



IBM enterprise analytics for the intelligent e-business

by Colin J. White
DataBase Associates

Table of Contents

1	The importance of enterprise analytics
1	<i>The first three generations of business information systems</i>
3	<i>The emergence of fourth-generation systems</i>
5	<i>Building the intelligent e-business</i>
6	The IBM enterprise analytics solution
7	<i>IBM enterprise analytics strategy</i>
7	<i>Supporting on-demand enterprise analytics and realtime decision-making</i>
8	<i>Integrating BI and enterprise analytical processing into DB2</i>
9	<i>Simplifying the building of an integrated system for delivering analytics</i>
10	<i>Partnering with business intelligence and enterprise analytics vendors</i>
10	<i>IBM enterprise analytics product set</i>
11	<i>Database management</i>
13	<i>Federated data access</i>
14	<i>Data warehouse development</i>
19	<i>Analytic application development</i>
23	<i>Analytic application packages</i>
24	<i>Portal development</i>
26	<i>Conclusion</i>
27	Appendix: summary of IBM enterprise analytics products
27	<i>Database management</i>
28	<i>Federated data access</i>
29	<i>Data warehouse management</i>
29	<i>Analytic application development</i>
30	<i>Analytic application packages</i>
30	<i>Portal development</i>



Highlights

First-generation systems were based on centralized batch processing

Second-generation data warehousing systems based on client/server computing were a major step forward.

The importance of enterprise analytics

Given the increasing competition in today's tough business climate, it is vital that organizations provide rapid, cost-effective access to business information for executives, managers, business analysts and other key decision makers. The solution here is a system that provides an integrated set of technologies and products to supply business users with the enterprise analytics they need to react rapidly to trends and changes in the marketplace.

Many enterprise analytics concepts are not new, but have evolved and been refined based on experience gained from early centralized business information systems, and more recently, from data warehousing and business intelligence applications. This paper provides an introduction to enterprise analytics, reviews requirements for an enterprise analytics system and takes a detailed look at IBM's enterprise analytics software strategy and products.

The first three generations of business information systems

Technology, products and requirements for enterprise analytics have changed dramatically since the days of mainframe computers and centralized computing. Early business information systems employed batch applications that reported on operational data to provide business users with the information they needed. The output from these applications typically involved large volumes of paper that users had to wade through to get the answers they needed to business questions. Also, these first-generation systems could only be used by experts, such as business analysts, who had an intimate knowledge of the operational data and extensive computer experience.

The second generation of business information systems came with data warehousing, which offered a giant leap forward in capability.

- Information extracted from operational systems into a data warehouse is clean and consistent, and is stored in a form that business users can understand
- Unlike operational systems, which contain only detailed current data, a data warehouse supplies both historical and summarized information
- The use of client/server computing in second-generation systems provided easier-to-use tools and improved user interfaces.



Highlights

Third-generation systems saw the introduction of OLAP and data mining and Web deployment

Packaged analytic applications in third-generation systems speed deployment.

Data warehousing, however, was still not a complete answer to the needs of business users. Most of the focus of data warehousing applications was on building the data warehouse, rather than accessing it. Many organizations thought that if they built a data warehouse and provided business users with the right query and reporting tools, the job was done. In fact, it was just beginning. Organizations discovered that unless the information in the data warehouse is thoroughly documented and easy to access, complexity limits data warehouse usage to the same experts as in first-generation systems.

The third generation of business information systems emerged with the introduction of business intelligence products that focused on improving access to information in a data warehouse, and on offering a rich set of information analysis capabilities. These third-generation systems saw the advent of online analytical processing (OLAP) and data mining technologies. These systems also began to exploit the Web as a cost-effective way to deliver business information applications to a large user audience.

While there is no doubt that data warehousing, OLAP and data mining products provide powerful capabilities for building and accessing a data warehouse, these products often require a significant amount of implementation effort. This issue, coupled with the use of the Web as a sales, marketing and support channel, has led organizations to demand solutions that enable them to quickly deploy applications that give business users the information and analyses they need to react rapidly to changing business conditions in the new world of e-business. To satisfy this requirement, vendors have begun offering packaged enterprise analytic applications that exploit existing data warehousing and business intelligence technologies. These application packages are out-of-the box solutions whose analytics can be customized as required to suit an organization's business requirements.



Highlights

Realtime closed-loop processing is becoming a requirement.

The emergence of fourth-generation systems

Business information systems are usually viewed as being stand-alone *strategic* decision-making applications that are separate from the operational systems that manage day-to-day business operations and supply the raw data for use in a data warehouse. The information and analyses provided by these business information systems, however, have also become vital to *tactical* day-to-day decision making, and many companies can no longer operate effectively without them. As a result there is a trend away from stand-alone decision-making systems toward integrating them into the overall business process. The increasing use of e-business is also encouraging this integration, since organizations need to be able to react much faster to changing business conditions in the e-business world than they do in the traditional *bricks-and-mortar* environment.

The integration of business information systems into the overall business process can be achieved by building a *closed-loop* decision-making system in which the output of business intelligence applications is routed to operational system users in the form of recommended actions (product pricing changes and marketing campaign modifications, for example) for addressing specific business issues. In the e-business environment, many companies are looking toward extending this closed-loop process to the *automatic* adjustment of business operations and marketing campaigns, based on decisions made by the business intelligence system. In fact, some companies would like this automated closed-loop processing to occur in *realtime*. A closed-loop enterprise analytics system that can support realtime processing represents a fourth-generation of business information systems.



Highlights

Realtime data warehouse

The objective of a realtime closed-loop system is to supply on-demand analytics for realtime decision making. Such a system involves three main capabilities, each of which can operate and be used independently of the others:

- A *realtime data warehouse* that contains information captured from operational and e-business data as it is created or updated. Other industry terms here include active data warehouse and *zero-latency* data warehouse. A typical use of such a facility is to build a realtime data warehouse that provides a single view of each customer by tracking every single customer interaction across all touch points including the Web. A single customer database is a key requirement for a CRM enterprise analytic solution. A realtime data warehouse has similar characteristics and requirements to an *operational data store*. To support such a data warehouse, vendors need to provide an *event-driven hub* that can capture data from operational and e-business transaction applications, messages from enterprise application integration (EAI) systems, etc., and stream the captured data into a warehouse as it is created or updated.
- An *analysis engine* that can generate and/or provide access to current business analyses from any place and at any time. Performance here will largely depend on the amount of data to be analyzed and the complexity of the analyses to be performed. To support *on-demand* analytics, vendors are adding parallel processing and integrating analytic capabilities into their database engines. They are also providing packaged enterprise analytic applications that pre-compute business metrics for rapid delivery via a portal.
- A rules-driven *decision engine* for making automated and realtime decisions. An example would be an application that can automatically grant or deny an e-business user's request for a loan or a new credit card. The requirement here is a decision engine that can employ business analytics and rules to automate the closed-loop decision-making process. One approach being offered by vendors is the ability to dynamically run a predictive data mining model in realtime, to score an e-business customer, for example. Realtime scoring is particularly valuable in the e-business environment where rapid and automated decision-making is a key competitive advantage.

