DB2® Information Management Software

The DB2 Framework for Business Intelligence

By Jon Rubin
IBM Software

Table of Contents

Highlights

Why a framework for BI?

The IT community is by now well acquainted with the value proposition of the data warehouse—indeed, the focus of most discussions has shifted beyond the warehouse itself to the larger issues surrounding an overall enterprise architecture for business intelligence and realtime analytics. The warehouse remains a vital foundation for any strategic approach to leveraging information assets: it establishes consistent, predictable levels of data quality, breadth and depth within a managed environment.

However, the scope of BI architecture has expanded beyond traditional core warehousing disciplines such as data extract, transform and load (ETL) and reporting. Today, business intelligence requires *Information On Demand*— actionable business insights *combining* data mining and multidimensional data analysis with advanced statistical and analytical functions in a realtime, integrated environment. OLAP and mining, in particular, have brought with them specialized tools, APIs, data structures—and sometimes purpose-built engines—spanning all levels of the system stack, from front-end client tool to middle-tier server cache to backend data warehouse. The multitier, enterprisewide nature of BI tools and infrastructure demands a high-level architectural model—a framework for BI.

A BI framework provides an overall conceptual model for understanding, planning and managing this complex topology. A vendor's BI framework should express its strategy and vision, ideally in a way that differentiates the vendor's approach from that of its competition. The proposed framework should tell customers what to expect in terms of future direction, technology and interfaces. It should guide system architects and integrators in selecting, designing and deploying BI tools and applications.
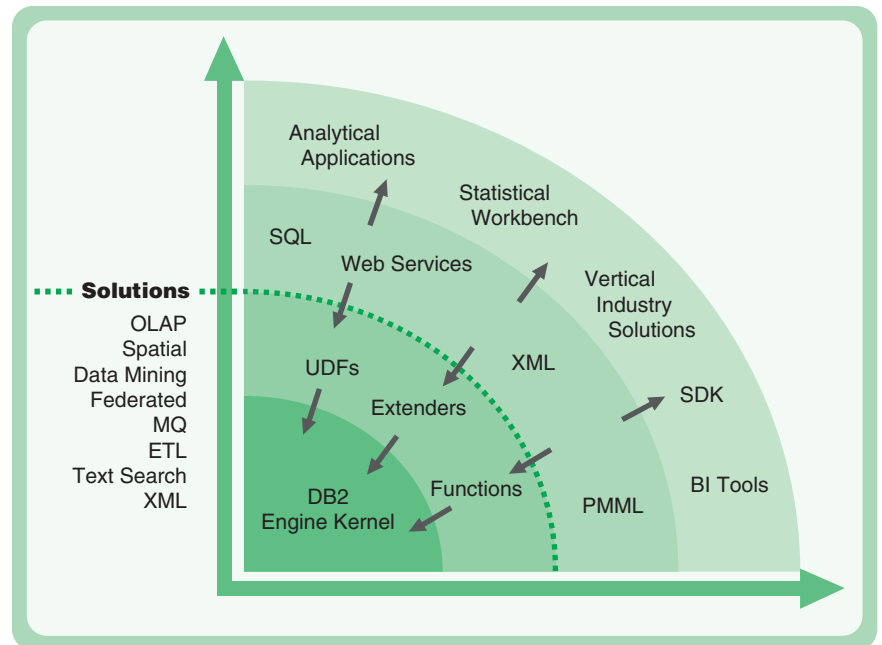
Highlights

A simple statement of direction for BI.

The DB2 Framework for BI

A cogent vision for BI is one that can be reduced to a single statement, bereft of technical jargon or buzzwords. IBM's vision for BI, as a cornerstone of Information On Demand, can be so expressed: "Build BI function into the database; make it accessible solely through open, standard interfaces, as part of an integrated BI platform; collaborate with partners for other layers of the architecture."

Let's take a closer look at the individual components of this vision statement. As we do, we'll use real-world scenarios to show how customers can apply IBM's vision for BI to achieve business value.

BI function has pervaded almost every level of the application stack — leaving customers wondering about data consistency, security and application performance.

Highlights

DB2 has been moving steadily down a path of supporting BI functions inside the database.

