

## THE INFORMATICA® SOLUTION FOR MAINFRAME DATA ACCESS



“Mainframe systems remain the heartbeat of our business. Thanks to global expansion and the web, we’ve become a true 24 x 7 enterprise. We have more data, but less time to extract it. How can we build analytic applications without disrupting our source systems?”

Regardless of the growth of client/server and the Internet, vast amounts of data still reside on mainframe systems. Today, mainframes continue running many of the world’s largest transaction databases because no platform yet rivals them for dependability and power.

Beyond sheer scalability, there are numerous other reasons why mainframe systems continue to thrive. In some organizations, legacy applications proved too complex to migrate. In some vertical industry sectors, ERP packages did not provide adequate solutions. Even for those enterprises that were able to migrate, many have continued to operate legacy systems for functions beyond the scope of ERP or other enterprise packages.

Yet, regardless of platform, virtually all enterprises require better information, not simply on how their businesses are performing, but why. The demands of intensified global competition and e-commerce are driving the need for business intelligence. Data warehouses and data marts have emerged, allowing business analysts to submit complex queries without disrupting transaction systems with complex queries, data transformations or multiple table joins.

Within the past 12 to 18 months, packaged analytic applications have emerged to make business intelligence from data warehouses even more accessible. By providing solutions instead of tools, analytic applications promise to spread business intelligence systems in much the same way that packaged applications did for ERP.





## BUILDING ANALYTIC SYSTEMS FROM MAINFRAME SOURCES.

If your company relies on mainframe systems, extracting, transforming, and moving data from source to target can be extremely challenging. For instance, the data must be converted from EBCDIC to ASCII. If the data was stored in a non-relational file or database, it must be mapped into relational table structures, requiring creation of flat files, a step that consumes significant mainframe computing cycles and DASD. Furthermore, if the source data comes from older systems, data structures (and the programs associated with them) may be inconsistent and poorly documented.

### *Data Volumes are Overwhelming.*

Compounding the problem, most organizations are dealing with more data than ever. A 1999 Survey.com (<http://www.survey.com>) database solutions survey shows that the amount of data moved into data warehouses will almost triple between 1999 and 2002.

There are many factors that are driving the need to integrate mainframe-based transaction systems with the rest of an organization's information architecture. A few examples include:

*The need to draw closer to customers.* Because customers have greater choices today, organizations must compete harder for their business. That often includes developing a wider array of products and services to meet customer needs, mass customizing products, adding new customer touch points or integrating more closely with the customer's supply chain. The end result generates more transactions to core business systems.

*The explosion of eBusiness.* Building e-commerce Web sites, providing on-line ordering, self-service customer information, or on-line trading communities generates more Web site hits that inevitably touch back-end systems.

*The requirement to expand business hours.* The growth of global business means that enterprises are now often 24 x 7 operations. Mainframes still provide one of the most stable platforms, providing the necessary up-time for global systems.

Regardless of the factor, the results are the same. Increased activity and increased sales place more stress and more data into core back-end systems. The need to become more customer-driven increases the need for analytic systems which allow organizations to know more about their products, operations and customer preferences. And the need to stay open longer hours, processing more transactions and managing more data forces IT organizations to do more with less. There is more data, more need to integrate established back-end and analytic systems and less time for batch windows to migrate the data.

### *The Need for Data Access.*

According to a recent Meta Group survey, data warehouse users are demanding more frequent data access than ever. The survey, conducted among 2800 attendees to Data Warehouse World conferences, pointed to daily data replenishment as a high priority across all industry sectors. Survey respondents attributed the need for frequent updates to the demands of operating 24 x 7 businesses and the need to manage global logistics as part of supply chain management challenges.

Where does this data come from? The same study also asked respondents where they maintained their operational systems and data warehouses. Of the companies running operational systems on mainframes, 79% of them ran their data warehouses on other platforms. And, according to IDC, mainframe databases remain alive and well. They estimate that the non-relational database market will continue averaging \$1.4 billion annually through 2004. The result is the need for an efficient approach to move data from mainframes to NT or UNIX servers.

The bottom line? In today's business climate, most enterprises have *less* time to move *more* data. There is little room for error. Customer service levels mandate that you cannot afford to take the system offline for long, limiting the batch window available to accommodate the previous day's orders. Therefore, high-speed data propagation has become mandatory, raising a huge challenge for data extraction programs.



