

# **Business Objects**

## **Data Integration**

---

- ▶ *A Technical Overview*

---

Authors: Darren Cunningham, Philip On

Contributors: David Barkaway, Tom Le, Jennifer Meegan, Awez Syed, David Townley

Audience: ETL developers, data warehouse managers, database administrators, business intelligence architects, business intelligence practice managers.

---

# Contents

---

<b>Executive Summary</b> .....	<b>iv</b>
<b>Business Objects Data Integration Products</b> .....	<b>1</b>
BusinessObjects Data Integrator .....	1
BusinessObjects Rapid Marts .....	1
<b>Data Integrator Architectural Overview</b> .....	<b>2</b>
<b>Delivering Trusted Information</b> .....	<b>3</b>
Unification with the BI Platform .....	3
End-to-End Lineage and Impact Analysis .....	4
Universe Creation and Updates within the ETL Design Environment .....	5
Rapidly Create a Data Mart from Historical Crystal Report Instances .....	5
Data Profiling within the ETL Design Environment .....	6
Data Validation .....	7
Data Auditing .....	7
Data Cleansing .....	8
<b>Maximizing Developer Productivity</b> .....	<b>9</b>
A Single Design Environment .....	9
Ease of Use .....	9
Graphical Workflow .....	10
Extensive and Reusable Transformations and Functions .....	10
Interactive Debugging and Error-Trapping .....	11
Multi-User Development .....	11
Portability .....	12
Centralized Management and Administration .....	12
<b>Delivering Extreme Scalability</b> .....	<b>13</b>
Parallel Performance Architecture .....	13
Degree of Parallelism, Database Partitioning, and Parallel Aggregation .....	13
Distributed Processing .....	14
Grid Computing Support .....	15
Comprehensive Changed Data Capture Support .....	15
Batch and Real-Time Data Movement .....	15
Comprehensive Platform, Source, and Target Support .....	16
Deep Integration with ERP and CRM Applications .....	16
Metadata Management .....	17
<b>BusinessObjects Rapid Marts</b> .....	<b>19</b>
<b>Conclusion</b> .....	<b>21</b>
<b>Appendix – Data Integrator Technical Specifications</b> .....	<b>22</b>

---

## Executive Summary

---

- ▶ Corporate compliance
- ▶ BI standardization
- ▶ Data credibility
- ▶ End-user trust

Do you have initiatives focused on meeting objectives in any of these areas?

Has business intelligence (BI) become strategic to your organization's performance?

Are you seeking ways to turn your tactical and departmental BI projects into mission-critical implementations that deliver "one version of the truth" to end users across the enterprise?

With data volumes and complexity growing as fast as end-user expectations for the right information at the right time, you must ensure that your approach to data integration is as well thought out as your approach to meeting their needs.

Only Business Objects delivers enterprise-class data integration that is unified with the market-leading BI platform. With BusinessObjects™ Data Integrator, you can:

- ▶ Reduce your total cost of ownership (TCO) through superior integration with the industry's leading BI platform
- ▶ Accelerate BI implementations
- ▶ Maximize developer productivity
- ▶ Ensure information accuracy, integrity, and trust

This paper will provide an overview of the Data Integrator architecture and summarize the ways that Data Integrator delivers trusted information, maximizes developer productivity, and ensures extreme extraction, transformation, and load (ETL) scalability.

---

# Business Objects Data Integration Products

---

Business Objects data integration products allow you to easily explore, transform, and move data anywhere, at any frequency. Industrial-strength ETL and packaged data integration solutions for enterprise applications such as SAP, PeopleSoft, Siebel, and Oracle ensure that you always deliver accurate, timely, and integrated data that BI users can trust.



*Business Objects provides a comprehensive data integration solution to accelerate and improve BI deployments.*

## BusinessObjects Data Integrator

Data Integrator is a productive and scalable data integration platform. With it, developers can easily explore, extract, transform, and deliver data anywhere, at any frequency through a single, graphical design environment. Data Integrator lets your organization ensure data integrity, maximize developer productivity, and accelerate reporting, query and analysis, and performance management projects.

## BusinessObjects Rapid Marts

BusinessObjects Rapid Marts™ are built with Data Integrator and provide packaged ETL for operational reporting and end-user query and analysis. With Rapid Marts you can generate value from a business intelligence project sooner by leveraging prebuilt data models, transformation logic, and data flows for extracting data from enterprise applications like SAP R/3, Siebel, Oracle, PeopleSoft, and J.D. Edwards applications.

# Data Integrator Architectural Overview

## Data Integrator consists of four primary components:

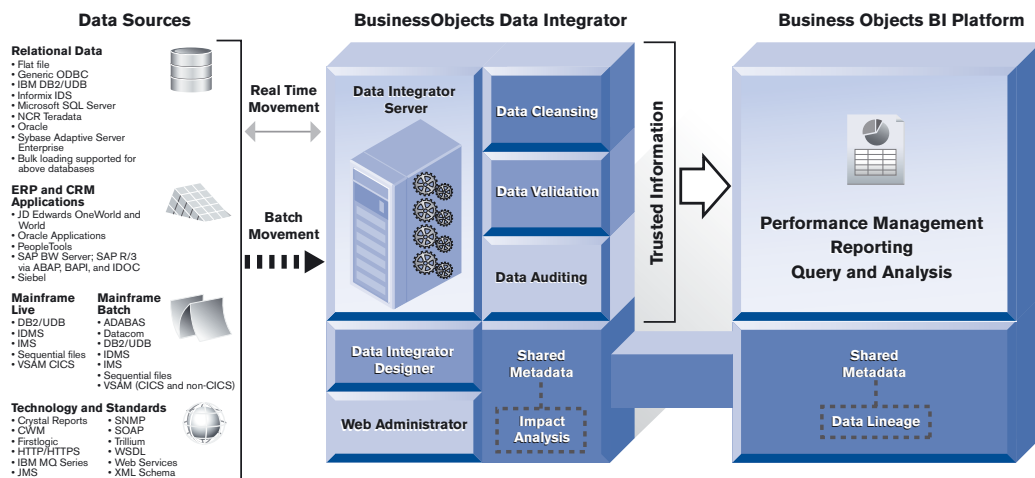
- ▶ A graphical designer
- ▶ A web-based administrator
- ▶ A data server
- ▶ A metadata repository

Data Integrator Designer is the single tool that provides a graphical interface for performing all the tasks involved with building, testing, and managing an ETL job. Through its easy-to-use graphical interface, you can manage projects; profile data; create ETL jobs; cleanse, validate, and audit data; set parallel job execution; build workflows; and test, debug, and monitor your ETL jobs.

The web-based administrator allows you to start, stop, schedule, and monitor ETL jobs independent from the design environment.

The Data Integrator data server is the data movement engine that integrates data from multiple heterogeneous sources, performs complex data transformations, and manages extractions and transactions from ERP systems.

