White paper
August 2007

Information Management software

Which DB2 Warehouse is right for you?

Contents

Introduction

Comprising the IBM foundation of dynamic warehousing, IBM DB2® Warehouse software provides a complete platform for delivering functional and scalable reporting and data warehousing solutions. A fully integrated environment, DB2 Warehouse is built around IBM DB2 9 server technology on Linux®, UNIX® and Microsoft® Windows® platforms. It provides common development and management user interfaces and supports application development, data modeling and mapping, SQL transformation, OLAP and data mining functionality.

"IBM, with these announcements, becomes the premier data warehousing appliance vendor, in terms of the range of targeted solutions they provide."

—James Kobielus, data management analyst, Current Analysis

The completeness, sophistication and simplicity of DB2 Warehouse make it a compelling offering for any organization. It, for the first time, brings to the marketplace a warehouse deployment workbench that integrates enterprise data modeling, OLAP design and deployment, data mining and integration with external transformations (such as IBM Information Server software), data quality and development metadata management.

Highlights

Integrated and easy to use, DB2 Warehouse solutions address the full data warehouse life cycle.

DB2 Warehouse provides support for exchanging data mining models with workbench providers such a SPSS or SAS, as well as the capability to interchange OLAP models with leading business intelligence (BI) vendors such as Microstrategy, Cognos and Business Objects. Additional features include operational warehouse metadata management and embedded analytical application capabilities. Many vendors claim to deliver these capabilities. However, when you look closely at the solutions, you'll see that only DB2 Warehouse truly addresses the full data warehouse life cycle and is integrated, sophisticated and easy to use.

One key focus point for any data warehouse is the time it takes to get up and running. To this end, IBM provides automated installation options, starting with an integrated application that installs and configures your environment the way you want it. For the DB2 Warehouse Starter, DB2 Warehouse Intermediate and DB2 Warehouse Advanced offerings, IBM provides a prebuilt run-time image for building IBM Balanced Warehouse environments C- and D-class offerings.

IBM provides automated installation options that allow you to rapidly configure your environment the way you want it.

DB2 Warehouse has been packaged to provide you with configuration options that enable your organization to start small and iteratively grow its investment in its warehouse or reporting environment in line with growing business requirements. This, therefore, cost-effectively supports your business needs without compromising functionality or performance.

DB2 Warehouse software is available in six packages:

- *DB2 Warehouse Starter*
- *DB2 Warehouse Intermediate*
- *DB2 Warehouse Advanced*
- *DB2 Warehouse Enterprise Base*
- *DB2 Warehouse Enterprise*
- *DB2 Warehouse Developer*

Highlights

Figure 1 illustrates how these packages are positioned in terms of growth in user and data volumes, but of course these are not the only decision factors in choosing the right edition. Organizations' analytical and data mining requirements, as well as their future growth requirements, may also drive this decision.

IBM offers a spectrum of DB2 packages, designed to meet a range of business needs, and user and data volumes.

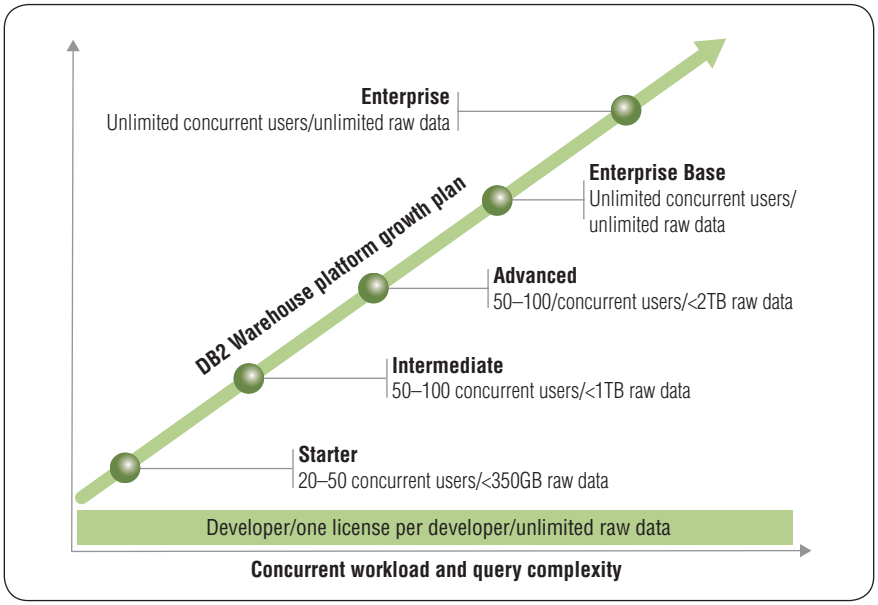

Figure 1. DB2 Warehouse growth path

DB2 Warehouse provides a common development and administration interface and teaming capabilities to assure an integrated development cycle.

The DB2 Warehouse platform creates an environment in which database administrators (DBAs), data architects, BI designers and BI deployment specialists can collaborate in a cohesive and integrated manner on a warehouse project. DB2 Warehouse achieves this by providing a common development interface, administration interface and collaboration and teaming capabilities to assure an integrated approach from design through to deployment.

Highlights

But DB2 Warehouse is more than a data warehouse platform, it also has data modeling, in-line data mining, OLAP design and deployment, workload management and other robust capabilities.

Over and above being an industry-leading data warehouse platform, DB2 Warehouse stands out for its capabilities in the following areas (see figure 2):

- *Data modeling*
- *Database transformation*
- *In-line data mining*
- *OLAP design and deployment*
- *Performance optimization feature (IBM DB2 Query Patroller software)*
- *Deep data compression*
- *Embedded analytics*

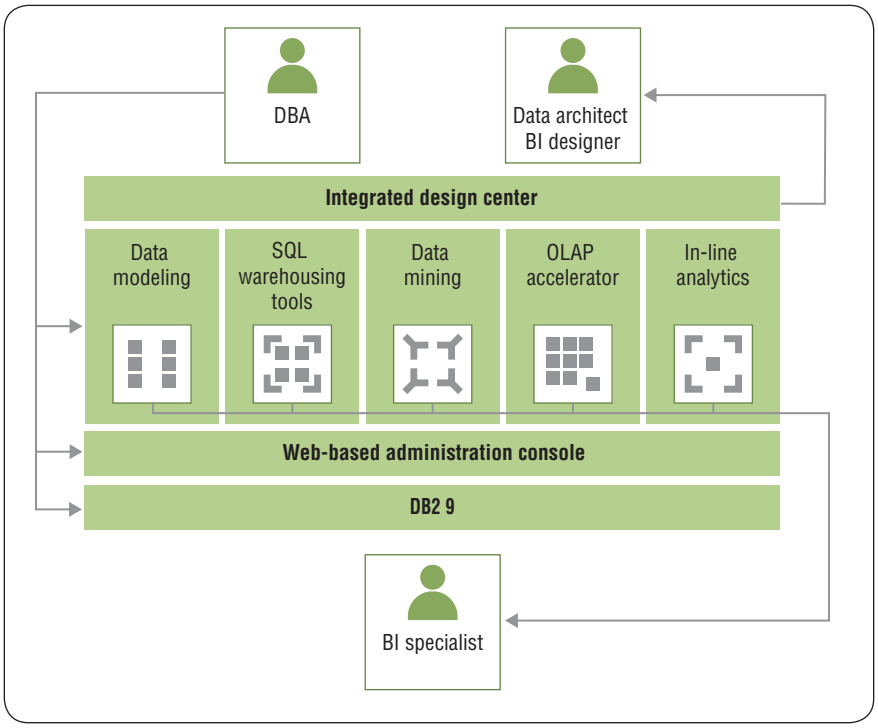

Figure 2. DB2 Warehouse layered architecture

Highlights

This paper will take you through a detailed overview of DB2 Warehouse, illustrating the different DB2 Warehouse technical and functional components, capabilities and packaging options.

What's in IBM DB2 Warehouse?

The following section describes the components that make up the DB2 Warehouse offering and outlines their functions and capabilities.

DB2 Warehouse architecture

Modular and component based, DB2 Warehouse uses open technologies such as Eclipse, Java and WebSphere solutions.

The DB2 Warehouse architecture is component based and uses best-of-breed open technologies including Eclipse, DB2 software, Java™ Platform, Enterprise Edition (Jave EE) technology and IBM WebSphere® Application Server middleware.