### Obstacles to Access.

Mainframes can house many different types of data. They include flat files, relational databases such as DB2, non-relational databases such as IMS or Adabas, and specialized file systems such as ISAM or VSAM.

Those are just a few examples of the varying forms by which data is stored on mainframes. Because each of these data types requires different methods for accessing data, extracting mainframe data has traditionally required highly specialized skills.

Furthermore, until recently, the only way to get at mainframe data was to write COBOL batch programs, an option that is not viable for most organizations. Programming is labor-intensive to develop and maintain—a burden that could easily limit the ability of data warehouses to serve the changing competitive needs of your organization.

The problem is compounded for analytic systems because they change more frequently than transactional counterparts. For instance, as new competitors enter a market, companies often alter their product mixes, service offerings, pricing structures and so on. Business users, in turn, are likely to demand new data views or query functions from their analytic systems. Those changes require COBOL programmers to make changes or write new code—a step that can become a serious bottleneck.

Another serious drawback is the narrowing availability of mainframe sources and databases. For instance, companies that embrace e-commerce are learning that their customers expect them to stay open around the clock. That heavily impacts mainframe data extraction programs, because they are usually run in batch windows which often require that the source systems be taken off-line to ensure that the latest data is extracted. The result: as data volumes snowball, programs have less time to extract the data. Ideally, the solution to this problem is to make extraction programs smart enough to limit their activity to data that has changed since the last extraction. However, most COBOL programs lack the intelligence to filter out unchanged data.

Furthermore, many companies are having difficulties attracting and retaining COBOL developers—a problem which was exacerbated by recent Y2K compliance activity. With most young developers gravitating toward languages such as Java or Visual Basic, the COBOL talent pool is slowly shrinking.

The bottom line? There are many kinds of mainframe data sources, each of which requires specialized skills. Organizations still rely on scarce COBOL programmers to get at those data sources. With analytic needs becoming moving targets, organizations with mainframe data are often fighting losing battles with time.

### Informatica: Providing a Choice of Solutions for Mainframe Data Access.

There are alternatives to building large, hard-to-maintain COBOL extraction programs. If your organization relies on mainframe transaction systems, Informatica delivers a complete set of solutions that can access DB2 or flat file data.

Our solutions are built around PowerCenter®, Informatica's engine-based approach to data extraction, transformation, and loading. Informatica's PowerCenter accesses data from all major server platforms—including UNIX, NT, and mainframes. PowerCenter delivers critical benefits to analytic system designers, including:

- **High productivity:** PowerCenter's object-oriented, visual data transformation engine provides faster, more maintainable solutions to moving data from source to target systems, compared to programming solutions.
- **Flexibility:** Thanks to PowerCenter's global metadata repository, you have complete flexibility in designing analytic system architectures. PowerCenter supports all analytic system configurations, from "hub-and-spoke" enterprise data warehouses and satellite data marts, to federated peer-to-peer groupings of marts.
- **Maintainability:** PowerCenter's metadata repository eases the task of change data capture, where individual parts of a data warehouse or data marts must be selectively updated when a piece of data changes.
- **High performance:** PowerCenter offers native database drivers for all major relational data sources, and supports key performance capabilities including multi-threading, parallel processing, pipelining and streaming. With these capabilities, PowerCenter ensures that your organization can migrate data to analytic systems within the shorter batch windows that are characteristic of today's 24 x 7 enterprises.



### Sourcing Data from DB2 UDB.

IBM's DB2 UDB is a product family that spans all of IBM's platforms. Informatica supports this diverse product line, providing the ability to source data from any IBM platform, including Windows NT, UNIX, OS/400, and the mainframe (OS/390, or MVS). This provides enormous flexibility. For instance, you can source data from DB2 on the mainframe and migrate data to any major open systems relational database target, including DB2 for NT or UNIX. Additionally, you can use DB2 on NT or UNIX as a target, or as the platform for Informatica's meta data repository.

Informatica provides several methods for accessing mainframe DB2 data, including:

- Integration with IBM's DB2 Connect
- PowerConnect™ for DB2
- Integration with BMC's Enterprise Data Propagator

### PowerCenter and IBM DB2Connect.

Informatica also supports IBM's native DB2Connect for providing seamless access into DB2. While DB2Connect is IBM's generic API for accessing all DB2 databases, it has been optimized for DB2 UDB on Unix platforms. There are some performance issues with performance against DB2 on AS/400 and the mainframe. While it is adequate for old-style client/server applications, its access speed can be problematic for the volumes of data needed for data warehouses.