The Data Integrator metadata repository holds user-created and predefined system objects, source and target metadata, and transformation rules. It is set up on an open client/server platform to facilitate the sharing of metadata with other enterprise tools. When using Data Integrator with the Business Objects BI platform, you have complete metadata integration between your BI tools and the ETL environment. This shared metadata provides advantages such as end-to-end impact analysis and data lineage, a key requirement for delivering trusted information.

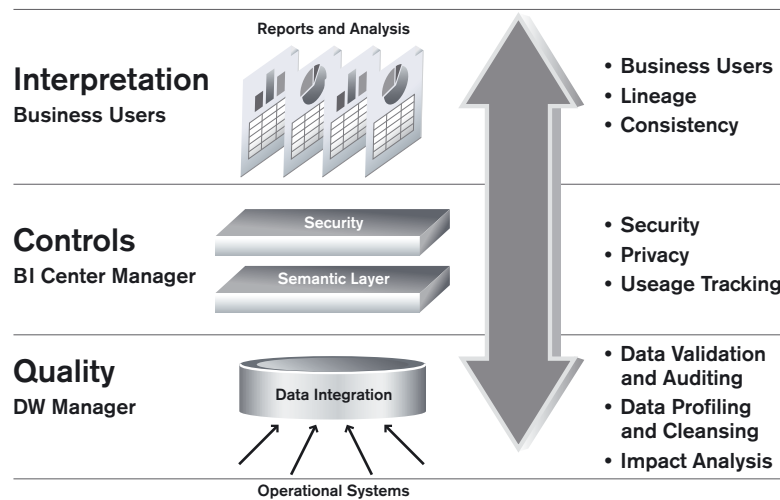


*Business Objects provides a comprehensive business intelligence solution that includes enterprise data integration for delivering trusted information .*

# Delivering Trusted Information

To deliver trusted information, you need a technology framework that includes data integration to ensure data quality, controls for security and privacy, and BI reporting and analysis that provides meaning and lineage of information. You need a data integration and business intelligence solution that is integrated down to the metadata level.

## Trusted Information Framework



Data quality is no longer simply an IT challenge. It is one that affects the overall success of your BI implementation. Data Integrator has many unique capabilities to help you deliver information that is of high-quality and trustworthy. Unification with the BusinessObjects XI platform also means that you can easily manage and understand the impact of source changes in the environment and ensure that BI end users are always working with the right data, at the right time. Data Integrator delivers trusted information through:

- ▶ Unification with the BI platform
- ▶ Data profiling
- ▶ Data validation
- ▶ Data auditing
- ▶ Data cleansing

## Unification with the BI Platform

Until now, data warehousing and the BI market in general has consisted of many best-of-breed players, with each vendor focusing on selling their component within the end-to-end “stack.” No one has focused on true integration from source systems to end-user reports and executive dashboards.

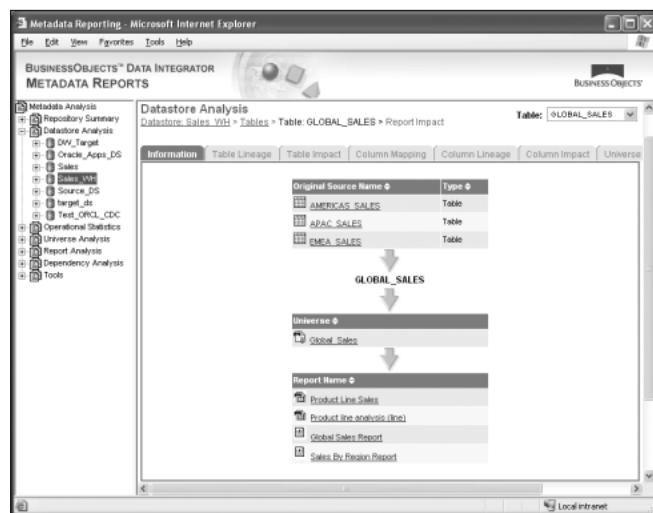
By deeply integrating the entire ETL process with the BI platform, both IT and business users gain measurable benefits that include easy metadata management, simplified and unified administration, lifecycle management, and—ultimately—trusted information. Data Integrator delivers the following key points of unification with the BusinessObjects Enterprise BI platform:

- ▶ End-to-end data lineage and impact analysis
- ▶ Create semantic layer (universe) and manage change within the ETL design environment
- ▶ Rapidly create a data mart from historical Crystal Reports® instances

## End-to-End Lineage and Impact Analysis

Change is inevitable. Data sources will be updated, added, or deleted. You need insight on the impact of source system changes to your entire BI system. Conversely, your BI end users need visibility to trace the chain of truth back to the source systems. Data Integrator is unique in its ability to exchange metadata up and down the BI platform. End-to-end data lineage and impact analysis means that Data Integrator allows developers and system administrators to communicate directly with the BusinessObjects XI platform.

Whether it is a Crystal report or a BusinessObjects Web Intelligence document based on a universe, with Data Integrator you can find the relationships between sources (tables or columns), targets, universes, and documents. This capability helps to significantly reduce maintenance costs by knowing the impact of changes to source and target systems. Knowing which sources provide data for which reports also increases end-user trust, and allows you to gain a better understanding of the overall data architecture. This means you can track data from an end-user report right back to its original data source.



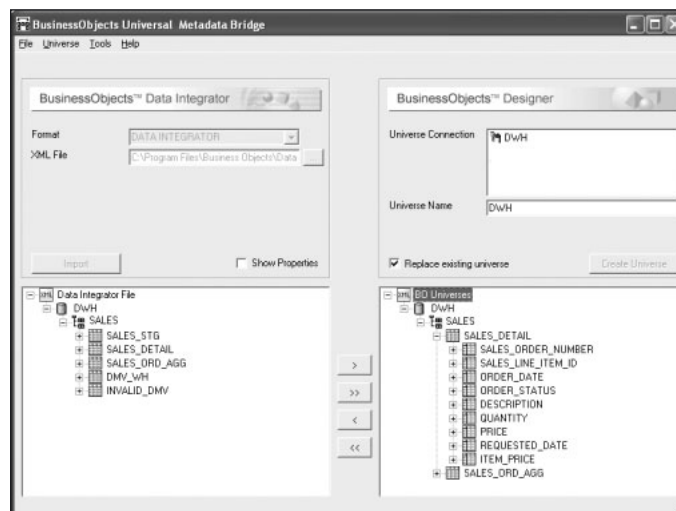
*Metadata reports allow you to see the impact of change all the way from the data sources to the end-user reports.*

---

## Create semantic layer (universe) and manage change within the ETL Design Environment

Within the Data Integrator Designer interface, you can simply create and update Business Objects universes with the click of a button saving time and avoiding errors commonly associated with manual effort. The universe is based on the metadata present in the Data Integrator repository. Information that is not normally available in the BusinessObjects Designer tool, such as column descriptions and primary and foreign key information from database catalogues, is automatically loaded into the universe. This means you can quickly build a universe for the target warehouse database and easily transfer ETL metadata such as data lineage, mapping expressions, and descriptions. Easy metadata sharing also lets universe developers know exactly where the data is coming from and helps them create better universes for end-user query and analysis.