IBM DB2 Warehouse offerings consist of four primary technical components:

Four primary architectural components make up the DB2 Warehouse solution: IBM DB2 9 server, IBM WebSphere Application Server software, IBM DB2 Warehouse Design Studio software and IBM DB2 Alphablox embedded analytics.

- *IBM DB2 9 server. Provides the foundation for all warehouse operations. It provides the execution platform for in-warehouse data transformations, data mining and OLAP. In addition, DB2 9 underpins these capabilities with DB2 workload management, partitioning and DB2 deep compression capabilities. It is also extendable to include non-DB2 relational data through the IBM pureXML™ option. This can be purchased in addition to DB2, although it is not part of the DB2 Warehouse packaging.*
- *IBM WebSphere Application Server software. Provides administration server capabilities for the warehouse environment, where OLAP, mining and data transformations can be administered and monitored. Plus it provides the deployment platform for IBM DB2 Alphablox® reporting applications.*

Highlights

- *IBM DB2 Warehouse Design Studio software. Delivers an Eclipse platform-based workbench where all design and build activities take place for the data warehouse, including data model design and deployment, data transformations, OLAP design and data mining development and deployment. It also includes deployment and integration capabilities for the Information Server software suite.*
- *IBM DB2 Alphablox embedded analytics* ***(integrated development environment [IDE]) workbench.*** *Provides a stand-alone environment, based on Eclipse. Developers can use it for developing DB2 Alphablox analytical applications for BI reporting.*

As shown in figure 3, these DB2 Warehouse application components are built upon proven technologies and are integrated using metadata.

The four primary components of DB2 Warehouse are built on proven technologies and integrated using metadata.

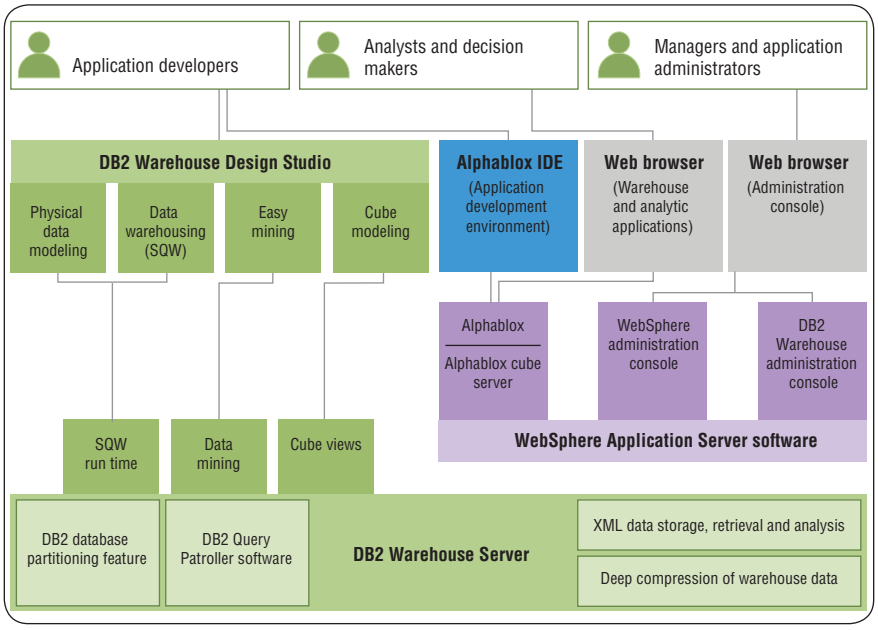

Figure 3. DB2 Warehouse application architecture

Highlights

DB2 Warehouse contains three primary functional components: DB2 9, DB2 Warehouse Design Studio and DB2 Administration Server software. The solution also includes optional technologies that provide additional functionality.

DB2 Warehouse functional components

Summarized in the table below, the functional components contained in DB2 Warehouse include DB2 9, IBM DB2 Warehouse Design Studio and IBM DB2 Administration Server software, plus a range of optional technologies that deliver advanced capabilities.

DB2 Warehouse functional components at a glance	
DB2 9	• Data server
DB2 Warehouse Design Studio	• Data modeling • Warehouse building (SQL warehousing) • OLAP design and deployment • Data mining interface • Information Server integration
DB2 Administration Server	• Administration console
Optional features	• Range partitioning and MDC tables • Database partitioning feature • Performance optimization feature • Storage optimization feature • DB2 Alphablox and connectors • Data mining

Highlights

The heart of DB2 Warehouse is DB2 9 Enterprise Data Server software—the perfect foundation for small, midsize and large businesses that want to build high-availability, high-volume enterprise-wide solutions.

While the DB2 9 architecture and advanced optimizer allow users to consolidate and migrate physical datamarts into a single warehouse environment, DB2 data partitioning allows businesses to support large databases, complex workloads and increased parallelism.

DB2 9 server

The core engine of DB2 Warehouse is the industry-leading DB2 9 Enterprise Data Server, designed to meet the needs of small, midsize and large businesses. DB2 9 can be deployed on servers of any size—from one CPU or server to hundreds of CPUs and servers. In addition, DB2 9 Enterprise Data Server is an ideal foundation for building on demand, enterprise-wide solutions, such as multiterabyte data warehouses, highly available operational systems, high-volume online transaction processing (OLTP) systems and Web-based BI solutions. It achieves this through its highly scalable and robust architecture combined with a comprehensive set of availability capabilities.

DB2 9 data partitioning provides businesses with the scalability to support very large databases, common in the data warehousing environment, as well as complex workloads and increased parallelism for administration tasks. DB2 provides advanced workload distribution through partitioning and data awareness, ensuring optimized query execution and resource utilization.

The DB2 architecture and advanced optimizer enable the consolidation and migration of physical datamarts throughout the organization into a single warehouse environment and model. This environment enables logical or physical implementations of the datamarts, thus placing it ahead of its rivals, which typically need dependent datamarts or highly denormalized data models aided by SQL hints for high-performance reporting. The DB2 approach minimizes storage, management and processing overhead, while preserving user functionality and performance.

Highlights

Relational databases like DB2 are making their way to the forefront because workloads are getting more complex and transactions are becoming more voluminous.

One key reason why relational databases, such as DB2, are taking the lead by managing both the analytical and operation BI areas is articulated in Gartner's 2006 Data Warehouse Magic Quadrant: "As a direct effect of the mixed workload, with continuous loading and the increase in automated transactions from the functional analytics in OLTP, the transactional DBMSs [database management servers] have an edge that challenges the DW [data warehousing] DBMSs."[1]

Further to this, DB2 9 has demonstrated and proven performance as shown by a Transaction Processing Performance Council (TPC) benchmark leader position for both TPC-H and TPC-C and customer benchmark records. For more information on DB2 performance leadership, visit the DB2 benchmark performance Web page.

DB2 Warehouse Design Studio can help ensure that the DB2 Warehouse deployment aligns with model development and refinement.

DB2 Warehouse Design Studio

The end-to-end capabilities of DB2 Warehouse Design Studio make it the most complete warehousing workbench in the marketplace today, with the breadth and depth you need in a warehousing environment, whether you are deploying a small reporting solution or a real-time enterprise data warehouse. Built on an enterprise-strength data modeling capability in IBM Rational® Data Architect software, DB2 Warehouse Design Studio helps ensure that the DB2 Warehouse deployment aligns with model development and refinement, and provides the ability to analyze the impact of changes to the warehouse. It also reduces the number of tools required for a warehouse deployment—thereby reducing the time to deploy and incrementally change the warehouse.

Highlights

Because it integrates with IBM Rational and IBM Information Server software, DB2 Warehouse Design Studio can link into enterprise information and application architectures.

In addition to the physical modeling and mapping functionality of Rational Data Architect software, DB2 Warehouse Design Studio deploys a set of plug-in tools for performing data transformation using SQL, OLAP design and deployment, and data mining. And its integration points with the Rational and Information Server product sets ensure it is linked into enterprise information and application architectures.

