AberdeenGroup

Profile

IBM Corporation

www.ibm.com
Route 100
Somers, NY 14454
(716) 243-2445

IBM DB2 Information Integrator: Scope, Power, Services

Preface

Aberdeen research shows that EII (enterprise information integration) solutions can deliver not only cost savings but also a new kind of competitive advantage. As it becomes more difficult for e-enterprises to achieve differentiation via packaged or custom applications, the new frontier of competitive advantage becomes IT's ability to leverage proprietary information that its competitors cannot access. Much of this information is scattered across multiple databases, spreadsheets, and rich-media Web sites, making it difficult if not impossible to detect the connection between a customer's existing product preferences and new products arriving via the supply chain and R&D. EII's ability to amass metadata and carry out queries across data sources allows the EII user to detect and display these key "relationships across data frontiers."

EII can also be a powerful enabler for "on-demand" computing. An enterprise's ability to respond quickly to changes depends strongly on its adeptness at correlating information not only across an enterprise's systems, but also systems of suppliers in its value chain. EII facilitates accessing and integrating this diverse and distributed business information in real time, across and beyond the enterprise.

EII's cost savings are nothing to sneeze at, either. Many of today's tasks in BI (business intelligence), CRM (customer relationship management), portals, and consolidation involve integrating business processes and information. EII simplifies the developer's job in these cases by giving one interface and one set of metadata for all data sources, speeding development in applications like enterprise portals. EII simplifies the administrator's job by exposing the metadata referencing key data that must be kept in synchronization across data sources.

In order to use EII effectively, IT buyers need suppliers with comprehensive solutions and strong services that can help the customer decide how best to use EII in

the short and long term, saving costs. This *Profile* describes the IBM DB2 Information Integrator EII product, an exceptionally strong EII solution from a supplier unmatched in computer services experience and resources.

Executive Summary

IBM's EII products include DB2 Information Integrator 8.1 (offering an SQL programming model focused primarily on relatively structured data, but supporting XML/content data sources) and DB2 Information Integrator for Content 8.2 (offering the IBM Content Manager programming model focused primarily on unstructured data, but supporting relational data sources). This *Profile* focuses on the DB2 Information Integrator 8.1 product. Although it is based on DB2 technology, enterprises need not have DB2 Universal Database installed to take advantage of its features. It includes a "federated data server" that provides an integrated view across data sources, supporting querying (and update of relational data from a single source) plus a "replication server" that can move data among major popular relational databases. DB2 Information Integrator allows access to all major relational databases; Excel; XML; message queues; Web services; flat files; life sciences data sources; content repositories; and (via IBM Lotus Extended Search) Lotus Notes, Lotus and Microsoft e-mail applications; WebSphere Portal indexes; LDAP directories; Web search engines (Yahoo, Google, Lycos, CNN, Business Wire, etc.); and similar unstructured or semi-structured data sources. Access to additional mainframe sources — Including VSAM, IMS, ADABAS, CA-IDMS, and CA-Datacom/DB - is available through partner add-ons.

IBM's EII products can be used to extend a data warehouse by allowing the analysis of historical/warehouse and real-time/OLTP data as well as structured and unstructured data. This, in turn, allows users to make decisions more quickly based on fresher, more comprehensive information. IBM's EII products also have begun to demonstrate application-development speed-up in real-world applications (such as Crystal Decisions) that involve access to multiple data sources. By allowing the use of one high-level interface, IBM's EII products reduce both the code and the training required for data-accessing tasks.

IBM differentiates its EII offering especially by its open standards support, its scalability and performance, its range of data sources supported, and its "deployability" due to IBM's services strength. DB2 Information Integrator is easily combined with WebSphere Studio, WebSphere Portal, WebSphere MQ, and WebSphere Business Integration as part of an overall IBM "business integration infrastructure" solution set that also supports application and business process integration.

Bluntly put, Aberdeen believes both in the enormous potential of EII and in the utility of DB2 Information Integrator and DB2 Information Integrator for Content, effectively implemented, in a wide range of enterprises and situations. The time to

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

begin using EII is now, and IBM's solid EII solution should be on every major enterprise's short list.

A Brief Introduction to EII

EII is software that provides a "database veneer" to make disparate data sources appear as one database, allowing users to apply data management queries to the pooled data, at a corporate or enterprise level, in order to support applications that present or analyze the data in new ways. EII has the following three key technical characteristics:

- It is *software infrastructure*. It is not an application; it supports the creation of applications such as enterprise portals that display and analyze combined data. As software infrastructure, EII forms a key part of any overall e-business architecture.
- It *links the second and third tiers* of a typical enterprise architecture in a "many-to-one-to-many" approach. In other words, EII gives a wide variety of second-tier Web applications running on the second tier of an e-Business architecture access to multiple back-end and legacy databases and file systems running on the third tier.
- It provides an *enterprise-scale, integrated* approach. Unlike today's point-to-point application-to-database links, EII provides a common infrastructure on which all links are built. This common infrastructure can deliver more rapid development, more cost-effective centralized administration of data from multiple data sources, and more flexible presentation and analysis of more data sources.