On-demand analytics

Automated decision engine.



Building the intelligent e-business

Corporations now recognize that rapid access to sound business intelligence and enterprise analytics is vital if they are to compete effectively and become an *intelligent e-business*. Figure 1 illustrates an IT infrastructure for supporting an intelligent e-business. The main components of this infrastructure are:

- A **portal** for providing business users with an integrated and personalized Web-based view of the business content (information, applications and collaborative services) they need.
- **Data warehouse development tools** to extract and transform data from operational systems, e-business systems and external information providers for loading into a data warehouse. As discussed earlier, fourth-generation systems need to be able to capture and stream this data to a data warehouse in near realtime.
- A **database management system (DBMS)** to manage near-realtime, historic, detailed, summarized and multidimensional OLAP cube data. The industry trend here is to integrate business information systems functions into a DBMS to enhance the performance, scalability and manageability of fourth-generation applications. Figure 1 summarizes a list of main requirements here.
- **Analytic application development tools (query, reporting, OLAP, data mining, etc.) and packaged solutions** that analyze data warehouse information and deliver enterprise analytics to both business users and decision engines in support of on-demand analytics and realtime decisions.

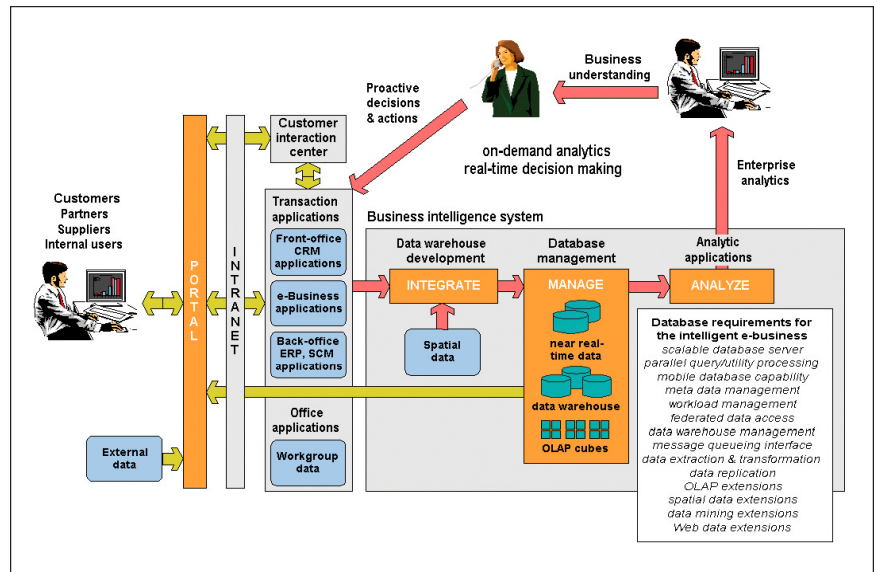


Figure 1. Intelligent e-business framework

© Copyright DataBase Associates, Inc. 2001

Now that we have defined the role and requirements for enterprise analytics in an intelligent e-business system, we can move on to look at IBM’s strategy and products for supporting enterprise analytics.

The IBM enterprise analytics solution

IBM has a long history of supplying solutions for the business information system marketplace. In the early 1960s, for example, one of the first products on the market to support analytic processing was IBM’s A Programming Language (APL). IBM was also one of the first vendors to offer data warehousing solutions. An early IBM source that detailed the architecture and business case for a data warehouse was the 1988 *IBM Systems Journal* article by Barry Devlin and Paul Murphy entitled, “An Architecture for a Business and Information System.” This paper described the use of data warehousing in IBM Dublin and ultimately led to the development of IBM’s Information Warehouse concept and architecture in the early 1990s.



Highlights

IBM is putting significant resources into providing an end-to-end enterprise analytics solution

More recently, IBM has put significant resources into both the development and marketing of its business intelligence and enterprise analytics products. As the industry moves toward the use of fourth-generation business information systems, IBM intends to maintain this development and marketing momentum by providing an integrated end-to-end enterprise analytics solution. This section of the paper discusses the IBM strategy for this solution from a software perspective and reviews the products that comprise this solution.

IBM enterprise analytics strategy

IBM's business intelligence and enterprise analytics strategy has four main objectives:

- To support on-demand enterprise analytics and realtime decision-making
- To integrate business intelligence (BI) and enterprise analytical processing capabilities into IBM DB2®
- To simplify the building of an integrated system for delivering enterprise analytics
- To form partnerships with leading business intelligence and enterprise analytics vendors.

IBM is developing packaged analytic applications for e-business.

Supporting on-demand enterprise analytics and realtime decision-making

IBM's aim here is to provide on-demand analytics for optimizing all aspects of e-business, and front- and back-office operations. This support involves not only facilities to integrate business data from CRM, ERP, SCM and legacy transaction systems into a data warehouse for analysis, but also tools that capture and store geospatial data in a data warehouse so that enterprise analytics can be displayed visually in a geospatial context. IBM and its partners are also developing analytic application packages (IBM WebSphere® Commerce Analyzer Advanced Edition, for example) that provide predefined analytics for analyzing e-business operations. These packaged solutions enable organizations to reduce development costs and rapidly deploy enterprise analytic applications.



Highlights

IBM sees the need for a near-realtime data warehouse

Some enterprise analytic applications require a data warehouse to contain near-realtime information. The result is sometimes called an *active- or zero-latency* data warehouse. This type of warehouse is used for tactical, rather than strategic, decision making. To support this style of processing, IBM is integrating its message software, IBM WebSphere MQ, into its business intelligence solution, so that operational and e-business applications can rapidly feed transaction data to the data warehouse. The database trigger and replication facilities of DB2 could also be used for maintaining such a warehouse.

