IBM Balanced Warehouse™

Simplifying Business Intelligence with a Hybrid Appliance

Haider Rizvi, Joyce Coleman and Sam Lightstone
IBM Toronto Lab, Canada

ICDE / SMDB Conference
April 2008

Problem Statement

- Data warehouse systems have a need for:
  - High performance
  - Scalability
  - Ease of use, including setup, and ongoing operations
- Typically data warehouses have been built on traditional relational database systems, with the customer doing the systems integration and long-term maintenance
  - Setting up and configuring the systems, software
  - Performance tuning
  - Applying maintenance for firmware, software updates, etc.
- Very large databases in these environments exacerbate the problem of administration and performance tuning
  - Data warehouses quite often start “small” (a few TBs) but grow very fast if successful
  - Successful data warehouses for large enterprises quite often reach 50+ TB!

Traditional strategies for data warehouses

Relational database servers

- Fundamental strategies include query optimization, caching, and physical database design (e.g. indexes, view materialization)
- Examples: DB2, Oracle, MS SQL Server, Teradata

- + Flexibility
- + Excellent performance for query processing with many updates and high concurrency requirements
- - Complexity of design, configuration, and administration

Data warehouse appliances

- Fundamental strategies include massively parallel I/O, table scans, preconfigured components, and no tuning / configuration / physical database design
- Examples: Netezza

- + Simplicity of design, configuration, and administration
- + Excellent performance for query processing with few updates and low concurrency requirements
- - No flexibility
- - Scalability may need a rip-and-replace operation

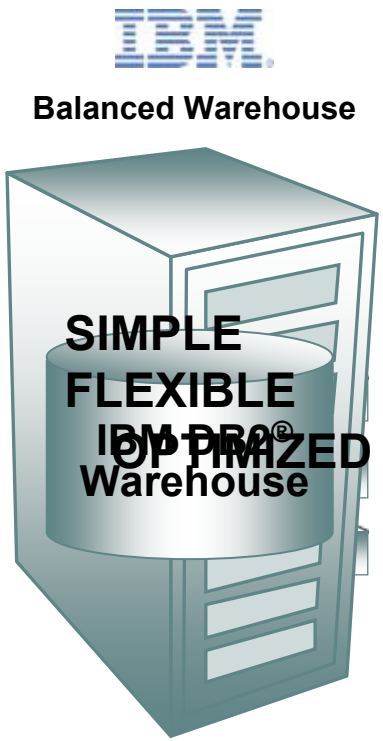
The Balanced Warehouse: The best of both approaches!

- Relational database server
  - DB2's relational features provide performance advantages over table scans alone
  - Excellent concurrency control
  - Can be customized according to customer needs

- Data warehouse appliance
  - Uses shared nothing architecture to achieve massively parallel I/O
  - Simplified implementation (pre-configuration of hardware; pre-installation of software; initial configuration settings will work for most customer environments)
  - Simplified administration (autonomic features)
  - Easy to scale up

IBM Balanced Warehouse™

A Hybrid Appliance



Better than an appliance

Simplicity

- Predefined configurations for reduced complexity
- One number to contact for complete solution support

Balanced Configuration Unit (BCU)

Preconfigured, pretested allocation of software, storage and hardware to support a specified combination of function and scale

Flexibility for growth

- Add BCUs to address increasing demands
- Multiple on-ramps for different needs
- Reliable, nonproprietary hardware for reusability

