

Inside Today's Data Warehouses: They're Fast, Flexible, and Essential

New innovations in database technology make data warehouses easier to launch and more valuable than ever

Table of Contents

Analytics on the Rise

The Right Database Foundation Reduces Complexity

Give Your Warehouse Room to Grow

Eliminate Integration Headaches with Embedded Analytics

Next Steps—How to Launch a Successful Warehouse

Analytics on the Rise

Once again, market analysts expect that spending for business intelligence technology will surpass most other IT categories this year, continuing an upward trend that began earlier this decade. How significant is the interest in BI? Last year, analytical applications and the underlying databases and related foundational technologies posted a 12.5 percent upsurge, and researchers are calling for continued healthy growth rates through 2011.

Fueling this activity is the rise of pervasive or real-time analytics, a new generation of easy-to-use tools that is spreading BI beyond small groups of statistical experts to help operations people like call center representatives and inventory managers do their jobs more efficiently. These new tools include electronic dashboards and business scorecards that are embedded within traditional business applications or work in pop-screens alongside them to give end users up-to-the-minute summaries of business conditions for better decision making.

While important, easily accessible end-user tools alone aren't enough to give organizations a competitive edge. Without ready access to organization-wide data that's accurate and complete, even the best analytical programs will produce dubious results. BI applications must be built on a reliable framework, which at its core consists of a data warehouse.

As data warehouses increase in importance, a number of new technology developments are making them easier to launch, maintain, and grow as an organization's needs evolve.

A data warehouse acts as a central repository for information gathered from across an organization, ranging from historical records to online transactional processing (OLTP) summaries to real-time operational data from CRM, ERP, and spreadsheet applications. Having a central warehouse means organizations can create a common view of data for a "single source of truth" that eliminates nagging inconsistencies, as when the sales and finance departments use different definitions for "revenues."

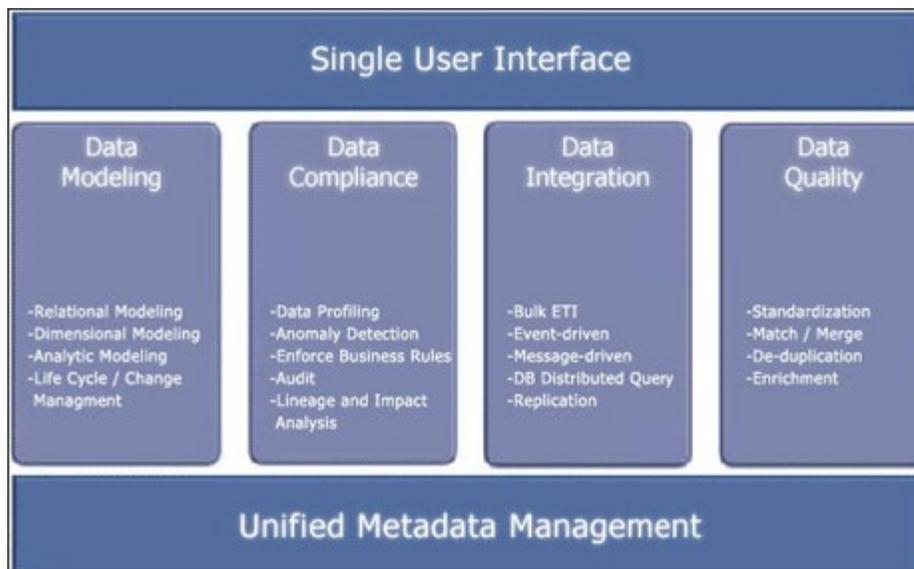
Having access to information that's accurate and complete isn't important just for business operations, it's become essential for regulatory compliance as a host of government rules continue to spring up in the Sarbanes-Oxley era.

Fortunately, as data warehouses increase in importance, a number of technology developments are making them easier to launch, maintain, and grow as an organization's needs evolve.

The Right Database Foundation Reduces Implementation Complexities

To meet the demands of today's BI applications a data warehouse must be built on a database foundation that provides scalability, high performance, and reliability. Oracle Database 11g offers these capabilities plus a wide range of additional features designed for the specific needs of data warehouses.

For example, included as a standard component of Oracle Database 11g is Oracle Warehouse Builder, which provides a data ETL (extract, transform, and load) tool, a metadata repository, and data-modeling capabilities that together help organizations design and quickly build large data warehouses. The implementation efficiencies result from having these tools all integrated within a common repository and user interface. This eliminates the historical data-warehouse challenge of forcing organizations to piece together separate tools from various vendors, a process that used to add excessive time and cost to projects.



