

# GreenplumDB Sales Training

Diane Hardman  
Product Manager

# Training Outline

- Overview of Big Data
  - Industries and Use Cases
  - Current RDBMS competitors
- Industry Trends and Unmet Needs
- What is GreenplumDB?
  - Key Features and Benefits
  - MPP Architecture
- What about 'other' Big Data Needs?
- Fit within the Pivotal Big Data Suite

# Big Data Market Overview

## What is Big Data?

*Big data* is all about getting value from the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis.

## Industries

- Financial Services
- Retail (online and brick and mortar)
- Manufacturing
- Government
- Health Care (EMRs, Genomics)
- IoT (personal and industrial sensors)

# Big Data Market Overview

## **Data Analytics drive business decisions:**

- Financial risk and wealth management services
- Mfg logistics planning
- Identify new revenue opportunities (product or services)
- Improved customer service
- More effective target marketing
- Fraud detection

# Leading installed RDBMS Systems

1. Oracle
2. Microsoft SQL Server
3. MySQL
4. IBM DB2
5. IBM Informix
6. SAP Sybase
7. Teradata

Source Gartner 2008

# Big Data Trends and Unmet Needs

## Exponential Data Growth challenges

- Ability to store data (scalability)
- Performance of queries and analysis

## Maintaining 24/7 online services

- Managing Fault Tolerance

## Data security

## Preserving legacy systems and expertise

- Rapidly scale using commodity HW

# What is GreenplumDB?

- Pivotal's flagship analytical data warehouse
- Relational database system for big data
- Used for mission critical analytics for large industrial systems
- Implements leading research in database technology
- Fully open source
- Performance tuned for multiple workloads

# GreenplumDB Key Features/Benefits

- Shared nothing, MPP architecture
  - Supports large data sets distributed across many servers (petabytes)
  - Scalable and fault tolerant
- Full SQL and MapReduce support
  - leverage existing analytic tools
- Parallel processing and Query Optimizer
  - improves performance for complex queries over large data sets
- Polymorphic Data Storage
  - Partitions organized differently to maximize query performance
  - access external databases
- Proven Open Source technology
  - Software hardened over 10 years
  - Open source available under Apache 2.0 license



# Why Open Source?

- More Enterprise clients demand it
  - Can customize to their needs
  - Proof of industry acceptance
- Accelerates adoption of GreenplumDB technology (hardened, proven standard)
- Accelerates innovation on the GreenplumDB framework