Data mining deployment models support realtime decision making

Another key IBM goal is to enable corporations to react rapidly to trends and changes in the e-business marketplace by providing tools that permit users and applications to make decisions in realtime while interacting with a customer. An example of IBM product support here is IBM DB2 Intelligent Miner™ Scoring, which allows e-commerce and operational transaction applications to dynamically execute scoring models stored in a DB2 database. This facility would enable a loan officer, for example, to determine a customer's credit-worthiness using a predictive model while the applicant is connected to a Web site or on the phone. Another product example is IBM WebSphere Application Server and IBM WebSphere Portal Server, which include a rules engine that enables automation of recommendations and actions in the e-business environment.

Data mining and OLAP capabilities have been integrated into DB2.

Integrating BI and enterprise analytical processing into DB2

The IBM DB2 Universal Database™ engine is designed to provide high performance and scalability. By integrating facilities such as parallel query and utility processing, online analytical processing (OLAP) and data mining facilities into this database engine, IBM is able to provide organizations with enhanced performance and high scalability for enterprise analytical application processing.



Highlights

DB2 provides federated data access to non-IBM data stores

IBM's strategy is to partner with the industry's key front- and back-office vendors.

Business information systems involve many different types of data, and DB2, like other relational database products, handles a wide range of different data types and functions. The product, however, also includes user-defined types and functions in support of an object-relational database capability. These user-defined types and functions are employed by IBM development organizations to build database *extenders* that enable spatial, text, multimedia and XML data to be efficiently managed by DB2. IBM DB2 Intelligent Miner Scoring also uses this extender technology to manage data mining models in a DB2 database. These models can be built using IBM's data mining tools or any third-party mining product that supports the industry-standard XML-based Predictive Model Markup Language (PMML). The benefit of storing these models in a DB2 database is that the models can be executed in a production environment using simple SQL statements.

Simplifying the building of an integrated system for delivering analytics

As illustrated in Figure 1, a business intelligence system requires a wide range of products and tools to create and deliver enterprise analytics. To reduce development costs, provide usability and minimize the total cost of ownership of a business intelligence system, an organization needs to easily integrate these products and tools into its IT infrastructure. IBM's strategy here is to provide an integrated DB2 Universal Database package whose functions and features work cohesively together. We have already discussed how IBM has integrated OLAP, data mining and performance features into the DB2 database engine in support of enterprise analytic processing. The DB2 package also contains integrated features for data extraction and transformation, data replication and federated data access to a variety of different database systems.

IBM's view is that business intelligence and enterprise analytics are no longer stand-alone technologies, but are now becoming integrated with a wide variety of operational and e-business systems. The company's strategy here is to develop leading-edge enterprise analytics technology, and to partner with the industry's key front- and back-office application vendors to integrate this technology into its products.



Highlights

IBM works closely with its partners to ensure that their products are well integrated with the complete DB2 package. An example of such integration is the IBM DB2 Data Warehouse Center, which can be used to manage and administer not only business intelligence operations performed by IBM's products, but also those from key partners.

Partnering with business intelligence and enterprise analytics vendors

To provide customers with an end-to-end solution, IBM's enterprise analytics framework is designed to be able to integrate and incorporate not only IBM products but also those from third-party vendors. To encourage support for its enterprise analytics offerings, IBM has created a Business Intelligence Partner Program for ISVs, VARs, systems integrators and consultants. The objective of the program is to have both joint marketing relationships with other organizations and joint development initiatives that enable other software vendors' products to be integrated with IBM products.

Proof that IBM is serious about tight integration between its products and those from other vendors can be seen in its current relationships with Ascential Software, Brio Technology, Business Objects, Evoke Software, Evolutionary Technology International (ETI), Hyperion Solutions, Trillium Software and Vality Technology.

IBM also has a *FastStart* program that is intended to help small- to medium-size companies get started with business intelligence and enterprise analytics. This program provides VARs and systems integrators with starter kits and application templates that they can use with their clients.

IBM enterprise analytics product set

We now move on to review the products provided by IBM for supporting an enterprise analytics software environment; these products are summarized in the Appendix. We will use *The intelligent e-business diagram* shown in Figure 1 to categorize and describe these products.

IBM partner programs aimed at product integration.



Highlights

DB2 parallel processing exploits both SMP and clustered SMP servers

Many extensions added to DB2 for enhancing the performance of analytical processing.

Database management

At the heart of IBM's enterprise analytics product set is **DB2 Universal Database**¹—an object-relational database system that operates on a wide range of platforms, including OS/390®, z/OS™, VSE and VM, AS/400®, UNIX® (AIX®, HP-UX, Solaris™), NUMA-Q®, Linux®, OS/2® and Microsoft® Windows® (2000, NT, 95, 98). Data maintained by third-party database products can be accessed by DB2 applications using IBM's database middleware products (see the section *Federated Data Access* for details). For mobile users, IBM provides **DB2 Everyplace™**, which supports the Linux, Palm OS, QNX Neutrino, Symbian EPOC and Windows CE/PocketPC operating environments.

One important element of IBM's database development strategy is to supply high performance and scalability through the use of parallel software and hardware. IBM markets several families of parallel computing hardware², and the DB2 offerings running on these machines all provide parallel SQL and utility processing for handling large databases and analytic workloads on both SMP and clustered SMP hardware configurations. Parallel processing on HP and Sun hardware platforms is also supported by DB2.

Another key component of IBM's database strategy is to aid enterprise analytics processing performance by adding extensions to the SQL language, and to the data management and SQL optimization components of DB2 Universal Database. Examples of facilities in this area include SQL ROLLUP and CUBE operators for multidimensional analysis, SQL aggregate functions for analytic processing, support for very large tables and bufferpools, partitioning of large tables for parallel processing, inter- and intra-query parallel processing, automated summary tables for creating and refreshing summarized data, index-only access, temporary tables, cost-based SQL optimizer with multiple-join algorithms, optimization of star schemas, index ANDing and ORing, caching of optimized SQL statements for repeat queries, optimizer rewrite of inefficient SQL queries, etc.

¹The term *DB2 Universal Database* will be used throughout this paper to refer to the DB2 product set. The reader should be aware, however, that certain features of DB2 Universal Database documented in this paper may not be supported on every operating platform.

