



The IBM Business Intelligence Software Solution

Prepared for IBM

by Colin J. White

DataBase Associates International, Inc.

Version 3, March 1999



TABLE OF CONTENTS

WHAT IS BUSINESS INTELLIGENCE?	1
The Evolution of Business Information Systems	1
First-Generation: Host-Based Query and Reporting	1
Second-Generation: Data Warehousing	1
Third-Generation: Business Intelligence	2
Business Driving Forces	3
Business Intelligence Requirements	4
IBM'S BUSINESS INTELLIGENCE SOFTWARE STRATEGY	5
An Historical Perspective	5
Business Intelligence Structure	5
Business Intelligence Partner Initiative	7
Technology Initiatives	8
Advanced Decision Support Tools	8
Meta Data Integration and Interchange	8
Database Performance and Scalability	9
Database Extensibility	10
Heterogeneous Database Support	10
Web Enablement	10
Summary: The IBM Strategy	11
THE IBM BUSINESS INTELLIGENCE PRODUCT SET	12
Business Intelligence Applications	12
Business Intelligence Tools	12
Access Enablers	14
Data Warehouse Modeling and Construction	15
Data Management	18
CONCLUSION	19

Brand and product names mentioned in this paper may be the trademarks or registered trademarks of their respective owners.

WHAT IS BUSINESS INTELLIGENCE?

Given the increasing competition in today's tough business climate, it is vital that organizations provide cost-effective and rapid access to business information for a wide range of business users if they are to survive into the new millennium. The solution to this issue is a business intelligence system¹ that provides a set of technologies and products for supplying users with the information they need to answer business questions, and make tactical and strategic business decisions.

Many of the concepts of business intelligence are not new, but have evolved and been refined based on experience gained from early host-based corporate information systems, and more recently, from data warehousing applications. This paper provides an introduction to business intelligence, compares and contrasts it with data warehousing, reviews requirements for a business intelligence system, and takes a detailed look at IBM's business intelligence strategy and product set.

THE EVOLUTION OF BUSINESS INFORMATION SYSTEMS

Three generations of systems

Inevitably the first question that arises when describing the objectives of a business intelligence system is, "Doesn't a data warehouse have the same objectives and provide the same capabilities as a business intelligence system?" A similar question arose when data warehouses were first introduced, "Isn't a data warehouse similar to the corporate information systems and information centers we built in the past?" Although a quick and simple answer to both questions is *yes*, closer examination shows that in the same way that there are important differences between a warehouse and early corporate information systems and information centers, there are also important differences between a business intelligence system and a data warehouse.

First-Generation: Host-Based Query and Reporting

First generation systems difficult to use

Early business information systems employed batch applications to provide business users with the information they needed. The output from these applications typically involved huge volumes of paper that users had to wade through to get the answers they needed to business questions. The advent of terminal-driven time-sharing applications provided more rapid access to information, but these systems were still cumbersome to use, and required access to complex operational databases. This first generation of business information systems could, therefore, only be used by *information providers*, such as business analysts, who had an intimate knowledge of the data and extensive computer experience. *Information consumers*, like business executives and business managers, could rarely use these early systems, and instead had to rely on information providers to answer their questions and supply them with the information they needed.

Second-Generation: Data Warehousing

¹ Database Associates calls such a system a *decision processing system*, but to avoid confusion we will use the IBM term *business intelligence system* in this paper.

Warehousing designed for the business user

The second generation of business information systems came with data warehousing, which provided a giant leap forward in capability. Data warehouses have several advantages over first-generation systems:

- Data warehouses are designed to satisfy the needs of business users and not day-to-day operational applications.
- Warehouse information is clean and consistent, and is stored in a form business users can understand.
- Unlike operational systems, which contain only detailed current data, warehouses can supply both historical and summarized information.
- The use of client/server computing provides data warehouse users with improved user interfaces and more powerful decision support tools.

Third-Generation: Business Intelligence

Warehouse vendors need to provide packaged application solutions

A data warehouse is still not a complete solution to the needs of business users. One weakness of many data warehouse solutions is that the vendors often focus on technology, rather than business solutions. While there is no doubt that data warehouse vendors provide powerful products for building and accessing a data warehouse, these products can require a significant amount of implementation effort. The issue here is that warehouse products rarely come prepackaged for specific industries or application areas, or address particular business problems. This is very much like the situation in the early days of client/server computing when vendors initially provided the technology for developing operational applications, but then quickly realized that organizations were looking for application and business solutions, and not yet more technology. Vendors fixed this problem, with the result that today many operational client/server applications are built using application packages, rather than being handcrafted by developers. The same evolution has to happen in business information systems – vendors must provide application packages, and not just more technology. One distinguishing factor of business intelligence systems is that they focus on providing prepackaged application solutions in addition to improved technology.

Business intelligence systems focus on easy access to information

Another issue with data warehousing is that much of the focus is still on building the data warehouse, rather than accessing it. Many organizations seem to think that if they build a warehouse and provide users with the right tools, the job is done. In fact it is just beginning. Unless the information in the warehouse is thoroughly documented and easy to access, complexity will limit warehouse usage to the same information providers as first-generation systems. Business intelligence systems focus on improving the access and delivery of business information to both information providers and information consumers. They achieve this by providing advanced graphical- and Web-based online analytical processing (OLAP) and information mining tools, and prepackaged applications that exploit the power of those tools. These applications may need to process and analyze large volumes of information using a variety of different tools. A business intelligence system must, therefore, provide scalability and be able to support and integrate products from multiple vendors.

