

From Operational to Analytical Reporting with BusinessObjects Data Integrator

Definitions

A **data warehouse** is a database dedicated for decision support. It collects and integrates data from disparate systems across departments and delivers a single source of truth.

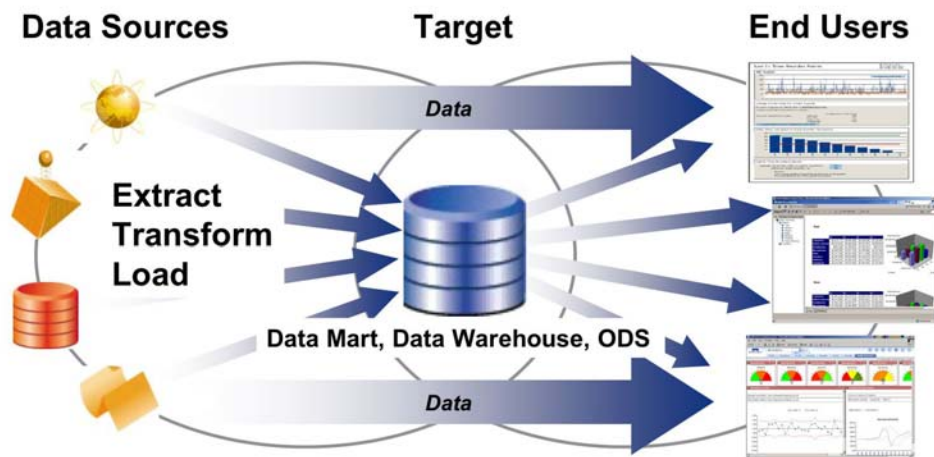
A **data mart** collects data on a specific subject area such as sales or customer information. It can be a subset of data warehouse data and it is intended to meet the needs of a single department.

An **operational data store (ODS)** is an updatable set of integrated data used for enterprise-wide tactical decision making. It contains live data, not snapshots, and has minimal history retained.

Helping to Meet the Needs of All BI Users

Reporting directly from operational systems is natural and common. However, as data volumes, data sources, and users of business intelligence (BI) tools grow, organizations must determine how to unify disparate departmental initiatives and ensure that these users are working with information that is consistent, accurate, and relevant. BI tools help turn operational data into information that enables better decision making and improved visibility. But this is only possible if users are working with the right tools, and more importantly, the *right data*.

As BI end-user populations grow, analytical requirements typically do as well. Users begin to want to interact with information, analyze trends, and potentially perform their own ad hoc queries. For this to be possible, the right infrastructure must be established. Whether it is subject-area data marts, a centralized data warehouse, or an operational data store, organizations need to determine the right strategy that will support operational reporting as well as ad hoc query and analysis, and performance management requirements across the enterprise.



End-user BI tools can deliver data directly from operational systems such as legacy apps, mainframes, databases, enterprise applications, web applications, and XML files, but most organizations develop a target system that serves the sole purpose decision support.

This Information Sheet will summarize why and when integrating data is the right approach for an enterprise information infrastructure. It will also outline what is required in a data integration platform to help organizations build and maintain this environment and summarize the key aspects of BusinessObjects Data Integrator that make it a truly enterprise-class solution.

From Operational to Analytical Reporting with Data Integrator

Definitions

Extract, transform, and load (ETL) tools play a critical part in creating data warehouses, which form the bedrock of BI. ETL tools sit at the intersection of myriad source and target systems and act as a funnel to pull together and blend heterogeneous data into a consistent format and meaning and populate data warehouses.*

Integrating Data

So why do organizations develop a target system such as a data warehouse or data mart? Here are a few of the key reasons:

- ▶ For organizations that have inconsistent data in different systems a data warehouse strategy can help clean up these inconsistencies by providing a single source of truth.
- ▶ Hitting your operational systems with query requests or reporting jobs can drain its resources from doing the job for which it was intended. Building a data mart or warehouse alleviates performance hits on your transactional systems while improving query and report performance for decision support.
- ▶ Although end-user tools and applications can join data from different sources, it is often a complex job requiring time and performance.
- ▶ Bad data leads to bad decisions. A data warehouse approach provides the opportunity for organizations to analyze and correct bad data such as name and addresses.
- ▶ Most operational systems such as a CRM system provide only customer transactions and inquiries that last up to one to two years. Old data is archived. If you want to analyze customer behavior over a long period of time, you will need a system that can support large amounts of data.
- ▶ A data warehouse provides the infrastructure for better security protocols.
- ▶ Operational systems are frequently changing. Running reports at different times can result in different results. With BI users accessing integrated data, they will know when it has been refreshed and they will have access to historical information.
- ▶ A data warehouse allows organizations to integrate data from outside sources so it can be combined, rationalized, cleansed, and aggregated in a single environment that can provide a comprehensive organizational view.

How to Integrate Data

While a data warehouse provides tangible benefits, developing and maintaining one is not easy. Analyst firms estimate that up to 80% of the effort of building and maintaining a BI solution is in managing the data extraction, transformation, and loading (ETL) process. Failing to properly manage this difficult aspect of a BI deployment is one of the primary reasons some data warehousing projects fail. ETL processes bring together and combine data from multiple source systems into a data warehouse, enabling all BI users to work off a single, integrated set of data—a single version of the truth. The result is an organization that no longer spins its wheels collecting data or arguing about whose data is correct, but one that uses information as a key process enabler and competitive weapon.

The ETL process includes designing a target, mapping sources to target, extracting data from sources, transforming data for the target, scheduling and monitoring processes, and managing the overall BI environment.

- **Extraction** is the physical collection of data. This step begins by designing and creating the target data schema. Then the appropriate data sources must be identified and its format, reliability, and value must be determined.

* Source: TDWI Report Series: Evaluating ETL and Data Integration Platforms, Wayne Eckerson & Colin White, 2003.

From Operational to Analytical Reporting with Data Integrator

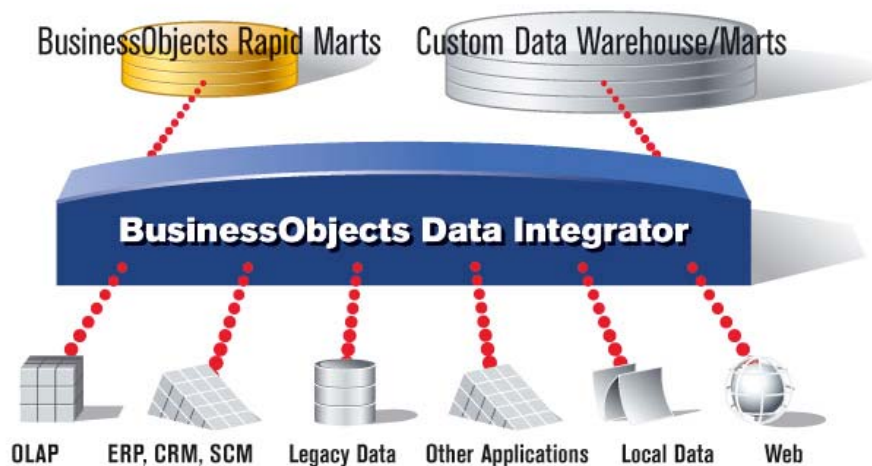
Organizations on average have no less than eight operational systems and accessing them requires an understanding of the business model and meaning of data elements and semantics.

- **Transformation** involves turning raw data into more business-descriptive content that is appropriate for end user BI. It includes a strategy for changed data capture or other warehouse refreshing. It can also include cleansing: eliminating duplicates, correcting wrong values, replacing missing values, and improving overall data quality. Some transformation examples are slowly changing dimensions, table comparison, hierarchy flattening, surrogate key generation, heterogeneous joins, user-defined functions, data-type conversion, and user-defined functions.
- **Loading** is the process of moving the transformed data into the target relational database. It includes administration steps of scheduling and monitoring of jobs that are being processed in shorter and shorter batch windows and in some cases real-time. Data can also be loaded into a staging area or operational data store.