But creating a metadata-rich universe is only the beginning. With Data Integrator, you can update universes based on source changes without manually having to make changes within the universe designer. You can also do a gap analysis between the technical and business metadata so you can easily see what has changed and decide if you want to propagate some or all of those changes. For example, you may want to transfer selected information into the universe such as lineage, but not descriptions. Easy metadata sharing and change management improves end-user understanding of information and gives users greater trust in their data because they can see where it came from and how it was transformed and calculated.



*Easily pass metadata from Data Integrator to the BI platform.*

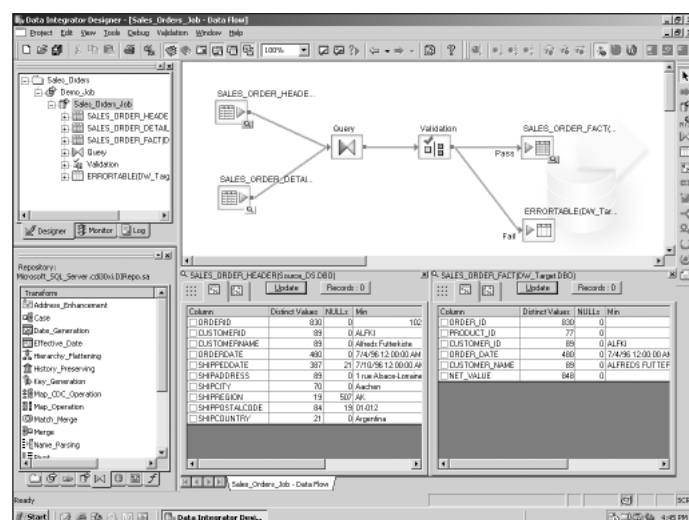
## Rapidly Create a Data Mart from Historical Crystal Report Instances

Many Crystal Reports customers are using the market-leading reporting tool to directly access information within operational systems. Because they are scheduling highly formatted reports broadly, they often do not create a data warehousing infrastructure that would be optimized for end-user ad hoc query and analysis. Instead, within the BI platform, each time a report is run, the

information is captured and saved as a historical “instance.” The Data Mart Accelerator for Crystal Reports feature of Data Integrator allows you to automatically build an ETL job that extracts data from this Crystal Reports instance and loads it into a series of aggregate-aware tables. The Data Mart Accelerator for Crystal Reports also dynamically creates a time dimension that allows BI users to perform trend-based analysis over time. And optionally, the wizard interface even allows you to automatically create a BusinessObjects universe that dynamically links back to the source report instance so you can instantaneously perform query and analysis on this data.

## Data Profiling within the ETL Design Environment

With the Data Integrator Designer you can define data mappings, transformations, and ETL logic. Designer is a single, graphical design environment for the majority of data integration tasks and is used to create workflows (job execution definitions) and data flows (data transformation definitions). Designer allows developers to create objects, then drag, drop, and configure them by selecting icons in flow diagrams, table layouts, and nested workspace pages. The objects in Designer represent metadata and the interface allows you to manage metadata stored in the Data Integrator metadata repository. Unique to Data Integrator—you can view and profile data before and after it is extracted and transformed without leaving the design environment. By helping to reduce design and data validation time and increase developer efficiency and productivity, Data Integrator allows you to profile data to get an understanding of its range and quality before the ETL process. Preview and profile physical data in source, target, and transform objects, profile and analyze column characteristics, analyze: # of distinct values, nulls, min, max, and column distribution, filter and manipulate rows in data flows, preview for all qualifying elements, preview, filter, and sort data before and after a transform in the dataflow, and view data side-by-side from heterogeneous objects.

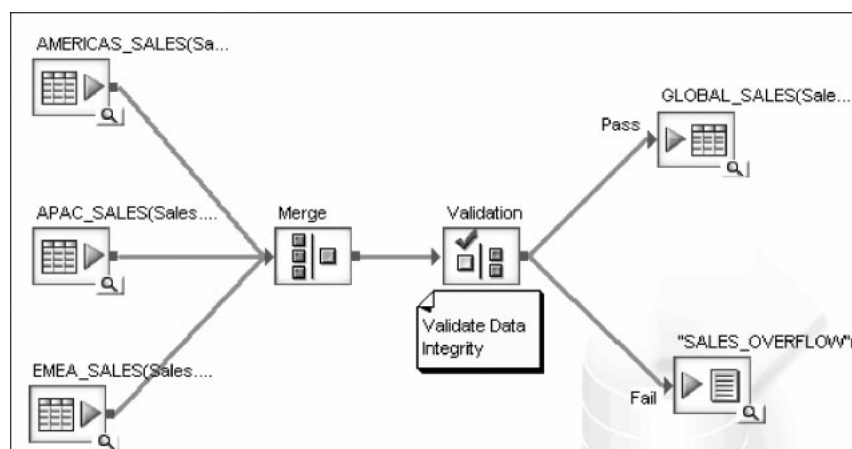


Preview and profile data within the ETL design environment.

---

## Data Validation

Data validation addresses the needs for delivering trusted information through a productivity enhancing process that ensures the accuracy of your data. A common challenge for ETL developers is exception handling: out of range data, fields with NULL value, or incorrect data. Data Integrator offers an easy and flexible way to identify and correct or reject erroneous data during the ETL process. Using a validation transform, you can define a reusable business rule to validate each record and column. For example, if you want to load only sales records for the month of October 2004, you may want to set up a validation rule that states: "Sales Date" is between "10/1/04" to "10/31/04". Data Integrator will look at this date field in each record to validate if the data meets this requirement. If it doesn't, you can choose to pass the record into a "Fail" table, correct it, or do both.