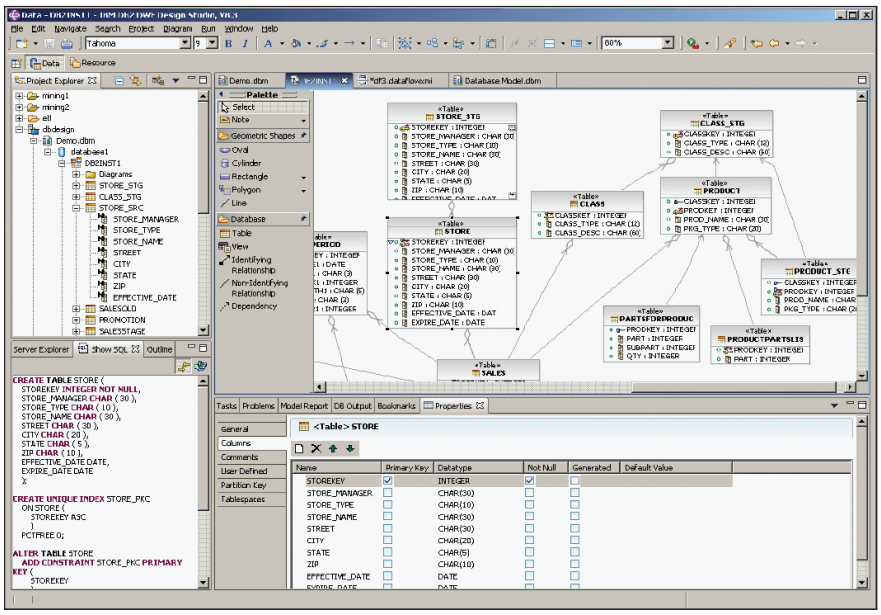

Figure 4. DB2 Warehouse Design Studio interface

Highlights

DB2 Warehouse Design Studio delivers essential data modeling capabilities that allow you to uncover and define data relationships, perform model and database compliance analysis, analyze the impacts of changes before you make them, and compare and synchronize data models.

Data modeling

With the underlying capabilities inherited from Rational Data Architect software, you can use DB2 Warehouse Design Studio to create physical models from scratch or from your databases using reverse engineering. Key data modeling capabilities delivered in DB2 Warehouse Design Studio include:

- *Design. The data mapping capabilities within the DB2 Warehouse Design Studio help uncover and define relationships within the data warehouse model and across other source or target models that have been imported into DB2 Warehouse Design Studio. It uses a unique mapping discovery capability to identify relationships and enables users to visualize the relationships and gain a cohesive view of the relationships.*
- *Model and compliance. The DB2 Warehouse Design Studio can perform model and database compliance analysis. This allows warehouse developers to analyze their model conformance to enterprise standards, norms and rules, thus ensuring the models they are deploying comply with enterprise standards, and, in a collaborative development environment, it ensures everyone is modeling to the same standard.*
- *Impact analysis. DB2 Warehouse Design Studio provides impact analysis functionality to identify the impact of a change before the change is actually implemented. This function lists the dependencies on a specific element and visually represents the results and provides the capability to generate a report for easy viewing.*
- *Compare and synchronize. DB2 Warehouse Design Studio can compare and synchronize data models. This assists in both day-to-day warehouse modifications as well as in major upgrades to the warehouse environment. By comparing the new and old models, DB2 Warehouse Design Studio then generates a differentiation script, including one that creates and updates Data Definition Language (DDL) for execution of the DDL on the database server.*

Highlights

Provided by DB2 Warehouse Design Studio after models have been defined and mapped, a SQL warehousing design tool can be used to build data and control flows for transformations. These can then be compiled and run within the warehouse or in an ETL tool.

SQL warehousing tool

Once models have been defined and mapped, DB2 Warehouse Design Studio provides a SQL warehousing (SQW) design tool with a graphical interface for data warehouse transformation design. Using a design canvas, the tool includes a palette with data, control and transformation operators that can be used to build data and control flows for transformations. These are then compiled into DB2 technology-specific SQL operations and can be run within the warehouse or optionally in an extract-transform-load (ETL) tool such as Information Server. With or without an IBM or third-party ETL tool, the SQL warehousing tool enables you to more efficiently prepare and populate data within the warehouse. Typically, this is done to support the building of analytic structures necessary for data mining, multidimensional and embedded analytics. It also can be integrated with IBM WebSphere DataStage software for Information Server software–built jobs by executing WebSphere DataStage software from within DB2 Warehouse control flows or by exporting DB2 Warehouse data flows to WebSphere DataStage software.

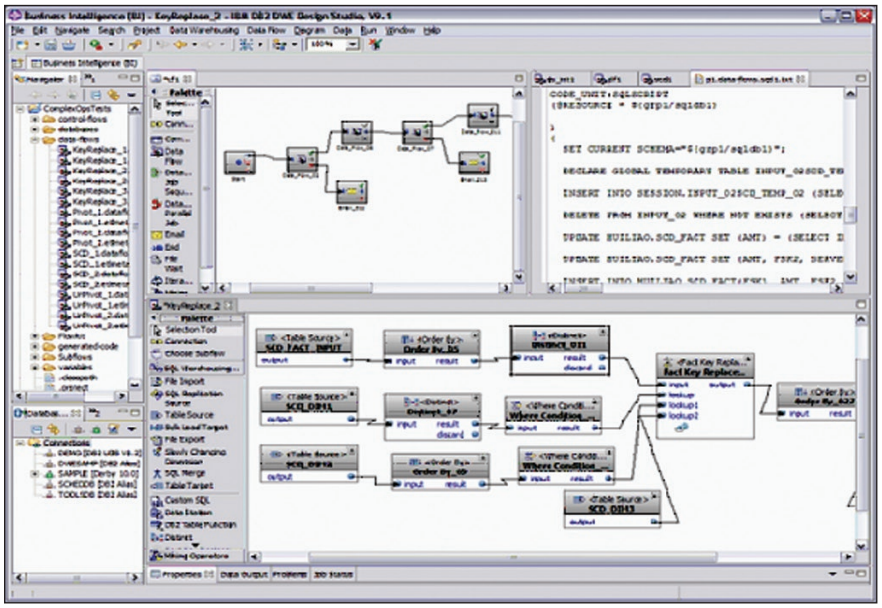

Figure 5. Illustration of SQW tool design interface

Highlights

In addition to the SQL warehouse tool capabilities, DB2 Warehouse Design Studio provides database development capabilities that enable users to import, compare, edit, test and debug SQL, and choose from a range of deployment options for exporting and batch deployment of generated SQL code, including stored procedures and user-defined functions.

The SQL warehousing tool allows you to create, edit and manage logical flows of high-level operations and translate them into optimized SQL code for execution.

Building transformation jobs with the SQL warehousing tool

The SQL warehousing tool quickly and easily solves data integration problems in a DB2 Warehouse environment and takes advantage of the scalable DB2 parallel architecture. Users can model logical flows of higher-level operations, which generate units of code that are organized inside execution plans. The tool provides a metadata system and a design interface. Within the interface you can create, edit and manage these flows and translate them into optimized SQL code for execution. When flow development is complete, the run-time engine moves packaged jobs and associated artifacts into a data warehouse application that can then be deployed into various target run-time systems.

Highlights

You can create two kinds of flows using the SQL warehousing tool: data flows and control flows. Data flows represent the movement of data from a source through a transformation and into a target database.

Within the SQL warehousing tools, you can create two types of flows:

- *Data flows. As shown in figure 6, data flows represent the movement of data from sources through transformation steps and into target databases. These include SQL data flows that use the SQL-processing power of DB2 for warehouse building operations that work on data from relational tables and flat files; mining flows that integrate key data mining operations into a SQL-based model; and SQL data flows that incorporate WebSphere DataStage ETL jobs as subflows.*

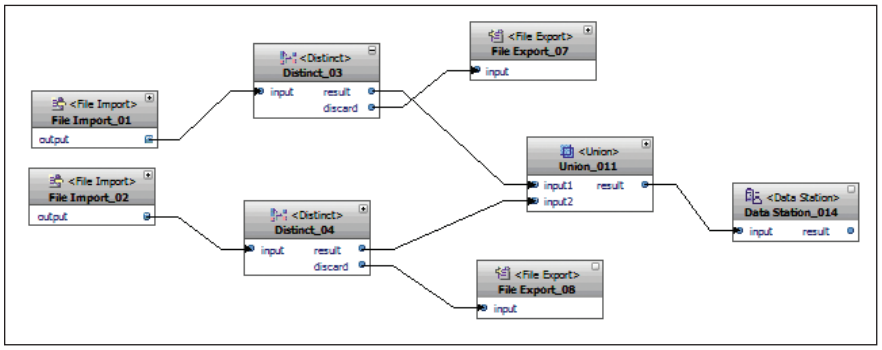

Figure 6. Example of a data flow

Highlights

Control flows initiate the movement of a set of related data flows and define the processing rules for those flows. DB2 9 and WebSphere Application Server can be used to manage various steps in a control flow.

