

Solutions
Review

Data Integration Buyer's Guide



Includes a Category Overview
The Top 10 Questions to Ask
Plus a Capabilities Reference of
the Leading 28 Providers for
Enterprise Data Integration



INTRODUCTION

Data Integration tools are perhaps the most vital components to take advantage of Big Data. Enterprise organizations increasingly view Data Integration solutions as must-haves for assistance with data delivery, data quality, Master Data Management (MDM), data governance, and Business Intelligence and Data Analytics. With data volumes on the rise and with no real end in sight, businesses are leaning on integration tools more and more to meet all of the data consumption requirements for vital business applications. The main function of Data Integration is to give organizations the ability to gain consistent access to their most important data, no matter where it lies, whether virtually or physically, whether on-premise, in the cloud, or in some other disparate location.

The migration, organization, and delivery of key organizational data assets is done in such a way that allow business teams to easily pull what they need for use within other business systems. The most comprehensive Data Integration tools available to companies today will include data quality, data governance and MDM capabilities, providing security that whatever data sources that are plugged in deliver only the most relevant data for analysis. Integration tools also give companies the ability to ensure data consistency across applications.

Accessing data doesn't just mean having a unified view of it all, however. For practical purposes of crunching all the data, it needs to be in one place where an analytics program can reach it. That involves moving data from one place to another, usually from storage systems into a data warehouse capable of analyzing it. Methods for doing that include ETL (for Extract, Transform and Load) and Data Replication, the latter of which, while often used for tasks like Disaster Recovery and migration, in relation to Big Data offers a high-performance data movement tool that should be able to quickly synchronize large quantities of data.

TRENDS

In recent years, open source Big Data frameworks such as Hadoop, Spark, and NoSQL have really emerged to give organizations more of a choice in where they store data and run applications on clusters of hardware, giving them additional storage options for all kinds of structured and unstructured data and enough processing power to handle virtually an infinite number of simultaneous tasks. In addition, data lakes have gained notoriety for their ability to continuously collect data and store it in a lightly structured repository. This is a positive development because it can help deliver data to stakeholders, business processes and applications with swiftness and ease.

Hadoop and friends have been able to thrive due to the demand that real-time and streaming data has put on organizations. No longer is it enough to store data in traditional relational databases, especially considering the pure volumes of data that now come from embedded sensors, computer applications, social media activity, and mobile devices. As more of these types of data sources become prevalent, open source storage frameworks should continue to grow in popularity, and with the Internet of Things expected to revolutionize virtually every consumer product with the addition of data capturing, this looks like a trend that will stick around for the long haul, making integration solutions that much more vital. The main consideration for buyers of integration software are features and functions of course, but Data Integration is not so much a product as it is a process, and there a variety of ways to move the needle.

Tim King
Editor
Solutions Review

“In this Buyer's Guide there's a solution for everyone, from small groups to multi-national organizations.”



5 Questions You Should Ask Yourself Before Selecting a Data Integration Solution

QUESTION #1 Why is a Data Integration tool necessary to my organization?

What your company needs a program to do will determine many of the technical requirements of the Data Integration solution that best suits you. Will you require real-time data access and transfer? How much data will you need to move and how quickly? Can you afford some downtime on source/target systems, or do you need them running at all times? Note that all these data requirements are based on your technical and business needs so that you can compare them with what specific vendors have to offer.

QUESTION #2 What kind of data do I need to analyze?

Is the majority of your company's data transactional in nature? Is it all structured? If so, a traditional or "legacy" tool may be right for you. If the bulk of your data streams to your storage framework in real-time via embedded product sensors, social media, computer applications and online customer feedback, then a solution that can integrate with the likes of Hadoop, Spark, and NoSQL is going to be the best fit for you. Be sure to take into account the types of data that run through your business and then match that up with the appropriate provider.

QUESTION #3 What are my data sources and where are they located?

The basic elements of Data Integration revolve around moving data from sources (applications) to targets (data warehouses, Hadoop, etc.). Much of what is powering the Big Data movement is the massive data being collected in the cloud through very large Software as a Service (SaaS) solutions like Salesforce.com. Some solutions listed in this guide specialize in the integration of cloud application data with on-premise systems to ensure that your users can access complete, current, and accurate data. Make sure the kind of integration you require is being offered by the vendors you are considering.