*Built-in data validation allows you to ensure data accuracy at the column and record level.*

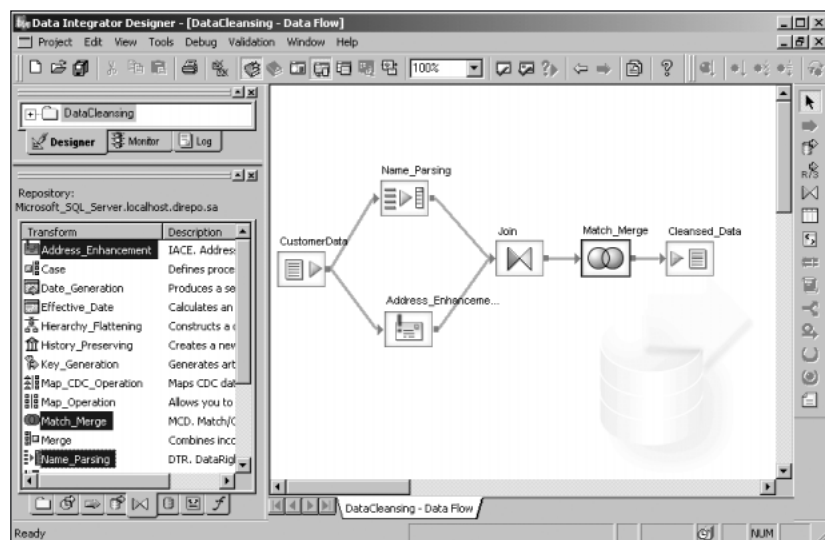
## Data Auditing

Another challenge for developers is auditing the integrity of the ETL job against operational rules. Data Integrator has built-in auditing capabilities that allow you to collect audit statistics such as "row count", "sum", "average", and "check sum" and verify it against user defined operational rules. With data auditing, you can verify if the expected data is read, processed, and loaded successfully. For example, if you were extracting tables from flat files and wanted to verify that all 100,000 records successfully loaded into the data warehouse. You can also verify the successful execution of a join. Auditing helps you see if any rows are missing and helps you determine if any joins have been improperly configured. Use audit statistics to specify rules to ensure that the correct amount of data is processed. In the case of errors, users can generate a notification of audit rule(s) failure. You can set the notification to email notification, raise exception, or custom audit script. In addition, the audit statistics persist in the Data Integrator repository. This provides you with an audit trail for all your data integration jobs.

---

## Data Cleansing

Most data quality issues are the result of poor customer contact information. Data Integrator addresses this challenge with data cleansing transforms that allow you to perform name and address parsing, matching (house-holding), and merging (de-duplication). Through tight integration with Firstlogic, the market leader in information quality, these transforms can be added to work flows and data flows allowing you to parse, correct, standardize, match, and consolidate data in the same ETL design environment. The user is shielded from the complexities of the underlying Firstlogic technology.



*Data cleansing transforms are available within the ETL design environment.*

---

# Maximizing Developer Productivity

If you are currently building and maintaining SQL scripts to extract, transform, and load data from disparate sources into your data warehousing infrastructure, you simply must consider the productivity benefits that Data Integrator delivers. Data Integrator allows you to develop and maintain complex ETL routines without writing or maintaining custom code. It means BI can be deployed faster and the ongoing maintenance challenges of growing data volumes and continuous change can be managed in an easy-to-use and administrative environment. It means your developers are more productive and can focus on delivering greater value to your organization at a lower total cost of ownership. Data Integrator ensures maximum developer productivity through:

- ▶ A single design environment
- ▶ Ease of use
- ▶ Graphical dataflow and workflow
- ▶ Extensive transformations and functions
- ▶ Interactive debugging and error-trapping
- ▶ Multi-user development
- ▶ Portability
- ▶ Centralized management and administration

## A Single Design Environment

Whether you are moving data in batch or real time or accessing and integrating data from SAP R/3 or mainframe systems—Data Integrator delivers maximum productivity through a single design environment. This enables faster time to deployment and easier maintenance, both of which result in a lower TCO. The Data Integrator Designer tool is a graphical, Windows-based development environment where data mappings, transformations, and control logic are defined. Within Designer, ETL jobs are created that contain workflows (job execution definitions) and data flows (data transformation definitions). Easily create objects, then drag, drop, and configure them by selecting icons in flow diagrams, table layouts, and nested workspace pages. The objects in Designer represent metadata and within the easy-to-use interface, you are able to manage the metadata stored in the Data Integrator metadata repository.

Data Integrator Designer is unique in allowing users to work from a single tool to manage projects, profile data, create ETL jobs, cleanse and consolidate data, set parallel job execution, build workflows, and test, debug, and monitor your ETL jobs.

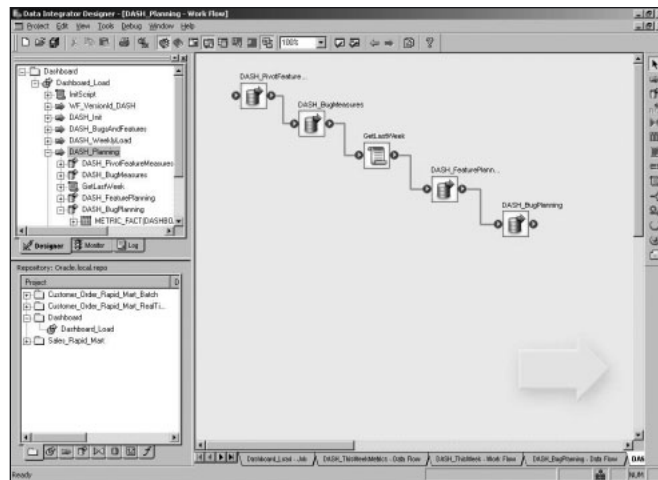
## Ease of Use

Data Integrator is a highly visual development environment. The vast majority of mappings are performed visually using the graphical interface by dragging and dropping. More complex mappings use drag-and-drop in combination with function wizards. Data Integrator includes many built-in functions including conversion functions, date functions, string functions, validation functions, math functions, if-then-else, lookup, and while. Data Integrator also includes more than a dozen powerful and flexible transforms for slowly changing dimensions, hierarchy flattening of XML content, table comparison, merge, pivot, history preserving, data cleansing, data matching and consolidation, and data validation.

---

## Graphical Dataflow and Workflow

Data Integrator projects are built graphically from data flows, the smallest reusable objects that load tables. Data flows can include complex logic, like the ability to read an unlimited number of data sources in one flow. They can be nested or combined into work flows, which dictate process order and conditions for processing. Work flows are combined into schedulable units, called jobs. And all Data Integrator graphical objects are self-documenting. This capability, along with the ease of building complex business logic data flows, makes ETL design faster and allows you to easily maintain your work within the single design environment. You can perform multiple transformation steps without writing code and easily build and combine complex data movement steps in reusable units of work that contain conditional logic based on your unique business rules.



*Workflow management is built into the Data Integrator design environment.*

## Extensive and Reusable Transformations and Functions

Data Integrator has a library of prebuilt transformations that make common data warehousing tasks easy. Reusable transformations for slowly changing dimensions, data validation, data cleansing, history preserving, table comparison, surrogate key generation, hierarchy flattening, and case statements are available to accelerate development. And many of Data Integrator's powerful transformation capabilities can be found within the query transform. Functions such as sorting and aggregating are performed within the query transform. While the vast majority of mappings are performed visually—using drag-and-drop in the graphical interface—more complex mappings use drag-and-drop in combination with function wizards. Data Integrator includes many built-in functions including conversion functions, date functions, string functions, validation functions, math functions, if-then-else, lookup, and while. In the rare case when scripting is required, Data Integrator provides a powerful scripting language that includes all basic programming constructs such as if then else, while, try, catch blocks, and global and local variables. Developers can use this scripting language to write reusable custom functions.

