho ETL 設定

1. ETL Kettle Spoon GUI では、起動のために gtk ライブラリが必要です。サーバー上にこれがない場合は、代わりに外部の Microsoft Windows クライアントで ETL Spoon データ変換を作成してください。
2. クライアントに Pentaho Data Integration (pdi-open-3.1.0-826.zip) をダウンロード・解凍して、spoon.bat を起動してください。ドキュメントに従ってレポジトリとデータベース接続を作成します。

3. Table Input、Filter Rows、Text Output アイコンで各データウェアハウステーブルにデータ変換を作成してください。Filter Rows で、フィルター条件を設定してください。
4. データ変換を保存して Sun Fire X4540 サーバーにコピーしてください。
5. Sun Fire X4540 サーバーで以下のコマンド使って pan ツールを実行してください。このコマンドはデータ変換XMLファイルのパス名を提供します。Pan ツールはETL Kettle Spoon で作成したデータ変換を実行するコマンドラインプログラムです。

```
# ./pan.sh -file=<ktr file> -level=Basic
```

### 6.1.6 Pentaho BI-CE 2.0.0 サーバー設定

1. ドキュメントごとに BI-CE 2.0.0 サーバーをインストール・設定・起動してください。
2. BI-CE サーバーでレポートが開けることを確認してください。

### 6.1.7 Pentaho Report Designer 設定

1. Pentaho Report Designer ( prd-open-1.7.1.tar.gz ) を解凍してツール (./startDesigner\_osx.sh) を起動してください。
2. MySQL データベースに対して Java Naming and Directory Interface™ (JNDI) 接続を設定してください。
3. QueryDesigner をクリックしてレポートを構築し、安定したクエリーを提供してください。
4. 前もってインストールしておいた BI-CE サーバー上にレポートを保存・パブリッシュしてください。

### 6.1.8 Mondrian OLAP サーバー設定

1. Mondrian OLAP サーバー (mondrian-3.0.4.11371.zip) を解凍してください。lib/mondrian.war を<tomcat>/webapps
2. 必要なデータソース名とパラメータを提供して、datasources.xml、mondrian.properties、web.xml を設定します。
3. デモデータベースクリエーションスクリプトと foodmart\_mysql.sql をインストールしてください。UNIQUE KEY 制約を廃止します。
4. データソース接続を確立するためにデモファイルを設定して、Mondrian サーバーにア

クセスしてください。: <http://<server:port>/mondrian/>

## 7. 謝辞

Pentaho Corporation のエンジニアたちのタイムリーな手助けとサポートに感謝します:

- James Dixon, Founder and CTO
- Matt Casters, Chief Architect, Data Integration, Kettle Project Founder
- Brian Hagan, Senior Support Engineer

## 8. 製品情報

### 8.1 Sun 製品

企業では大容量のデータを素早く、確実に、そして経済的に伝える必要があります。Sun のサーバー、ストレージコンポーネント、革新的な技術はエンタープライズクラスのデータサービスを提供しながら、強力なスケーラビリティとパフォーマンスはもちろん大幅にコストの削減を行うことが可能です。

Sun の Solaris 10 OS はマルチプラットフォームで、スケーラブルです。そしてデータベース、Web、Java™テクノロジーベースのサービスで高いパフォーマンスを発揮します。OS の高度な機能の中にはシステムリソース使用率、ネットワークスタックの最適化、データ管理、システム可用性、インターオペラビリティ（相互運用性）ツールといったものが含まれています。

Sun プラットフォーム 16 Sun Fire x64 サーバー上での Pentaho オープンソースビジネスインテリジェンススイートのサイジングでは AMD Opteron™ プロセッサまたは Intel® プロセッサが使われており、またこれらは x86/x64 ベースのシステムとしても知られています。これらでは Solaris OS、Microsoft Windows、Linux、VMware を実行できます。サーバータイプにはラックマウントサーバー、ブレードサーバー、ワークステーションが含まれます。これらは高い効率性、スマートなスケーリング、信頼性、パフォーマンスを提供しています。買収した Sun MySQL データベースはもっとも一般的なオープンソースデータベースです。マルチスレッドで、SQL サーバーを構成し、クライアントプログラム、ライブラリ、管理ツール、そして API から構成されています。JDBC™ 接続を使用する Java クライアントプログラムは MySQL コネクタ/J インタフェースから MySQL サーバーにアクセスすることができます。

## 8.2 Pentaho 提供

Pentaho OSBI スイートはお客様に対してビジネスインテリジェンスを提供するオープンソースの製品です。これにはレポート、アナリシス、ダッシュボード、データマイニング機能があり、データ統合（ETL: extract, transform, and load）のようなツールを含んでいます。これらの機能によって売り上げ分析、顧客と製品の収益性、HR（人事）レポート、財務分析といったさまざまな情報へのアクセスを可能にし、そして経営陣に複雑な情報を伝えることも可能にします。