QUESTION #4 Cloud, on-premise, or both?

A hybrid approach to Data Integration is a growing trend in the enterprise market as it provides organizations the ability to execute integration in both on-premise and cloud environments. Thus, organizations are able to interchange data to and from either framework as a way to gain business agility to their integration infrastructure and manage cloud data delivery and address the need for data sharing between cloud environments. On-premise integration is certainly not dead, but a hybrid approach will set your organization up nicely for the future, even if cloud exposure is currently limited.

QUESTION #5 Are data quality and Master Data Management feature considerations?

It's one thing to use an ETL tool to move data from one storage medium to another. It's an even different thing to replicate a data source for use elsewhere. Most importantly, though, you'll want to make sure that the appropriate data gets moved to the right place so it can be analyzed in a fashion that will yield actionable insights. As mentioned earlier, the most comprehensive Big Data solutions today include data quality and MDM capabilities in one form or another. With MDM as the umbrella, data governance and management tools are going to be essential to you if you capture data from a wide variety of sources.



And 5 Questions You Should Ask Your Potential Data Integration Solution Provider

QUESTION #6 Is your tool compatible with my existing infrastructure?

Data Integration is of central importance to the data infrastructure of any modern digital business. Organizations want a solution that can be flexible in order to fit into an existing company framework. Synergy between any integration tool and the applications, business processes, and data sources a business already has in place should be the first topic addressed, and should act as qualifying question number-one.

QUESTION #7 Is flexibility a strength of your integration solution?

Enterprise organizations are increasingly interested in improving the flexibility of their data infrastructures. Thus, stakeholders are becoming aware of the kinds of integration activities that are being enacted upon vital business information. As a result, managers now understand how important Data Integration is to their company's overall data management strategy. With data democratization all the rage, decision-makers are needing to deploy integration tools that enable cross-department sharing. Fortunately for customers, vendors are taking notice of a need for flexibility, and the providers that embrace it and provide the most comprehensive tools will flourish in the years ahead.

QUESTION #8 Do you integrate with the Hadoop, Spark, and other NoSQL repositories?

Chances are that if you are on the hunt for a Data Integration solution, you capture at least a moderate amount of data for analysis. The chances are also high that you use one of these storage mediums to collect, organize, and qualify data for further use. Be sure to see if potential vendors integrate with these open-source Big Data frameworks, and although most do, find out to what degree each does. There are many integration vendors out there, and even more worthy solutions. However, some are pure-play, meaning that they support traditional tools such as ETL and replication, though the move to open-source is certainly upon us.

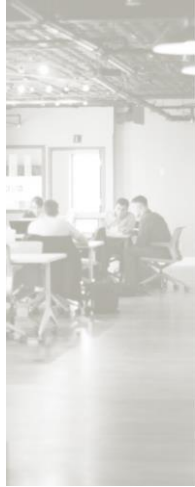
QUESTION #9 Does your solution come packaged with any support services?

Consumers are increasingly expecting vendors to provide more than just a software solution. Given the lack in talent surrounding the complexities of many of the best integration tools out there, stakeholders are asking providers to include more in terms of customer support and services in the form of experts and professionals who can assist with deployment, integration best practices, and issues.

QUESTION #10 What is your main area of focus what does your tool do best?

Not all solutions providers are created equal. Oftentimes, vendors will key in on a niche in the market. By focusing on a specific part of the space, such as cloud integration, for example, solutions providers can become experts at developing a narrower piece of software that although vocational, really works for its users. As a result, the vendor increases efficiency and overall quality and decreases the time it takes to deploy the solution organizationally.

Solution Provider Profiles



Action	7
Adeptia	8
Altova	9
Attivo	10
Attunity	11
CloverETL	12
Dell Boomi	13
HVR Software	14
IBM	15
Informatica	16
Information Builders	17
Jitterbit	18
Liaison Technologies	19
Microsoft	20
MIOsoft	21
Oracle	22
Pentaho	23
Progress Software	24
SAP	25
SAS	26
Software AG	28
Stone Bond Technologies	27
Syncsort	29
Talend	30



Action Corporation

Based in California, Action delivers an end-to-end analytics platform that runs entirely, and natively, in Hadoop software. Action's analytics platform helps to speed up the organizational value chain by connecting with large amounts of raw Big Data and converting between formats. The company's open-source solution also may be used with Spark and Hive, and enables customers to leverage existing investments, extend business ecosystems, and provide protection from vendor lock-in.



