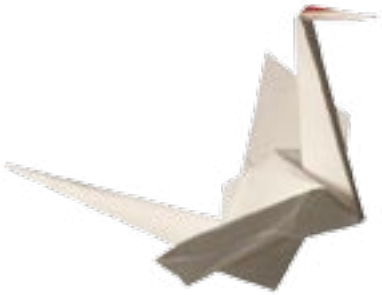
Data warehousing
White paper
August 2007

Information Management software

IBM DB2 Warehouse software: unlock enterprise information with embedded analytics and a central workbench.



Contents

Introduction

Information is the lifeblood of business processes and successful operations. And as information volumes and sources continue to grow rapidly, businesses that can deliver accurate, timely and contextual information from across the enterprise to users and business processes will be better poised to drive innovation rather than simply react to threats. To respond to this new business need, IT departments must address a host of challenges that range from siloed and widely distributed information and departmental datamarts; to rigid and non-scalable data warehouses and datamarts that can't cost-effectively and efficiently support today's growing data volumes' information management needs; to solutions that are difficult to use and to maintain.

IBM designed DB2® Warehouse software to help businesses overcome the limitations of current data warehousing approaches and to support next-generation data warehousing—or what IBM calls dynamic warehousing. IT departments can use DB2 Warehouse to:

- *Simplify enterprise data warehouse development, deployment and maintenance with a comprehensive integration solution.*
- *Streamline business intelligence processes and enhance flexibility with powerful, analytic solution components.*
- *Boost data warehouse performance and improve productivity.*

The foundation for DB2 Warehouse is the IBM DB2 9 platform, which underpins your warehouse environment with a highly parallel architecture that is massively scalable and provides high performance for mixed workload query processing. Standard features in DB2 9, such as data partitioning, row compression, multidimensional clustering and materialized query tables (MQTs), provide a powerful engine for overcoming evolving warehousing challenges.

Highlights

Unlike most currently available offerings, IBM DB2 Warehouse incorporates extended design analytics and visualization capabilities, making it easier to provide more users with improved access to different types of information.

DB2 Warehouse is unique because it provides a fully integrated "appliance-like" environment, with a variety of configurations that can be matched to your business requirements. Standard features and capabilities include modeling and design tools, embedded data movement and structured query language (SQL) transformation, data mining and visualization, enhanced online analytical processing (OLAP) design and optimization, embedded analytics and workload control. And, although deployment, configurability and scalability benefits are appliance-like, compared to most appliance offerings DB2 Warehouse has a greater depth and breadth of functionality. For example, it includes a fully integrated development and management interface, analytics and extract, transfer, load (ETL) capabilities.

Unlike most offerings available in the marketplace, DB2 Warehouse incorporates extended design, analytics and visualization capabilities, making it easier for you to give more users improved access to different types of information. Additionally, IBM created the DB2 Warehouse Design Studio tool to support the integration of OLAP and data mining with IBM Alphablox® software. The result is a robust, flexible platform that can support anything from departmental to enterprise warehousing needs—and full integration across the business.

In addition to fully integrated analytics capabilities, the IBM dynamic warehousing solution supports all of the leading business intelligence tools.

Given business users' rapidly evolving needs for near real-time information, analytics capabilities are especially important to today's data warehousing environments. Unfortunately, IT teams often struggle to keep up with users' needs because of poor integration between analytics and data warehouse tools. To address this challenge, IBM has created a vertical analytics strategy, which has been incorporated into DB2 Warehouse. Moreover, IBM not only offers fully integrated analytics capabilities, it also supports integration with all of the leading business intelligence tools.

Highlights

Traditional approaches to data warehousing offer limited, if any, connectivity between business analytics and process tools, requiring redundant copies of the data and increasing the time spent searching for information.

In DB2 Warehouse, the analytics capabilities are embedded with the database, so you can run multiple processes from a single copy of the data to both reduce storage costs and enable faster searches.

This white paper explores the analytics capabilities that are embedded in DB2 Warehouse and the associated benefits. It also examines the Design Studio component of DB2 Warehouse and how it simplifies the process of designing powerful, easy-to-use analytic structures.

Simplified searching and maintenance through embedded analytics

Historically, connectivity between business analysis tools and business process tools has been very limited or nonexistent. Management could only look back at operational data. And to get the information needed to make educated choices, decision makers had to spend large chunks of time searching the data warehouse. Today, everyone from managers to frontline decision makers needs right-time information to do his or her job effectively, so status quo approaches that separate analytics tools from the database are no longer viable.

What's unique about DB2 Warehouse analytics capabilities?