- ***Control flows.*** *As shown in figure 7, control flows order a set of related data flows and define processing rules for the execution of those flows. This function assembles workflows of one or more data flows, external applications, WebSphere DataStage jobs and control logic for warehouse applications. Control flow processes can be executed and managed using DB2 9 as the run-time engine and WebSphere Application Server warehouse administration software for control and scheduling, as well as for the run-time deployment environment for data warehouse applications. Administrators can access the administration console through a browser-based interface.*

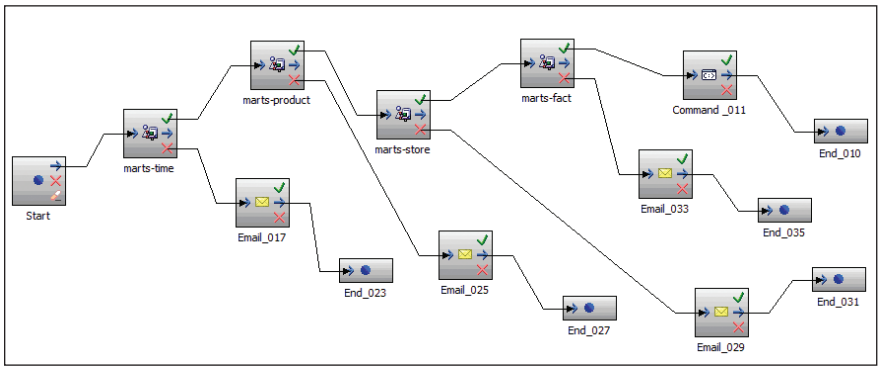

Figure 7. Example of a control flow

When a transformation is ready to deploy, the user can generate a deployment package that contains all of the files and parameters that form a complete application. Once this is done, the application is ready for deployment and execution through the administration console and server.

Highlights

The DB2 Warehouse Design Studio OLAP acceleration feature allows you to build and deploy hybrid OLAP structures within the data warehouse, and it makes optimization recommendations to improve the performance of integrated OLAP tools.

After an initial round of optimization recommendations, the OLAP acceleration feature can be reapplied to the warehouse after refinements are made.

OLAP design and deployment (OLAP acceleration feature)

DB2 Warehouse dramatically improves OLAP application deployment and query performance through the OLAP design and acceleration feature. A core component of data warehousing, OLAP gives users the ability to analyze business data using intuitive multidimensional and hierarchical navigation from summary levels to detailed data (slice, dice and drill).

The DB2 Warehouse Design Studio provides a tool for building and deploying hybrid OLAP structures within the warehouse based on DB2 cube views technology—tools that enable you to create, edit, import, export and deploy OLAP models over the relational warehouse schema. It also makes optimization recommendations to dramatically improve the performance of integrated OLAP tools. Newly enhanced features provide deep integration with the embedded analytics capabilities, as well as relational cubing for optimized browser-based embedded OLAP analytics.

Once OLAP structures are designed over the top of the data warehouse, an optimization function can be run over the top of the warehouse to optimize OLAP performance for the deployed models within the warehouse. The metadata from the OLAP structures is also exchanged with tools such as the Information Server MetaServer, Cognos ReportNet and Business Objects XI. Users of those tools are able to immediately get up and running with the OLAP models. If necessary, users can refine the OLAP structures and feedback changes for reoptimization to the data warehouse environment. This reduces time to delivery of OLAP and dimensional reports, and helps ensure consistency between OLAP structures in the warehouse and the various reporting tools that access the cubes.

Highlights

Integration between DB2 Warehouse Design Studio and Rational products enables dispersed development teams to collaborate on projects and reduces the number of work environments and applications developers need to run on their desktops.

Data mining interface

DB2 Warehouse Design Studio provides data mining visualization and development capabilities for integrating third-party (SAS, SPSS and others)-generated mining models into the data warehouse, as well as the ability to develop new models. This capability comes as an optional feature of DB2 Warehouse.

Information Server integration

DB2 Warehouse Design Studio enables developers to browse WebSphere DataStage servers and use DataStage jobs within control flows. It also has the ability to export DB2 Warehouse data flows for execution by WebSphere DataStage software.

Points of integration with other IBM products

DB2 Warehouse Design Studio plug-ins can be directly integrated into the full version Rational Data Architect and IBM Rational Software Architect software, allowing broader development teams to work together on projects and minimizing the number of Eclipse work environments and applications individuals need to run on their desktops. When you integrate DB2 Warehouse Design Studio with licensed Rational Data Architect software, you can access user requirements stored and managed in IBM Rational RequisitePro® software and then associate them to corresponding modeling elements and synchronize them with user-selectable rules. Plus, modeling files can be managed by IBM Rational ClearCase® or Concurrent Versions System (CVS), to provide seamless versioning, branching and synchronization of changes and provide full support for a collaborative development environment.

Highlights

The DB2 Warehouse administration console enables you to deploy and monitor warehouse objects and applications, and administer SQL warehousing, OLAP, mining and DB2 Alphablox functionality.

DB2 Administration Server

The DB2 Warehouse administration console provides an interface whereby warehouse administrators can deploy and monitor data warehouse objects and applications. The following are some of the functions provided by the administration console:

- *SQL warehousing. Deploy, schedule and monitor data warehouse applications that were created in the DB2 Warehouse Design Studio. View statistics and logs associated with processes and troubleshoot run-time failures.*
- *OLAP. Import and export cube models, use the OLAP optimizer and display metadata content for the cube model, such as its tables, joins, measures and attributes.*
- *Mining. View, export, update and delete models in the mining database. You can also import mining models into the database and load models into the cache. The mining visualization tool provides graphic representations of the results of the mining model.*
- *DB2 Alphablox. Launch the native DB2 Alphablox administration environment for configuring and administering Alphablox functions.*

Highlights

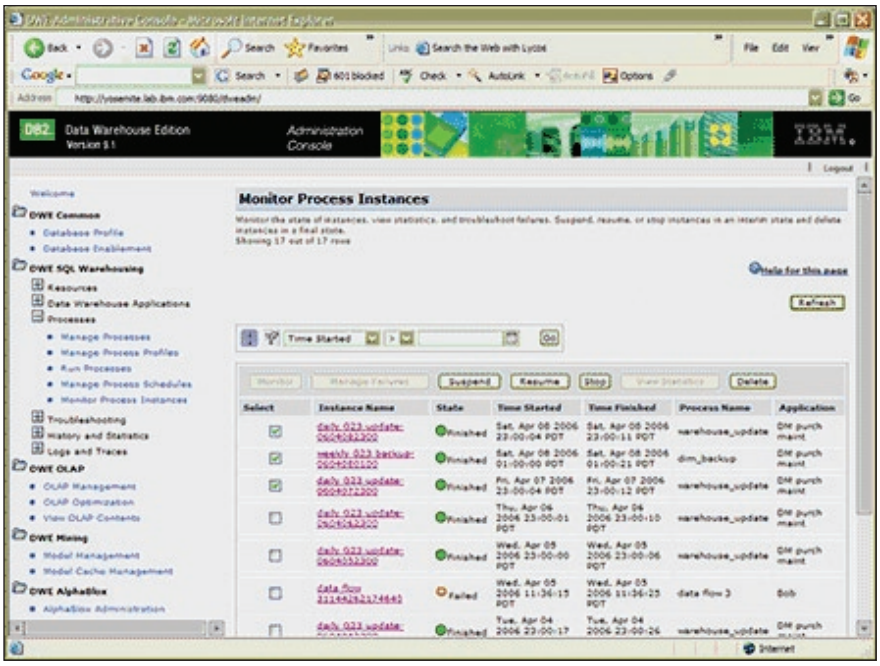

Figure 8. DB2 Warehouse Administration Console

For DB2 Warehouse Center users, a migration tool is available to migrate your jobs to SQW. For more information, visit the DB2 Warehouse Center Migration page on the IBM developerWorks® Web site.