From the Company

"Action Corporation enables organizations to transform Big Data into business value with data management solutions to connect, analyze, and take automated action across their business operations. Action's innovative portfolio of solutions for Big Data include cloud integration for frictionless connection to any data; analytical solutions. In January 2013, Action acquired Pervasive Software - a database integration software solution.

Whether designing integration processes for data warehouse loading, converting data between formats, deploying complex application integration scenarios, or integrating applications both on-premises and in the cloud, Pervasive Data Integrator includes numerous foundation features that accelerate the creation of flexible integration solutions. Pervasive Data Integrator includes locally installed as well as cloud-based visual designers and drag-and-drop Link-Style Mapper, RIA Browser-Based User Interface, Lifecycle Management, SOA Platform, Cloud-to-Cloud Computing Interchangeable and Reusable Metadata."

Key Features

Vortex – Provides elastic data preparation for rapid processing through analytic engines, ingestion technology, and KNIME user interfaces. Also provides SQL and predictive analytics to detect trends and patterns with hyper-parallelized operators via Hadoop

Matrix – A parallel, columnar, compressed, compiled approach that increases performance and algorithms, in-database analytics, loading capacity, and accelerates processing power to provide faster delivery of MPP analytics

Vector – Enables concurrent business owners to explore data faster, and can be connected to the SMP analytics database to any application through standard SQL

Bottom Line

Pervasive Data Integrator includes locally installed as well as cloud-based visual designers as well as Drag-and-Drop Link-Style Mapper, RIA Browser-Based User Interface, Lifecycle Management, SOA Platform, Cloud-to-Cloud Computing Interchangeable and Reusable Metadata.



Adeptia

Adeptia is a software company based in Chicago that enables users to connect with their customers and partners fluidly. Adeptia's self-service software allows for digital business connectivity and enables quick response time between clients and company. The company's solution assists IT teams with management and security, and provides a hybrid offering of both cloud and on-premise enterprise integration. Adeptia's software platform is entirely web-based and can be accessed remotely.



From the Company

"Adeptia was founded in June 2000, with the aim of taking business process management and integration to a broader market. Adeptia Extract, Transform & Load (ETL) Suite combines data transport with metadata management and data transformation capability. It uses Service-Oriented Architecture (SOA) and a process-based model to create flexible and loosely-coupled data flows that suit any aggregation or data warehousing solution. ETL Suite offers a middleware application to document, automate and execute critical data processes especially for creating and populating data marts from various internal or external data sources. This allows data to be received from multiple sources in different formats and mapped to a common database schema. This enterprise-class software allows companies to centrally manage their data schema and mapping rules in a central repository and directly archive those rules in PDF documents for easy sign-off."

Key Features

ETL (Extract, Transform, Load) – Extracts data from one database, converts it, and loads it into another for business analytics. Allows users to access data in real-time and transform it to suit business operations while simultaneously allowing IT teams to manage processes and security

Functionality – Provides the ability to transform any number of files, such as flat files to XML, XML to CSV, Excel to Database, HIPAA to HML, HL7 to XML, etc., which enables simple aggregation and more rapid reporting and overall productivity

In-Built Job Scheduler – Real-time and batch triggers that may be built directly into any OEM solution, providing Data Integration capabilities, including a job scheduler and polling events

Bottom Line

Adeptia makes it easy for customers to deliver data to you, speeding up the pace of business, allowing for automation of integration processes end-to-end, eliminating the need for manual intervention. Adeptia ETL Suite supports ANY TO ANY conversion: Flat file to XML, XML to CSV, Any Database to XML, Excel to Database, XML to EDI, HIPAA to XML, HL7 to XML, Date format, string, numerical conversions etc.



Altova

Altova, a company with headquarters located in Massachusetts and Austria, provides Data Integration capabilities accessible through a graphical, drag-and-drop mapping design. Their MissionKit is a software development suite of enterprise-class XML, SQL, and UML tools for information architects and application developers. This suite is offered in two editions that cater to business needs in order to deliver the functionality organizations require.