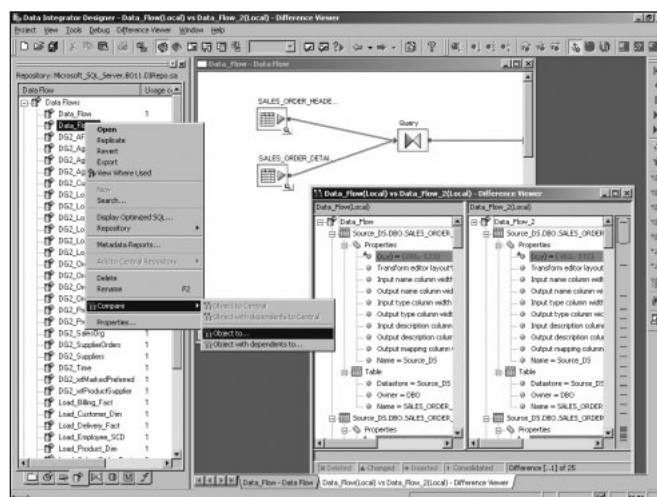
## Interactive Debugging and Error-Trapping

Data Integrator provides built-in capabilities for managing data and improving the quality of the ETL design. Interactive debugging and error-trapping allows you to easily track problems and identify errors in your ETL design. The interactive debugger allows you to debug one or more records by setting a breakpoint (i.e., conditions) and/or a filter. The breakpoint will suspend the ETL job once the condition is met and allow you to get only the next row of data, step to the next transform, or continue debugging. You can edit data on the fly and see how it evolves through the ETL process as well as analyze a specific transform by viewing a range of data that passes through it. Interactive debugging helps ensure that you are bringing the right data into your environment and the ability to view and modify data within the single design environment helps maximize developer productivity.

## Multi-User Development

While Data Integrator is suitable for any size data warehousing initiative, it is designed to support large-scale, enterprise deployments. Multi-user, multi-project, check-in and out repositories enables reuse of objects and helps organizations standardize on one data integration platform across multiple projects. With Data Integrator, developers can work in parallel and maintain version control as they check their work in and out. Data Integrator also provides options for managing secure access and tracking for objects in central repositories. The mechanisms for managing this security include authentication of valid users, granting authorization and permissions to different objects and levels, and the ability to maintain an audit history of changes made to an object including user names. These security procedures allow you to establish groups, users, and object-level permissions so you can securely share and collaborate between users in a multi-developer environment.

Additionally, for team-based development environments, the Data Integrator Difference Viewer allows you to compare the metadata for similar objects and their properties. It allows developers to proactively compare differences between the current version of the object with the previous version to see if the required changes were applied. You can also use the Difference Viewer to identify problems if you are encountering errors in an updated ETL job.



*The Difference Viewer simplifies maintenance of the ETL processes in a multi-user development environment.*

---

## Portability

A typical challenge that developers have is editing their ETL jobs to support different database environments. This is a common problem when you want to port jobs from a testing to production environments. In addition, if you are embedding ETL capabilities into another application through another OEM partnership, portability is necessary to easily deploy across different database environments. With Data Integrator XI, you can have a single datastore that supports multiple database configurations—regardless of database type, instance or version.

This feature allows you to decrease end-to-end development time in a multi-source, 24x7, enterprise data warehouse environment by allowing you to more easily port jobs among different database types, versions, and instances. For example, if you had a job that you developed using Microsoft SQL Server source tables, you would not need to rebuild or configure the job to support an Oracle or Teradata environment. This feature provides greater ease-of-use for job portability scenarios, such as:

- ▶ OEM (different connections for design and distribution)
- ▶ Migration (different connections for development, test, and production environments)
- ▶ Multi-instance (databases with different versions or locales)
- ▶ Multi-user (databases for central and local repositories)

## Centralized Management and Administration

The Data Integrator Administrator provides browser-based management and administration of Data Integrator resources including:

- ▶ Scheduling, monitoring, and executing batch jobs
- ▶ Configuring, starting, and stopping real-time services
- ▶ Configuring the ETL job server and repository usage
- ▶ Configuring and managing interfaces
- ▶ Managing users

As well, while Data Integrator has a built-in job scheduler that enables a user to schedule jobs, you can also launch jobs based on events. Event-driven scheduling is usually achieved through a third-party scheduler and the Data Integrator simple network management protocol (SNMP) support allows it to integrate with network management products such as Tivoli, HP OpenView, and BMC Patrol.

---

## Delivering Extreme Scalability

---

Performance and scalability are critical to your overall data warehousing and BI success. You must be able to manage growing data volumes and complexity while leveraging all of your data assets so that you can deliver timely and trusted information. Data Integrator offers a parallel performance architecture to meet your scalability requirements for massive data movement. Features include table partitioning, native bulk loading for most databases, parallel processing, degree of parallelism, grid computing, aggregation, and intelligent caching. To optimize performance, Data Integrator minimizes excess data movement and allows you take advantage of your database hardware by distributing processing to where the data resides. Data Integrator delivers extreme scalability through:

- ▶ A parallel performance architecture
- ▶ Degree of parallelism, database partitioning, and parallel aggregation
- ▶ Distributed processing
- ▶ Grid computing support
- ▶ Comprehensive changed data capture (CDC) support
- ▶ Batch and real-time data movement
- ▶ Comprehensive platform source and target support
- ▶ Deep integration with ERP and CRM applications
- ▶ Metadata management

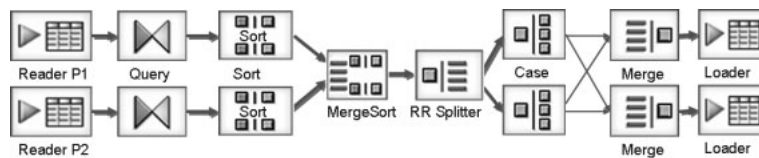
### Parallel Performance Architecture

The Data Integrator data server can move data in either batch or real-time mode and uses distributed query optimization, multi-threading, in-memory caching, in-memory data transformations, and parallel pipelining to deliver high data throughput and scalability. Once a data integration project has been designed, the designer accesses the data server to run, test, debug, and monitor a job. The data server has real-time capabilities built into its engine, which allows for evolutionary enhancements to Data Integrator, such as support for HTTP and web services. The Data Integrator architecture is based on the Performance Optimizer, which automatically routes tasks to the appropriate resource—like the underlying database system—while automatically spawning new threads for complex data flows and ensuring optimal CPU utilization. The ability to read application data and translate it as it is being read means there are fewer steps needed to transform the data. Having many choices for methods of loading data targets—for example using a bulk loader or parameterized SQL—also adds to the superior performance that Data Integrator delivers.