All EII products usually offer certain basic components:

- A *directory*. This "database about databases" contains information (called metadata) about the data stored in the databases it accesses — often, the same information that is stored in each database's data dictionary, as well as information about the data that is shared across databases. A directory may also include information about "normalized" versions of the data — i.e., data converted into a common format that is particularly suitable for passing on to users, Web sites, or enterprise portals.
- *Conversion* capabilities. An EII solution must often convert incoming data from non-relational to relational format, or from relational to object or XML format.
- A *database veneer*. The EII solution should look like a database to the user, the developer and, if possible, the administrator. Therefore, the EII solution should at least support standard SQL and query operations, such as joins.

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

- *Front- and back-end communications*. The EII solution should have a way of communicating with front-end programs and end-users via programmatic and communications invocations, and of sending queries to and receiving results from back-end databases.

The IT buyer should carefully distinguish an EII solution from an ETL (extract, transform, load) engine for data warehousing or an EAI (enterprise application integration) solution. An ETL engine (sometimes called a "data integration" solution) allows the conversion of multiple data streams into a common, data warehouse data format that is loaded into the warehouse "in bulk" or in short "bursts," thereby centralizing the data for consolidated access. However, it does not deliver that data to external targets, such as an enterprise portal; it does not make the source data appear as one database; and it does not typically emphasize real-time access to the data.

An EAI solution passes event data between multiple applications, matching the programming interface and model of one environment to those of multiple other environments. It often operates at the level of business processes rather than data. Tools associated with EAI focus on modeling, monitoring, and managing the business process.

EII, EAI, and ETL are distinct, complementary technologies. All of them belong in an enterprise's strategic architecture.

The Business Benefits of EII

Aberdeen clients find that the data connectivity projects for which EII is a new tool are not small tasks. The project scope is typically greater than anticipated and the data connectivity problems typically involve heterogeneous environments. IT usually must perform these integration projects in a complex environment and must rely on a rapidly changing technology infrastructure. Other frequently encountered issues include the following:

- Increased industry and government regulation
- Ever-increasing need for speed to market
- A hodgepodge of legacy systems
- Multiple legacy databases from multiple suppliers
- Disparate operating platform environments — MVS, Unix, Windows NT and XP, wide area network (WAN), local area network (LAN), and the Internet
- Heterogeneous and geographically dispersed database management and file systems
- Many communications protocols — SNA, TCP/IP, HTTP, and XML/SOAP

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

Finally, when the project is done, there is no guarantee that the data-level integration will have integrated the applications' business processes as a seamless whole. Today's data connectivity solutions deal with queries on a per-application, per-database basis, and do not make it easy to handle business processes that involve different databases or different applications.

By contrast, EII does the following:

- It supports queries across databases and applications, making implementation of business processes much easier
- It makes the reuse of integration techniques across heterogeneous environments much easier
- It gives users the ability to implement integration solutions without needing detailed knowledge of the application infrastructure's specific, underlying technologies

EII delivers other immediate benefits to the enterprise. The ability to cross database boundaries, used by applications, gives wider end-user access to key existing data for data mining, enterprise portals, or Web searches from inside or outside the enterprise. For example, insurance companies can provide auto claims images for customer service together with the client's account data. Establishing relationships between data sets across frontiers and types of data source empowers deeper data analysis, leverages proprietary information more effectively for competitive advantage, and provides a more comprehensive view of the enterprise's information assets. Development is speeded when programmers need only code for access to one data source rather than several disparate ones. Administrators often cut costs by managing more easily across data-source frontiers.

EII may also be useful in many typical enterprise situations (Table 1).

Table 1: Ways to Use EII in Strategic Initiatives

Aim	Ways in Which EII Can Help
Integrate data for complex business activity monitoring	Write code in the applications to access an EII rather than the databases, to simplify process development and maintenance. For example, an enterprise could use EII to gather all relevant information to respond to an alert: "A big order came in from our best customer but inventory on hand is committed. I need to gather information from supply chain, distribution points, and my customer database in order to evaluate order commitments"
Implement Web services interfaces to existing information	Use the EII tool's Web service provider interface rather than a Web service interface to each data source. EII can leverage service-oriented architectures that put Web services over application interfaces, thus providing flexible integration across disparate application sources.