# GreenplumDB MPP Architecture

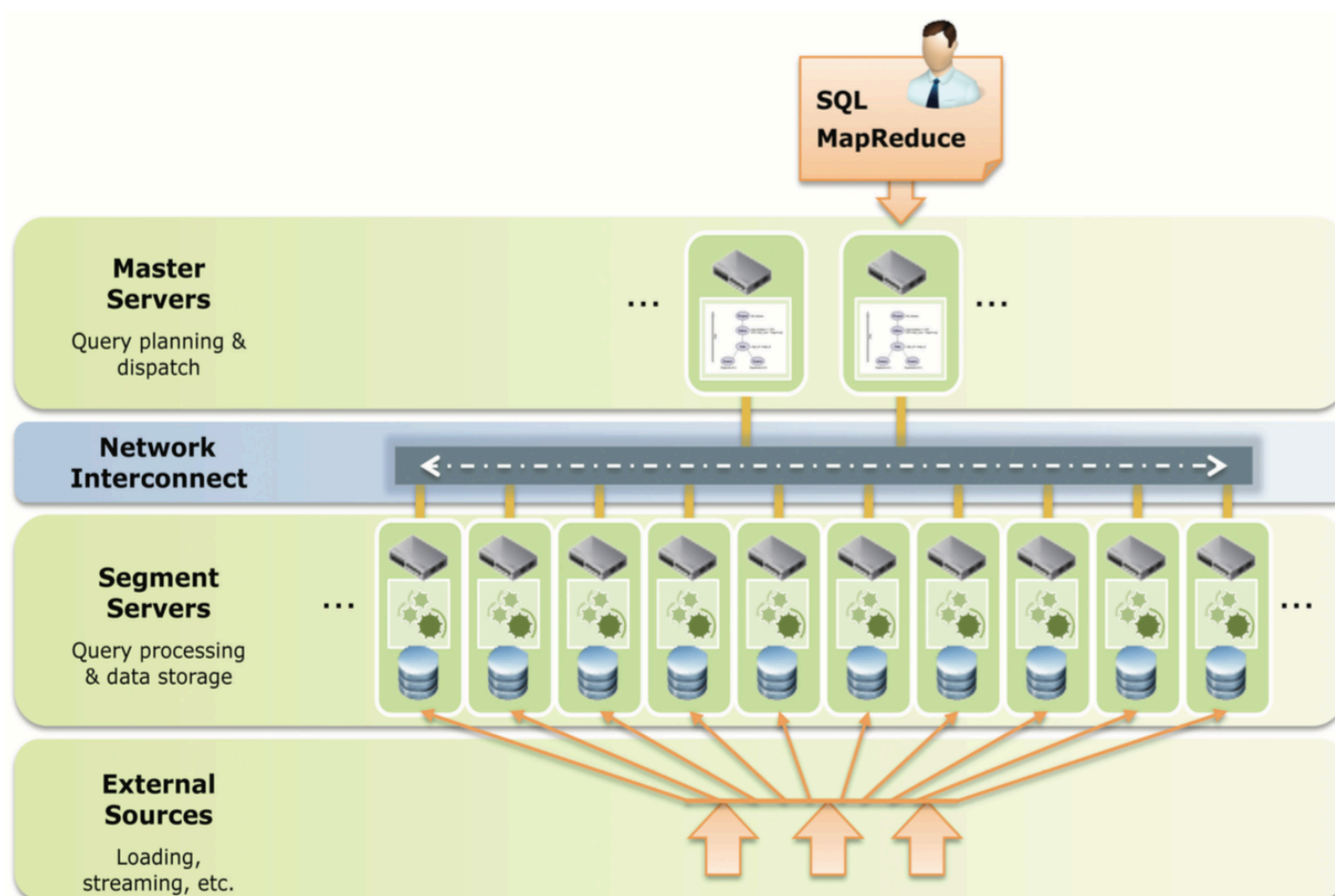


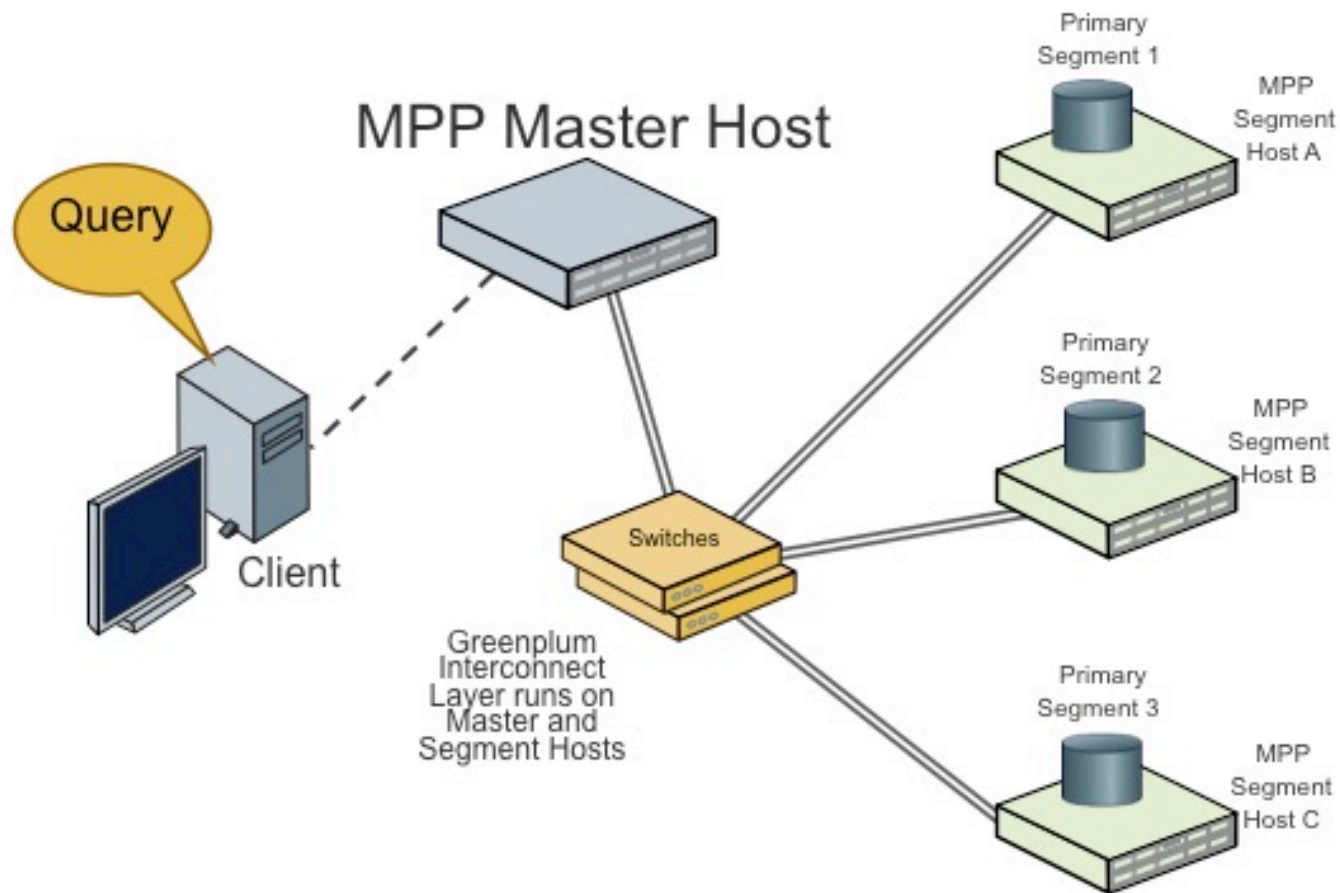
FIGURE 2. GREENPLUM'S MPP SHARED-NOTHING ARCHITECTURE

# Example Data Storage

Example records  
to be stored

- 1 Apple MacBook Pro
- 2 Apple Iphone 4s
- 3 Apple Ipad 3
- 4 Samsung Galaxy Tab 10.1
- 5 Blackberry Bold
- 6 Amazon Kindle Fire
- 7 Google Android Tablet 9
- 8 Nook Color
- 9 Lenovo IdeaPad

# Data Distribution



## Data

1	Apple MacBook Pro
4	Samsung Galaxy Tab 10.1
7	Google Android Tablet 9"
2	Apple Iphone 4s
5	Blackberry Bold
8	Nook Color
3	Apple Ipad 3
6	Amazon Kindle Fire
9	Lenovo IdeaPad

# Fault Tolerance



FIGURE 5. GREENPLUM DATABASE MULTI-LEVEL SELF-HEALING FAULT TOLERANCE PROCESS

1. Segment server fails
2. Mirror segments take over with no loss of service
3. Segment server is restored or replaced
4. Mirror servers restore primary via differential recovery (while online)

# What about 'other' Big Data

- Big Data requires support for unstructured data too! (email, documents, pictures)
- Hadoop is the predominant open source framework for unstructured data.
  - Distributed storage (HDFS)
  - Supports raw, unstructured data storage
  - Several distros available: Cloudera Impala, Hortonworks
- **BUT** unstructured data is difficult to analyze; requires expertise in writing MapReduce algorithms

# Pivotal's Solution

- Pivotal HDB
  - Brings ANSI-compliant SQL to Hadoop using open source Apache HAWQ
- Pivotal Greenplum Data warehouse system
  - MPP, ANSI-compliant SQL architecture for data warehousing and analytics
- Pivotal GemFire
  - Open source, scale out, in-memory application for high performance transactional data processing

# GreenplumDB

- Delivers industry leading performance and scalability for structured data
- Part of the Pivotal Big Data Suite which offers a complete solution for high performance data ingest, storage, and analytics for structured, semi-structured, and unstructured data needs.