レポートはオンデマンドまたはスケジュールを選択可能で、一般的なフォーマットでパブリッシュできます。アナリシスはピボットテーブルビュー、グラフィカルディスプレイ、そしてワークフローインテグレーションを提供します。ダッシュボードにはアプリケーションに組み込むとき再利用可能なディスプレイウィジェットがあります。データマイニングではエンドユーザーにスマートなデータアナリシスを提供するための OLAP 技術を備えたアルゴリズムを提供します。そしてデータ統合はデータ処理のためのデザイン GUI と高いスケーラビリティとフレキシビリティを持っています。

## 9. 参照

以下はこの文書に関するリソースです。:

- Pentaho Open Source Business Intelligence web site: <http://www.pentaho.com>
- Pentaho Spoon and Kettle: <http://etl-tools.info/en/pentaho/kettle-spoon.htm>
- TPC-H data warehouse workload kit and DBGEN data generator tool:  
<http://www.tpc.org/tpch>
- TPC-H data warehouse workload kit and DBGEN data generator tool:  
<http://www.tpc.org/tpch>
- Mondrian FoodMart demo database:  
<http://business-intelligence.phi-integration.com/2008/04/mondrian-mysql-setup.html>
- Tsung open source multiprotocol distributed load testing tool:  
<http://tsung.erlang-projects.org/>
- VNC software on the Solaris 10 OS:  
[http://www.salixtraining.co.uk/index\\_files/vncsol10.htm](http://www.salixtraining.co.uk/index_files/vncsol10.htm)
- Solaris OS web site: <http://www.sun.com/software/solaris/10/index.jsp>
- Sun Fire X4540 web site: <http://www.sun.com/servers/x64/x4540/>
- MySQL 情報:
- On sun.com: <http://www.sun.com/software/products/mysql/index.jsp>
- On mysql.com: <http://www.mysql.com/>
- Sun 製品情報:
- Documentation at <http://docs.sun.com>, such as Sun Fire X4540 Server documents, Solaris 10 System Administrator Collection, and MySQL documents
- MySQL documents at mysql.com: <http://dev.mysql.com/doc/>
- Sun Documentation Center: <http://www.sun.com/documentation/>
- BigAdmin に関するリソース:
- MySQL Resources for System Administrators:  
<http://www.sun.com/bigadmin/topics/mysql/>
- Database resource collection (includes community submissions):  
<http://www.sun.com/bigadmin/collections/database.jsp>
- Storage resource collection (includes community submissions):  
<http://www.sun.com/bigadmin/collections/storage.jsp>
- Feature Article: Pentaho Business Intelligence and MySQL on Sun Storage 7000 Unified Storage System:  
[http://www.sun.com/bigadmin/features/articles/pentaho\\_7000.jsp](http://www.sun.com/bigadmin/features/articles/pentaho_7000.jsp)
- BigAdmin wiki at <http://wikis.sun.com/display/BigAdmin/Home>

- Storage Tech Tips page:  
<http://wikis.sun.com/display/BigAdmin/Storage+Tech+Tips>
- Databases page: <http://wikis.sun.com/display/BigAdmin/Databases>
- Servers page: <http://wikis.sun.com/display/BigAdmin/Servers>
- Sun BluePrints™ papers (registration required) at the Sun BluePrints wiki (<http://wikis.sun.com/display/BluePrints/Main>), for example:
- Running MySQL Database in Solaris Containers:  
<http://wikis.sun.com/display/BluePrints/Running+MySQL+Database+in+Solaris+Containers>
- Improving MySQL Database Scalability:  
<http://wikis.sun.com/display/BluePrints/Improving+MySQL+Database+Scalability>
- Sun download site: <http://www.sun.com/download/>
- Sun training courses at <http://www.sun.com/training/>, for example:
- Intermediate System Administration for the Solaris 10 Operating System (WSB-200-S10)
- Sun Fire X4540 Server Administration (WET-6179)
- MySQL training courses on sun.com:  
<http://www.sun.com/software/products/mysql/training.jsp>
- MySQL training courses on mysql.com: <http://www.mysql.com/training/>
- Discussions:
- Sun forums: <http://forums.sun.com/index.jspa>
- BigAdmin Discussions collection: <http://www.sun.com/bigadmin/discussions/>
- Support:
- Register your Sun gear: <https://inventory.sun.com/inventory/>
- Services: <http://www.sun.com/service>
- SunSolveSM: <http://sunsolve.sun.com>
- Community system administration experts:  
<http://www.sun.com/bigadmin/content/communityexperts>
- Events of interest to users of Sun products:
- Worldwide Developer Events and Sun Tech Days:  
<http://developers.sun.com/events/>
- Current Events: <http://www.sun.com/events/index.jsp>

## 10. ライセンス情報

他に特別の指定がない限り、ソフトウェアの使用はライセンスの条項に従って権限が与えられています。

参照: [http://www.sun.com/bigadmin/common/berkeley\\_license.html](http://www.sun.com/bigadmin/common/berkeley_license.html).