When analytics tools and the database are separate, the analytics tool has to pull data from the database to complete a process, resulting in several disadvantageous side effects. The challenges include slower searches and the creation of redundant copies of data for different systems, leading to higher storage and maintenance costs. In DB2 Warehouse, the analytics capabilities are embedded or integrated with the database, so data mining processes and multidimensional cubing are performed in the database. As a result, you benefit from the ability to run multiple processes from a single copy of data, lower data latency, easier maintenance and a reduced total cost of ownership (TCO). And if your data warehousing strategy is built around a business intelligence tool from another industry-leading provider like Business Objects, Cognos or Microstrategy, DB2 Warehouse is designed to integrate easily with your tool of choice.

Highlights

Managers and frontline decision makers can access analytics from the tools they use all the time.

Why should you care about embedded analytics?

When end users need information, the last thing they want to do is launch a business intelligence tool or sift through a list of reports or cubes for the data they need. It isn't their job, they often don't have the expertise, and either way it stifles productivity. That's why embedded analytics are so valuable. Users can access the analytics from the tools they use all of the time—such as a Web browser or spreadsheet.

In DB2 Warehouse, Alphablox provides the window through which business users can look at right-time, in-context information from across the enterprise—including structured and unstructured data from relational and multidimensional databases, transaction system and other content feeds—to support more informed decisions. Alphablox is used to develop and support information dashboards and enable interactive multidimensional data analysis for:

- *Self-service reporting and analysis.*
- *Operational analysis.*
- *Financial reporting and analysis.*
- *Planning.*
- *Business performance and key performance indicators (KPIs) for interactive information.*

Alphablox is provided with DB2 Warehouse to simplify development and support of information dashboards and enable interactive multidimensional data analysis.

A closer look at the differences in Alphablox

Alphablox and all analytics applications enabled with Alphablox run as Java™ Platform, Enterprise Edition (Java EE) technology–compliant applications in an application server, and they are accessed by using a Web browser. Unlike traditional query and reporting tools that interact with application servers, Alphablox leverages the application services, portal services and integration broker services provided by the application server. Moreover, Alphablox uses the common foundation of Java EE architecture for developing, deploying and maintaining distributed applications.

Highlights

Alphablox offers a component-based approach to building Web-based analytics applications and scales to support large numbers of users querying many different OLAP cubes.

Alphablox also addresses performance considerations related to an increasing number of users leveraging analytics capabilities. The Alphablox Relational Cubing Engine can provide high-performance OLAP analysis functionality for large numbers of users querying many different OLAP cubes. It enables quick, multidimensional analysis to relational data stored in a data warehouse or data-mart, facilitating the creation of multidimensional data sources in cases where full-featured OLAP data sources are cost prohibitive or impractical. The Relational Cubing Engine also supports a component-based approach to building Web-based analytics applications, making it ideal for custom business intelligence portals and integrated analytics for general applications.

From customizable business dashboards or business intelligence portals, Alphablox equips users with a range of sophisticated calculations and capabilities for performing analysis. Some of the key capabilities include ranking, derived calculations, ordering, filtering, percentiles, variances, standard deviations, correlations, trending and statistical functions.

Designed for flexibility, Alphablox can be customized to fit every user's specific job role and business needs.

Given the range of different business users' needs, IBM designed Alphablox to be highly flexible and customizable. In fact, you can customize Alphablox for every user's specific job role and business needs. Business dashboards that can deliver business KPIs that are relevant to users' jobs—in near real time—provide intuitive user access to business intelligence information. For example, using the DB2 Warehouse Design Studio tool, you can develop dashboards with easy-to-read maps, pie charts and other visual indicators that provide sales users with a breakdown of the top five best-selling products or most profitable regions. And you can provide finance users with monthly summary figures for sales, cost of goods, marketing, payrolls and profits. Best of all, users can customize Alphablox based on their preferences.

Highlights

Based on the open source Eclipse platform and IBM Rational Data Architect software, the DB2 Warehouse design workbench speeds development and deployment of analytics capabilities.

Standard design capabilities include data mapping, model and database compliance, impact analysis and data model comparison and synchronization—making Design Studio one of the most comprehensive development tools available today.

Streamline the development process with a central workbench

Given the variety of players involved, the development life cycle for business intelligence solutions has long been laden with collaboration challenges. With different job roles—including architects, warehouse administrators, OLAP developers, application developers and data stewards—often using different tools, development cycle time was prolonged by burdensome project coordination tasks.

How does Design Studio help streamline development?