Business intelligence systems provide an information catalog for managing information

Most business intelligence systems provide an information catalog that helps organizations organize, manage, and find enterprise business information. A publishing facility allows both technical and business users to document the business information that exists in an organization. To find information, business users enter a description of the type of information they are looking for, and the tool searches its catalog looking for information objects (documents, reports, analyses, etc.) that potentially satisfy the user's request. A list of these information objects is returned to the user, who can then select the ones of interest to be retrieved. A subscription facility enables the user to have information delivered to them on a regular basis via a corporate intranet or e-mail.

Business intelligence systems support many types of information

The information stored in a data warehouse is typically sourced from operational databases (and in some cases external information providers). There is, however, also a considerable amount of business information kept in office and groupware systems, on Web servers on intranets and on the Internet, and in paper form on people's desks. To solve this issue, business intelligence systems are designed to support access to all forms of business information, not only the data stored in a data warehouse. A business intelligence system does not negate the need for a data warehouse – a data warehouse is simply one of the data sources that can be handled by a business intelligence system. We see then that a business intelligence system is a third-generation business information system that has three key advantages:

Three key advantages over warehousing

1. Business intelligence systems not only support the latest information technologies, but also provide prepackaged application solutions.
2. Business intelligence systems focus on the access and delivery of business information to end users, and support both information providers and information consumers.
3. Business intelligence systems support access to all forms of business information, and not just the information stored in a data warehouse.

BUSINESS DRIVING FORCES

So far we have seen that many of the driving forces behind business intelligence come from the need to improve ease-of-use and reduce the resources required to implement and use new information technologies. There are, however, also three important business driving forces behind business intelligence:

Three business driving forces

1. *The need to increase revenues, reduce costs, and compete more effectively.* Gone are the days when end users could manage and plan business operations using monthly batch reports, and IT organizations had months to implement new applications. Today companies need to deploy informational applications rapidly, and provide business users with easy and fast access to business information that reflects the rapidly changing business environment. Business intelligence systems are focused towards end-user information access and delivery, and provide packaged business solutions in addition to supporting the sophisticated information technologies required for the processing of today's business information.

Business intelligence systems also help to reduce IT costs

2. *The need to manage and model the complexity of today's business environment.* Corporate mergers and deregulation means that companies today are providing and supporting a wider range of products and services to a broader and more diverse audience than ever before. Understanding and managing such a complex business environment and maximizing business investment is becoming increasingly more difficult. Business intelligence systems provide more than just basic query and reporting mechanisms, they also offer sophisticated information analysis and information discovery tools that are designed to handle and process the complex business information associated with today's business environment.
3. *The need to reduce IT costs and leverage existing corporate business information.* The investment in IT systems today is usually a significant percentage of corporate expenses, and there is a need not only to reduce this overhead, but also to gain the maximum business benefits from the information managed by IT systems. New information technologies like corporate intranets, thin-client computing, and subscription-driven information delivery help reduce the cost of deploying business intelligence systems to a wider user audience, especially information consumers like executives and business managers. Business intelligence systems also broaden the scope of the information that can be processed to include not only operational and warehouse data, but also information managed by office systems and corporate Web servers.

BUSINESS INTELLIGENCE REQUIREMENTS

Five main business intelligence requirements

Summarizing the two previous sections we see that the main requirements of a business intelligence system are:

1. Support for prepackaged application solutions.
2. A cost-effective solution that provides a quick payback to the business and enables an organization to compete more effectively.
3. Fast and easy access to an organization's business information for a wide range of end users, including both information providers and information consumers.
4. Support for modern information technologies, including information analysis and discovery techniques like online analytical processing (OLAP) and information mining.
5. An open, extensible, and scalable operating environment.

Now that we have defined what a business intelligence system is, and have also identified its key requirements, we can move on to look at IBM's business intelligence strategy and products.

IBM'S BUSINESS INTELLIGENCE SOFTWARE STRATEGY

AN HISTORICAL PERSPECTIVE

In the first part of this paper, we saw that business intelligence represents a third-generation business information system that has evolved from early host-based information systems and more recently, data warehousing. IBM has provided products for all three generations of business information system. In fact, APL running on IBM's time-sharing systems was one of the first commercial OLAP tools.

IBM an early pioneer of data warehousing

Although many people claim to have invented data warehousing, one of the initial sources of the data warehouse concept was a 1988 *IBM Systems Journal* article by Barry Devlin and Paul Murphy entitled "An Architecture for a Business and Information System." This paper documented 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. Although IBM was one of the first vendors to enter the commercial data warehouse arena, the company was slow to exploit its early lead from both a development and marketing perspective, and was rapidly challenged by competing vendors. More recently, IBM has put significant resources into both the development and marketing of its warehousing products, with the result that it has made up much of the ground lost. As the industry moves toward the use of third-generation business information systems, IBM intends to maintain its development and marketing momentum by providing an integrated end-to-end business intelligence solution. This part of the paper discusses the structure, technologies, and strategy of this solution from a software perspective.

BUSINESS INTELLIGENCE STRUCTURE

The IBM business intelligence structure is an evolution of IBM's earlier Information Warehouse architecture, and is illustrated in Figure 1. (The products that support this structure are shown in Figure 2 and are discussed in the next part of this paper.) The structure consists of the following components:

Business applications a key component

- **Business intelligence applications.** These applications are complete business intelligence solution packages tailored for a specific industry and/or application area. These packages use products from other components of the business intelligence structure. IBM's key business intelligence applications are marketed under the DecisionEdge brand name.