From the Company

"Altova is a software company specializing in tools to assist developers with data management, software and application development, and Data Integration. The creator of XMLSpy and other XML, SQL, and UML tools, Altova is a key player in the software tools industry and a leader in XML solution development tools. Altova MissionKit Data Integration tools provide powerful Data Integration capabilities that are easy to access via a graphical, drag-and-drop data mapping design. All of today's prevalent data formats, including all major relational databases, are supported in one, affordable solution. Generation of royalty-free C#, C++, Java, XSLT, and XQuery code for automating Data Integration projects, plus optional integration with Visual Studio and Eclipse, give you flexible development and deployment options at no extra cost."

Key Features

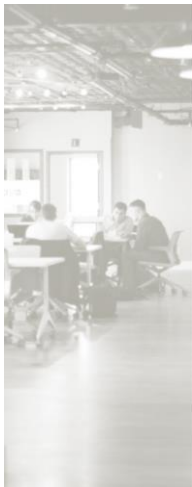
MapForce – Map data graphically with support for XML, databases. EDI, JSON, flat files, Excel/OOXML, XBRL, and web services

DiffDog – Allows users to compare and merge text, XML, MS Office (OOXML), directories, XML and database schemas

StyleVision – Integrate, combine, and publish XML, database, and XBRL data with Altova's visual design tool. This allows customers to drag-and-drop data to design reports with the ability to publish in HTML, PDF, Word, and e-Forms

Bottom Line

Altova MapForce is an advanced integration application tool with support for integrating data with virtually any file format. Drag-and-drop functionality allows users to easily associate target and source data structures. Altova MissionKit database tools include support for all major relational databases in their native interface language: Microsoft® SQL Server® 2000, 2005, 2008, IBM DB2® 8, 9, IBM DB2 for iSeries® v5.4, 6.1, Oracle® 9i, 10g, 11g, Sybase® 12, MySQL® 4, 5, PostgreSQL 8, Microsoft Access™ 2003, 2007.



Attivo

Attivo, "the Data Dexterity Company" is based in Massachusetts and offers an information access platform that uses a single query and can analyze structured data and unstructured content. Their Active Intelligence Engine (AIE) indexes data so IT teams can dynamically access the information from any location or format. Attivo's flagship solution also provides a predictive analytics module to more efficiently analyze, search, and manipulate data. Attivo prides themselves on empowering users to get their hands on the right data and work to get it to the point of informing decisions.



From the Company

"Attivo's Active Intelligence Engine (AIE) unifies the worlds of structured and semi-structured data, Big Data and unstructured content. Unlike traditional data warehouses or other information systems, AIE does not require defining data structures and relationships between data and/or content in advance. AIE's schema-less universal index eliminates the process of modifying data models to accommodate new sources of information."

Attivo's Active Intelligence Engine (AIE) can ingest both structured and semi-structured data, Big Data and unstructured content, from a wide variety of databases, document repositories, content management systems, email systems, websites, social media and file servers, without the need for extract, transform and load (ETL) processing."

Key Features

Semantic Enrichment – Delivers a semantic view of all data and information across an organization by extracting data and making it discoverable

Universal Cataloguing – Provides immediate visibility to all of an organization's information through keyword search, recommending what is most relevant

Automated Correlation – Allows users to generate data models, correlate both structured and unstructured data, and simplify provisioning to BI and analytic tools

Bottom Line

Attivo includes out-of-the-box connectors to relational databases, file content, XML and CSV data. Optional application-specific connectors are available for Microsoft SharePoint, Microsoft Exchange, Active Directory, EMC Documentum Content Server, website harvesting, and Hadoop.



Attunity

Attunity is operated in the U.S., but also has headquarters in Israel. They provide a Big Data software solution that enables access, management, sharing and distribution of data, as well as the ability to access heterogeneous enterprise platforms, organizations, and the cloud. Attunity can be used either directly or indirectly through partners such as Microsoft, Oracle, IBM, and HP. The vendor also provides data replication and flow management, test data management, change data capture (CDC), managed file transfer and cloud data delivery.