In addition to Oracle Warehouse Builder's standard features, organizations can also choose three options to tailor the technology to their specific requirements. The Enterprise ETL option increases performance and productivity with advanced data-mappings impact analysis, change propagation and multi-configuration management support. Similarly, organizations can add the Data Quality option for extended data profiling and information compliance. Finally, Oracle offers connectors to integrate the

data warehouse with leading business applications, including the Oracle e-Business Suite, Peoplesoft Enterprise, Siebel (CRM), and SAP R/3.

New Strategies for Maintaining Performance

Once an organization builds and launches its data warehouse, Oracle Database 11g provides a range of additional features to guarantee the storehouse keeps running effectively. With the Database Resource Manager, a DBA can create groups of end-users according to their job functions and data needs and then allocate resources accordingly. For example, groups that routinely send complex queries with demanding response-time requirements can receive greater CPU resources than only occasional users of the warehouse.



In addition, a patented read-consistency model keeps data loads from slowing query performance. Oracle Database 11g eliminates the conflicts that formerly resulted from these two processes by using a technology called multi-version read consistency. This capability maintains the version of the data that existed when a user submitted the query, even if someone else updates the underlying information while the query is in progress. This allows organizations to keep warehouse data up to date without disrupting the analytical operations that business users require.

Reap the Value of Cubes

One of Oracle Database 11g's most important innovations for data warehousing is the ability to pack materialized views into OLAP (Online Analytical Processing) cubes. In the past, organizations created summary tables, also known as materialized views, to fine tune the performance of the star schemas that organized data in data warehouses. This approach improved performance by pre-computing one or more of these summaries and storing them in the data warehouse. Unfortunately, this strategy often forced

organizations to store thousands or perhaps millions of summaries for each star schema, which drained performance and storage resources.

Oracle Database 11g solves this problem by storing thousands of summaries in a single and easily managed OLAP cube that can be tightly compressed and efficiently updated. Thus, DBAs can optimize their data-warehouse resources while still achieving the higher performance of a star schema with thousands of relational materialized views. In addition, Oracle Database 11g allows data-warehouse users to access the OLAP cubes through standard SQL tools rather than more complicated OLAP-based APIs (Application Programming Interfaces).

Give Your Warehouse Room to Grow

Oracle Database 11g provides other important scalability features to assure data warehouses keep pace with growing organizations.

Oracle Partitioning breaks up large tables up into smaller pieces, which keeps query response times fast even as the size of the warehouse expands. The latest release of the Oracle database increases the options DBAs have for partitioning tables, including the ability to design custom partitioning schemes. Also included is a “partition adviser” tool that helps DBAs create effective partitions.

Building on existing algorithms that can compress information in ratios ranging from 2-4x, Oracle Database 11g adds compression support for frequent, online updates and inserts. In previous releases, the Oracle database only compressed large blocks of data as they were inserted into a table during bulk operations in which thousands of rows were being added. The new capability in Oracle Database 11g brings the benefits of compression to data warehouse tables where organizations update their warehouses frequently.

For scalability, Real Application Clusters (RAC) can run a single database across multiple servers, which means organizations can keep adding CPUs to maintain performance at desired levels as the data warehouse grows. RAC’s failover capabilities keep a data warehouse operational even if a server node fails.

RAC can also flexibly manage multiple workloads within a single database. For example, organizations may assign different applications to separate services, which can run on separate cluster nodes. Thus, queries may use one set of processing nodes while ETL operations tap another set. Administrators also have the choice of dynamically adjusting node workloads to boost ETL processing overnight and reallocate CPUs to query processing during the workday, for example.

Eliminate Integration Headaches with Embedded Analytics

Oracle Database 11g comes with OLAP, data mining, and statistics tools integrated inside the database engine, which eliminates the extra step of having to move data out of the data warehouse to these analytic engines. In addition, Oracle Database 11g provides SQL

access to all of its analytics, so users can perform queries with any standard SQL-based tool or application environment.

By having analytics integrated within the database, organizations don't have to worry whether or not their standalone OLAP server can scale to hundreds of CPU's or clusters of servers as their data warehouse increases in volume. In addition, this assures close integration between the data-mining engine and the user-authentication server. Similarly, organizations can easily integrate the results of statistical analyses with data warehouse data.

The tight integration of Oracle Database 11g's embedded OLAP-based materialized views brings the advantages of OLAP cubes, which provide scalable and compressed storage of dimensional data, fast incremental updates and query performance, and the ability to compute advanced analytical calculations.