Speed time to react to current supply chain data	Provide an EII interface to supply chain and internal data for query purposes, so that the enterprise can cross company boundaries in a single query.
Speed time to develop new applications, such as portals, leveraging proprietary information	Use an EII to integrate the portal with both the data warehouse and other proprietary information, ensuring one set of code for all information access Use an EII's XML support to avoid coding XML-to-relational translations to access unstructured data Let EII engine deal with the differences in connection and access logic, reducing skill requirements. Leverage the EII engine's join and transformation logic to reduce error-prone hand coding.
Audit all information in the enterprise for government requirements	Create an EII metadata repository by importing (or generating) metadata from the data warehouse, ETL tools, and file systems and rich-media sites, then query the EII interface
Cut IT costs	Cut administrative costs by consolidating data access through a hub architecture, reducing maintenance complexity.
Merge companies	Create integrated reports about customers, suppliers, and financial position, e.g., over multiple data marts or data warehouses.
Increase the amount of information available to end users	Create an EII solution across multiple databases that expands the range of information available to an enterprise portal via the same interface Enhance the contextual value of enterprise reports by using an EII to add unstructured data — e.g. add marketing images/messages to executive reporting systems Increase the tactical-decision-making value of existing warehouse systems by using an EII to span historical and real-time data and deliver the results to enterprise reporting systems Use EII to facilitate enterprise-wide reporting where reports must be generated from disparate data warehouses or data marts Use EII to combine customer data spread across divisions and separate repositories to deliver a holistic view to customers

Source: Aberdeen Group, May 2003

EII As a Keystone of Strategic Information Management

EII can also play a key role as a foundational "information aggregation" tool for *strategic information management*. Currently, the obvious place to search for a cost-effective, long-term advantage is in proprietary information that cannot be readily duplicated by competitors. However, this information is usually locked in widely differing and often aged data sources — for example, much Wall Street investment research depends on data ranging from Excel spreadsheets to reports,

up-to-the-minute news articles, Web pages, and historical stock databases. Simply writing new code for each source without combining the information, or buying a packaged business intelligence (BI) application that accesses one type of data source, will not do the trick. The IT strategist also needs to focus on combing the information and making it more widely available.

The long-term job of strategic information management is to identify, manage, leverage, and add to the information resources of the enterprise. A collection of tools and technologies has arrived over the last seven years that allow users to combine data across data-source frontiers for strategic information management. This necessarily includes:

- Tools to move or copy data from multiple "source" data sources to a "target" aggregated-data repository, which can carry out operations such as queries across differing data types
- Tools for accessing data from multiple data sources and combining it temporarily for specific purposes, such as cross-database queries

Presently, representative tools include data migration and bulk-download tools, replication and ETL tools, data warehouses and data marts, enterprise and mid-tier operational data stores, federated databases, and EII tools.

When IT wishes to manage information strategically, EII can play a key role in leveraging existing information. EII's ability to establish relationships between data sets in widely different data sources helps the IT strategist identify key data and key "data connections" by placing that "information about information" in one place. Moreover, typical "side-effects" of EII use are the gathering of cross-data-source metadata and the ability to examine relationships between various databases by monitoring aggregated metadata. Thus, the information management strategist can also identify and manage information across frontiers more effectively.

The culmination of these benefits is when IT decides to implement full-fledged strategic information management. By applying EII and the other tools noted above appropriately and strategically, the IT strategist can potentially attain an "almost enterprise-wide" metadata repository, an "almost-enterprise-wide" data-source administration tool, and an "almost enterprise-wide" "information-leveraging tool" to constantly search the enterprise's ever-growing information store for new ways to combine the information for competitive advantage.

Criteria for Buying an EII Solution

EII solutions differ markedly in their ability to support developers, end-users, and administrators. Aberdeen recommends that IT buyers use different criteria according to the "main constituency" for an e-business architecture. Key criteria for developers are as follows:

Telephone: 617 723 7890
260 Franklin Street
Fax: 617 723 7897
Boston, Massachusetts 02110-3112
www.aberdeen.com

- *Performance/Scalability* of the applications built over the EII solution
- *Programmer productivity and simplified maintenance/upgrade*, by componentizing key EII operations at the business level, allowing the developer to swiftly write EII-using applications such as enterprise portals. Developers should not need to know and access multiple different types of data sources, thus simplifying the coding of connection and access logic; should be able to replace hand-coded operations, such as joins or transformations with ones provided by the EII engine; should be able to use a familiar development environment with EII operating transparently under it; and should be able to code to an abstracted view of the data source tailored to the application
- *Flexibility* to support a wide variety of platforms and architectures