Build BI function into the database

Noting that IBM builds BI functionality into the database is not trivial, because today, BI function resides all over the system map. Many BI applications use simple SQL syntax to extract data from the warehouse in bulk, stage it within the application tier, and filter it (or aggregate it) to produce the desired granularity. Only then do they apply the functionality, which is their true value-add. Client tools often take the same approach, whether the temporary cache is a spreadsheet, cube or local file system. User applications often mimic this model, because it is the paradigm with which developers are most familiar.

IBM DB2® Universal Database™ has been moving steadily down a path of supporting BI functions *inside* the database. These BI capabilities include data mining; OLAP; ETL; spatial and advanced statistical and analytical functions for regression, covariance, sampling, ranking, moving windows and much more—hereafter referred to in the broad sense as "BI function."

OLAP tools are very helpful to analyze the business in a flexible way. For example, the effectiveness of a marketing campaign depends on many variables, and a predefined report might not show the right information. But where do you start a multidimensional analysis? To surf the cube and drill into the right cells, you first need to know how to define the attributes of the cube. For example, if geography is an influential dimension in measuring campaign effectiveness, should you base its hierarchy on cities, ZIP codes or Metropolitan Statistical Areas? Should these roll up to counties, states or regions?

DB2 mining functionality lets you discover and rank important dimensions and attributes automatically. If population density is found to have a significant correlation with the measures for marketing cost and sales, you can invoke DB2 analytical function to produce an equi-height histogram of population density ranges. Once these ranges are defined as an attribute of the cube, client tools can drill across population density to gain insight into the effectiveness of marketing programs.

IBM

BI functions based on the same data, in the same DB2 data warehouse, are more likely to express a "single version of the truth" across the enterprise—regardless of end-user tool or application

Information Integration extends the BI functions of DB2 across heterogeneous data sources, while preserving a consistent view of the enterprise warehouse to upstream tools and applications.

The benefits of pushing BI function into DB2 are significant and varied:
- It allows the database to deliver refined data to BI endpoints (applications, users and tools) at the desired level of granularity.
- It shifts more of the "heavy-lifting" work of scanning, sorting, joining and aggregating data to the layer of the architecture designed and optimized for precisely this role—the warehouse server.
- It reduces the amount of data flowing over the network.
- It exposes less data to less secure areas outside the firewall.
- It lowers the risk of data inconsistency wrought by separate engines applying separate algorithms for common business measures. BI functions based on the same data, in the same DB2 data warehouse, are more likely to express a "single version of the truth" across the enterprise—regardless of end-user tool or application.

Here, it is important to note that due to the complex, heterogeneous nature of enterprise information systems, a "single version of the truth" is often best achieved as a conceptual architecture rather than a physical one. Information On Demand needs to encompass selective, transparent access to distributed data sources as required, or appropriate, for a given data topology and application. *Information Integration* extends the BI functions of DB2 across heterogeneous data sources, while preserving a consistent view of the enterprise warehouse to upstream tools and applications. In this way, the extended data warehouse becomes a *virtualized* realization of a "single version of the truth."

Make BI part of an integrated platform, made accessible solely through open, standard interfaces
Given the readily apparent benefits of leveraging the warehouse platform for BI function, why do so many tools and applications opt to do more of the work themselves? Posed another way, why do BI endpoints act as data wholesalers rather than consumers of refined analytical results?

IBM

Highlights

*The pursuit of analytical advantage
has compromised the underlying
value proposition of data warehousing.*

Traditional development practices hinder BI development

Part of the answer stems from a developer culture based on a time when even simple "top n" rankings could not be expressed in SQL. Today, functions such as plotting a regression slope for customer churn or a smoothed moving average of stock prices may not be widely supported by mainstream database management systems (DBMS) (although DB2 does support such functions). Multidimensional analytics have historically favored a specialized structure, the multidimensional online analytical processing (MOLAP) cube, with a proprietary engine and interface. Data mining algorithms have traditionally operated on files outside the database. In all these examples, data is staged from the warehouse to some intermediary platform and re-structured before the BI function is applied.