Includes OLAP and information mining

- **Decision support tools.** These tools range from basic query and reporting tools to advanced online analytical processing (OLAP) and information mining tools. All these tools support GUI-driven client interfaces. Many can also be used from a Web interface. These tools are designed to handle structured and unstructured information managed by a variety of different database and file system products. IBM provides its own tools here, but also has marketing and development

relationships with leading third-party vendors including Brio Technology, Business Objects, Cognos, and Hyperion Solutions.

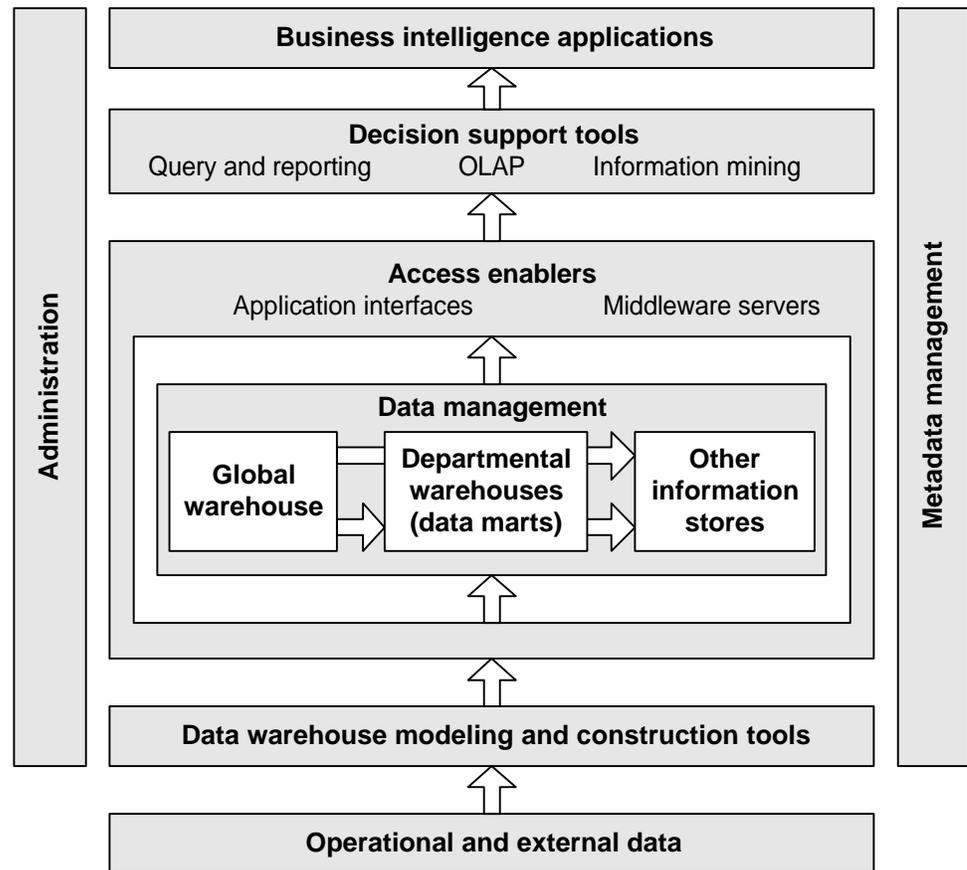


Figure 1. IBM business intelligence structure

Middleware access to heterogeneous databases

- **Access enablers.** These consist of application interfaces and middleware that allow client tools to access and process business information managed by database and file systems. Database middleware servers enable clients to transparently access multiple back-end IBM and non-IBM database servers — this is known as a *federated* database. Web server middleware allows Web clients to connect to this federated database.
- **Data management.** These products are used to manage the business information of interest to end users. Included in this product set are IBM’s DB2 products for the OS/390, VSE and VM, AS/400, UNIX (AIX, HP-UX, Solaris, UnixWare), Linux, OS/2, and Windows (NT, 95, 98) environments.² Business information can

² With the exception of DB2 for VSE and VM, IBM uses the term *DB2 Universal Database* to refer to this set of DB2 products. For simplicity, we will use the term DB2 Universal Database throughout this paper, but the reader should be aware that certain features of DB2 Universal database documented in this paper may not be supported by every DB2 product on every operating platform.

also be accessed and maintained by third-party relational database products through the use of IBM's database middleware products.

Three levels of information store

IBM sees up to three levels of information store being used to manage business information. This three-level architecture is based on existing data warehousing concepts, but as has already been mentioned, other types of information, for example, multimedia data, can also be supported. At the top level of the architecture is the global warehouse, which integrates enterprise-wide business information. In the middle tier are departmental warehouses that contain business information for a specific business unit, set of users, or department. These departmental warehouses may be created directly from operational systems, or from the global warehouse. (Note that these departmental warehouses are often called *data marts*.) At the bottom of the architecture are other information stores, which contain information that has been tailored to meet the requirements of individual users or a specific application. An example of using this latter type of information store would be where financial data is extracted from a departmental information store and loaded in a separate store for modeling by a financial analyst.