²Servers here include S/390®, zSeries™, RS/6000®, pSeries™, AS/400, iSeries™, NUMA-Q, Netfinity® and xSeries™.



Highlights

The OS/390 Workload Manager can be used to dynamically manage analytical processing workloads.

Organizations employing enterprise analytic solutions require not only high performance and scalability, but also management facilities for handling large-scale data warehouses and complex analytic application workloads. IBM provides several capabilities here. To simplify the administration of large-scale data warehouse environments, all DB2 database administration, utility, extract, transformation and loading operations associated with data warehouse construction are managed from a single interactive interface known as the **DB2 Control Center**. Resource consumption by analytic application workloads can be tracked, analyzed and governed using the **DB2 Query Patroller**.

The OS/390 and z/OS **Workload Manager (WLM)** also provides a powerful capability for managing and balancing data warehouse update and analytical processing operations. The WLM allows administrators to control operations and performance, based on business-oriented targets for different types of workloads. The WLM uses these targets to dynamically manage dispatching priorities, use of memory and other resources to try to achieve the target objective for each workload. It can be used to control mixed transaction and business intelligence processing workloads, or business intelligence workloads involving a variety of different types of work, such as basic queries, complex analytical processing, database loading, workloads from a specific department and so forth. Targets are expressed in terms of transaction response time ranges (for example, 90 percent in less than 0.5 of a second), batch run times, etc. Targets can also be broken down into multi-period definitions that allow long-running SQL queries to be downgraded the longer they run, to make way for shorter queries. The WLM is particularly useful in realtime enterprise analytic environments where data warehouse updates may need to occur concurrently with analytic processing.



Highlights

IBM Teraplex Centers are available for customers to validate large-scale business intelligence projects

DB2 DataJoiner supports federated data access to database and file data.

To aid in testing the integration, scalability and manageability of its business intelligence and enterprise analytics hardware and software solutions, IBM has invested millions of dollars in creating a set of Teraplex Integration Centers that are used by customers, business partners and IBM development organizations to stress-test proposed solutions involving very large databases and business intelligence workloads³.

Federated data access

Federated data access facilities allow DB2-based applications to access both DB2 and non-DB2 data stores. These facilities also enable IBM's data warehouse construction tools to construct DB2 data warehouses from non-DB2 source data, and, if required, to use non-IBM databases as a target for storing data warehouse information.

IBM has supported federated data access for several years via its **DB2 DataJoiner**[®] product, which allows DB2 client applications both to query and update data managed by one or more backend database servers. A DB2 DataJoiner server running on either Windows NT[®] or UNIX can connect to backend servers running IBM or non-IBM database relational database products, such as IBM DB2, IBM Informix[®], Microsoft SQL Server, NCR Teradata, Oracle and Sybase, for example. Support for non-relational data is provided by **DataJoiner Classic Connect**, which is a separately orderable component of DataJoiner that provides read-only access to data stored in IMS[™] databases and VSAM files on OS/390 and z/OS.

Features of DB2 DataJoiner that are worthy of note include:

- Transparent and heterogeneous database access using a single dialect of SQL.
- Global optimization of distributed queries with query rewrite capability for poorly coded queries.
- A stored procedure feature that allows a global DB2 DataJoiner procedure to transparently access data or invoke a local procedure on any DB2 DataJoiner-supported database. This feature includes support for Java[™] and Java Database Connectivity (JDBC[™]).
- Heterogeneous data replication (using **DB2 DataPropagator**[™], which is integrated with DB2 DataJoiner) between DB2, Informix, Oracle, Sybase and Microsoft relational database products.

³Colin White. *IBM's Teraplex Integration Centers: The Key to the Intelligent Business*. DataBase Associates report, October 2000.



Highlights

DB2 DataJoiner capabilities are being added to DB2

Access to DB2 data from Java applications is a key objective.

IBM is now beginning to migrate the federated data access capabilities of DB2 DataJoiner into DB2 itself. DB2 Universal Database currently supports distributed query access to DB2 databases and OLE DB data sources. The optional **DB2 Relational Connect** feature extends this capability to include native SQL access to Oracle databases. Additional database products will be supported in future releases of DB2 Relational Connect.

Like many vendors, IBM sees that Web technology can significantly reduce the cost of deploying enterprise analytic solutions to a broad end-user audience. IBM's strategy here is to provide Web applications with federated data access to both DB2 and non-DB2 data. IBM **Net.Data**[®] Web middleware tool, which is included with DB2, can be used to access relational and flat file data on a variety of platforms, including DB2, DB2 DataJoiner-enabled databases and ODBC data sources. Net.Data supports client-side and server-side processing using applications written in Java, REXX, Perl, C++ or its own macro language. Web applications written in Java access DB2 using either the JDBC or SQL/J application interfaces. Data retrieved using the Net.Data macro language can be transformed into XML and formatted using an XSL style sheet. For Lotus[®] Domino[™] applications, IBM provides **DB2 for Domino** application connectors for accessing DB2 data.

Data warehouse development

IBM provides several integrated DB2 facilities and optional data warehouse development products for extracting data from transaction and e-business systems, and transforming and loading it into enterprise and departmental data warehouses. IBM also has marketing and development relationships with several third-party data warehouse product vendors, including ETL, Evoke Software, Hyperion Solutions, Trillium Software and Vality Technology.



Highlights

Data Warehouse Center used to design and build a data warehouse

Data transformations can be written in C++, Java and Visual Basic

Data Warehouse Center can extract data from many different data sources.

Two of the main facilities integrated into DB2 for data warehouse development are the **Data Warehouse Center** and **DB2 DataPropagator**. The Data Warehouse Center (an extension to the **DB2 Control Center** administration facility) is used to design a data warehouse, register and access data sources, define data extraction and transformation steps, populate data warehouses, automate and monitor Data Warehouse Center operations, and manage and exchange meta data using a DB2-based technical meta data store.

Included with Data Warehouse Center are the **Warehouse Schema Modeler** and the **Process Modeler**. These tools are used to define and generate a data warehouse star schema database design and graphically link together the steps needed to build and maintain a data warehouse. The Process Modeler supports dependency relationships, conditional processing, notification, and one-time, repeated and triggered scheduling.

Data extract and transformation requests in the Data Warehouse Center are defined using SQL. The **SQL Assist** facility aids in the development of these requests, and some 100 transformations are available. More sophisticated transformation requests can be built by the **DB2 Stored Procedure Builder**, which allows custom routines to be developed using languages like C++, Java, and Visual Basic.