Data Mining Made Easy

Oracle Data Mining, analytics software embedded in Oracle Database 11g, helps business users uncover business trends, market opportunities, and other insights hidden in their data. For example, Oracle Data Mining can help organizations target their best customers, find and prevent fraud, and discover the most influential attributes that affect Key Performance Indicators (KPIs). Oracle Data Mining also helps technical professionals find patterns in the data and develop new clusters and associations.

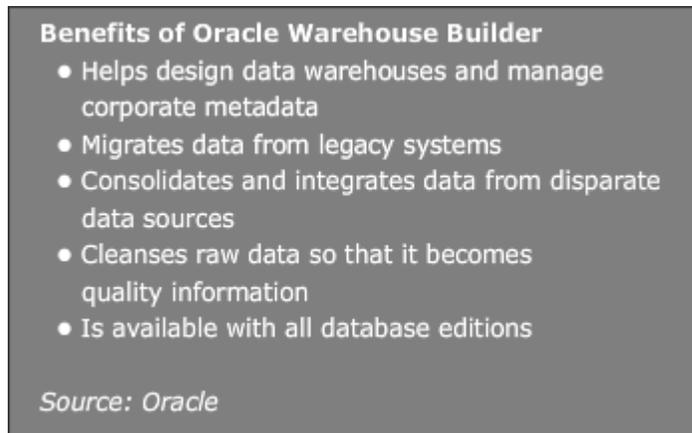
Underlying these capabilities are a host of sophisticated data-mining algorithms that use machine-learning techniques to analyze information. Experts divide data-mining algorithms into two broad categories, "supervised learning" and "unsupervised learning." Supervised learning requires the data analyst to identify a target attribute or dependent variable (for example, customers who bought a specific product). The supervised-learning technique then sifts through data trying to find patterns and relationships between other attributes and the target attribute (for example, the characteristics that indicate whether a prospective customer is likely to buy a specific product). Supervised learning algorithms with Oracle Data Mining include Naïve Bayes, Decision Tree, General Linear Models, and Support Vector Machines.

By contrast, with "unsupervised learning," target attributes don't exist. Instead, the data-mining algorithms look for associations and clusters in the data independent of any defined business objective. These algorithms include Enhanced k-Means Clustering, Orthogonal Partitioning Clustering, Association Rules (market basket analysis), and Nonnegative Matrix Factorization.

Oracle Data Mining includes Oracle Data Miner, a graphical user interface for data analysts to build, evaluate, and apply data mining models. Oracle Data Miner guides the data analyst through the data mining process with complete flexibility and presents results in graphical and tabular formats. Oracle Data Miner can generate the PL/SQL code associated with a mining activity.

Next Steps—How to Launch a Successful Warehouse

A data warehouse can quickly begin returning value to analysts and business users when it's built on a database designed with embedded analytics and enhancements for performance, scalability, and reliability. To get the implementation up and running smoothly, organizations can take advantage of vendor-sponsored support programs. For example, with its Optimized Warehouse Initiative Oracle joins forces with a number of key server, and storage partners to deliver a choice of solutions for easy-to-implement, high performance and scaleable data warehouses.



Benefits of Oracle Warehouse Builder

- Helps design data warehouses and manage corporate metadata
- Migrates data from legacy systems
- Consolidates and integrates data from disparate data sources
- Cleanses raw data so that it becomes quality information
- Is available with all database editions

Source: Oracle

The initiative offers two types of solutions depending on an organization's requirements. The Reference Configurations option consists of a best practices guide for selecting the right hardware components to match with Oracle databases. The best practices incorporate years of collective configuration experience from the various vendors. A range of configurations are available, from entry-level implementations run by a single server to multi-terabyte data warehouses using server clusters to support thousands of users.

A second choice, known as Oracle Optimized Warehouses, delivers a pre-built, optimized data warehouse with Oracle software, servers, storage and networking components ready to load data and run queries and reports. In addition to being pre-fabricated, the data warehouse has passed a full testing and validation process.

Both the reference configurations and the pre-built warehouses are designed for modular scalability to accept additional storage and processing resources as organizations grow.

Click here to learn more about data warehousing with Oracle Database 11g :
http://www.oracle.com/solutions/business_intelligence/dw_home.html

Flexibility Yields Results

As technical analysts and business-department end users both increase their reliance on up-to-the-minute information for decision making, it's easy to see why BI and data warehouses will continue to be hot technology categories into the next decade. Thanks to warehouse building blocks like Oracle Database 11g, organizations will spend less time

launching and maintaining these analytical applications and more time reaping their rewards.