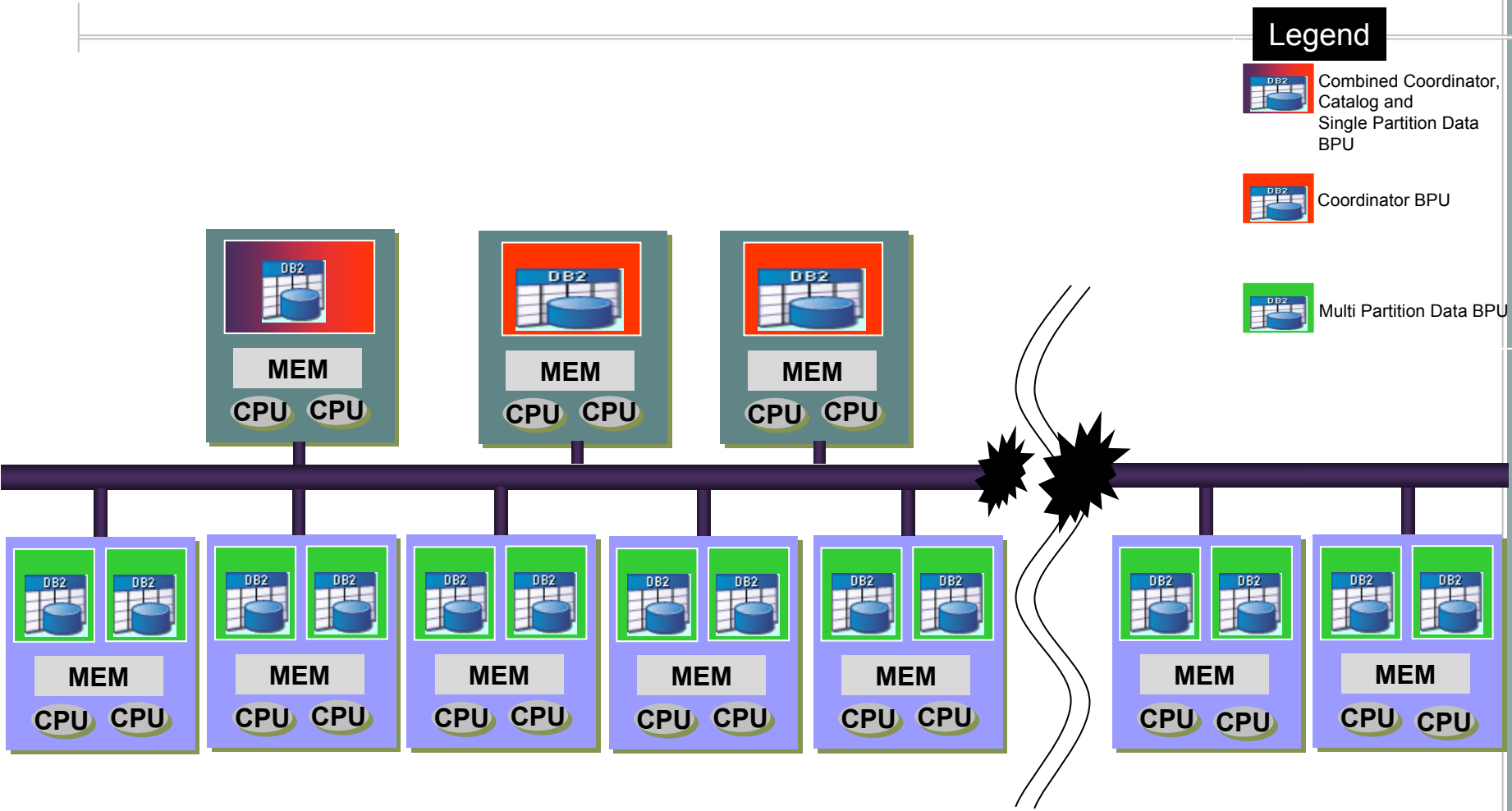
Optimized performance

- Preconfigured and certified for *guaranteed* performance
- Based on best practices for reduced risk