From the Company

"Attunity Ltd. is a leading enterprise-class software provider empowering some of the world's largest companies with the solutions to seamlessly and efficiently connect, transfer and join to and from virtually any data source. Attunity solutions identify critical changes in real-time, and subsequently service-enable that data for use in modern, rapid-to-deploy and re-configurable business applications.

The Attunity Integration Suite (AIS) consists of three primary products: Attunity Connect, a real-time connectivity software, Attunity Stream, a log-based, real-time change-data-capture software, and Attunity Federate, virtual data federation for Enterprise Information Integration (EII)."

Key Features

Replicate – Provides high-performance data replication and loading through one interface while utilizing TurboStream change data capture (CDC) technology combined with in-memory transaction streaming

CloudBeam – Automates and accelerates the process of loading data from multiple sources via Amazon Web Services and Azure environments continuously

Attunity Compose – Automatically designs, generates, and populates enterprise data warehouses and data marts, thereby adding data modeling and structuring capabilities

Bottom Line

Attunity Replicate's "Zero Footprint" architecture means that no agents must be placed on the source or target, eliminating overhead for mission-critical systems. In addition, every step of the replication process is automated and the Click-2-Replicate user interface offers drag-and-drop functionality.



CloverETL

CloverETL, developed in 2002 by Javlin, was one of the first open-source ETL software platforms available. Their solution provides Java-based Data Integration framework and can be used in either standalone mode or embedded within other applications. Clover's commercial solution includes a designer and server platform to add automation and workflow orchestration.



From the Company

"Javlin is a provider of Data Integration software and solutions for mass data processing. CloverETL began as an open source project by founder and Javlin president David Pavlis in 2002. CloverETL provides users the ability to manage data solutions such as integration, migration, cleansing, audit, synchronization, consolidation, Master Data Management and data warehousing. The CloverETL product is platform independent and scalable with a smooth upgrade path. It is also easily embeddable thanks to its small footprint. The CloverETL Open Source Engine is a core Java library that powers all CloverETL products. It provides the execution base for data transformations, a core set of transformation components, and a multi-threaded execution model. Commercial editions come with an enhanced engine with additional components and Server support."

Key Features

CloverETL Designer – Enables users to develop and manually execute data transformations without extensive setup, and connect to text and Excel files, web services, and NoSQL databases

CloverETL Server – Allows organizations to deploy internal data pipelines via enterprise Data Integration runtime, orchestration, and monitoring platform

CloverETL Cluster – Provides parallel processing for large datasets on multiple nodes through partitioning and transformation design

Bottom Line

CloverETL is Java-based. Supported platforms include: Windows, Unix, Linux, OSX, and many others. It utilizes XML-based resources - all resources, such as graphs, metadata, shared connections, etc., are stored in an XML format. Thus, transformations may also be generated (by anything able to create a XML file. Clover's flagship solution reduces manual workload time, enables communication amongst incompatible systems, creates a single, consistent view for business critical data and increases the efficiency of organizational development.



Dell

Dell provides a Data Integration solution that allows users to connect any combination of cloud and on-premise applications without software or appliances. Their software allows for quick integration processes through an intuitive drag-and-drop interface, and can suggest mappings and the mapping of data fields between enterprise applications. Organizations can also easily monitor in-house applications, as well as leverage existing SOA software.



From the Company

"Dell Boomi, a business unit of Dell, is an integration solution built in the cloud to fully exploit the value of the cloud. Boomi AtomSphere delivers enterprise-class integration capabilities without the complexity. Complete application integration requires a lot more than just "data mapping". Boomi AtomSphere delivers high-end integration capabilities such as content-based routing, process flow control, exception handling, and messaging that take you well beyond just data mapping. Using our Visual Integration Technology, we take care of the complexity that would typically require developers: deciphering complex APIs, developing web services, deploying SOAP or REST technology, or even deploying a service-oriented architecture (SOA). It's all handled through one simple to use, visual interface."

Key Features

AtomSphere – Supports common transport methods and enterprise integration scenarios, including B2B, EDI, and web services, while also offering universal translation capabilities for non-standard data formats

Boomi Assure – An optional feature that allows customers to submit organizational integration processes and date to Dell for testing in order to improve the quality of their performance capabilities

Molecules – An enterprise-grade version of Atom technology that can be deployed across multiple servers to accelerate load balancing and deliver better availability for mission-critical integration processes

Bottom Line

Dell Boomi provides navigation between Master Data Management, Data Integration, and data quality services (DQS) in a single interface. Centralized user and role management and single sign-on bring ease of use to the data management and Data Integration process.