Key criteria for administrators are as follows:

- *Ability to easily identify data sources*, configure their access, and build integrated views
- *Ability to monitor that data mappings remain valid*
- *Ability to monitor and tune queries* within the EII system in order to detect bottlenecks, preferably via a remote console
- *Comprehensive cross-database administration* features that allow a database administrator to monitor and maintain multiple databases via EII as if they were one database, as well as coordinate with the database administration tools of back-end and legacy databases — no supplier presently offers these

Key criteria for end-users are as follows:

- *Comprehensive support* for a wide variety of back-end databases and other data sources
- *Easy "plugability"* into front-end e-commerce, B-to-B, and especially enterprise portal implementations, as well as popular Web and application servers and BI tools
- Powerful *cross-database query features,* such as sophisticated querying for data mining

IT buyers should also consider the following overall criteria:

- The supplier's ability to serve as a "one-stop store front" for system design and integration.
- The ability of the EII solution to be a "good foundation" for a Web services architecture. Over the next two years, Web services implementations will become much more common. EII must provide the standards

Telephone: 617 723 7890
260 Franklin Street *Fax: 617 723 7897*
Boston, Massachusetts 02110-3112 *www.aberdeen.com*

support to integrate well with other upcoming Web services architecture components, such as a UDDI-based directory, plus the ability to access a wide variety of intra- and extra-enterprise data sources that these Web services promise.

Technology Overview

IBM DB2 Information Integrator serves three masters: the developer, the administrator, and the end-user. In order to serve the developer, DB2 Information Integrator provides a bridge between the components and business logic residing in the second tier of the e-business architecture and back-end data sources — that is, it allows the developer to treat multiple, disparate databases, data management systems, and files as one gigantic enterprisewide database. It fits transparently beneath common development environments, including WebSphere Studio and Microsoft Visual Studio .Net.

To aid the administrator, DB2 Information Integrator provides a directory of data and object types across multiple data sources, based on DB2 Universal Database, plus specific tools that allow the administrator to maintain, and implement security for the "federated database server."

To aid the end-user, DB2 Information Integrator supports SQL transactions, including add/remove/update (for single-source relational databases) and querying (for decision support across all data sources) with access to a wide range of structured an unstructured data sources. Users can continue to use the standard SQL-based tools from suppliers such as Crystal Decisions and Business Objects. IBM positions the tool primarily for read access. These allow support for Web-analytics queries across databases, and abet enterprise portal display and combination of information from various data sources for employees, customers, and partners.

Scalability

IBM's particular performance strength is its long experience in query optimization, first captured in DB2 and now in the cross-data-source querying of its EII products. Querying across disparate data sources is not easy to load-balance, and care in query optimization can make dramatic order-of-magnitude differences in the time such queries take. IBM's cost-based query optimizer is smart enough to optimize at the database copy level where appropriate and at the EII level otherwise. The SQL queries generated are optimized for a particular database.

IBM's EII engine also supports updates for a single relational database copy, with two-phase commit for extending to multiple sets of related data coming in a future release. Developer-enabled heterogeneous caching allows administrators to store frequently accessed or performance-critical information in Materialized Query Tables locally for faster access.

Flexibility

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

As noted, IBM supports an exceptionally broad range of data sources. IBM converts these to a common format via wrappers, with a simple interface allowing extensible "wrapping" of additional types of data source, as needed. Particularly notable is the Web data accessible via IBM Lotus Extended Search. Extended Search includes brokering for efficient, parallel searching, aggregation and ranking of result sets, development support that allows relatively easy incorporation into applications, and XML document generation for transmission of results to a wide range of end-users and applications.

IBM's query processing engine adapts well to variations in data sources, with the ability to compensate for the lack of a particular function, such as a SQL query engine at the database-copy level, and with the ability to invoke functions unique to the database-copy level when needed. Built-in functions (based on the User Defined Function architecture) allow users to extend SQL statements for data sources such as WebSphere MQ, Web services, and OLE DB-accessed data sources.

IBM's EII products support composition, validation, and storage of highly flexible XML documents. An EII access point can be a full-fledged Web service provider, with WebSphere Studio tools for converting SQL into WSDL. These Web-service access points can also be accessed in SQL statements.

Robustness and Cost of Ownership

IBM provides EII administrative tools, storage of environment characteristics for query optimization, and wrapper administration. A Control Center allows users to configure and administer standard wrappers, as well as plug-in administration components for custom-built wrappers. Replication administration includes replication controls and monitoring.

A notable feature of IBM's EII administration is the ability to dynamically discover and configure new data sources.