The Data Warehouse Center supports the extraction of data from DB2, ODBC and OLE DB data sources. It includes ODBC drivers for Informix, Oracle, Microsoft SQL Server and Sybase. The **DB2 Relational Connect** facility adds a native interface to Oracle source data. Installing **DB2 DataJoiner** further extends the number of data sources that can be used to create the data warehouse. The Data Warehouse Center provides full refresh and incremental update of data warehouse information.

The Data Warehouse Center is also used to control and manage other data warehouse tasks including the running of IBM data replication jobs, the execution of third-party data cleansing jobs, the loading of Hyperion Essbase multidimensional databases and the triggering of DB2 Intelligent Miner services.



Highlights

The data replication facility of DB2 can also be used to populate a data warehouse

DB2 Warehouse Manager extends the capabilities of Data Warehouse Center.

Data replication technology can also be used to populate and update data warehouse information. For this task, IBM offers its **DB2 DataPropagator** facility, which is integrated into DB2 Universal Database for homogeneous data replication and into DB2 DataJoiner for heterogeneous data replication. The replication facility captures data changes from DB2 database logs and applies those changes to a DB2-managed data warehouse. Data changes are transported from the source to the target warehouse using staging tables. SQL is used to retrieve and transform data from the staging tables and apply it to the DB2-based warehouse at user-defined intervals. DB2 DataJoiner can also act as a data source or target for the replication facility, which means it can be used to replicate data from a third-party relational DBMS to a DB2-based data warehouse, or to replicate data from a DB2 data source to a data warehouse managed by a non-IBM relational DBMS. The Data Warehouse Center can be used to specify and manage the target databases of data replication.

One of IBM's main add-on products for DB2 data warehouse development is the **DB2 Warehouse Manager**, which extends the capabilities of Data Warehouse Center outlined earlier. These extensions offer:

- Improved scalability using distributed *warehouse agents* that manage the flow of data between warehouse data sources and warehouse target databases.
- The ability to collect and send data using WebSphere MQ (formerly known as IBM MQSeries®). This facility is particularly useful for creating and maintaining a near-realtime data warehouse.
- Optional *connectors* that enable data to be extracted from SAP® R/3®, i2 Trade Matrix and Web data sources for processing by DB2 Warehouse Manager and loading into a data warehouse.
- Advanced data transformations for cleansing data, generating keys and performing statistical calculations.
- Integration with third-party products (Evoke Software, Trillium Software and Vality Technology) for analyzing and transforming name-and-address data.
- An integrated business information catalog to help users determine the contents of a data warehouse and run queries, reports and analyses against warehouse data.



Highlights

Agent technology eliminates the need for an intermediate data transformation server.

- The IBM Query Management Facility (QMF™) for query and reporting on data warehouse data.
- The DB2 Query Patroller for analyzing, tracking and governing data warehouse queries and analyses.

Warehouse agent technology provides improved performance and scalability for loading large data warehouse information stores. Data is extracted and loaded into a data warehouse information store by warehouse agents that move information directly from one or more data sources to one or more warehouse databases. The use of warehouse agents eliminates the need for information to pass through an intermediate data transformation server that might otherwise become a performance bottleneck as data volumes grow. Agents additionally enable transformation logic to be distributed across the source and target systems. The source data to be captured, transformed and integrated into the data warehouse by one or more agents is defined in a *data warehouse process*. The definition, scheduling and monitoring of warehouse processes is handled by the Data Warehouse Center.

Warehouse agents run on OS/390, z/OS, AS/400, OS/2, UNIX and Windows NT/2000, and, depending on the volumes of data being moved, any given implementation may have one agent or many agents running concurrently. The Windows NT agent is included in the Data Warehouse Center, while the remaining agents are provided with the DB2 Warehouse Manager.

The Data Warehouse Center and the DB2 Warehouse Manager play a key role in managing the meta data associated with the IBM enterprise analytics environment. In such an environment, two types of meta data must be managed—technical meta data and business meta data. Technical meta data is associated with the design, building and operation of a data warehouse, whereas business meta data is used in conjunction with the business intelligence tools and analytic applications used to access and analyze data warehouse information.



Highlights

Technical meta data store supports CWM meta data interchange standard

DB2 Warehouse Manager adds a business information catalog.

The Data Warehouse Center employs its own DB2-based meta data store for managing the technical meta data associated with the building and managing of a data warehouse. IBM has developed interfaces to products from Computer Associates (ERwin), ETL, Hyperion Solutions, Trillium Software and Vality Technology for meta data interchange. The Object Management Group's (OMG) XML-based Common Warehouse Metamodel (CWM) standard is also supported. CWM employs the OMG Meta Object Facility (MOF), an UML-based facility for defining metamodels, and XMI, an XML specification for meta data interchange.

The DB2 Warehouse Manager adds an information catalog to the data warehouse environment, which is used to document and manage business meta data. This information catalog helps business users navigate data warehouse information and business intelligence objects and run associated queries, reports and analyses. Business users can browse the information catalog using both graphical and Web-based interfaces.

Meta data in the DB2 Warehouse Manager information catalog is stored in a DB2 database and can be accessed and maintained using supplied SQL and application APIs, and can be imported and exported using files formatted in a documented tag language. Meta data interchange capabilities are available for Brio, Business Objects, Cognos, Hyperion and IBM QMF. IBM also supplies a variety of sample applications that use these interfaces to exchange meta data with third-party desktop products (Lotus 1-2-3 and Microsoft Excel, for example). Technical meta data managed by the Data Warehouse Center can also be selectively published and synchronized with the information catalog.

Other IBM products for DB2 data warehouse construction include **IMS DataPropagator**, for capturing data changes from IMS database logs, and **DB2 DataRefresher™**, for capturing and transforming data stored in non-relational databases and files such as IMS and VSAM.



Highlights

DB2 extenders add support for additional data types, including XML documents

DB2 Spatial Extender enables analytic applications to display information in a geospatial context.

Analytic application development

Enterprise analytic applications involve more complex business information and processing compared with traditional query and reporting. We discussed earlier how IBM has built a DB2 extensibility framework that enables both IBM and non-IBM developers to add support for more complex data types and functions to the DB2 database engine. DB2 comes with extenders for handling image, text, audio, video and XML data. The **DB2 XML Extender**, for example, allows an XML document to be stored in a single DB2 column or collection of columns. The extender can index stored XML data and perform both field and full text searches on this data. Regular data stored in a DB2 table can also be rendered into an XML document when retrieved. For organizations that need to index and search large amounts of text data, IBM provides the optional **DB2 Net Search Extender**. Net Search supports word, phrase, fuzzy and wildcard searches.