Different DB2 Warehouse versions include a variety of optional features that can help you adapt the warehouse to your business needs.

Optional features

Optional features for the different DB2 Warehouse versions provide extensions to the base components that can help you build and adapt your warehouse in line with your business. The following section provides insight into various optional capabilities and how they can assist you in creating your desired warehouse environment.

Highlights

DB2 range partitioning and MDC tables

The DB2 table (range) partitioning and multidimensional clustering (MDC) table options are available in all DB2 Warehouse packages except the Starter package. These two capabilities enable more effective management and access to data within the warehouse, thus reducing the amount of administration required and increasing the performance and usability of the data warehouse.

Table partitioning improves query processing, simplifies administration and enables you to easily roll in and roll out table data.

Table partitioning

Table partitioning is a data organization scheme in which table data is divided across multiple storage objects, called data partitions or ranges, according to values in one or more table columns. Each data partition is stored separately. These storage objects can be in different table spaces, in the same table space, or in a combination of both. As a result, table partitioning offers easy roll-in and rollout of table data, simplified administration, flexible index placement and better query processing.

Consider the numerous benefits of table partitioning if:

- *You have a data warehouse that would benefit from easier roll-in and rollout of table data.*
- *You have a data warehouse that includes large tables.*
- *You are considering a migration to a DB2 9 database from a previous release or a competitive database product.*
- *You need to use hierarchical storage management (HSM) solutions more effectively.*

Highlights

MDC tables, which block data based on dimensional combinations, can help improve query performance and reduce data maintenance overhead.

MDC tables

MDC provides an elegant method for automatically clustering data in tables along multiple dimensions in a flexible, continuous manner. MDC can significantly improve query performance, especially for OLAP and dimensional-style data models. In addition, MDC can significantly reduce the overhead of data maintenance, such as reorganization and index maintenance operations during insert, update and delete operations.

MDC blocks data based upon dimensional combinations. For example, figure 9 shows a cube whereby data is blocked by two criteria: nation and color. As the data for each unique dimensional combination is kept in one contiguous block, only one index entry point is required for the block, not for each of the individual rows contained within it. Therefore, MDC tables contain fragmentation to within a block and minimize the size of each table's index. It also means fewer input/output (I/O) and paging operations, leading to greater performance under query workload.

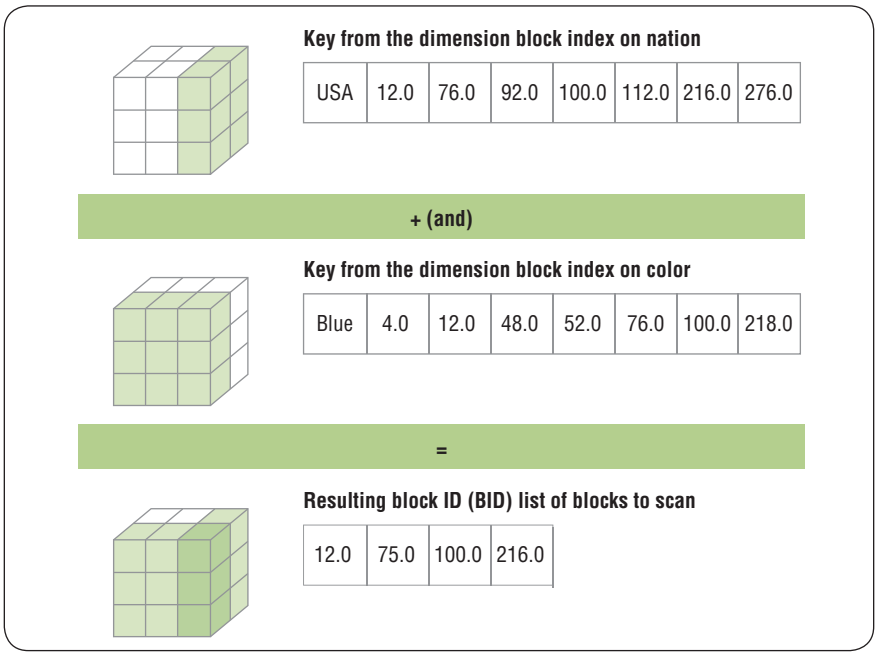

Figure 9. Illustration of how MDCs block data

Highlights

The DB2 database partitioning feature allows you to partition your database across multiple servers, providing a scalable, high-performance and easy-to-manage environment.

Database partitioning feature

The DB2 database partitioning feature delivers the ability to partition your database across multiple servers. It can provide organizations with the scalability to support very large databases and complex workloads, common to the data warehousing environment. To achieve this end, the DB2 database partitioning feature provides parallel query execution and increased parallelism for administration tasks. It enables advanced workload distribution through partitioning and data awareness, ensuring optimized query execution and resource utilization. In combination with MDCs and data partitioning, it provides a perfect architecture for managing data within your warehouse.

Figure 10 shows an example of how the data partitioning feature manages the distribution of data across the database partitions. The range-partitioned tables with each database partition organize the data into tables by *month,* and the MDCs block the data in those tables by *region.* This, in turn, leads to fewer indexing requirements, more parallel querying and more directed queries. It therefore provides a higher-performance and easier-to-manage environment.

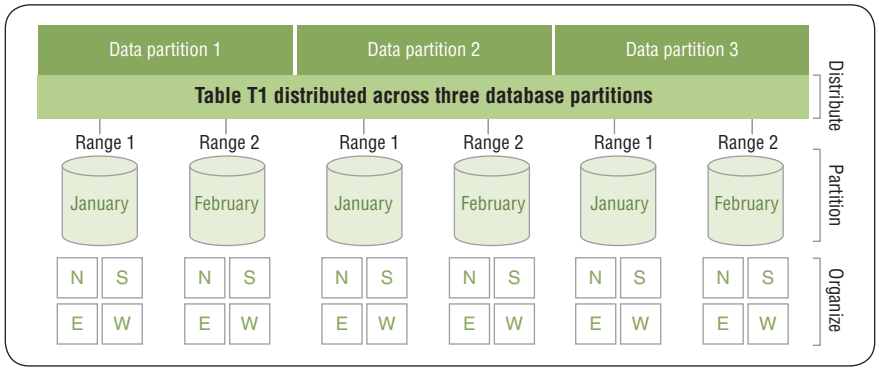

Figure 10. Illustration of three layers of data partitioning in DB2 Warehouse

Highlights

The DB2 Query Patroller feature, which provides SQL query performance management and monitoring, helps you effectively control your data warehouse workloads.

Using the DB2 storage optimization feature, you can compress your data to optimize performance and reduce memory and disk storage requirements.

Performance optimization feature (DB2 Query Patroller only)

DB2 Warehouse packages that include the DB2 performance optimization feature include only the DB2 Query Patroller component of the feature. The full performance optimization feature package includes DB2 Performance Expert for performance monitoring and DB2 Query Patroller for workload management. This section will cover only the DB2 Query Patroller.

DB2 Query Patroller provides SQL query performance management and monitoring, thus enabling more effective control of the workloads operating within the data warehouse. This feature helps reduce the impact of heavy users on light users, and helps ensure business service-level requirements are met. It can also provide information that can be used for charge-back accounting on warehouse use. To achieve this, it combines user/group and query classification with query cost threshold profiles, which then dictate how queries can be prioritized, scheduled and executed—in a manner that optimizes available system resources and throughput.

DB2 storage optimization feature

Made up of row-level compression and backup compression services, the DB2 storage optimization feature provides storage compression services to optimize the performance and footprint of your data. This feature can reduce the memory and disk storage requirement by up to 80 percent for a traditional data warehouse table, compared to other non-DB2 platforms.[2] The storage optimization feature not only helps reduce the amount of storage you require, but can also help achieve a substantial performance improvement, reducing I/O times and CPU cycles for the same workload, especially for I/O-bound queries. This can help reduce your overall expenditure on hardware, software, environment management and computing center power, and cooling requirements.

Highlights

DB2 9 customers have reported space savings in excess of 50 percent using the DB2 row compression feature.

The amount of space savings achieved when using the DB2 data row compression feature can vary depending on the data. DB2 9 customers have reported savings in excess of 50 percent and up to 80 percent for certain large database installations.[3] For one customer data set, a table was reduced by 76.4 percent. Another report has shown an initial saving of over US$2 million in disk savings alone by using the DB2 storage optimization feature.

