

SAP Solution Brief

SAP Solutions for Enterprise Information Management

SAP Data Integrator

SAP® Data Integrator: Delivering Trusted Data for Your Enterprise



Quick Facts

Summary

SAP® Data Integrator software enables organizations to gain deeper insight with a single view by accessing and integrating structured and unstructured data from data sources across the enterprise. By making it easier to profile, extract, transform, and deliver data, SAP software provides your organization with trusted, integrated data to support the information needs of your enterprise.

Business Challenges

- Overcome data integration challenges such as information silos, data inconsistencies, and the enormous volume of data being generated
- Tap into unstructured data, such as e-mail, Web logs, social media, blogs, customer surveys, and other text content

Key Features

- **Comprehensive data integration functionality** – Access and integrate from any structured or unstructured data source, and design an efficient, reliable data integration process
- **Extract, transform, and load (ETL) functionality** – Move and integrate data in near-real time or at any interval
- **Push-down extract, load, transform (E-LT) functionality** – Take advantage of the power of the database, since SAP Data Integrator pushes down processing to the database layer whenever possible, including into SAP HANA® appliance software, to leverage in-memory transformation processing and performance
- **“Big Data” support** – Move and load data to and from Apache Hadoop data store

Business Benefits

- **Compliance, governance, and real-time decision making** with a single, trusted view of your data and faster access to all sources
- **Rapid deployment with prebuilt integration** with other key SAP software and components
- **Flexibility** with a wide variety of packaged adapters for databases, data warehouses, and data stores

For More Information

Call your SAP representative or visit us online at www.sap.com/eim.



The Best-Run Businesses Run SAP™

As data sources and volumes proliferate, organizations that manage and harness their data hold a strategic, competitive advantage. To gain control over your data, you need to build an agile and trusted data foundation that can rapidly respond to the information demands of operational or analytical processes and systems. Such a foundation requires data integration capabilities that can help you [deliver trusted information across the enterprise](#).

SAP® Data Integrator software allows organizations to extract, transform, and deliver any type of data anywhere, at any frequency. The functionality delivered in SAP Data Integrator is a core component of SAP Data Services software – the first single offering for data integration, data quality management, data profiling, meta-data management, and text analytics.

With SAP Data Integrator, you can:

- Integrate data from a variety of data sources for a complete view of enterprise information
- Deliver timely, integrated, and trusted data across the enterprise
- Deploy with faster time to market
- Access a variety of data sources, both structured and unstructured
- Deliver enterprise-class data integration performance

A COMPLETE VIEW OF ENTERPRISE INFORMATION

As volumes of data rise and acceptance of delays drops, you need to be able to rely on a high-performance, scalable data integration solution. SAP Data Integrator delivers extreme scalability

with support for the Apache Hadoop data store, near-real-time data movement, parallel processing, grid computing, and broad source and target support. This open and services-based architecture allows you to integrate with third-party products using industry-standard protocols such as CWM, XML, HTTP/HTTPS, JMS, SNMP, and Web services.

Currently, SAP Data Integrator is the only data integration technology that is certified and optimized to load data into SAP HANA® appliance software. SAP Data Integrator delivers a bulk loading feature that enables you to replicate millions of records quickly into SAP HANA by setting up the connection between target and source, selecting the required tables, and executing the move of data with a single mouse-click. In addition, whenever possible, SAP Data Services will push down processing to the database layer via extract, load, transform (E-LT) functionality to extract and load data quickly into your target database system. With SAP HANA, SAP Data Integrator takes advantage of in-memory transformation processing for even greater performance. SAP Data Integrator

has also been optimized to natively read, write, and process data in SAP Sybase® Adaptive Server® Enterprise and SAP Sybase IQ server.

To support “Big Data,” SAP Data Integrator supports efficient read, write, and processing capabilities with Hadoop. Developers can submit queries against a Hive data warehouse to search for and filter data in Hadoop before it is extracted and loaded into SAP HANA or SAP Sybase IQ for deeper analysis. Developers can also submit text data-processing operations as MapReduce functions to provide structure to unstructured data in Hadoop.

SAP Data Integrator maximizes data integration performance by intelligently pushing the processing down to the database or mainframe platforms (otherwise known as E-LT). For example, the software will leverage your database server to aggregate functions at the source level without moving the data to the SAP Data Integrator server. With the ability to distribute processing to your source systems, the software minimizes unnecessary data movement across your network to optimize data integration performance.

With its text data-processing feature, SAP Data Integrator not only extracts content from unstructured data sources; it also understands the context of those data elements. The software understands that “Abraham Lincoln” could be a person’s name, and also understands that the name refers to a U.S. president. The ability to understand the phrase and apply correct context to the data element is critical to accurate downstream analytical activities. Additionally, SAP Data Integrator can

also interpret the sentiment represented by key phrases found within unstructured data sources – for example, whether there are strong positive or negative statements within the sentence or usage.

SAP Data Integrator can help you connect to your data whether it resides in a relational database, mainframe system, or enterprise application. No other product provides deeper metadata-based connectivity to packaged applications from SAP, Siebel, PeopleSoft, J.D. Edwards, and Oracle applications. Native connectivity is available for virtually all database types or through Open Database Connectivity (ODBC), and you can access legacy mainframe applications. SAP Data Integrator also supports flat files, XML, and Web services. If your application is proprietary, you can connect to it using a Java software development kit (SDK).