> *Finding the critical business anomalies hidden as cells in a large cube can be tedious, but mining algorithms for the detection of deviations can highlight unusual figures automatically. In some cases, such anomalies might indicate a new business opportunity; in other cases, they might point to a problem that needs to be fixed. DB2 mining functions in the database are scalable to the entire warehouse population. You won't miss a critical customer record because it was added after the "mining extract" was cut. That's almost impossible to achieve with external mining tools relying on autonomous files.*

Enterprises commonly perform advanced statistical analysis and mining in a specialized workbench with its own local files extracted from the warehouse. Then they use the results to populate cubes in a MOLAP server and query the warehouse on an exception basis for detail drilling into the original source data. Consequently, they obtain at least three different inter-dependent data structures reflecting three distinct points in time spread across three server platforms, each with its own APIs, administrative controls and optimization techniques. While each analytical silo performs its own role admirably, *the pursuit of analytical advantage has compromised the underlying value proposition of data warehousing.*

Highlights

DB2 consolidates BI function to preserve consistency and improve performance.

DB2 solves this by basing all BI function, including mining and OLAP, on the *same* data structures in the *same* database on the *same* warehouse tier. Even in a federated DB2 warehouse, all BI functions "see" a single logical schema, which is abstracted from the details of distributed data topology.

Data miners often use specialized tools for data analysis and visualization. These statistical workbenches typically lack integration with other BI tools and data in the enterprise. However, with DB2 you can feed mining results into any other reporting or OLAP tool. For example, your mining expert can use a favorite workbench to produce an associative model for customer segmentation and a predictive model for risk of attrition. You can export those models in the standard PMML format and store them in the database, where the integrated mining functions in DB2 can invoke the models from a SQL procedure. Next, you can update the entire customer population in the database using the segmentation clusters as an attribute and the predicted risk as a measure.

A line-of-business manager surfing the current sales cube containing these measures is able to pinpoint particular customer segments where high sales are offset by high risk of attrition. A preemptive retention campaign can target those segments, reaching only those customers with the optimal combination of profitability and risk—even some whose risk score alone does not yet reflect a higher potential for attrition (based on cluster association).

DB2 exposes BI functions with the most ubiquitous standard interface available—SQL. Extensions are submitted to the standards body, where they may ultimately be incorporated in formal revisions to the SQL standard. IBM helped champion the CUBE, ROLLUP and GROUPING SET extensions—now standard SQL syntax—allowing relational queries to return multidimensional results as in viewing a cube.

The SQL extensions for BI in DB2 are often complemented with XML for extended description and *shaping* of non-tabular data, such as mining models or cube views, and for providing meta data *bridges* to external tools. The preceding

scenario involving mining, OLAP analysis and detailed warehouse drill-down
can be accomplished entirely within DB2 (or through DB2 Information
Integrator™ to any federated data source) using SQL and XML exclusively.

Obscuring value-add to pursue portability

There is another reason BI endpoints act as data wholesalers: application or
tool vendors may make an explicit choice to use lowest common denominator
SQL syntax for portability across relational DBMS (RDBMS). This leaves them
performing a lot of heavy lifting and intermediate staging of raw data (clearly not
part of their value proposition) just to reach the point where they can apply their
unique value-add. By leveraging common warehouse data with open interfaces
exposing OLAP, mining and statistical function, DB2 encourages BI partners to
exploit the backend for greater data reduction. This improves performance while
allowing application and tool vendors to focus on delivering customer benefit at
the front end of the value chain.
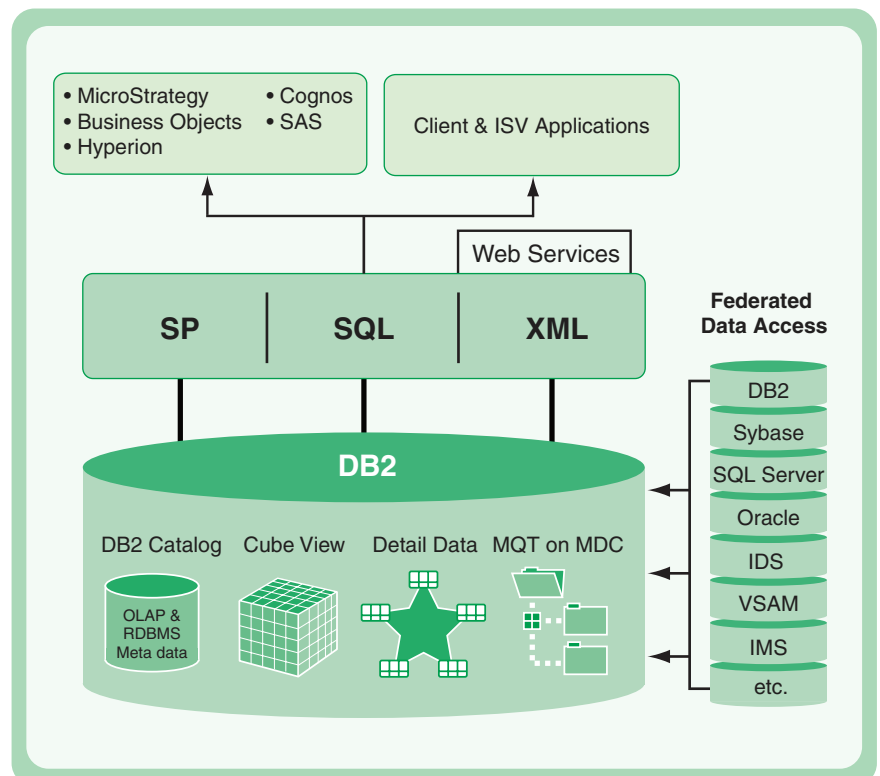
*Supporting a wide range of data types, functions and access options, DB2 makes it easy for BI vendors
to deliver powerful analytical solutions.*