Compression in databases has traditionally been seen as a performance inhibitor. For highly I/O bound systems, however, DB2 compression can substantially reduce the query wait time. Figure 11 illustrates the impact of compression on an average warehouse workload.

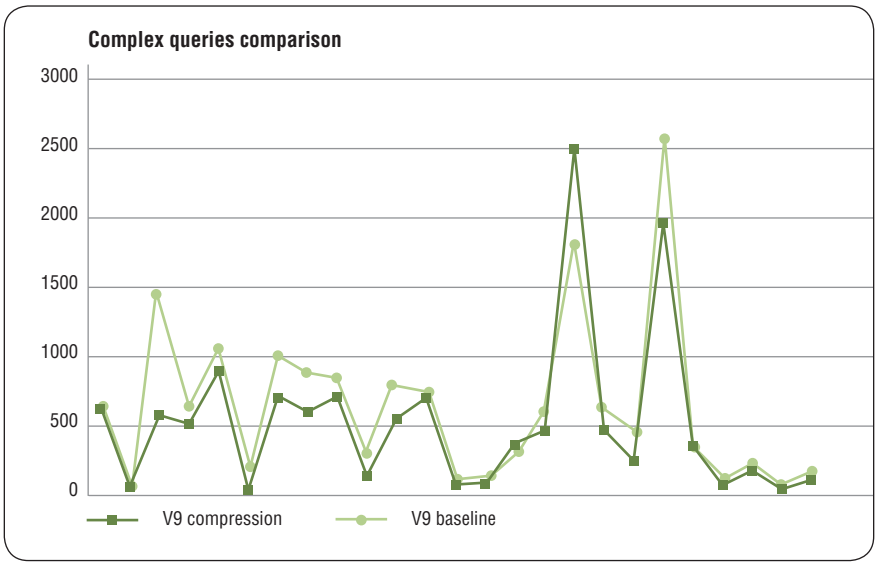

Figure 11. Query performance with and without compression

Highlights

Figure 12 shows examples of space savings that can be achieved with the DB2 data row compression feature on different tables.

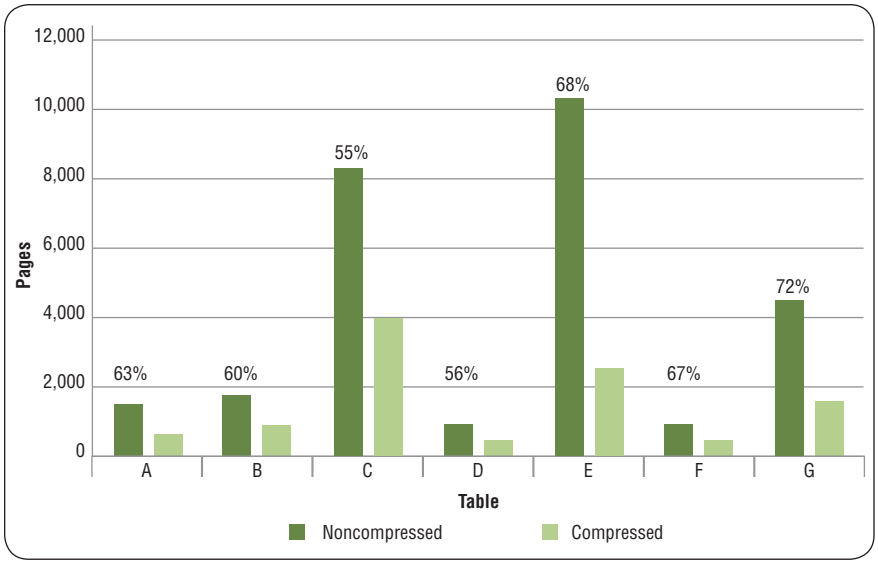

Figure 12. The degree of compression achieved on a sample set of tables

Integrated with DB2 cube style views technology, DB2 Alphablox software enables you to build analytical applications on top of the data warehouse.

DB2 Alphablox and connectors

DB2 Alphablox software and its connectors provide the capability to build analytical applications over the top of the data warehouse. Integrated with DB2 cube views technology that enables BI developers to rapidly build and deploy OLAP-style reporting, it has a rich set of reporting components that can be built into a BI portal or integrated into other Web-based applications. DB2 Alphablox is deployed in a three-tier architecture running WebSphere Application Server software as the application delivery platform. This approach drives the user-delivered interface through dynamic Hypertext Markup Language (DHTML), resulting in a thin, browser-based client.

Highlights

DB2 Alphablox components can be embedded into any Web-based development environment, making them ideal for use with custom BI portals.

DB2 Alphablox capabilities include dynamic OLAP reporting through to charts and balanced scorecard configurations. Because of its component-based approach to building analytical applications, DB2 Alphablox components can be easily embedded into any Web-based development environment. This makes DB2 Alphablox ideal for custom BI portals and integrated analytics for broader applications like the one shown in figure 13.

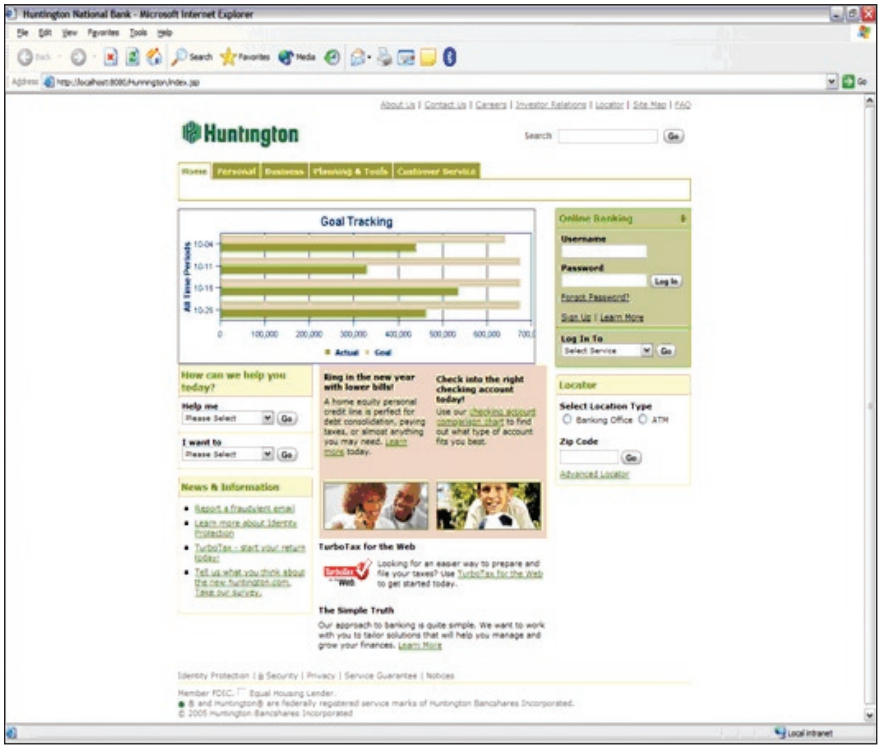

Figure 13. DB2 Alphablox application example

Highlights

Data mining capabilities included in DB2 Warehouse enable integrated data analytics.

Data mining

Powerful data mining capabilities are included in DB2 Warehouse, enabling integrated analytics of the data in the enterprise data warehouse. Standard data mining model algorithms (clustering, associations, classification and prediction) are supported and can be developed in the DB2 Warehouse development environment. Rich presentation components are also provided to enable visual analysis of data mining results, which, like DB2 Alphablox, can be embedded into Web-based applications like the one shown in figure 14.