A DB2 extender particularly applicable to enterprise analytical processing is the **DB2 Spatial Extender**, which gives organizations the ability to store geospatial data (maps and location information, for example) in a DB2 data warehouse. Managing geospatial data in a DB2 database provides high performance and scalability, advanced query functions and the ability to display the results of analytical processing visually in a geospatial context.

The DB2 Spatial Extender provides industry-standard spatial data types and functions, a spatially aware indexing scheme and SQL optimizer, and utilities for importing, exporting and geo-encoding spatial data. The spatial data in a DB2 database can be queried using either an SQL API, or an ESRI-compatible GIS API. The benefit of the SQL API is that enterprise analytic applications can access both spatial and non-spatial data in the same database query. The advantage of the ESRI API is that existing GIS applications (ArcView, ArcExplorer, etc.) can be used to access and analyze the DB2 spatial data.



Highlights

DB2 Intelligent Mining Scoring supports realtime scoring of customers.

Another IBM enterprise analytics product that employs DB2's extensibility architecture is **DB2 Intelligent Miner Scoring**, which enables in-house applications and third-party tools to use simple SQL statements to invoke a data mining model stored in a DB2 database. This facility is used to score (segment, classify or rank) records based on a set of predetermined business rules expressed in a model. A data mining deployment model stored in a relational database acts as a realtime decision engine that uses predictive techniques to help make a decision or recommendation. The predictive model contains a set of business rules that have been created in a separate data mining development environment by analyzing existing business operations using information from an operational system or data warehouse. A data mining deployment model is particularly useful in CRM applications for doing one-to-one marketing. Isolating the model from an application allows an organization to enhance the model as trends change, without the need to modify application code. Key features of this product include:

- The ability to use SQL to score an entire database, or to score a single record in realtime in response to a live transaction⁴. Scoring results can be placed in a DB2 database for further analysis.
- Support for industry standards—data mining deployment models are stored in XML format and can be imported into DB2 using files in PMML⁵ format.
- Support for all scoring functions offered by DB2 Intelligent Miner for Data, including decision trees, center-based and distribution-based clustering, polynomial regression and neural networks.
- Implementing the product as a DB2 extender or as an Oracle cartridge in a Windows, UNIX or Linux operating environment.

⁴The SQL used to do this is based on the ISO draft SQL standard concerning SQL multimedia and application packages (including data mining).

⁵The Predictive Model Markup Language (PMML) is an XML-based industry standard developed by the Data Mining Group (DMG) for enabling the interchange of predictive models.



Highlights

DB2 Intelligent Miner for Data can be used to create industry-standard XML-based mining models

DB2 OLAP Server provides multidimensional analysis for a wide range of platforms

Both DB2 and Essbase can be used as database managers.

One product that can be used to develop models for deployment (using either native or PMML formats) into DB2 Intelligent Miner Scoring is IBM **DB2 Intelligent Miner for Data**, which runs on OS/390, z/OS, AS/400, UNIX and Windows NT. DB2 Intelligent Miner provides several data mining algorithms, has good data visualization capabilities, and can exploit the parallel processing capabilities of DB2 Universal Database. The product is one of the few data mining tools on the market to support an external API, allowing result data to be collected by other products for further analysis (by an OLAP tool, for example). IBM also offers its **DB2 Intelligent Miner for Text** product, which provides the ability to extract, index, analyze and categorize information from text sources such as documents and Web pages.

For developing enterprise analytic applications that need to perform significant multidimensional analysis, IBM offers **DB2 OLAP Server™**. This server provides a comprehensive set of financial, mathematical and statistical functions, and runs as an analytical engine in the middle tier of a three-tier client/server configuration. The product was developed in conjunction with Hyperion Solutions, and supports OS/390, z/OS, AS/400, UNIX and Windows. It has the same client API and calculation engine as Hyperion Essbase, and any multidimensional analysis application or tool that supports Essbase can act as a client to DB2 OLAP Server.

Depending on application requirements, the developer can choose one of two backend database servers for use with DB2 OLAP Server (both are included with the product): Hyperion Essbase for optimal performance, or DB2 Universal Database for more flexible data access. When using DB2 Universal Database as the backend data store, the product will automatically generate an optimal relational star schema based on the dimensional design the developer specifies. The Data Warehouse Center can then be used to populate this relational star schema with data extracted from external data sources such as DB2, Oracle, Informix, IMS or VSAM.



Highlights

DB2 OLAP Server Analyzer provides an analysis front-end for Web devices

QMF is used for query and reporting on all DB2 operating platforms.

DB2 OLAP Server Analyzer is an extension to DB2 OLAP Server that supports the development of analyses and reports for display using an HTML or Java Web-based interface. The Analyzer is used to process DB2 OLAP Server multidimensional data and to provide drill-through reporting to detailed DB2 relational data. New analyses are created by dragging and dropping graphical controls such as a chart, spreadsheet or form object onto a graphical working surface. Developers can extend DB2 OLAP Server Analyzer using the Analyzer API toolkit, which enables the rapid assembly of custom, Web-based analytical applications.

DB2 Universal Database comes with an integrated subset of DB2 OLAP Server. This subset, known as the **DB2 OLAP Starter Kit**, enables users to quickly create multidimensional applications and connect and populate those applications with data warehouse information. More advanced analytical processing, support for large numbers of users, and Web capabilities can be obtained by upgrading the starter kit to DB2 OLAP Server.

IBM's strategic offering for query and reporting is the QMF family of tools. DB2 customers have used the OS/390 version of QMF for many years as a query, charting and reporting tool. The High Performance Option (HPO) of the OS/390 product enhances QMF's query governor and administration capabilities, and allows queries and procedures to be compiled and run as static COBOL programs. **QMF for Windows** provides query and reporting capabilities for the Microsoft Windows environment. It supports Java-based SQL queries and integrates with most desktop tools, such as Lotus Notes®, spreadsheets and local databases via a Microsoft OLE automation interface. Both QMF products support access not only to DB2 databases, but also to any heterogeneous data source supported by DB2 DataJoiner. Report output from QMF can be published to HTML for viewing by a standard Web browser.