### PowerConnect for DB2.

This is the product of a partnership with NEON Systems. Using PowerConnect, DB2 on MVS or OS/390 can be used as a source or a target.

PowerConnect for DB2 combines NEON's DB2 drivers with PowerCenter's high-performance data extraction and transformation solution, and Informatica's Source Analyzer mainframe code conversion utility. The result: mainframe DB2 can be accessed just as easily as any major open systems relational database. PowerConnect for DB2 makes all the necessary data conversions, such as transforming EBCDIC to ASCII, Packed Decimal to Character, Big Endian to Little Endian, and 370 to IEEE floating point numbers. Additionally, PowerConnect for DB2 provides detailed tracing options to make troubleshooting easy.

PowerConnect for DB2 delivers extremely high performance through features such as:

- *DB2 Static SQL support:* This provides superior performance for DB2 compared to generic SQL, and allows the use of DB2 "plan-based security," which is far more flexible than conventional access control methods that are user, table, or query-based.
- *Optimized data communications:* Data moved by PowerConnect for DB2 is compressed to make more efficient use of processor and channel resources. This is supplemented by advanced buffering, call anticipation and other optimizations that further improve performance. If wide-area networks are used, data can be encrypted.
- *Optimized DB2 catalogs:* While DB2 catalog tables were not designed for fast access, PowerConnect for DB2 can build optimized catalogs that are pre-indexed and pre-joined for better performance.

The architecture is simple, eliminating the need for gateway software or gateway machine. PowerConnect for DB2 operates using a driver, based on the target system, which runs Source Analyzer, and a host-based task on the mainframe, which is administered by a collection of ISPF panels under TSO. This means lower complexity, higher reliability and faster performance than competitive solutions that require gateways.

### PowerCenter and BMC's Enterprise Data Propagator.

Informatica and BMC have partnered, integrating PowerCenter with BMC's Enterprise Data Propagation (EDP) products. The result is a powerful combination, joining BMC's high-speed, mainframe streaming capabilities with Informatica's efficient, high-performance, engine-based data integration solutions.

BMC's products specialize in extracting mainframe data and propagating it into UNIX or Windows NT relational databases. These products provide high-performance, bulk data propagation from mainframe sources, including DB2 and non-relational sources such as IMS, VSAM batch and CICS VSAM systems, supporting relational targets including Oracle (UNIX), SQL Server 7.0 (NT), DB2 (OS/390), and Sybase (Unix) databases. BMC's mainframe data propagation products deliver optimum performance

thanks to their native support of TCP/IP. They will soon offer the ability to support open storage subsystems for storing and staging data. The advantage of this approach is that data warehouse refreshes can be scheduled and executed without requiring any mainframe cycles.

BMC's data propagation products include:

- *DataMove*: To initially populate the data warehouse
- *ChangeDataMove*: To provide data updates and keep the warehouse data sources synchronized

PowerCenter works with both BMC products, performing transformations on the staged data and moving it into one – or more – data warehouses or data marts. When used with PowerCenter, BMC's DataMove is used to initially populate the data warehouse. PowerCenter then performs transformations on the staged data and moves it into one – or more – data warehouses or data marts. When BMC's ChangeDataMove is used, PowerCenter is configured to operate in update mode by using Informatica's unique data-driven Update Strategy Transformation object.

The integrated Informatica/BMC solution provides an efficient approach that eliminates:

- Intermediate flat files, which drain CPU resources and consume disk space
- Time-consuming steps to write extraction transformation, and loading programs from scratch
- Batch windows and availability hits to the OS/390 sources

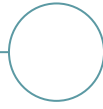
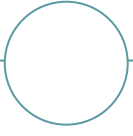
The integration of PowerCenter with BMC's Enterprise Data Propagation system provides Informatica customers with a high-performance, scalable, extensible solution for capturing and propagating mainframe data to business intelligence and analytic systems running on mainframe, open systems, and NT environments. Together, both products provide a solution that reduces impact on mission-critical transaction systems while improving developer productivity.