DB2 Warehouse helps to streamline the development life cycle through the DB2 Warehouse Design Studio tool. Design Studio is the central, comprehensive workbench developers can use to model, design and package data warehouse components, including design and deployment of analytics. Based on the open source Eclipse platform and IBM Rational® Data Architect software for enterprise design and modeling tooling, Design Studio is one of the most complete and flexible warehousing workbench solutions available in the marketplace. Some of the key standard capabilities in Design Studio include:

- *Data mapping. Helps you identify and define relationships between data in the data warehouse model as well as in other sources or target models that you import into Design Studio.*
- *Model and database compliance. Enables developers to better ensure the models they create comply with enterprise standards.*
- *Impact analysis. Allows you to determine how a change to your model will impact functionality before you make the change.*
- *Data model comparison and synchronization. Enables you to synchronize new and old data models as you modify or upgrade them.*

Highlights

Design Studio enables a project-based approach and teaming capabilities that support enterprise warehouse consolidation.

With a full set of teaming capabilities tooling, Design Studio provides an intuitive, scalable and coherent platform for developing data warehouses. The project-based approach and teaming capabilities in Design Studio also support enterprise warehouse consolidation, so you can move toward more dynamic warehousing that delivers a real-time, single version of the truth.

The advantages of integrated tooling

Traditional data warehouse environments include the enterprise data warehouse and siloed physical datamarts. As a result, companies must pull data out of the various databases into multidimensional databases or third-party analytic tools, creating redundant copies of data and management challenges.

You can combine all of your datamarts and your enterprise data warehouse under one umbrella for a real-time, single version of the truth.

The integrated tooling in Design Studio enables you to consolidate all of your datamarts and your enterprise data warehouse under a single umbrella, so you can use the same infrastructure to build analytic structures for data mining and to create data mining flows and models. In other words, you can build multidimensional structures and deploy them to the tools of your choice or Alphablox from a single location. Moreover, using the Alphablox relational cubing engine, you can perform true multidimensional cubing without removing data from the database. The benefits include:

- *Fewer copies of data.*
- *Lower data latency.*
- *Quicker time to value.*
- *Lower TCO.*

Highlights

By adopting DB2 Warehouse and Design Studio, you can enable developers to work on data warehouse design and dashboard design for Web-based analytics in a common development environment. And you can also embed analytics components into data warehouse life-cycle management.

Developers can use the Design Studio graphical data mining flow editor to explore and visualize data prior to developing a physical data model.

Developing mining models in Design Studio

Using the Design Studio workbench, developers can create physical data models from scratch or reverse engineer them from existing databases. Moreover, they can explore and visualize data to better understand it prior to developing a model. DB2 Design Studio provides a graphical data mining flow editor that provides a palette of data transformation and data mining operators. The Design Studio editor provides a palette of data transformation and data mining operators and sequencing connectors that enable the data miner to define a data mining algorithm, data inputs and outputs and any data preparation transformations needed to create an executable data mining model. Users can verify and test data mining flows in the Design Studio environment the same way that they can verify and test SQL warehousing (SQW) data flows and control flows.

Highlights

Verifying and testing data mining flows in the Design Studio environment is the same as verifying and testing SQW data and control flows.

The data mining component comprises three modules, including modules for modeling, scoring and visualization. Scoring is a database-level execution of a data mining model. Scoring can use models built in the modeling module or can import a Predictive Modeling Markup Language (PMML) model from other data mining vendors for execution against data in DB2 Warehouse. The modeling module contains data mining algorithms designed to address a wide spectrum of data insights, including:

- *Clustering.*
- *Association.*
- *Classification.*
- *Prediction.*

Users can leverage the visualization module to view a graphical analysis of the analysis performed by the model. The visualization module helps the requester better understand the data, rather than simply accepting the answer provided by the algorithm. Users can package and deploy completed data mining models for execution in the run-time data warehouse environment using the DB2 Warehouse Administration Console. Users can also review data mining results through features in Alphablox.

Highlights

Built to integrate readily with existing technologies and to support next-generation warehousing requirements, DB2 Warehouse can help you meet your enterprise business needs for years to come.

Meet current and future data warehousing needs

The ability to effectively manage information and provide it to users and applications on demand is an increasingly critical function of IT. And the choices you make for your data warehouse environment today will affect your information management strategy and your ability to meet enterprise business needs for years to come.

Recognizing the importance of flexibility and greater simplicity in data warehousing environments, IBM built DB2 Warehouse to integrate readily with existing investments and to support next-generation dynamic warehousing requirements. And IBM can provide it in easy-to-install appliance-like packages with full worldwide support. So when you go with DB2 Warehouse, you can be confident that you are choosing a reliable solution that meets your current needs and can easily scale as business demands grow.

For more information

To learn more about IBM DB2 Warehouse software, contact your IBM sales representative or visit:

ibm.com/bi

IBM Corporation
Software Group
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
08-07

IMW10920-USEN-00