### Degree of Parallelism, Database Partitioning, and Parallel Aggregation

For linear scalability, Data Integrator is able to split a data stream into multiple pipelines using a "degree-of-parallelism" setting. A single stream flow can be executed by simply changing the setting and taking advantage of SMP and MPP machines. Data Integrator can read database partitions in parallel, process the data according to the number of available CPUs, and load data to a number of target partitions in a database. Partitioning information (e.g. Oracle) is imported as part of the table metadata, and used to extract and load in parallel. Data Integrator instantiates a

separate thread for each partition at runtime and executes them in parallel. Data Integrator also provides the ability to specify custom range partitions (DB2, SQL Server, etc.) by modifying the table metadata.

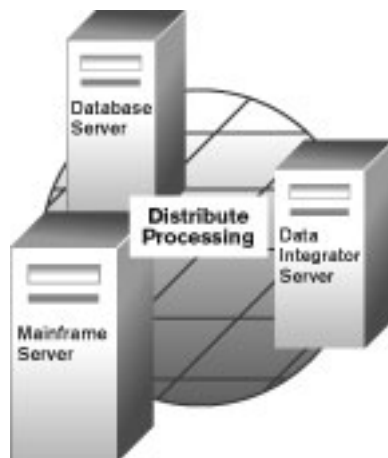


*Data Integrator supports degree of parallelism and database partitioning simultaneously.*

Additionally, you can achieve linear scalability because Data Integrator can perform aggregation in parallel in order to use additional CPUs. There is no need to serialize data for sorting and aggregating as aggregation keys and partial aggregate values are cached, reducing overall memory use and making it possible to handle larger data sets.

## Distributed Processing

It is no longer sufficient for enterprise data integration platforms to simply support parallelism. Data Integrator takes data integration to the next level and offers a parallel performance architecture that can intelligently push processing down to source database or mainframe systems. By leveraging the power of your existing infrastructure, Data Integrator maximizes ETL performance by letting your source or target systems perform the transformation. This also frees up your network from excessive data movement. For example, if you have an ETL job that joins different data formats on a mainframe such as VSAM and IDMS, you can push this transformation to the mainframe server instead of moving the data to the Data Integrator server for processing. Supported distributed processing operations include: aggregations, distinct rows, filtering, joins, ordering, projection, and functions that exist in the underlying database.



*Data Integrator's distributed processing minimizes unnecessary data movement across your network to optimize ETL performance.*

---

## **Grid Computing Support**

Data Integrator helps you increase efficiencies and reduce costs by balancing workloads among a group of servers—or grid—without a single point of failure. Taking advantage of available computing resources, Data Integrator intelligently sends new job requests to the least used machine, thereby ensuring maximum performance. Intelligent load distribution across multiple servers is based on CPU usage, available memory, and algorithms to determine the fastest possible execution within a group. Data Integrator provides a hot back-up by continuously polling the availability of other nodes. Detailed tracing, workload, and log reports are generated to help with system tuning and configuration. Throughput and fault tolerance are critical to extreme scalability. A Data Integrator server grid can be heterogeneous or homogeneous (Windows, Sun, HP, IBM, Linux) and managed via the web administrator and central repository.

## **Comprehensive Changed Data Capture Support**

Changed data capture support is a requirement for organizations with large data volume movement. Since only a small percentage of records change (update, insert, delete) in a day, you want to easily move only the changed data instead of the entire data set from your source systems. Data Integrator offers a log-based CDC approach that leverages native CDC technology from database and mainframe vendors. The advantage of a log-based approach is lower impact on source systems as Data Integrator reads from a database log instead of having source systems track and handle changes directly. For example, with Microsoft SQL Server, changed data is put into a distribution database that Data Integrator is able to read. Data Integrator supports mainframe CDC using IBM and Attunity technology.

## **Batch and Real-Time Data Movement**

Data Integrator provides batch and real-time data integration within the single tool and interface. Common data definitions across batch and real-time processes ensures data consistency and ease of development, and bi-directional, real-time interfaces and web services support allows for metadata integration with a wide array of tools and applications. Real-time data flows can be designed graphically that include logic to pull data from ERP and other enterprise systems to supplement a request and to construct a reply. These real-time data flows process requests and return messages in the form of XML. The real-time message processing server can process incoming messages and trigger outgoing messages in real time from any application. Data Integrator can also easily transform hierarchical documents—such as XML or EDI documents—to a relational format. Hand coding these transformations is highly complex and difficult because breaking data down into a flat format is cumbersome and can cause some loss in meaning and context. It can also significantly degrade performance. Data Integrator deals with transformations on a nested relational data model (NRDM) from within the graphical user interface mitigating the need for hand coding. Hierarchical data can be combined with relational data or flat files in the same data flow.

Data Integrator offers comprehensive support for web services and allows you to publish any batch or real-time ETL jobs as a web service called from another application. Data Integrator can also call web service-enabled applications to easily access virtually any data.

### Comprehensive Platform, Source, and Target Support

The Data Integrator server can run on Windows, Red Hat Linux, 64 -Bit Itanium, IBM, HP, and Sun UNIX platforms. And with an extensive list of native database, application, mainframe, and technology interfaces, it is truly an enterprise-class data integration platform upon which you can standardize.

<p><b>Database Interfaces</b></p> <ul style="list-style-type: none"> <li>• Oracle</li> <li>• DB2</li> <li>• Sybase</li> <li>• SQL Server</li> <li>• Informix</li> <li>• Teradata</li> <li>• ODBC</li> </ul>	<p><b>Mainframe Bulk Interfaces</b></p> <ul style="list-style-type: none"> <li>• ADABAS</li> <li>• DataCom</li> <li>• DB2/UDB</li> <li>• IDMS</li> <li>• IMS</li> <li>• Sequential Files</li> <li>• VSAM</li> </ul>	<p><b>Technology Interfaces</b></p> <ul style="list-style-type: none"> <li>• IBM MQ Series</li> <li>• HTTP/HTTPS</li> <li>• JMS</li> <li>• XML</li> <li>• CWM</li> <li>• SOAP</li> <li>• Firstlogic</li> <li>• Trillium</li> <li>• Web Services</li> <li>• COBOL Copybooks</li> <li>• SNMP</li> <li>• WSDL</li> <li>• Crystal Reports</li> </ul>
<p><b>Application Interfaces</b></p> <ul style="list-style-type: none"> <li>• SAP R/3 and BW</li> <li>• J. D. Edwards</li> <li>• Oracle Applications</li> <li>• PeopleSoft</li> <li>• Siebel</li> </ul>	<p><b>Mainframe Live Interfaces</b></p> <ul style="list-style-type: none"> <li>• DB2/UDB</li> <li>• IDMS</li> <li>• IMS</li> <li>• VSAM C ICS</li> </ul>	

*Data Integrator provides broad, standards-based source and target support.*

### Deep Integration with ERP and CRM Applications

Data Integrator significantly reduces the amount of time and expertise needed to access, integrate, and deliver data from ERP and CRM applications. You can import live ERP descriptive metadata from packaged applications from SAP, Oracle, and Siebel and then capture that data and use it to design integration projects offline. ERP metadata is presented in a business process-related way and built-in search capabilities make it faster to extract from ERP systems. Within Data Integrator Designer, ERP data appears in the same table and column format as any other source, and native application interfaces deal directly with data dictionaries so you can concentrate on building ETL jobs and not becoming experts in the ERP system itself. And Data Integrator can automatically capture customizations made to ERP and CRM systems when metadata is imported. Alternatively, you can import single tables, functions, or other objects on an individual basis. Data Integrator also has built-in utilities to easily determine changes that have been made to data sources.