Highlights

Brio and Business Objects are key IBM partners

WebSphere platform is used to deliver analytic application solutions.

To increase the scope of its query and reporting offerings, IBM has also forged relationships with Brio Technology and Business Objects. IBM intends the relationships with these tool vendors to be more than mere joint marketing deals—they also involve agreements to integrate the products from these companies with IBM's business intelligence offerings in the area of meta data interchange and administration, for example.

Analytic application packages

Organizations need to be able to quickly deploy enterprise analytic applications if they are to remain competitive in today's dynamic business world. To support this requirement, IBM is developing partnerships with leading analytic application package vendors to make their products available on the IBM enterprise analytics platform. For e-business users, IBM offers packaged analytic applications under the WebSphere brand name. WebSphere is a software platform for building and deploying Web-based enterprise applications. It consists of:

- A set of WebSphere e-business tools and applications that are integrated into a Java-based application server. Key capabilities here that are relevant to the business intelligence and enterprise analytics environment are the WebSphere Portal Server (discussed in the section *Portal Development*), a suite of e-business applications known as the WebSphere Commerce Suite and the WebSphere Site Analyzer and WebSphere Commerce Analyzer Advanced Edition e-business analytic application packages.
- VisualAge[®] development tools for building and testing Java-based Web applications.
- WebSphere MQ for developing message-based applications. One important facility here is the WebSphere MQ Integrator, which transforms, routes and distributes message-based data between systems using XML-based adapters. One use of this product is to feed data from new and existing applications to a near-realtime data warehouse using the services of DB2 Warehouse Manager.



Highlights

WebSphere Commerce Analyzer Advanced Edition builds a data warehouse for analyzing e-business data

The WebSphere Site Analyzer provides analysis Web site usage. It can be used to detect visitor trends and preferences, and manage Web site content and structure. The WebSphere Commerce Analyzer Advanced Edition is a component of the WebSphere Commerce Suite that enables organizations to create e-business analytics for measuring the effectiveness and success of e-business marketing campaigns. It uses the DB2 Warehouse Manager to extract Web data into a DB2 data warehouse for analysis by DB2 OLAP Server, DB2 Intelligent Miner, and Brio analysis and reporting tools.

Portal development

A portal provides an integrated and personalized Web-based view of the business content users need to perform their jobs. The main objectives of a portal are to make business content easier to find, make business users more efficient, and to improve collaboration between internal business users and external clients and business partners. A portal may provide access to business intelligence and enterprise analytics, operational and e-business applications, office documents, Web HTML pages and multimedia, collaboration services, and so forth. Portal products vary in the types of services and business content they provide and the types of users they support. IBM markets three portal products: WebSphere Portal Server, IBM Enterprise Information Portal and Lotus K-station™.

WebSphere Portal Server is the strategic IBM platform for portal development.

WebSphere Portal Server (WPS) forms the core of IBM's strategy for supporting the building of a Java-based enterprise-level portal. It is integrated into the WebSphere e-business environment, and offers pervasive device support (using WebSphere Everyplace Suite), personalization, categorization, collaboration services (through integration with Lotus Notes or Microsoft Exchange), workflow services (using WebSphere MQ Workflow), user group authorization and security, federated search (using Enterprise Information Portal), a set of *portlets* for accessing a variety of different information and application content stores and servers, and a development platform for building customized portlets.



Highlights

K-station is a packaged portal for supporting unstructured data and collaboration.

Enterprise Information Portal (EIP) extends the capabilities of the WebSphere Portal Server by providing a federated search capability, and access to a broader set of structured and unstructured information sources such as IBM Content Manager, Lotus Notes databases, DB2 Universal Database, DB2 DataJoiner-enabled data sources, DB2 Warehouse Manager, etc. IBM's direction is to integrate the capabilities of EIP into WebSphere Portal Server.

Lotus K-station is a collaborative portal that organizes and manages multilingual unstructured business content by community, interest, task or job focus. It is designed to maximize business user information sharing and collaboration. K-station can be used in conjunction with the **Lotus Discovery Server™** to locate information and expertise. To do this, the Discovery Server extracts, analyzes and categorizes structured and unstructured information to reveal relationships between content, people, topics and user activity in an organization. It automatically generates and maintains a knowledge map (K-map) that can be searched or browsed to determine relevant content categories and their appropriate hierarchical mappings. Unlike WebSphere Portal Server and EIP, which both provide portal development platforms, Lotus K-station is an out-of-the-box portal application. IBM's direction is to integrate portlets from K-station and Discovery Server into the WebSphere Portal Server.



Highlights

A sound and scalable infrastructure is required for enterprise analytics

IBM is ideally positioned for supporting scalable enterprise analytics.

Conclusion

Fourth-generation business information systems provide organizations with the ability to integrate business intelligence solutions into their business processes to reduce costs and increase profits. They also deliver on-demand enterprise analytics in support of realtime decision making. To be successful, however, such systems must be built on top of a sound and scalable infrastructure.

In this paper we have presented the strategy and development initiatives of IBM's enterprise analytics software solution. The five main thrusts of this solution are:

1. The rapid deployment of data warehouse and business intelligence applications that support near-realtime data warehouses, on-demand enterprise analytics and realtime business decisions and actions
2. An integrated end-to-end enterprise analytics solution involving products and services from IBM and its business partners
3. Support for leading-edge business intelligence technologies and evolving industry standards
4. High performance, scalable hardware and software
5. A multitiered and heterogeneous business information environment that supports both IBM and non-IBM database products.

This solution provides an ideal platform for building a scalable and integrated fourth-generation business information system, and IBM is well positioned to be one of the leaders in supplying a new generation of business intelligence solutions to provide users with the information and analytics they need to manage their business operations.



Appendix: summary of IBM enterprise analytics products

Documented below is a summary of the main IBM enterprise analytics products discussed in this paper.

Database Management

DB2 Universal Database is an object-relational database that runs on many different operating platforms including OS/390, z/OS, AS/400, AIX, Sun Solaris, HP-UX, NUMA-Q, Linux, OS/2 and Windows. It comes in a variety of different packages, the key ones being:

- *DB2 Satellite Edition* is a single-user database system for Windows-based mobile computers
- *DB2 Personal Edition* is a single-user database system for PC-based desktop computers
- *DB2 Workgroup Edition* for LAN-based workgroup database servers
- *DB2 Enterprise Edition* for large uniprocessor and symmetric multiprocessor database servers
- *DB2 Enterprise-Extended Edition* for large massively parallel and clustered database servers
- *DB2 Universal Database for AS/400* for AS/400 database servers
- *DB2 Universal Database for OS/390* for OS/390 database servers
- *DB2 Personal and Universal Developers' Editions* provide all the tools a software developer needs to build applications for DB2 Universal Database.