Non-IBM warehouses supported

- **Data warehouse modeling and construction tools.** These tools are used to capture data from operational and external source systems, clean and transform it, and load it into a global or departmental warehouse. IBM products use the database middleware of the Access Enabler component shown in Figure 1 to access and maintain warehouse data in non-IBM databases. IBM markets its own warehouse construction tools under the Visual Warehouse brand name, and also has marketing and development relationships with third-party vendors such as Evolutionary Technologies International and Vality Technology.

Handles technical and business meta data

- **Meta data management.** This component manages the meta data associated with the complete business intelligence system, including the technical meta data used by developers and administrators, and the business meta data for supporting business users.
- **Administration.** This component covers all aspects of business intelligence administration, including security and authorization, backup and recovery, monitoring and tuning, operations and scheduling, and auditing and accounting.

BUSINESS INTELLIGENCE PARTNER INITIATIVE

Partners support the business intelligence structure

IBM's business intelligence structure is designed to be able to integrate and incorporate not only IBM's business intelligence products, but also those from third-party vendors. To encourage support for its business intelligence structure, IBM has created a Business Intelligence Partner Program for ISVs, VARs, systems integrators, and consultants. Over 250 companies have joined this program to date. The objective of the program is to have not only joint marketing relationships with other organizations, but also joint development initiatives that enable other software vendors' products to be integrated with IBM's business intelligence 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 Brio Technology, Business Objects, Cognos, Evolutionary Technology International, Hyperion Solutions, and Vality Technology. The next part of this paper on IBM's business intelligence product set reviews the level of integration that has been achieved to date with products from these vendors.

TECHNOLOGY INITIATIVES

IBM has always had a solid reputation for its technology, and in this section we briefly review some of its ongoing business intelligence software technology initiatives.

Advanced Decision Support Tools

Emphasis on OLAP and information mining

IBM has for several years put significant research into its information mining technology and its Intelligent Miner product is an industry leader here. More recently, IBM developed with Hyperion Solutions its DB2 OLAP server, which supports sophisticated online analytical processing of information stored in relational and multidimensional databases. IBM's direction with both its information mining and OLAP products is to provide prepackaged business applications that allow organizations to rapidly deploy advanced decision applications.

Meta Data Integration and Interchange

Metahubs facilitate meta data interchange

For years IBM and other vendors have struggled to solve the problem of meta data integration. The key issue here is that every product requires different meta data and different meta data models (known as *metamodels*). There have been enough failed efforts over the years to clearly demonstrate that it is simply not possible to have a single meta data store that implements a single metamodel for all the meta data in an organization. The only solution is, instead, to improve meta data interchange between products, and to automate and synchronize this interchange wherever possible. This is IBM's strategy for its business intelligence structure, and it intends to achieve this by employing a *metahub* to manage the flow of meta data between products.

At the center of IBM's business intelligence structure is its Visual Warehouse product family, which is used for building global and departmental data warehouses. Visual Warehouse provides two meta data stores — one for the technical meta data used by the Visual Warehouse Manager. The other is the Visual Warehouse Information Catalog (formerly known as DataGuide), which is used for handling the business and technical meta data associated with the complete business intelligence environment.

Visual Warehouse manages warehouse construction

One of the objectives of the Visual Warehouse Manager is to act as a central control point for managing warehouse construction operations done by both IBM products and by products from business partners like Hyperion Solutions, Evolutionary Technologies International, and Vality Technology. To facilitate this management process, Visual Warehouse handles the interchange of meta data between these partner products and Visual Warehouse Manager's technical meta data store.

Visual Warehouse provides a meta data interchange hub

IBM's meta data strategy is for the Visual Warehouse Information Catalog to act as a metahub for meta data interchange between products in the business intelligence environment. The equivalent metahub for operational systems is VisualAge Data Atlas, which can exchange meta data with the Visual Warehouse Information Catalog.

IBM is working on an Object Management Group (OMG) initiative known as Common Warehouse Metadata Interchange (CWMI). The CWMI initiative is driven by IBM, Oracle, and Unisys, and has the objective of establishing meta data standards for data warehousing. At present OMG has issued a Request for Proposal (RFP) for vendors to offer technology for supporting CWMI. It is expected that CWMI will employ CORBA, and the OMG Meta Object Facility (MOF), which is a UML-based facility for defining metamodels. IBM will respond to this RFP with technology and tools that support meta data interchange based on XML, and a common set of UML-based data warehouse metamodels. As this meta data interchange technology evolves it will be incorporated into Visual Warehouse.

Database Performance and Scalability

Through all three generations of business information systems, it has been the case that the number of users, amount of data, and complexity of processing always exceeds predictions. The building and use of business intelligence solutions requires careful planning and administration, but organizations also need to employ products that can scale to deal with the growth in the use of a business intelligence system. Scalability covers many areas, including the ability to granularly add hardware processors and disk drives, the availability of parallel processing hardware and software, support for a large number of users, and the ability to manage large databases.

Support for parallel processing

IBM's strategy is to provide relational database products that operate on a wide range of platforms, and which can handle small to very large information stores. One key element of this strategy is scalable hardware and software that supports parallel processing. IBM markets four families of parallel computing hardware – System/390, RS/6000, AS/400, and Netfinity. The DB2 offerings running on these machines all provide parallel query and utility processing for handling large decision-support workloads and for managing large databases. To aid in testing the integration and scalability of its business intelligence hardware and software solutions, IBM has invested \$47 million in creating a set of Teraplex Integration Centers that can be used by IBM, its business partners, and selected customers to stress test proposed solutions involving very large databases and workloads.