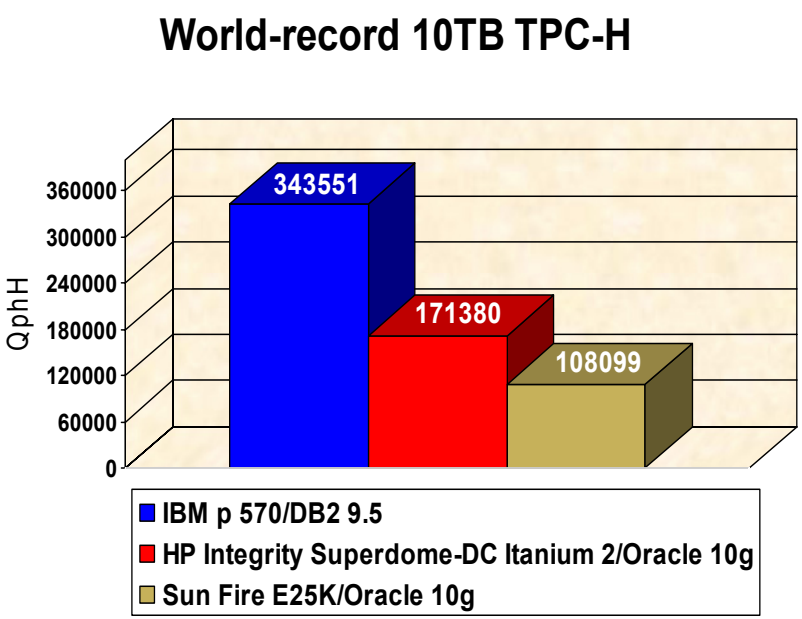
Self-managing characteristics of Balanced Warehouse

Pre-configured	Adaptive
Storage	Memory configuration parameters
Servers	Statistics for tables, indexes, etc.
Initial memory configuration	Physical design (using design advisors) for Indexes, Mat Views, MDC, Hash partitioning
Configuration parameters for OS / DB (process model, locking, data prefetching)	Throttled utilities for backup, reorg, and runstats
	Integrated workload management

Typical scale up growth path for Balanced Warehouse



Leading TPC-H result on IBM Balanced Warehouse E7100



Significant proof-point for the new IBM Balanced Warehouse E7100

DB2 Warehouse 9.5 running on POWER6 servers and DS4800 storage

DB2 Warehouse 9.5 takes DB2 performance on AIX to new levels

Highest per-core performance levels ever!

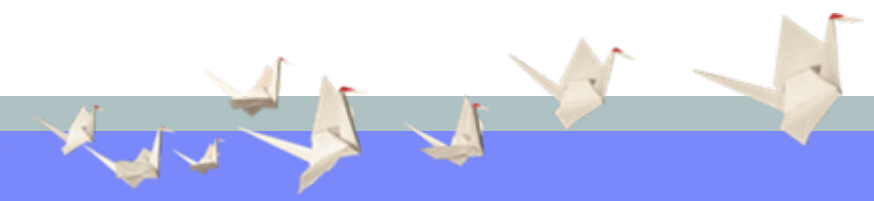
Loaded 10TB data @ 6 TB / hour (incl. data load, index creation, runstats)

TPC Benchmark, TPC-H, QphH, are trademarks of the Transaction Processing Performance Council. For further TPC-related information, please see http://www.tpc.org. Data correct as of Feb 28, 2007

DB2 Warehouse 9.5 on IBM System p 570 (128 core p6 4.7GHz), 343551 QphH@10000GB, 32.89 USD per QphH@10000GB available: April 15, 2008
Oracle 10g Enterprise Ed R2 w/ Partitioning on HP Integrity Superdome-DC Itanium 2 (128 core Intel Dual Core Itanium 2 9040 1.6 GHz), 171380 QphH@10000GB, 32.91 USD per QphH@10000GB, available: April 1, 2007
Oracle 10g Enterprise Ed R2 w/ Partitioning on Sun Fire E25K (144 core Sun UltraSparc IV+ - 1500 MHz): 108099 QphH @53.80 USD per QphH@10000GB available: January 23, 2006

Future Work

- Enhancing autonomic features in the database engine
- Automating maintenance tasks such software and firmware updates
- Extending functionality for system-wide performance monitoring, health checking, and problem determination
- Improving the seamless addition of new modules into the environment
- Providing an automated sizing tool that facilitates arriving at the correct sizing of the Balanced Warehouse configuration, given the details of the customer workload and data set



Summary

- Present a hybrid appliance, the IBM Balanced Warehouse
  - merges the best of fully-configured Appliances and traditional data warehouse systems.
- Provide a building block approach using the shared-nothing database design that allows organic growth
- Provide a complete pre-configured system to improve on the pain-points of a DBA / sys admin