Programmer Productivity

Aside from simplifying the interface to disparate data sources, IBM provides extensive development support. Support for development tools such as Microsoft Visual Studio .Net, IBM WebSphere Studio, and BI tools such as those supporting Crystal Decisions- and Business Objects-based analysis solutions ensures easy integration of EII development with the rest of a user's programming efforts. IBM's development tools allows programmers to develop applications using common programming models, such as JSPs, EJBs, Web services, and ADO.Net. Stored procedures or user-defined functions allow developers to place data-access logic in the EII database for higher performance or more rapid coding. A variety of wizards and assistants such as SQL Assistant or view wizards aid in productivity and reduce the code developers need to write. In addition, IBM provides a wrapper SDK for creating custom wrappers.

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

Replication Server
DB2 Information Integrator's replication server complements its EII capabilities by allowing data movement between relational databases, with flexibility to load during batch windows or online with a user-chosen amount of overhead. Replication may also be used to maintain Materialized Query Tables that cache data from relational sources.

Consulting Services
Effective use of supplier services, especially "best practices" advice on what immediate and long-term applications to target, is key to users' EII success. IBM's consulting services aid customers who want to adopt best practices in every area of data integration, as well as integrate these efforts with ongoing consulting and outsourcing. IBM offers specific EII and data integration consulting with an unmatched breadth and depth of resources.

Where IBM's EII Solution is Most Effective
IBM's EII solution is especially effective where e-enterprises seek to achieve competitive advantage by extending BI solutions beyond the information that can be incorporated in the data warehouse. DB2 Information Integrator allows querying across a broader range of semistructured and unstructured data than typically included in the data warehouse, and allows reporting across multiple data marts. DB2 Information Integrator also allows querying new real-time data in OLTP systems, rather than waiting for end-of-day or end-of-week downloads into the data warehouse. Where mergers or acquisitions have left the company with multiple data warehouses, or line-of-business initiatives have driven the development of data marts in the absence of or disconnected from the enterprise warehouse, DB2 Information Integrator can facilitate enterprise-wide reporting.

IBM's EII solution is also especially effective when users want to achieve cost savings by speeding development and upgrading of key applications that access multiple back-end data sources, such as enterprise portals and business process integration solutions. DB2 Information Integrator allows applications, such as human resources portals, to access information such as personnel records and government requirements databases via a common, high-level transactional interface. DB2 Information Integrator aids creation of "composite applications" by allowing a single interface to synchronize access to linked data across applications, such as customer order entry and supply-chain ordering data.

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

DB2 Information Integrator is especially adept at aiding enterprises seeking to implement "on-demand computing." It supports relatively high-performance queries across a wide range of internal/external data types. This allows enterprises to adapt rapidly and effectively to changes in their environment as evidenced by changes in their — and their value chain's — information systems.

Aberdeen Conclusions

IBM deserves close attention from IT buyers not only for taking a leadership role among major software-infrastructure suppliers in perceiving the potential benefits of EII and crafting a product to meet that need, but also in its "follow through." IBM DB2 Information Integrator, right out of the starting gate, has exceptional scope and a strong foundation in IBM's long experience in database technology, plus the incalculable benefit of its combination with IBM's services. Services that allow users to leverage EII fully are key to the success of EII in the typical enterprise. Moreover, IBM DB2 Information Integrator can serve as a complement and enhancer to IBM's BI, EAI, and BPI solutions.

In addition, EII opens up a new range of strategic possibilities for competitive advantage for e-businesses as a class — not only for IBM's customers. These IT shops should also examine IBM's EII solution as an open product that is effective in non-IBM environments, as well.

It is an axiom of the computer industry that the pioneers take the arrows; but in this case, given IBM's solid record of "technology insurance," the pioneers are more likely to reap additional rewards. Aberdeen therefore recommends that IT buyers place IBM DB2 Information Integrator at the very top of their short lists.

To provide us with your feedback on this research, please go to www.aberdeen.com/feedback.

Aberdeen Group, Inc.
260 Franklin Street
Boston, Massachusetts
02110-3112
USA

Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

Aberdeen Group is a computer and communications research and consulting organization closely monitoring enterprise-user needs, technological changes and market developments.

Based on a comprehensive analytical framework, Aberdeen provides fresh insights into the future of computing and networking and the implications for users and the industry.

Aberdeen Group performs projects for a select group of domestic and international clients requiring strategic and tactical advice and hard answers on how to manage computer and communications technology. This document is the result of research performed by Aberdeen Group, which was underwritten by IBM. Aberdeen Group believes its findings are objective and represent the best analysis available at the time of publication.