Another key component of IBM's database strategy is to aid decision processing performance by adding extensions to the data management and SQL optimization components of its DB2 products. One example of this is the encoded vector index (EVI) feature recently added to DB2 for AS/400. EVI was developed by IBM research, and is the industry's first vector approach to bitmapped indexes. Other examples of this strategy include DB2 Universal Database support for SQL ROLLUP and CUBE operators for multidimensional analysis, and aggregate aware optimization of repeated SQL queries.

Database Extensibility

DB2 Universal Database supports extensibility

We mentioned in the first part of this paper that one key distinguishing factor of a business intelligence system was its ability to handle all types business information. To aid in the management of complex data such as documents, maps, multimedia, and so forth, IBM has enhanced its database products to support complex data via user-defined data types and functions. One of the first products to support these new facilities is the DB2 Spatial Extender, which gives users the ability to store and analyze geo-spatial data (maps and geographic information, for example). Many organizations are beginning to recognize that they need to manage geo-spatial data in addition to the standard business information handled by a data warehouse. Supporting geo-spatial data in a data warehouse provides users with advanced query functions and the ability to display results visually in a geo-spatial context.

There are several factors that need to be considered when determining if complex data should, or should not, be stored in a database system. The advantages databases offer include the ability to search and manipulate data, and data recovery and security. The disadvantages are that this creates a more complex environment, and the database system must be available to access business information. To help resolve this conflict, IBM has added the concept of *data links* to DB2 Universal Database. A data link enables data to be maintained outside of the database system, for example, in a flat file, but still allows database interfaces to be used for both accessing and updating data. This feature supports three ways of handling business information:

- Store the data in a database and access it via database APIs.
- Store the data in a flat file and access it via database or file APIs.
- Store the data in a flat file (for example, on a Web server) and access it via a file API.

Heterogeneous Database Support

DB2 Data Joiner supports access to non-IBM DBMSs

In the past, IBM has often justifiably been accused of supplying proprietary solutions that can only be used with its database products. One major step forward in resolving this issue comes with IBM's DB2 DataJoiner database middleware product. This product not only allows decision support tools to access and update data managed by both IBM and non-IBM database servers, but also permits warehouse construction tools to use non-IBM databases as both a data source for a warehouse and as a data target for storing warehouse information.

Web Enablement

Like many vendors, IBM sees that the use of Web technology can significantly reduce the cost of deploying business intelligence solutions to a broad spectrum of end users. IBM's strategy here is to integrate its Web offerings with its business intelligence products. For example, DB2 for Domino is a set of application connectors that provide Lotus Notes users and Lotus Domino applications with easy access to DB2 databases. IBM also allows Web users and applications to access DB2 data via its Net.Data Web middleware product, which is included with DB2 Universal Database and the IBM WebSphere Web application server.

SUMMARY: THE IBM STRATEGY

**IBM strategy
consistent with
business
intelligence
requirements**

The business intelligence structure and technology initiatives reviewed above clearly indicate the thrust of IBM's business intelligence strategy:

- The provision of business-focused applications that incorporate leading-edge information and decision support technologies.
- A complete and integrated end-to-end solution involving products from both IBM and its key business partners.
- The provision of scalable hardware and software that can handle a wide range of different types of business information.
- A federated database environment that supports both IBM and non-IBM database products that can manage a variety of different types of information store, including global and departmental data warehouses.

This strategy is consistent with the business intelligence system requirements outlined in the first part of this paper.

THE IBM BUSINESS INTELLIGENCE PRODUCT SET

This part of the paper reviews the products and tools provided by IBM (and its key partners) for supporting a business intelligence software environment — these products are listed in Figure 2. We will use the IBM Business Intelligence Structure shown in Figure 1 to categorize and describe these products.

Business Intelligence Applications

Applications for customer relationship management

IBM's business intelligence applications are marketed under the **DecisionEdge** brand name. DecisionEdge is a Customer Relationship Management (CRM) solution that allows organizations to analyze consumer behavior with the objective of increasing market share and customer profitability. To date, IBM has announced DecisionEdge packages for the finance, insurance, telecommunications, and utilities industries. Each DecisionEdge offering provides integrated hardware, software, consulting services, and business applications centered on an industry-specific data model. DecisionEdge for Telecommunications, for example, analyzes customer information measuring profitability, predicting customer behavior, analyzing attrition, and assists in the creation of tailored customer marketing programs. DecisionEdge for Finance, Banking, and Securities offers pre-defined solutions in the areas of marketing and sales, and risk and profitability analysis. All DecisionEdge packages support the OS/390, AS/400, UNIX, and Windows NT operating environments, and include the VALEX marketing automation and campaign management software developed by Exchange Applications.

Include information mining applications

DecisionEdge also capitalizes on IBM's heavy investment in information mining research. Utilizing the Intelligent Miner development environment, DecisionEdge provides the optional **Intelligent Miner for Relationship Marketing** application to help the business user obtain a better understanding of key business issues such as customer segmentation, and potential buying and loyalty behavior.

IBM is placing increasing emphasis on the use of business intelligence applications and is bringing applications to market in several industry areas including student administration, retail banking, local and state human services, and e-commerce.

OLAP application packages

Business intelligence applications are also available for the DB2 OLAP Server (see description below). This product (which was developed by IBM and Hyperion Solutions) employs the same API as Hyperion Essbase, and can, therefore, be used with the many industry-specific third-party application packages available for Essbase.