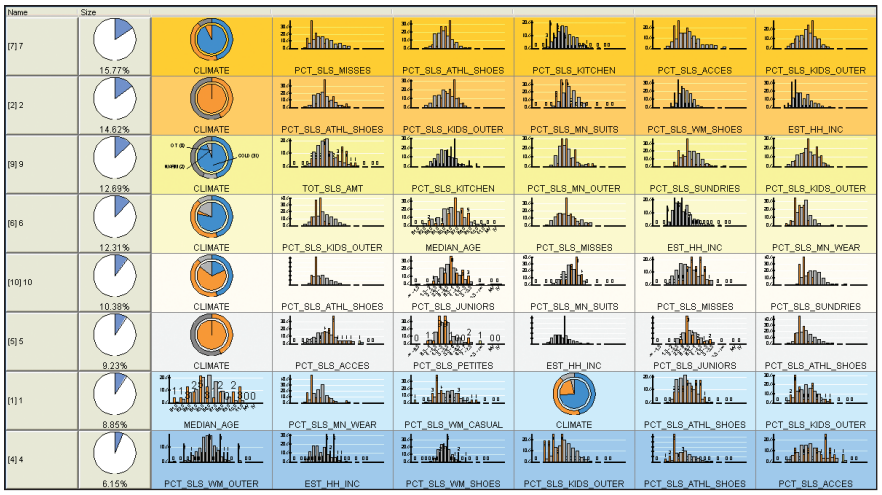

Figure 14. DB2 mining visualization

These algorithms also may be imported via industry standard Predictive Model Markup Language (PMML) format from third-party modeling tools. The data mining models can be executed in the production environment to provide real-time scoring of data records.

Highlights

Using SQL functions from BI products such as DB2 Alphablox, Cognos or Business Objects, you can execute DB2 data mining on the fly.

For organizations running mining workbench applications, such as SPSS or SAS, this functionality provides the ability to bring mining to the masses. To achieve this, the refined models from the mining workbench are exported into PMML format and imported into DB2 data mining. By doing this, organizations can realize real benefits in terms of timeliness of data scoring, because data can be scored using the models in real time. DB2 data mining allows you to execute these mining models on the fly via SQL functions from BI products such as DB2 Alphablox, Microstrategy, Cognos or Business Objects as well custom-built applications, as illustrated in figure 15. Some workbenches, such as SPSS, have integrated capabilities and can interoperate mining jobs with DB2 data mining functions.

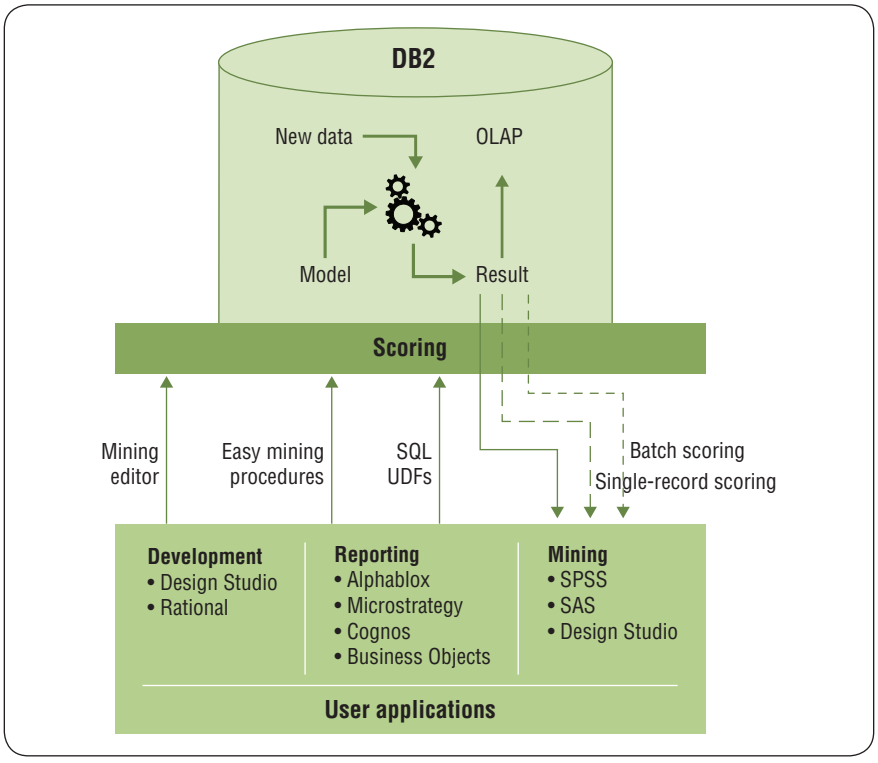

Figure 15. DB2 data mining overview

Highlights

Carefully analyze both your existing and future requirements before deciding which DB2 Warehouse is right for you.

This can reduce the cost of managing and maintaining batch-driven scoring of data, which can require a lot of time and hardware, depending on the volumes of data that need to be exported from the warehouse and run through a mining engine.

Which DB2 Warehouse is right for your business?

Choosing the right DB2 Warehouse for your needs has as much to do with what your requirements are today as where you are going in the future. Once you work through this, the decision should be straightforward. One thing should be clear by now: that there is a DB2 Warehouse package that is right for every organization, large or small.

The table below provides a guide to which DB2 Warehouse is right for your business, based on user and data volumes. Outside these two factors, you need to consider query complexity, analytic requirements and future requirements, when you make your final decision.

	Starter	Intermediate	Advanced	Enterprise Base	Enterprise
Small number of concurrent users (20–50)	●	●			
Small data volumes (<350GB)	●	●			
Medium number of concurrent users (50–100 users)		▲	●	●	●
Low to medium data volumes (<1TB)		■	●	●	●
Upper-medium data volumes (<2TB)			●	●	●
Large number of users (>100)				▲	●
Large data volumes (>2TB)				■	●
Data mining					●
Build analytical applications			★		●

▲ Yes, with performance optimization option
■ Yes, with storage optimization option
◆ Yes, with mining option
★ Yes, with DB2 Alphablox option
• Yes

Highlights

Starter and Intermediate versions of DB2 Warehouse are best for smaller businesses with small to medium data requirements, while the Enterprise Base and Enterprise versions can accommodate several terabytes to tens of terabytes of data.

For those smaller businesses or departments that are comfortable with their small to medium data and user requirements now and in the future, Starter and Intermediate solutions are likely the best choice.

The Intermediate or Advanced package is likely your best choice, if you:

- *Are not sure of your future growth requirements.*
- *Are starting off with less than one terabyte of data.*
- *Wish to choose a package that gives you the performance you need at the moment.*
- *Wish to start your warehousing strategy at a low cost.*

If you are looking at a data range starting at several terabytes to the tens of terabytes, Enterprise Base—with the view to moving toward full Enterprise—is likely the best choice for your organization.

The following section provides additional details to help you determine which DB2 Warehouse offering is right for your business. If you require further assistance, contact your local IBM representative or IBM Business Partner.

Highlights

There are numerous components available in each of the DB2 Warehouse offerings.

DB2 Warehouse versions

This section will take you through the DB2 Warehouse packages, positioning their functional and nonfunctional capabilities and licensing restrictions. The table below provides a summary of the components available in each of the DB2 Warehouse offerings.

Packaged options	Starter	Intermediate	Advanced	Enterprise Base	Enterprise	Developer
Integrated installer	●	●	●	●	●	●
DB2 9	●	●	●	●	●	●
DB2 Warehouse Design Studio	●	●	●	●	●	●
SQL warehousing tool	●	●	●	●	●	●
OLAP acceleration feature	●	●	●	●	●	●
DB2 range partitioning and MDC tables		●	●	●	●	●
Database partitioning feature		★	●	●	●	●
DB2 storage optimization feature		★	●	★	●	●
Performance optimization feature			■	★	■	■
Data mining					●	●
DB2 Alphablox and connectors			★	●	●	

• standard ★ optional ■ DB2 Query Patroller only

Highlights

Designed for SMBs, DB2 Warehouse Starter supports 20 to 50 concurrent users per server and up to 350GB of data.

DB2 Warehouse Starter

DB2 Warehouse Starter is designed for small and midsize businesses (SMBs) with 20 to 50 concurrent users per server, first-time data warehouse clients who require the ability to transform up to 350GB of data to reliable and consistent business insights in a single-node environment. It is also a great base to build small reporting applications for mass distribution. As highlighted in the table below, this configuration is similar to the Enterprise Base, except that it is limited in processors, available only in a Linux version and does not include database partitioning, range partitioning or multidimensional clustering.

IBM DB2 Warehouse Starter at a glance	
Licensing and limitations	• Limited to 200 Processor Value Unit (PVU) plus 4GB of memory* • Licensed either per PVU or per authorized user
Supported platforms	• Linux
Foundation components	Integrated installer, DB2 9, SQL warehousing tool, DB2 Warehouse Design Studio, DB2 Warehouse Administration Console, OLAP acceleration

* As of August 2007, memory is equivalent to a two dual-core processor AMD or Intel® machine.