HVR Software

HVR, a company based in the Netherlands, offers a Data Integration solution that can be applied to various environments and provide different capabilities as needed. Their solution can capture volumes of data from common sources in real-time, and is supported by platforms including Oracle, Teradata, Pivotal, Actian, and all Hadoop distributions. HVR's architecture enables data to be compressed over the network, and interruptions or communications challenges can be recovered.



From the Company

"HVR Software concentrates on developing HVR, an open and standard based Real-time Data Integration platform. HVR Software supplies three innovative products for mission-critical environments; HVR Real-time Data Integration, HVR Managed File Transfer and HVR Compare and Refresh. These products cover distinct functional areas but share important common elements; an efficient network streaming architecture, a flexible feature framework and an intuitive Graphical User Interface.

HVR's performance and scalability allows it to replicate very large databases in real-time. And its elegant design and rich functionality means difficult replication schemes can be implemented with minimum effort. Strong tools are provided for comparing and refreshing replicated tables and for monitoring replication scheduling."

Key Features

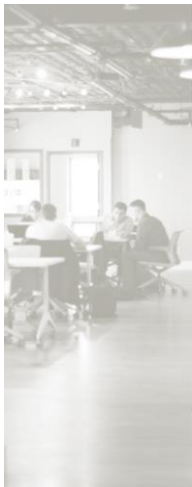
Database Replication – Uses log-based capture between databases within computing environments. Changes applied to the source are detected and transmitted to be copied to one or more target databases

HVR Refresh – Compares the contents of different databases, and if there are differences or inconsistencies, the tool can refresh or repair to synchronize

Support – Supports asynchronous log-based capture out of Oracle, SQL Server, IBM DB2, and Ingres

Bottom Line

HVR can connect different Online Transaction Processing (OLTP) Systems in real-time, such as a link between a back-office system and an internet database. HVR supports multiple database solutions. It is possible to replicate between Oracle, SQL Server and Ingres database as well as perform file replication and bridging to XML-based web services.



IBM

IBM's InfoSphere Information Server is a solution that helps to transform data in any style, and deliver it to any system. Their software provides built-in transformation functions, as well as a common metadata framework for increased efficiency. IBM provides two options for data delivery: ETL (extract, transform, load), virtual (federated), and incremental (data replication). InfoSphere also increases deployment time using a graphical interface, and allows integration with DMBS, Big Data sources, messaging queues, ERP, and other packaged applications, mainframe systems, and industry formats via API connectivity and parallelism.



From the Company

"IBM InfoSphere Information Server provides a rich set of information integration and governance capabilities that enable you to integrate Big Data with your traditional enterprise data and gain critical business insights. The end-to-end information integration capabilities of InfoSphere Information Server allow you to understand your data, cleanse, monitor, transform, and deliver data. Included component are the InfoSphere DataStage, which integrates data across multiple and high volumes of data sources and target application; InfoSphere QualityStage to ensure systems deliver accurate and complete information to business users across the enterprise; and InfoSphere Business Glossary, which enables users to create, manage and share an enterprise vocabulary and classification system."

Key Features

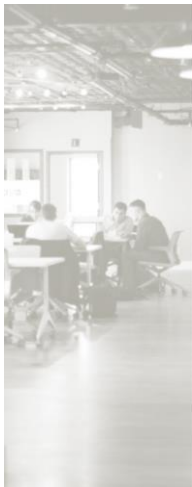
Data Delivery – Deliver Big Data with scalable integration platform to process the right workloads with the right tools and enable data governance through data lineage

Cloud – Enables cloud initiatives via private or public cloud, or to integrate on-premise data with a cloud environment

FastTrack – Streamlines data between analysts, data modelers, and developers in a common format, enabling collaborative development across products, user roles, and geographics

Bottom Line

IBM InfoSphere helps create and maintain trusted information to support strategic business initiatives involving Big Data, analytics, Business Intelligence, data warehousing, Master Data Management and application migration and consolidation. InfoSphere can integrate to databases such as its own InfoSphere Warehouse Enterprise (based on IBM DB2), IBM's Big Data Analytics solution Netezza, SQL, Oracle and other databases or the Enterprise Service Bus or message brokers such as MQ series.