Business Intelligence Tools

Business intelligence tools can be broken down into three categories: query and reporting, online analytical processing (OLAP), and information mining.

QMF for Windows product

Query and Reporting

The main IBM query and reporting offering is the **Query Management Facility (QMF)** family of tools. The System/390 version of QMF has been used for many years as a host-based query, and reporting tool by DB2 for S/390 users. More recently, IBM introduced QMF for Windows, a native Windows version of QMF that supports access not only to DB2 databases, but also any relational and non-relational data source supported by its DB2 Data Joiner middleware product (see description below). QMF host objects are compatible with QMF for Windows, extending the enterprise query environment to Windows and the Web. Output from QMF can be published to the Web, and can be passed to other Windows applications like Lotus 1-2-3, Microsoft Excel, and other desktop products via Windows OLE.

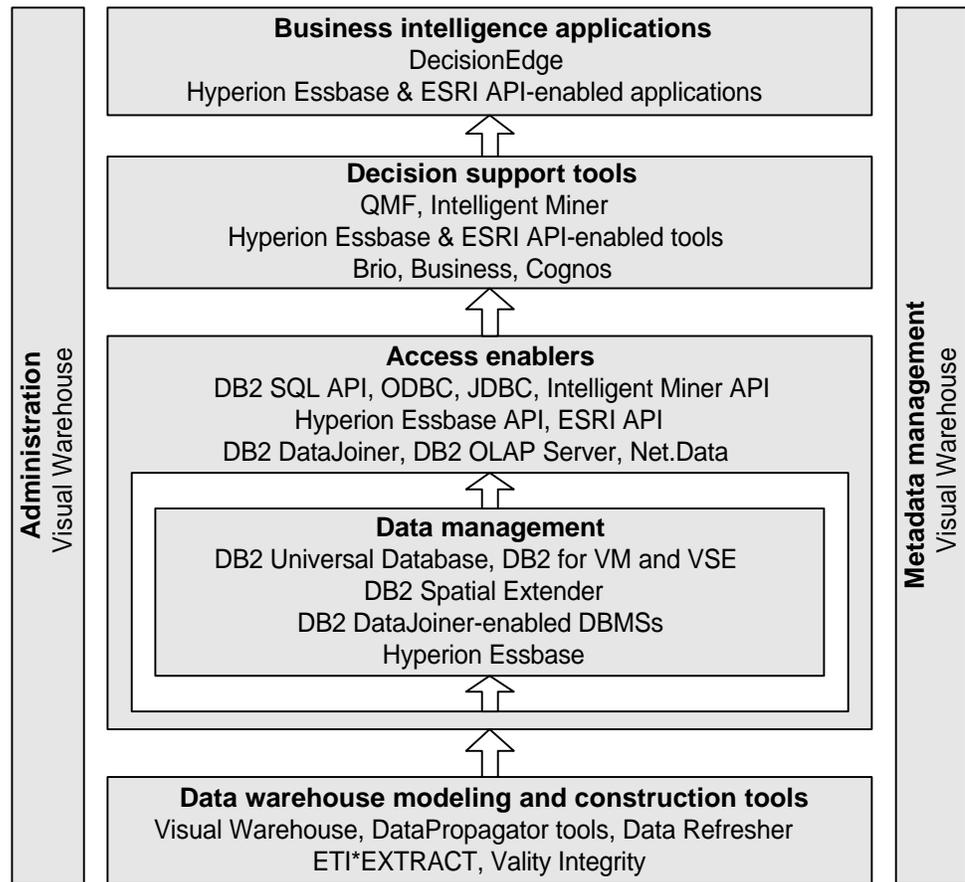


Figure 2. IBM business intelligence product set

Brio, Business Objects and Cognos are partners

To increase the scope of its query and reporting offerings, IBM has forged relationships with Brio Technology, Business Objects, and Cognos. 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, for example, in the area of meta data interchange.

Online Analytical Processing (OLAP)

DB2 OLAP Server a key product

IBM's key product in the OLAP marketplace is the **DB2 OLAP Server**, which implements a three-tier client/server architecture for performing complex multidimensional data analysis. The middle tier of this architecture consists of an OLAP analytical server developed in conjunction with Hyperion Solutions, which is responsible for handling interactive analytical processing and automatically generating an optimal relational star schema based on the dimensional design the user specifies. This analytical server runs on Windows NT or UNIX and can be used to analyze data managed by a DB2 Universal Database server. Support for Oracle servers is planned for a future release. The DB2 OLAP Server supports the same client API and calculation engine as Hyperion Essbase, and any of the many third-party GUI- or Web-based tools that support Essbase can act as clients to the DB2 OLAP Server.

Essbase applications work with the DB2 OLAP Server

The value of the DB2 OLAP server lies in its ability to generate and manage relational tables that contain multidimensional data, in the available Essbase applications that support the product, and features within Visual Warehouse for automating the loading of the relational star schema with information from external data sources such as DB2, Oracle, Informix, IMS, and VSAM.

Information Mining

IBM has put significant research effort into its **Intelligent Miner for Data** product, which runs on OS/390, OS/400, UNIX and Windows NT, and can process data stored in DB2 databases, any relational database supported by DB2 Data Joiner, and flat files. Intelligent Miner Version 1, released in 1996, enabled users to mine structured data stored in relational databases and flat files, and offered a wide range of different mining algorithms. Intelligent Miner Version 2 features a new graphical interface, additional mining algorithms, DB2 Universal Database exploitation, and improved parallel processing.