Highlights

Partner enthusiasm for DB2 can pay customer dividends in the form of improved support, availability and features

Collaborate with BI partners for other layers of the architecture

Major competitors to DB2 in the enterprise data warehouse space place a strong emphasis on their own tools and applications, sometimes overshadowing the database itself. Often, this approach brings them into direct conflict with partners. The DB2 Framework for BI differs fundamentally in this respect: its entire focus is on the infrastructure layer—the BI function built into DB2 itself—not on front-end user tools or applications. This benefits customers in two ways:

- It preserves and reinforces their existing investment in industry-leading BI tools, skills and practices.
- More importantly, it provides an incentive for BI tool and application vendors to view DB2 as their BI enabler, since IBM is not competing for their customers or segment. This is the dynamic that has seen the leading enterprise application vendors (such as SAP, PeopleSoft and Siebel) embrace DB2 with just such a preferred relationship. In the end, partner enthusiasm for DB2 can pay customer dividends in the form of improved support, availability and features.

Based on the IBM project code-named Aurora, this latest step in the evolution of DB2 as the OLAP engine enhances the interoperability between DB2 and leading BI tools.

A unified engine for OLAP

Since it is a pronounced area of differentiation, this aspect of the DB2 Framework for BI merits a bit more explanation. Framework architects faced a unique circumstance with regard to the key OLAP capability. Unlike rivals, IBM had never acquired a proprietary MOLAP company and technology. Its historical presence in the MOLAP space, called DB2 OLAP Server™, is an OEM of the Essbase product from IBM Business Partner Hyperion. DB2 OLAP Server integrates Hyperion Essbase with DB2 Universal Database. Without a strategic commitment to an external MOLAP technology, IBM has been free to develop and extend the DB2 engine to better support OLAP inside the database. Over a period of several versions, IBM has steadily beefed-up the native OLAP capability in DB2 with enhancements such as Automatic Summary. IBM later added Materialized Query Tables, CUBE and ROLLUP syntax, Star Join optimization and Multidimensional Clustering, culminating in the new IBM DB2 Cube Views, V8.1 technology.

Highlights

Based on the IBM project code-named Aurora, this latest (but not final) step in the evolution of DB2 as an OLAP engine provides XML-augmented shaping of the result sets obtained in querying cube views, along with import/export of XML-described multidimensional models, for enhanced interoperability between DB2 and leading BI tools. Aurora also introduces a multidimensional model-based tool to automatically build the optimal structures for boosting the performance of OLAP queries.

Thus, when considering OLAP within the DB2 Framework for BI, IBM faced a "clean-sheet" decision more than half a decade *after* its rivals had made their major MOLAP acquisitions and committed to the dual-architecture path (separate relational and MOLAP engines, data structures and APIs). For many analytical applications today, the enhanced ability of DB2 to describe and optimize a *virtualized* cube alters the rules of the OLAP game. While cube caching and serving outside the warehouse can still offer value, it is primarily a partner prerogative. The warehouse itself is the clear platform of choice for the heavy lifting of materializing cube data, whether for populating intermediate MOLAP structures or for direct viewing.