<p><b>SAP R/3</b></p> <ul style="list-style-type: none"> <li>• ABAP</li> <li>• BAPI</li> <li>• IDOC</li> <li>• Hierarchies</li> <li>• Functions</li> <li>• Pool and cluster tables</li> <li>• Z-tables</li> </ul>	<p><b>Oracle Applications</b></p> <ul style="list-style-type: none"> <li>• Open interface</li> <li>• FlexFields</li> <li>• Advanced queuing</li> </ul> <p><b>Siebel</b></p> <ul style="list-style-type: none"> <li>• Table and column description</li> <li>• Business Component</li> </ul>	<p><b>PeopleSoft</b></p> <ul style="list-style-type: none"> <li>• Effective dates</li> <li>• Domains</li> <li>• Trees</li> <li>• Panels</li> </ul> <p><b>J. D. Edwards</b></p> <ul style="list-style-type: none"> <li>• Numeric formats</li> <li>• Date and time formats</li> <li>• Module objects</li> </ul>
---	--	---

*Data Integrator provides industry- leading access to packaged applications.*

## Metadata Management

Each Data Integrator repository is stored on an existing relational database and must be associated with one or more instances of a Data Integrator data server. There are two types of repositories:

- 1) A local repository used by an application designer to store definitions of the Data Integrator objects (projects, jobs, work flows, and data flows) and source/target metadata.
- 2) A central repository (optional) that can be used to support multi-user development. The central repository provides a shared object library allowing developers to check objects in and out of their local repositories.

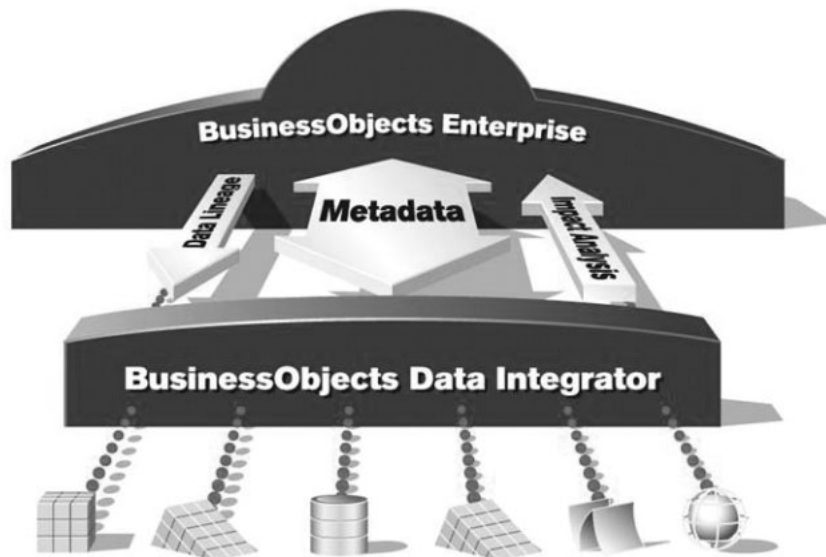
Data Integrator captures all metadata about the ETL process from source-to-target BI reports and analytics. This includes source-to-target transforms, propagation of business descriptions, operational, and mapping data in which BusinessObjects Web Intelligence or Crystal Reports uses the source data as well as the tool metadata. Data Integrator automatically manages and captures metadata within the metadata repository and notes and descriptions can be added to objects.

Data Integrator provides web-based metadata reports that provide impact analysis, lineage, where-used, execution statistics, and historical execution analysis reports that are customizable to specific customer requirements. Metadata is available externally to Data Integrator via an export utility to products like ERwin or via CWMI. All Data Integrator metadata is held within a relational database of the customer's choice. The metadata is open and the metadata repository schema is published. Data Integrator provides open support for the following standards and technologies: CWM, ERwin XML, XMI, and XML schema. You can also query metadata through straight SQL access against the source relational tables.

---

Data Integrator is also unique in its ability to share metadata up and down the Business Objects BI platform providing customers with a single-vendor, end-to-end solution. With this level of metadata integration, the combination of Data Integrator and the Business Objects BI platform provides you with :

- ▶ Source-to-target report impact analysis across all ETL processes
- ▶ Report-to-source data lineage
- ▶ Source-to-dashboard, real-time integration
- ▶ Source-to-target report auditability
- ▶ The ability to automatically create the semantic reporting layer for both Crystal Reports and Web Intelligence documents
- ▶ ETL auditing reports on execution statistics, trace and error metadata that are available within the BI platform



*Metadata integration with the BusinessObjects Enterprise platform provides end-to-end data lineage and impact analysis.*

---

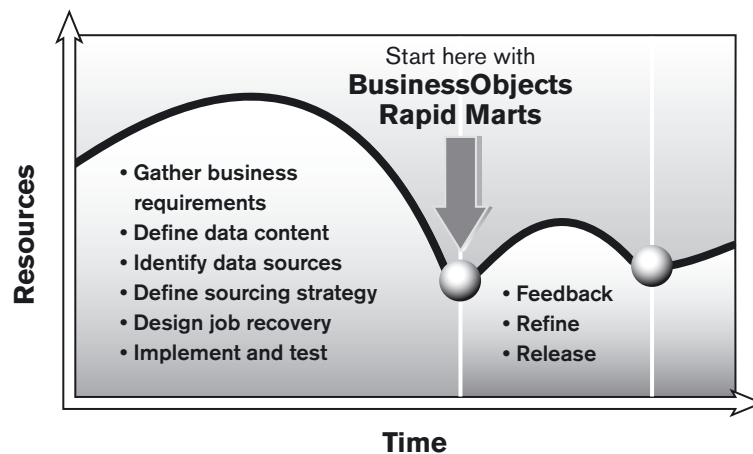
## BusinessObjects Rapid Marts

---

Accelerate your BI deployments with BusinessObjects Rapid Marts, a packaged data integration solution for ERP and CRM applications.