Most (but not all) DB2 packages contain the following:

- *DB2 Control Center* for managing DB2 databases
- *Data Warehouse Center* for designing and populating Windows NT-centric data warehouses
- *DB2 OLAP Starter Kit* for performing online analytical processing (OLAP) using Microsoft Excel and Lotus 1-2-3 clients. Both the IBM DB2 and Hyperion Essbase database managers are included in the kit, and can be selected on an application-by-application basis



- *DB2 Connect™* for supporting client-to-server and server-to-server network connections
- *DB2 Extenders™* for managing text, image, audio, and video
- *DB2 XML Extender* for managing XML data
- *DB2 DataPropagator* for replicating data between DB2 and third-party database servers
- *Net.Data* for accessing DB2 and other data stores from the Web
- *WebSphere Application Server Standard Edition*
- *WebSphere Studio and Visual Age for Java* for developing Web applications.

DB2 Everyplace is a relational database manager for handheld mobile devices.

DB2 Net Search Extender employs in-memory database technology to index and rapidly search large DB2-based Web sites.

Federated data access

DB2 Connect for supporting client-to-server and server-to-server network connections.

DB2 Relational Connect provides a native SQL interface to Oracle enabling DB2 applications to conduct distributed SQL queries across DB2 and Oracle databases.

DB2 DataJoiner supports distributed heterogeneous data access and update against a variety of different database systems.

DataJoiner Classic Connect works in conjunction with *DB2 DataJoiner* to provide read-only data access to IMS databases and VSAM files.

Data Links Manager enables DB2 applications to access and update data in standard operating system files.

Net.Data for accessing DB2 and other data stores from the Web.



Data warehouse management

DB2 Warehouse Manager extends the *Data Warehouse Center* facility of DB2 in the areas of scalability, data transformation capabilities and meta data management. The product also includes query, reporting and query governing using the services of IBM Query Management Facility (see description below) and the DB2 Query Patroller.

DB2 DataPropagator for replicating data between DB2 and non-IBM DBMSs.

IMS DataPropagator captures data changes from IMS databases for loading into a DB2-based data warehouse.

DB2 DataRefresher captures and transforms data stored in non-relational databases and files such as IMS and VSAM.

WebSphere MQ (formerly *IBM MQSeries*) is a family of products that provides messaging facilities that support data transformation, workflow, intelligent routing, assured message delivery to mobile devices and tools for building customized XML-based application adapters.

Analytic application development

DB2 Spatial Extender employs DB2 user-defined types and functions for managing and analyzing spatial data in DB2 databases. Visualization tools from ESRI can also be used to process the DB2 spatial data.

DB2 Intelligent Miner Scoring uses DB2 user-defined types and functions to store XML-based production data mining models in a DB2 database. This facility enables DB2 applications and tools to dynamically invoke a predictive mining model to analyze and score data in DB2 and third-party databases.

DB2 Intelligent Miner for Data and **DB2 Intelligent Miner for Text** offer a variety of different data mining algorithms for developing analytic applications that process business data and text managed by a variety of database and file systems.



IBM Query Management Facility supports the querying, reporting and charting of data in DB2 and third-party databases.

DB2 OLAP Server is an advanced version of the *DB2 OLAP Starter Kit* that is used for designing, populating and analyzing multidimensional databases managed by DB2 Universal Database and Hyperion Essbase.

DB2 OLAP Server Analyzer is an extension to the DB2 OLAP Server that supports the development of analyses and reports for display via an HTML or Java Web-based interface.

Analytic application packages

WebSphere Commerce Suite employs *WebSphere Application Server* to deliver a set of functions for managing and analyzing business-to-business, business-to-consumer and e-marketplace e-business operations. Included with the Commerce Suite are the **WebSphere Commerce Analyzer Advanced Edition** and **WebSphere Site Analyzer**. The Commerce Analyzer measures and analyzes the effectiveness and success of marketing campaigns. The Site Analyzer supports the analysis of Web site visitor trends and usage.

Portal development

WebSphere Portal Server provides a scalable portal development framework that exploits *WebSphere Application Server* for Web services, the Enterprise Information Portal for information access, Lotus software for collaboration and knowledge management, and the *WebSphere Everyplace Suite* for mobile and pervasive device support.

Enterprise Information Portal provides a set of development tools and services for building customized portals that can provide personalized access to a variety of different business information stores.

Lotus K-station is a packaged collaborative portal application that organizes and manages multilingual, unstructured business content by business community, interest, task or job focus.



About DataBase Associates, Inc.

DataBase Associates is a consulting company specializing in leading-edge technologies in the fields of data warehousing, business intelligence, analytic applications, corporate portals and database technology.



© Copyright IBM Corporation 2001

IBM Corporation
Silicon Valley Laboratory
555 Bailey Avenue
San Jose, CA 95141
U.S.A.

Reproduced with Permission of
DataBase Associates
09-01
All Rights Reserved

AIX, AS/400, DB2, DB2 Connect, DB2 DataJoiner, DB2 Extenders, DB2 OLAP Server, DB2 DataPropagator, DB2 DataRefresher, DB2 Universal Database, Everyplace, IMS, Informix, Intelligent Miner, iSeries, MQSeries, Net.Data, Netfinity, NUMA-Q, OS/2, OS/390, pSeries, QMF, RS/6000, S/390, VisualAge, WebSphere, xSeries, z/OS and zSeries are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries or both.

Lotus, Lotus Notes and Notes are registered trademarks and Discovery Server, Domino and K-station are trademarks of Lotus Development Corporation in the United States, other countries or both.

Linux is a registered trademark of Linus Torvalds.

Microsoft, Windows and Windows NT are registered trademarks of Microsoft Corporation in the United States, other countries or both.

SAP and R3 are registered trademarks of SAP AG.

Java, all Java-based trademarks and Solaris are trademarks of Sun Microsystems, Inc. in the United States, other countries or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Other company, product or service names may be trademarks or service marks of others.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates. Offerings are subject to change, extension or withdrawal without notice.