DB2 customers achieve the following benefits, whether they perform query and reporting, data mining, OLAP or advanced analytics:
- Simple standard APIs exposing common data in the same warehouse
- Faster end-to-end response by doing more data reduction on the backend
- No need for specialized servers and data structures to administer and tune
- No specialized languages to master
- A single virtualized version of the truth, regardless of data topology.

IBM

How do you "close the loop" with realtime analytics? Suppose that while interacting with the customer at the point-of-sale, your core operational system sends a message to the DB2 warehouse. A pre-defined table function makes the message look just like any SQL request. This one invokes a stored procedure using a predictive risk scoring model to compute a new up-to-the-instant propensity for attrition, based on juxtaposing the current transaction with the customer's historical risk profile stored in the database. A risk score above your specified threshold triggers an outbound message to your automated call center application, which leaves a personalized message on the customer's voice-mail.

Highlights

The DB2 Framework for BI incorporates the four basic tenets of IBM's vision for e-business on demand:
- *Open standards*
- *Integration*
- *Virtualization*
- *Autonomic computing.*

The DB2 Framework for BI and Information On Demand

Even as the DB2 Framework for BI enables Information On Demand within a business intelligence context, it must also fit into a larger picture—IBM's overarching vision for e-business on demand. The following threads of e-business on demand, aligned to quickly and simply provision, deploy, integrate and administer the business model, are woven throughout the fabric of the DB2 Framework for BI:
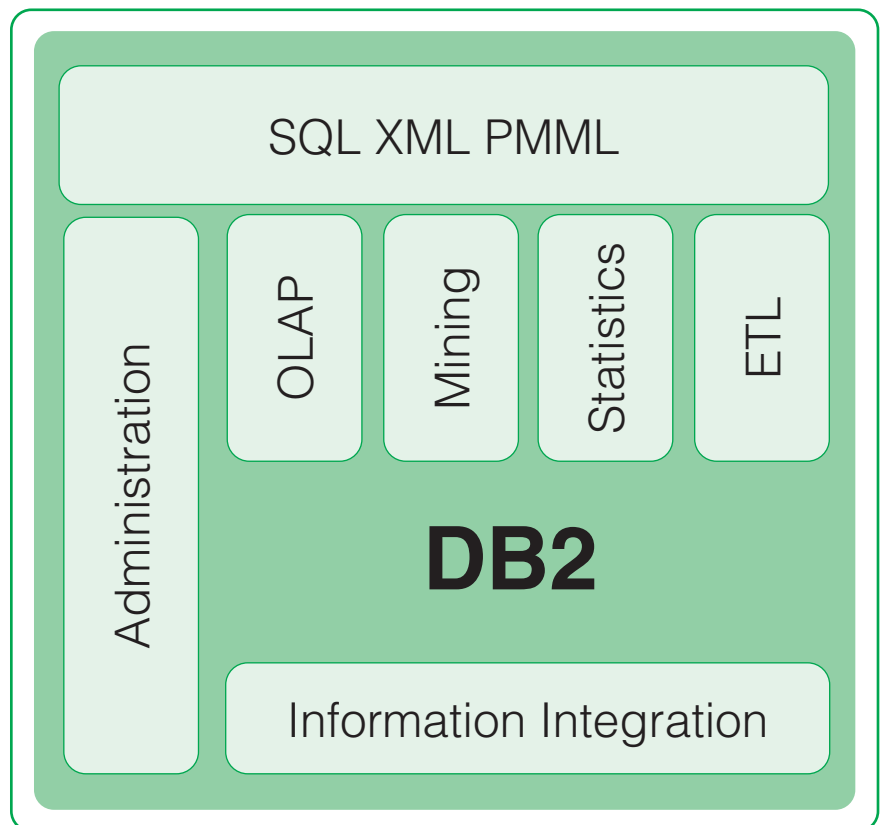
Open standards feature prominently in the role of SQL and XML as the sole interface to BI function in DB2.

Integration appears in multiple contexts: federated access to distributed data sources; Web services invocation of BI function; and integration of operational systems with the realtime warehouse through DB2 transactional features such as triggers, procedures, replication and messaging-based table functions.