Informatica

A software development company based out of California, Informatica's integration solution helps to refine fragmented data into complete assets. Their platform provides development tools to help prototype and operationalize data, and can simplify and accelerate integration with out-of-the-box purpose-built connectors. Informatica's visual development interface can increase productivity and help to ensure that open source platform innovations can be adopted without sacrificing maintainability or reusability.



From the Company

"Informatica is a leading provider of Data Integration software. Informatica's enterprise Data Integration and management solutions and both mainframe and cloud based and include data governance, data migration, data quality, data synchronization and data warehousing."

The Informatica enterprise Data Integration product family is scalable, high-performance enterprise Data Integration software. These products access and integrate data from virtually any business system, in any format, and deliver that data throughout the enterprise at any speed. The Informatica Platform enables a wide variety of complex enterprise-wide Data Integration initiatives through the following technologies: Enterprise Data Integration, Data Quality, Master Data Management ("MDM"), B2B Data Exchange, Application Information Lifecycle Management ("ILM"), Complex Event Processing, Ultra Messaging, and Cloud Data Integration."

Key Features

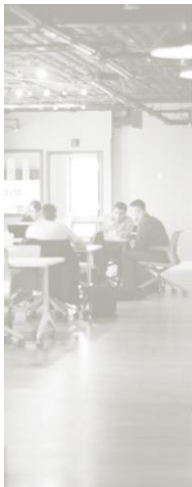
Universal Access – Provides accessibility to transactions, applications, databases, log files, social, machine, and sensor data

Data Parsing – Allows users to parse complex, multi-structured, unstructured, and industry structured data via Hadoop

Transparency – Offers complete transparency with end-to-end data lineage of all data activity from source data to target applications

Bottom Line

Informatica's mainframe Data Integration solutions are designed to interoperate with Informatica PowerCenter and Informatica PowerExchange on Linux, UNIX, and Windows platforms. Informatica Cloud connects to a wide variety of on-premise and cloud-based applications—including enterprise applications, databases, flat files, file feeds, and even social networking sites.



Information Builders

Information Builders is one of the largest privately held software firms in the world. IB's flagship integration solution, the iWay Integration Suite, allows organizations to speed up the deployment process and can help reduce potential risks that might be encountered during integration projects such as ETL (extracting, transforming, and loading), or enterprise information system integration (EII) initiatives. The company also offers end-to-end integration of sources such as cloud-based information, social systems, and Big Data.



From the Company

"Founded in 1975, Information Builders helps organizations transform data into business value. Our software solutions for business intelligence and analytics, integration, and data integrity empower people to make smarter decisions, strengthen customer relationships, and drive growth.

With tools for real-time data integration, data movement and delivery, and data federation, Information Builders enables organizations to manage both structured and unstructured information. Highly scalable, the company's iWay tools accelerate the deployment of data integration projects such as extract, transform, and load (ETL); enterprise information integration (EII) initiatives; and big data integration."

Key Features

DataMigrator Change Data Capture (CDC) – Provides a real-time capability using database logs to read only changes made in databases, and then delivers that information to DataMigrator

Universal Adaptor Framework – Contains off-the-shelf adapters to provide access to any information, and repurpose various forms of data, applications, any B2B format, and any service

Flexibility – Supports real-time and batch integration, ETL, and message-based styles of integration

Bottom Line

Instead of requiring physical data movement, Information Builders enables WebFOCUS users to issue distributed queries that correlate and manipulate data from disparate relational databases, packaged applications, structured data files such as XML and EDI documents, and legacy databases and files.



Jitterbit

Jitterbit provides a Data Integration suite that is comprised of cloud, on-premise, social, and mobile applications. Their solution offers an easy-to-use interface to simplify the process, whether your business requires data warehousing, BI, or legacy system integration initiatives. Jitterbit's provide native connectivity to enterprise applications, relational databases, flat files, XML, and SaaS/Cloud data, and employ parallel processing algorithms to synchronize disparate computing platforms efficiently.