### Non-Relational Sources.

PowerCenter provides access to any kind of mainframe data source, including IMS, VSAM and ADABAS through its ability to import fixed block sequential files. There are several ways to do this. For small files, PowerCenter supports extraction using standard ODBC techniques. For larger data blocks, PowerCenter has the ability to directly source fixed sequential flat files that can be generated from virtually any mainframe data source. This is accomplished using the following steps:

1. Analyzing and importing metadata from COBOL Copy Books or data divisions of COBOL files. (If the Copy Book is not available, PowerCenter can create copy specifications.) When importing metadata from a COBOL file, PowerCenter automatically creates a PowerCenter Source definition. The PowerCenter server then uses the metadata to source data from a physical sequential fixed block file produced by a COBOL extract program. With this feature, PowerCenter automatically reads data characteristics such as the order, size and type of each field within the COBOL record definitions; clauses such as OCCURS, USAGE, and REDEFINES; and any comments that are useful for source documentation.
2. Generating appropriate normalization transformations and targets. This is critical, because mainframe data sources such as VSAM, IMS or Adadbas often denormalize data when they compress several records together. This can be quickly performed using PowerCenter's drag and drop design capabilities.
3. Directly reading fixed sequential files.

Based on the metadata that is derived from COBOL Copy Books, PowerCenter can source fixed sequential block data in a single pass, even if there are REDEFINES clauses. The PowerCenter server automatically transforms common mainframe data types, including EBCDIC, COMP, COMP-3 (packed decimal), COMP-5, COMP-6 (binary coded decimal) and COMP-X into ASCII. PowerCenter extraction logic can reside in the most efficient point: on the mainframe or on the PowerCenter server. If extract logic is deployed on the PowerCenter server, built-in FTP capabilities can be used to initiate extract and manage sessions.



## **CONCLUSION.**

With PowerCenter, your company can build analytic systems without concern for data source. PowerCenter's flexible, high-performance solutions are developed using visual engine-based techniques that are far more powerful, maintainable and productive compared to raw coding. PowerCenter's proven, high-performance capabilities are helping organizations build analytic systems in an era where batch windows are shrinking.

This is a critical advantage, especially for larger enterprises, which maintain large volumes of data on mainframe systems. PowerCenter delivers powerful, more productive alternatives to writing extraction and transformation programs in COBOL. PowerCenter can access your mainframe data, regardless of whether it resides in DB2, or any other database or file system source. Informatica has partnered with leaders in the mainframe tools space to deliver the most powerful solutions for sourcing legacy data. The result is a winning combination: the versatility and power of PowerCenter, paired with proven mainframe data extraction technology.

With analytic systems becoming a critical linchpin for organizations seeking success in the eBusiness environment, organizations cannot afford to worry about where the data is, how to get it out, and how fast it can get to the right target. Informatica and its partners are delivering solutions to help your organization populate analytic systems easily and quickly from all of your major enterprise data sources.

### About Informatica.

Informatica provides eBusiness analytic solutions that enable companies to leverage business insight for competitive advantage. Over half of the Fortune 100 leverage the Informatica Insight Network to integrate, analyze and personalize critical business information to improve business performance and responsiveness across the entire business value chain. Informatica's products include a comprehensive suite of analytic applications, the industry's most powerful data integration platform and a wide range of supporting services and partnerships.

Our list of leading global customers includes 3Com, AMD, Boeing, Borders Group, Chevron, Cisco, CNET, Deutsche Bank, eBay, GE, Genencor, Hewlett-Packard, Inktomi, J. Crew, Merrill Lynch, MetLife, MCI WorldCom, Motorola, Mutual of Omaha, Philips, Polo Ralph Lauren, Preview Travel, Priceline.com, Sprint, and Zurich Insurance.

## INFORMATICA®

**Informatica Headquarters**  
3350 West Bayshore Road  
Palo Alto, California 94303  
USA  
Phone: 650.687.6200  
Fax: 650.687.0040  
Toll-Free: 800.653.3871  
Web: [www.informatica.com](http://www.informatica.com)

**Informatica Corporation**  
6 Waltham Park  
Waltham Road, White Waltham  
Maidenhead  
Berkshire SL6 3TN  
UK  
Phone: +44.0.1628.511311  
Fax: +44.0.1628.511411

**Informatica GmbH**  
Am Soeldnermoos 12  
85399 Hallbergmoos  
Germany  
Phone: +49.0.811.555.0.777  
Fax: +49.0.811.555.0.797  
Toll-Free: 0800.powermart

**Informatica Software (Schweiz) AG**  
Täferstrasse 4  
5405 Baden-Dättwil  
Switzerland  
Phone: 004156.483.38.38  
Fax: 004156.483.38.99

**Informatica International, Inc.**  
Asia Pacific Region Office  
16 Raffles Quay  
Level 36 Hong Leong Building  
Singapore 048581  
Phone: 65.322.8589  
Fax: 65.234.6456