Virtualization underlies the cube views of relational OLAP, and the abstraction of advanced analytics as SQL-invoked DB2 functions. Federation represents a virtualization of a "single version of the truth" across the enterprise data topology.

Autonomic computing is built into DB2 itself, in the form of automatic detection and handling of events such as error conditions; in automated setting of configuration parameters based on user-provided guidelines; and in design tools that recommend optimal data structures (such as indexes and summary tables) based on workloads and data models.

How do Web services extend realtime analytics for the Information on Demand e-business? Take the dynamic risk-scoring scenario discussed on the previous page. Let's say the acquisition of a new business unit requires the company to integrate the same model-based scoring capability into the new unit's legacy system. Rather than undertake a lengthy and complex development effort, the existing risk-scoring function in DB2 is exposed as a Web service by wrapping the SQL procedure in appropriate Internet protocol. The legacy system invokes risk-scoring over the Internet as an ordinary URL.

SQL XML PMML

Administration

OLAP

Mining

Statistics

ETL

DB2

Information Integration

DB2 Framework for BI.

IBM

Highlights

The complementary introduction of DB2 Data Warehouse Edition (DWE) and DB2 Information Integrator together establish the foundation for IBM's vision of Information On Demand.

Introducing DB2 Data Warehouse Edition

DB2 offerings aligned with the DB2 Framework for BI began rolling out in the first half of 2003. The complementary introduction of DB2 Data Warehouse Edition (DWE) and DB2 Information Integrator together establish the foundation for IBM's vision of Information On Demand. (Please refer to the DB2 Data Warehouse Edition and DB2 Information Integrator announcements and related collateral for details.)

DB2 Data Warehouse Edition combines a carefully selected set of IBM business intelligence products to provide the essential infrastructure needed to extend the enterprise data warehouse as a comprehensive platform for BI. Based on DB2, DB2 Data Warehouse Edition leverages the inherent strengths of the database for scaling, platform portability, availability and manageability. It also brings all the benefits of IBM's services, support and customer partnership capabilities.

With DWE, IBM has an offering that can open BI alliance partners to view DB2 as their BI enabler. By combining DWE with its portfolio of partner tools, applications and services, as well as its own software portfolio (including Rational®, Tivoli®, Lotus® and WebSphere®), IBM transforms DB2 from warehouse-as-repository into a full-function BI platform for customer and partner solution development.

Going forward, DB2 Data Warehouse Edition provides a point of focus and convergence for IBM and partner offerings in the BI space. An upcoming white paper will describe future direction in terms of a roadmap for the DB2 Framework for BI.

Highlights

The DB2 Framework for BI stakes out a clear and differentiated position in the business intelligence marketplace.

Summary

BI tools, applications and infrastructure span system tiers and architectural boundaries. The complexity and criticality of enterprise BI demands an overarching framework to augment traditional data warehousing architectures for high-level design and planning. IBM's key rivals in the enterprise database space also promote frameworks for BI, and they even speak of a role for partners. Yet, each of the competitors promotes its own OLAP tools, API and an end-to-end vertical solution stack. Under the covers, each of these BI vendors is locked into a dual-engine architecture predicated on legacy MOLAP acquisitions.

The DB2 Framework for BI stakes out a clear and differentiated position, based on a simple statement of direction: "Build BI function into the database; make it accessible solely through open, standard interfaces, as part of an integrated BI platform; collaborate with partners for other layers of the architecture."

This paper has shown how DB2 enables mining, OLAP and advanced analytics to be performed by a single engine operating on common data structures within a consolidated data warehouse residing on the server tier. Examples described the invocation of BI functions through SQL extensions, sometimes augmented with XML, and optionally expressed as Web services. A rationale was put forth for a no-compete partner strategy that gives BI tool and solution vendors real incentive to make DB2 their preferred platform. Finally, the paper looked at how the DB2 Framework for BI, together with DB2 Information Integrator and DB2 Data Warehouse Edition, share the underlying principles and form the foundation for Information On Demand.

For more information

Please contact your IBM marketing
representative or an IBM Business
Partner, or call 1-800 IBM CALL within
the U.S. Also, visit our Web site at:
ibm.com/software/data/integration

IBM

e business software

GC18-7763-00