A typical ERP or CRM data warehousing initiative requires an organization to go through the following steps:

- ▶ Interview the functional leaders of the company to determine metrics and analytic requirements
- ▶ Determine the structure of the data mart or warehouse data model needed to capture required analytical data
- ▶ Create the target database
- ▶ Look to the ERP/CRM system and find the best source tables and columns to pull from to get the required data
- ▶ Do the source-to-target data mappings
- ▶ Write custom code or use an ETL tool to execute the source-to-target mappings to move the data
- ▶ Write programs to do the change data capture or delta extract jobs, error recovery, etc.



*Required resources and time for building a data mart or data warehouse.*

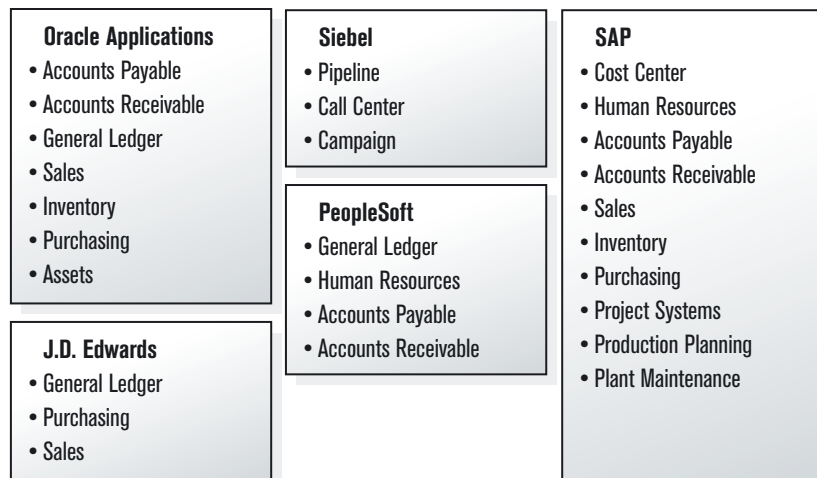
---

BusinessObjects Rapid Marts jumpstart this process by providing packaged data integration for enterprise applications such as SAP, PeopleSoft, Oracle, J.D. Edwards (PeopleSoft), and Siebel. They combine domain knowledge with ETL best practices to jumpstart BI and performance management initiatives, helping to accelerate your overall time to value. Rapid Marts speed deployment and lower costs by delivering:

- ▶ Prebuilt universe and reports
- ▶ Definition of data-content requirements
- ▶ Data flows to support standard business processes
- ▶ Data schema down to individual data lines and items
- ▶ Identification of source-to-target mappings and transformations
- ▶ Identification of specific data sources and sourcing strategy
- ▶ Design of change-data-capture processes
- ▶ Basis for quick implementation and testing

Rapid Marts even provide sample Business Objects universes and reports that illustrate the depth of analytical content that Data Integrator is able to deliver.

And there's more. Many organizations want to begin with source-specific operational reporting and query and analysis. You can implement Rapid Marts as phase one of a broader performance management application initiative. Using Rapid Marts as the foundation, you can feed the data model used by the Business Objects performance management applications for analytics that span across the enterprise.



*Rapid Marts deliver prebuilt data flows, transformation logic, and data models to accelerate BI projects.*

---

## Conclusion

---

Research shows that up to 80% of a BI project is spent building and maintaining the data infrastructure. Leading analyst firms have recognized Business Objects data integration as enterprise-class technology and our strategy is to continue to provide the best-of-breed platform that will satisfy the needs of today's most complex IT infrastructures. Proven real-time and batch data movement capabilities, built-in data validation and auditing, and a single, easy-to-use design environment make Data Integrator the right choice for ensuring maximum developer productivity. And a parallel architecture ensures optimized ETL performance and throughput. But Data Integrator really delivers extreme value to your organization through deep integration with the trusted BusinessObjects XI platform. Some of the benefits that our customers are seeing from unifying BI and enterprise-class data integration include: shortened implementation times, increased IT and end-user productivity, reduced deployment and system maintenance costs, and more timely and accurate data that end users can trust.

To find out more about Business Objects data integration, visit our website.

<http://www.businessobjects.com/products/dataintegration/>

---

## Appendix – Data Integrator Technical Specifications

### Designer Client

- ▶ Windows

### Data Integrator Server

- ▶ AIX
- ▶ HP-UX
- ▶ Solaris
- ▶ Windows
- ▶ 64 bit Itanium HP-IPF
- ▶ Linux, Red Hat

### Database Interfaces

- ▶ Flat file
- ▶ Generic ODBC
- ▶ IBM DB2/UDB
- ▶ Informix IDS
- ▶ Microsoft SQL Server
- ▶ NCR Teradata
- ▶ Oracle
- ▶ Sybase Adaptive Server Enterprise
- ▶ Bulk loading supported for above databases

### Application Interfaces

- ▶ JD Edwards OneWorld and World
- ▶ Oracle Applications; PeopleTools
- ▶ SAP BW Server; SAP R/3 via ABAP, BAPI, and IDOC
- ▶ Siebel

### Mainframe Bulk Interfaces

- ▶ ADABAS
- ▶ Datacom
- ▶ DB2/UDB
- ▶ IDMS
- ▶ IMS
- ▶ Sequential files
- ▶ VSAM (CICS and non-CICS)

### Mainframe Live Interfaces

- ▶ DB2/UDB
- ▶ IDMS
- ▶ IMS
- ▶ VSAM CICS

### Technology Interfaces

- ▶ Crystal Reports
- ▶ CWM
- ▶ Firstlogic
- ▶ HTTP/HTTPS
- ▶ IBM MQ Series
- ▶ JMS
- ▶ SNMP
- ▶ SOAP
- ▶ Trillium
- ▶ WSDL
- ▶ Web Services
- ▶ XML Schema

(The Adapter SDK is also available for unlisted products).

### Business Intelligence Tools

- ▶ BusinessObjects XI
- ▶ BusinessObjects 6.5
- ▶ Crystal Reports XI
- ▶ Crystal Enterprise 10
- ▶ Other BI tools through CWM

### Changed Data Capture (CDC)

- ▶ Microsoft SQL Server
- ▶ Oracle
- ▶ IBM DB2
- ▶ Mainframes

**Americas**

Business Objects Americas  
3030 Orchard Parkway  
San Jose, California 95134  
USA  
Tel: +1 408 953 6000  
+1 800 877 2340

**Asia-Pacific**

Business Objects Asia Pacific Pte Ltd  
350 Orchard Road  
#20-04/06 Shaw House  
238868  
Singapore  
Tel: +65 6887 4228

**Europe, Middle East, Africa**

Business Objects SA  
157-159 rue Anatole France  
92309 Levallois-Perret Cedex  
France  
Tel: +33 1 41 25 21 21

**Japan**

Business Objects Japan K.K.  
Head Office  
Yebisu Garden Place Tower 28F  
4-20-3 Ebisu, Shibuya-ku  
Tokyo 150-6028  
Tel: +81 3 5447 3900

For a complete listing of our sales offices, please visit our web site.

► [www.businessobjects.com](http://www.businessobjects.com)