New Intelligent Miner for Text product

Intelligent Miner is one of the few products 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). Intelligent Miner has good data visualization capabilities, and unlike many other mining tools, supports several information mining algorithms. IBM also offers its **Intelligent Miner for Text** product, which provides the ability to extract, index, and analyze information from text sources such as documents, Web pages, survey forms, etc.

Access Enablers

Client access to warehouse and operational data from business intelligence tools requires a client database API. IBM and third-party business intelligence tools support the native DB2 SQL API (provided by IBM's Client Application Enablers) and/or industry APIs like ODBC, X/Open CLI, and the Hyperion Essbase and ESRI APIs.

DB2 Data Joiner has heterogeneous data replication

Often, business information may be managed by more than one database server, and IBM's strategic product for providing access to this data is its **DB2 Data Joiner** middleware server, which allows one or more clients to transparently access data managed by multiple back-end database servers. This *federated* database server capability runs on Windows NT, OS/400, and UNIX, and can handle back-end servers

running IBM or non-IBM data products, for example, IBM DB2, Informix, Microsoft SQL Server, Oracle, Sybase, VSAM, IMS, plus any ODBC, IBI EDA/SQL or Cross Access supported data source. Features of this product 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.
- Stored procedure feature that allows a global DB2 Data Joiner procedure to transparently access data or invoke a local procedure on any DB2 Data Joiner-supported database. This feature includes support for Java and Java Database Connectivity (JDBC).
- Heterogeneous data replication (using IBM DataPropagator, which is now integrated with DB2 Data Joiner) between DB2, Informix, Oracle, Sybase and Microsoft relational database products.
- Support for Web-based clients (using IBM's Net.Data product).

**Net.Data provides
Web access to data**

IBM's **Net.Data** Web server middleware tool (which is included with DB2) supports Web access to relational and flat file data on a variety of platforms, including DB2, DB2 DataJoiner-enabled databases, and ODBC data sources. Net.Data tightly integrates with Web server interfaces, and supports client-side and server-side processing using applications written in Java, REXX, Perl, C++, or its own macro language.

Data Warehouse Modeling and Construction

IBM supports the design and construction of a data warehouse using its Visual Warehouse product family and data replication tools, and via third-party relationships with Evolutionary Technologies International (for its ETI•EXTRACT Tool Suite) and Vality Technology (for its Integrity Data Reengineering tool).

The **Visual Warehouse** product family is a set of integrated tools for building a data warehouse, and includes components for defining the relationships between the source data and warehouse information, transforming and cleansing – acquired source data, automating the warehouse load process, and managing warehouse maintenance. Built on a DB2 core platform, Visual Warehouse can acquire source data from DB2, Informix, Microsoft, Oracle, Sybase, IMS databases, VSAM and flat files, and DB2 Data Joiner-supported sources.

**Visual Warehouse
packaging**

Organizations have the choice of two Visual Warehouse packages, both of which are available with either Brio Technology, Business Objects or Cognos add-ins for information access. The base package, Visual Warehouse, includes:

- DB2 Universal Database for meta data storage.
- A Visual Warehouse Manager for defining, scheduling, and monitoring source data acquisition and warehouse loading operations.

- A Visual Warehouse agent for performing the data capture, transformation and load tasks.
- The Visual Warehouse Information Catalog (formerly known as DataGuide) for exchanging meta data between administrators and business users.

Visual Warehouse includes the DB2 OLAP Server

The second package, **Visual Warehouse OLAP**, adds the DB2 OLAP Server to the mix, allowing users to define and load a star schema relational database, as well as to perform automatic precalculation and aggregation of information as a part of the load process.

Provides agents for performance

Visual Warehouse provides several features that make the implementation and management of a data warehouse more efficient: its use of agent technology, its management capabilities, its handling of meta data, and its ability to invoke user-written and third-party tools to perform additional processing outside the scope of the product. The first of these, its use of agent technology, is intended to satisfy the performance requirements for loading large warehouse information stores. Data is acquired and loaded into an information store by warehouse agents whose job it is to move information directly from one or more data sources to one or more warehouse information stores. Unlike many competing products, information does not have to pass through a central intermediate server that might otherwise become a performance bottleneck as data volumes grow. Visual Warehouse agents run on OS/400, OS/2, UNIX, and Windows NT, and, depending on the volumes of data being moved, any given implementation may have one or many agents running concurrently. The source data to be captured, transformed and loaded into the warehouse information store by one or more agents is defined in a business view. The definition, scheduling and monitoring of business view operations is handled by the Visual Warehouse Manager, which runs under Windows NT.

Manages user-coded data transformation applications

In addition to initiating agent activities, the Visual Warehouse Manager can also be used to schedule user-written data capture and transformation applications, as well as applications available from IBM business partners. This facility is employed by Visual Warehouse to enable the loading of Hyperion Essbase multidimensional data, and to integrate other non-agent-driven processing such as ETI•EXTRACT programs, IBM data replication jobs, and Vality data cleansing processes.

Visual Warehouse also plays a key role in managing the meta data associated with the IBM business intelligence environment. In such an environment there are two types of meta data to 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 used to access and analyze warehouse data.