PROVIDE TRUSTED INFORMATION

As the amount and variety of information continues to explode, organizations are faced with new challenges for gaining a complete view of structured and unstructured data across their enterprise. SAP Data Integrator helps ensure the quality of your enterprise data with advanced data profiling, which allows you to understand the content, structure,

quality, and relationships between tables in your source data. The data validation feature helps you build a firewall between your source and target systems, so you can filter out unwanted data based on your business rules.

During the ETL process, SAP Data Integrator audits data to verify the integrity of the data; additional data quality functionality to cleanse data of any operational domain (such as customer, product, or service) is accessible.

By deeply integrating the ETL process with the BI product used in your data warehousing project, both IT and business users gain measurable benefits such as easy metadata management, simplified and unified administration, improved lifecycle management, and reduced maintenance costs. The software enables you to automatically create and update a semantic layer in SAP BusinessObjects BI solutions to accelerate BI deployment and simplify change management.

From an IT perspective, SAP Data Integrator enables you to see how a change in source data can affect downstream processes and systems through metadata analysis. The software is unique in its ability to offer smooth integration with metadata from SAP BusinessObjects™ business intelligence (BI) solutions. This integration provides end-to-end impact

analysis that allows you to see the effect of changes in the source systems down to the BI reports. As a result, you can more easily manage change in your entire ETL and BI environment.

From a business perspective, a huge benefit of metadata integration is data lineage, which allows information consumers to view the context of data in their reports. The software allows users to see when data was updated, how it was computed, and where it came from – all the way back to the original transactional source.

DEPLOY FASTER AND EASIER THAN EVER BEFORE

As demands on IT and the business users continue to increase, SAP Data Integrator helps by enabling you to deploy the software faster and more easily than ever before. An independent analyst study on ETL tools even rates SAP Data Integrator as number one in ease of use.¹ With SAP Data Integrator, you can significantly simplify and reduce your IT efforts by providing an intuitive development environment to handle all tasks associated with building a data integration process.

SAP Data Integrator helps accelerate and simplify development efforts by building and managing data integration jobs within a single graphical environment.

SAP Data Integrator delivers extreme performance with [support for “Big Data,”](#) real-time data movement, and broad source and target support.



50 093 604 (12/06) ©2012 SAP AG. All rights reserved.

SAP, R/3, SAP NetWeaver, Duet, PartnerEdge, ByDesign, SAP BusinessObjects Explorer, StreamWork, SAP HANA, and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP AG in Germany and other countries.

Business Objects and the Business Objects logo, BusinessObjects, Crystal Reports, Crystal Decisions, Web Intelligence, Xcelsius, and other Business Objects products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of Business Objects Software Ltd. Business Objects is an SAP company.

Sybase and Adaptive Server, iAnywhere, Sybase 365, SQL Anywhere, and other Sybase products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of Sybase Inc. Sybase is an SAP company.

Crossgate, m@gic EDDY, B2B 360°, and B2B 360° Services are registered trademarks of Crossgate AG in Germany and other countries. Crossgate is an SAP company.

All other product and service names mentioned are the trademarks of their respective companies. Data contained in this document serves informational purposes only. National product specifications may vary.

These materials are subject to change without notice. These materials are provided by SAP AG and its affiliated companies ("SAP Group") for informational purposes only, without representation or warranty of any kind, and SAP Group shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP Group products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.



The drag-and-drop user interface allows you to quickly build jobs that profile, validate, audit, cleanse, transform, and move data. SAP Data Integrator offers team-based development functionality. Users can check work in and out of a central metadata repository securely. Your users can share and version their work to accelerate development, and you can easily compare differences between objects. Advanced debugging features let you analyze problematic data through the entire process. The smart documentation functionality saves you from having to manually document your work.

The software lets you perform a complete range of data transformations. Choose from a library of powerful, extensible, and reusable transforms for operations, such as:

- Unstructured data
- Hierarchy flattening for XML files and pipelining

- Pivot and reverse-pivot of rows and columns
- Slowly changing dimensions
- Change data capture
- Data validation
- Data cleansing (via SAP Data Services software)

Implement Faster with Packaged and Rapid-Deployment Solutions

SAP Rapid Marts™ packages are packaged, powerful, and flexible data integration solutions for delivering prebuilt data marts from enterprise applications. SAP Rapid Marts packages provide domain knowledge with data integration best practices to deliver an application-specific data mart, a semantic layer (also known as a "universe"), and BI reports. SAP Rapid Marts packages are available in key domain areas: financials, human resources, order management, and purchasing.

SAP Rapid Deployment solutions take things one step further. These offerings include preconfigured software and implementation services – all up and running in just weeks. With a defined scope and predictable costs, there are no surprises. And you can expand the solution as you go. A variety of rapid-deployment solutions are available that include SAP Data Integrator, including offerings for data migration.

To learn more about SAP Data Integrator and SAP solutions for enterprise information management, visit our Web site at www.sap.com/eim.

FOOTNOTES

1. "Integrating Data in the Enterprise." The ETL product survey 2012 by Passionned International, February 2012.