Highlights

DB2 Warehouse Intermediate supports 50 to 100 concurrent users and up to 1TB of data, making it a perfect fit for larger SMBs.

DB2 Warehouse Intermediate

DB2 Warehouse Intermediate is designed for larger SMB and departmental clients with 50 to 100 concurrent users and data warehouse requirements up to 1TB, to work with IBM Business Partner BI applications and analytic tools. It supports a 400 PVU/4 CPU environment, single or multiple nodes with requirements for up to 32GB of memory. This is an excellent starting point for organizations or departments with more than 500GB of raw warehouse data that want to get started with BI solutions at a low cost point. As highlighted in the table below, this configuration is similar to the Enterprise Base, except that the database partitioning feature is optional and is limited in processors, memory and operating system. It therefore provides the foundation components and architecture of the Advanced and Enterprise versions of DB2 Warehouse. Thus it provides a solid starting point and clear and straightforward upgrade path to the DB2 Advanced and Enterprise Warehouse offerings.

IBM DB2 Warehouse Intermediate at a glance	
Licensing and limitations	• Limited to 400 PVU plus 32GB of memory* • Licensed either per PVU or per authorized user
Supported platforms	• Linux
Foundation components	Integrated installer, DB2 9, SQL warehousing tool, DB2 Warehouse Design Studio, DB2 Warehouse Administration Console, OLAP acceleration, DB2 range partitioning and MDC tables, database partitioning
Optional components	Performance optimization, DB2 storage optimization

* As of August 2007, memory is equivalent to a four dual-core processor AMD or Intel machine.

Highlights

For organizations that need to build a warehouse environment at a low entry cost, while keeping the option to upgrade to a larger environment later on, DB2 Warehouse Advanced supports up to 2TB of raw data and can be deployed to single or multiple hardware nodes.

DB2 Warehouse Advanced

DB2 Warehouse Advanced is targeted at midsize businesses, enterprises and large departmental clients with requirements for 2TB of raw data. It can be deployed in an environment with single or multiple hardware nodes. As shown in the table below, this configuration provides a step up from the DB2 Warehouse Intermediate offering in terms of capacity. Plus, it provides the added benefits of the DB2 storage optimization feature—enabling greater data volumes per processor in the warehouse—and the DB2 Query Patroller feature of the performance optimization option. This package provides the right starting point for any organization that needs to build a warehouse environment at a low entry cost, while maintaining the option to upgrade seamlessly to a larger data warehouse environment later on.

IBM DB2 Warehouse Advanced at a glance	
Licensing and limitations	• Limited to 1,000 PVU* • Licensed by PVU • Limited to 2TB of raw data**
Supported platforms	• Linux
Foundation components	Integrated installer, DB2 9, SQL warehousing tool, DB2 Warehouse Design Studio, DB2 Warehouse Administration Console, OLAP acceleration feature, DB2 range partitioning and MDC tables, database partitioning, DB2 storage optimization, performance optimization***
Optional components	DB2 Alphablox and connectors

* As of August 2007, this is equivalent to a 10 dual-core processors of AMD or Intel type.
** Raw table data size excludes MQTs and indexes.
*** DB2 Query Patroller only.

Highlights

Available on Linux, UNIX and Microsoft Windows platforms, DB2 Warehouse Enterprise Base provides midsize to large-scale enterprises with a robust data warehouse and datamart infrastructure.

DB2 Warehouse Enterprise Base

DB2 Warehouse Enterprise Base provides midsize businesses to large-scale enterprises with a data warehouse and datamart infrastructure. As shown in the table below, this configuration is available on Linux, UNIX and Microsoft Windows platforms, making it widely deployable. Plus, it has no ceiling on the number of processors upon which it can be deployed.

IBM DB2 Warehouse Enterprise Base at a glance	
Licensing and limitations	• Licensed by PVU
Supported platforms	• Linux, Windows and UNIX
Foundation components	Integrated installer, DB2 9, SQL warehousing tool, DB2 Warehouse Design Studio, DB2 Warehouse Administration Console, OLAP acceleration feature, DB2 range partitioning and MDC tables, database partitioning
Optional components	Performance optimization, DB2 storage optimization feature

DB2 Warehouse Enterprise delivers scalability and manageability, along with a complete range of analytic capabilities.

DB2 Warehouse Enterprise

This DB2 Warehouse package is ideal for midsize to large data warehouses and other BI solutions that require scalability and manageability along with a complete range of analytic capabilities. By including all the options in this version, IBM provides a complete enterprise offering at a cost that is substantially less than the sum of its components, providing real value on price and completeness of solution.

Highlights

The performance optimization and deep compression features in DB2 Warehouse Enterprise make it easier and more affordable to build and manage a large data warehouse.

With the benefit of the integrated data mining with the SQL data warehouse workbench and the DB2 Alphablox embedded analytics capability, the DB2 Warehouse Enterprise provides the perfect foundation for building enterprise real-time analytics, opening up sophisticated analysis to all warehouse users. It also delivers a scalable platform for building and managing an enterprise data warehouse and analytical applications. Plus, its performance optimization and deep compression features can make building and managing a large data warehouse more affordable and can significantly reduce the cost of ownership.

As shown in the table below, DB2 Warehouse Enterprise is available on Linux, UNIX and Microsoft Windows platforms, making it widely deployable on the platform of your choice. Moreover, it has no ceiling on the number of processors upon which it can be deployed.

IBM DB2 Warehouse Enterprise at a glance	
Licensing and limitations	• Licensed by PVU and has subcapacity licensing
Supported platforms	• Linux, Windows and UNIX
Foundation components	Integrated installer, DB2 9, SQL warehousing tool, DB2 Warehouse Design Studio, DB2 Warehouse Administration Console, OLAP acceleration feature, DB2 range partitioning and MDC tables, database partitioning feature, DB2 storage optimization feature, performance optimization feature (DB2 Query Patroller only), data mining, DB2 Alphablox and connectors

Highlights

Designed for developers, DB2 Warehouse Developer includes the same content and functionality as DB2 Warehouse Enterprise, but is licensed with terms and conditions for evaluating, testing and developing application programs.

DB2 Warehouse Developer

Targeted at developers, DB2 Warehouse Developer (see the table below) has the same content and functionality as the DB2 Warehouse Enterprise, licensed with terms and conditions for the purposes of evaluating, demonstrating, testing and developing application programs.

IBM DB2 Warehouse Developer at a glance	
Licensing and limitations	• Limited to development purposes only • Licensed by authorized user
Supported platforms	• Linux, Windows and UNIX
Foundation components	Integrated installer, DB2 9, SQL warehousing tool, DB2 Warehouse Design Studio, DB2 Warehouse Administration Console, OLAP acceleration feature, DB2 range partitioning and MDC tables, database partitioning feature, DB2 storage optimization feature, performance optimization feature (DB2 Query Patroller only), data mining, DB2 Alphablox and connectors

Highlights

IBM offers numerous resources to help you learn more about DB2 Warehouse products.

Free trial versions of DB2 products are also available.

Resources

Learn

- *Visit the DB2 Warehouse roadmap to learn more about the DB2 Warehouse components*
- *Search the developerWorks Web site for material on DB2 warehousing*
- *Visit the developerWorks resource page for DB2 for Linux, UNIX and Microsoft Windows to read articles and tutorials and connect to other resources to expand your DB2 skills*
- *Visit the developerWorks Information Management zone and find more resources for DB2 developers and administrators*
- *Stay current with developerWorks technical events and webcasts*

Get products and technologies

- *Download a free trial version of DB2 Enterprise 9*
- *Build your next development project with IBM trial software, available for download from the developerWorks Web site*

Discuss

- *Read developerWorks blogs and get involved in the developerWorks community*

End notes

1 Gartner RAS Core Research Note G00141484, Ted Friedman, Mark A. Beyer and Andreas Bitterer, 22 November 2006.

2 http://www-306.ibm.com/software/data/db2/9/editions_features_storage.html

3 http://www-306.ibm.com/software/data/db2/9/editions_features_storage.html

IBM Canada
8200 Warden Avenue
Markham, ON
L6G 1C7, Canada

Produced in United States of America
08-07

IMW10925-USEN-00