Supports meta data interchange with partner products

The Visual Warehouse Manager employs its own DB2-based meta data store for managing the technical meta data associated with the building and managing of a data warehouse. As mentioned earlier, IBM has developed interfaces to products from Hyperion Solutions, Evolutionary Technologies International, and Vality Technology for meta data interchange with Visual Warehouse. Meta data can also be exchanged with business intelligence tools from Brio Technology, Business Objects, and Cognos.

Included with Visual Warehouse is the Visual Warehouse Information Catalog (formerly known as DataGuide). The objective of this information catalog is to document and manage the business and underlying technical meta data that helps business users access and exploit the business intelligence environment. Business users can browse this meta data using both graphical- and Web-based interfaces.

**Meta Data Coalition
MDIS interface
provided**

Meta data in the Visual Warehouse 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. IBM supplies a variety of sample applications that use these interfaces to exchange meta data with third-party products (Hyperion Essbase, Bachman DBA, Microsoft Excel, for example). Visual Warehouse Manager's technical meta data can also be imported into the information catalog. With Visual Warehouse, IBM supports the Meta Data Coalition's Meta Data Interchange Specification (MDIS) for moving meta data into and out of the Visual Warehouse Information Catalog.

**IBM data replication
tools**

IBM's data replication capabilities are based on its **DataPropagator Relational** product, which has now been integrated into DB2 Universal Database (for homogeneous data replication), and DB2 Data Joiner (for heterogeneous data replication). The replication facility captures data changes from DB2 source databases, and applies those changes to a DB2-managed data warehouse. Data changes are transported from the source to the target warehouse via 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 Data Joiner 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.

Other IBM products for data warehouse construction include **DataPropagator NonRelational**, for capturing data changes from IMS databases, and **Data Refresher** for capturing and transforming data stored in non-relational databases and files such as IMS and VSAM.

**Evolutionary
Technologies is a
partner**

IBM partner Evolutionary Technologies International markets the **EXTRACT Tool Suite** for generating warehouse data capture and transformation applications. This consists of:

- A Data Conversion Tool for defining data cleanup and transformation rules and generating data acquisition programs.
- Pre-built Data System Libraries (DSLs) for key operating and database environments including SAP, IDMS, IMS, VSAM, and leading relational database products. A DSL defines the native access method to be used for processing data, the grammar for generating application programs, and the business rules available to the Data Conversion Tool.
- A Master ToolSet for extending, creating and maintaining DSLs.

IBM has been working with ETI to optimize the DB2 DSL (to support parallel loading, for example), and to integrate EXTRACT with Visual Warehouse in the areas of meta data interchange and EXTRACT program scheduling. One of the key benefits EXTRACT adds to Visual Warehouse is support for additional data sources and application packages such as SAP.

**Vality Technology
Integrity improves
data quality**

Vality's **Integrity** Data Reengineering tool complements both Visual Warehouse and ETI•EXTRACT by adding a capability to analyse the *content* of data extracted from operational systems and enhance the quality of data before it is loaded into a data warehouse. During the data reengineering process, unique data entities are identified in data from multiple systems, allowing the data to be merged, reconciled and consolidated, even when there is no common key to support the merge. Important meta data that is discovered in this process can be used to validate and adjust the data model for the data warehouse information store. As with ETI, IBM has worked with Vality to integrate Integrity with Visual Warehouse in the areas of meta data interchange and program scheduling.

Data Management

**DB2 Universal
Database supports
parallel processing**

Data management in the business intelligence environment is provided by **DB2 Universal Database**, which offers intelligent data partitioning and parallel query and utility processing on a range of IBM and non-IBM multiprocessor hardware platforms. DB2 Universal Database also supports both partition and pipeline parallelism, SQL CUBE and ROLLUP OLAP operations, integrated data replication, dynamic bit-mapped indexing, user-defined types, and user-defined functions.

**DB2 Spatial
Extender manages
geo-spatial data**

The **DB2 Spatial Extender** enables geo-spatial data to be incorporated into a relational DBMS. The product is a joint development effort between IBM and Environmental Systems Research Institute (ESRI), a leading GIS developer. IBM is initially delivering the DB2 Spatial Extender on DB2 DataJoiner, and plans to add this capability to the next release of DB2 Universal Database. GIS tools and applications can use either an ESRI or an SQL API to access and analyze geo-spatial data. Existing tools and applications that support the ESRI API will also work unmodified with the DB2 Spatial Extender.

CONCLUSION

In this paper we have presented the structure, development initiatives, and products of IBM's business intelligence software solution. The goals and strategy of this solution has five main thrusts:

1. The rapid and cost effective deployment of industry-specific business applications.
2. An integrated end-to-end solution involving products and services from IBM and its business partners.
3. Leading-edge information and decision support technologies.
4. Scalable hardware and software.
5. A multi-tiered and heterogeneous business information environment that supports both IBM and non-IBM database products.

With the advent of third-generation business information systems, IBM's business intelligence solution is ideally positioned to be one of the leaders in supplying a new generation of tools and applications for providing users with the information they need to manage their businesses.

About DataBase Associates International, Inc.

Database Associates International is a consulting and training company specializing in leading-edge technologies in the fields of database, distributed computing, data warehousing, and Web technology.



DataBase Associates International, Inc.

Post Office Box 310

Morgan Hill, CA 95038-0310

408-779-0436 (voice)

408-779-3274 (fax)

<http://www.dbaint.com>

dbaint@dbaint.com (e-mail)

The IBM Business Intelligence Software Solution

Version 3, March 1999

Copyright © 1999 by DataBase Associates International, Inc.

All rights reserved.