The right ETL tool can speed BI deployments, minimize the impact of systems changes and new user requirements, and mitigate overall project risk.

BusinessObjects Data Integrator

Business Objects provides best-of-breed ETL technology that is integrated with marketing leading end-user BI tools and analytic applications, helping to deliver greater IT and end-user productivity, faster time to results, a lower total cost of ownership (TCO), single vendor accountability, and reduced project risk.



BusinessObjects Data Integrator enables organizations to access, integrate, and deliver data across the enterprise. Data Integrator allows organizations to develop and maintain 98% of their complex ETL routines without writing or maintaining custom code. It means BI can be deployed faster and ongoing maintenance challenges of growing data volumes and continuous change can be managed in an easy to use and administer environment.

BusinessObjects Data Integrator Overview

Data Integrator can help:

- ▶ Accelerate your BI deployment time and reduce project costs
- ▶ Increase IT productivity and reduce your total cost of ownership
- ▶ Ensure your BI users are always working with credible, meaningful information
- ▶ Optimize your return on investment from enterprise applications

Data Integrator is capable of large-scale data movement in both batch and real-time. *Powerful* enterprise-class features include:

- One data integration environment for both batch (volume) and real-time (high frequency) data movement
- Parallel performance architecture to break large ETL jobs into smaller pieces that can run in parallel, leveraging all of your computing resources, and meeting ever shrinking batch windows
- Grid computing for multi-server load balancing, scalability, and fault tolerance
- Broad source and target support including native access to all major databases, flat files, XML, ODBC, and mainframe sources
- Native integration with real-time environments through XML
- Metadata level integration to enterprise applications like SAP, PeopleSoft, Oracle, J.D. Edwards, Siebel, and prebuilt data marts to accelerate BI projects on these systems
- Data cleansing via seamless, in memory, integration with First Logic and a native adapter to Trillium

Data Integrator provides a *productive* single, easy-to-use interface for designing, testing, debugging, running, and monitoring ETL jobs as well as a codeless, object oriented development environment. Key features include:

- Reuse with inheritance
- Automatic documentation as metadata
- Complete suite of powerful transformation objects and functions
- Collaborative development environment with multi-user check in / check out and version control
- Built in data preview and profiling
- Graphical workflow for scheduling jobs including external processes
- Interactive debugging and error trapping

While able to standalone as a best-of-breed data integration development platform, Data Integrator is also tightly integrated with the Business Objects information infrastructure, helping organizations to better manage and maintain their *complete*, end-to-end BI environment. Key features include:

- Ability to exchange metadata via CWM interface with over 65 other tools
- Automatic documentation of all development work in repository
- End-to-end impact analysis reports showing change impact on everything from sources to reports
- Automated change propagation: generation & incremental updates of Business Objects universes
- Data lineage from reports showing source systems, when updated, how metrics were calculated, filters applied, etc.

▶ www.businessobjects.com

Americas

Tel: +1 800 877 2340

Europe

Tel: +33 1 41 25 21 21

Asia-Pacific

Tel: +65 887 4228

Japan

Tel: +81 3 5720 3570

The Business Objects product and technology are protected by US patent numbers 5,555,403, 6,247,008, 6,578,027, 6,490,593, and 6,289,352. The Business Objects logo, WebIntelligence, BusinessQuery, the Business Objects tagline, BusinessObjects, BusinessObjects Broadcast Agent, Rapid Mart, Set Analyzer, Crystal Analysis, Crystal Enterprise, Crystal Reports, Crystal Applications, Crystal Analysis Holos, and Crystal Info are trademarks or registered trademarks of Business Objects SA in the United States and/or other countries. Various product and service names referenced herein may be trademarks of Business Objects SA. All other company, product, or brand names mentioned herein, may be the trademarks of their respective owners. Specifications subject to change without notice. Not responsible for errors or omissions. Copyright © 2004 Business Objects SA. All rights reserved.