From the Company

"Jitterbit creates open source integration software that aims to overcome the challenges of cost and complexity associated with connecting applications, data and systems. Jitterbit's open-source integration solution is a graphical platform for connecting legacy, enterprise and On-Demand applications, including Business Process Fusion, ETL, SaaS, and SOA.

The Jitterbit Open Source Integration Platform also makes use of sharable integration templates called Jitterpaks that allow the community to reuse common integrations with the most popular web services, application APIs, and databases. Jitterbit's graphical "No-Coding" approach simplifies the configuration and management of otherwise complex integration projects."

Key Features

Clicks, Not Code – Allows users to connect to almost any application and map fields using the Automapper

Deployment – Enables deployment via cloud, on-premise or hybrid approach. Their elastic cloud agent network scales to meet real-time business requirements

Integration – Provides wizard-based connectors for enterprise applications, and lets customers connect to any SOAP or REST web device, and complete ODBC and JDBC database connectivity

Bottom Line

Available in the cloud or on-premise, Jitterbit automatically discovers system configurations and allows non-technical users to point and click to define source and target systems, drag-and-drop to map data transformations, and run integration operations on batch, trickle or real-time schedules.



Liaison Technologies

Liaison offers a customizable Data Integration solution that can be tailored to specific business needs. Liaison's ALLOY platform can be scaled to future-proof your business and create custom data models and support, as well as ingest and transform data from a wide variety of sources. Their solution offers an interface for data stewardship, governance workflow, and data profiling.



From the Company

"Liaison Technologies offers a Data Platform as a Service (dPaaS) solution handling all your integration and data management needs in one holistic cloud-based multi-tenant platform. It integrates your data, both structured and unstructured, from every source imaginable – connecting, transforming, harmonizing, and managing it to lead you to discover valuable business insights. The ALLOY Platform consists of three modules: Data Orchestration, a set of functionalities to integrate applications and data in cloud as well as enterprise system; Data Persistence, which is Big Data technologies and usage of polyglot data management techniques with accessible APIs to manage data and support schema on read approach; and Data Visualization, which gives users transparency into data and data flow through customizable interfaces."

Key Features

LENS (Liaison Data Visualization Module) – Achieves visibility an organization's data flows and the data itself, which enables better access and control, fostering compliance, security, and data governance

Complete Integration View – Allows users to plug any data source into LENS to consolidate A2A and B2B integration information into a single view

dPaaS – Enables data-centric integration in addition to addressing data management use cases. All integration is delivered as managed services through metadata and business processing rules

Bottom Line

Liaison ALLOY is the world's first Data Platform as a Service for integration and management. It enables data-centric integration and delivers it as a managed service enriched through metadata and business processing rules. Liaison Technologies Enterprise Application Integration includes: B2B / A2A / EDI Integration, Managed Services, Cloud Services, Master Data Management, Data Integration, Data Mapping, Data Transformation, Data Security, Security.



Microsoft

Microsoft's SQL Server Integration Services can be used to perform a broad range of data mitigation tasks. The platform helps build enterprise-level Data Integration and data transformation services at its core. Microsoft's SSIS capabilities can extract, transform, and consolidate data from multiple recreational databases and sources such as XML data files and flat files, and then load the processed information into an enterprise data warehouse or other large target system. Their solution provides tools for developing and testing integration platforms as well.



From the Company

"Microsoft SQL Server is a cloud-ready, information platform that will help organizations unlock breakthrough insights across the organizations and quickly build solutions to extend data across on-premises and public cloud.

SQL Server Integration Services (SSIS) is Microsoft's Extract, Transform, Load (ETL) tool and is provided with SQL Server. SSIS provides a set of built-in tasks, containers, transformations, and data adapters that support the development of business applications. Without writing a single line of code, you can create SSIS solutions that solve complex business problems using ETL and Business Intelligence, manage SQL Server databases, and copy SQL Server objects between instances of SQL Server."

Key Features

Memory – Enhances in-memory performance, and provides faster transactions and queries than disk-based relational databases and real-time operational analytics

Encryption – Helps protect data both on-premises and in the cloud with master keys within the application with no altercations

PolyBase – Simplifies management of relation and non-relational data by querying with T-QSL, and addresses distributed data sets in Hadoop and cloud environments (i.e. Azure storage)

Bottom Line

The Microsoft runtime environment of SSIS consists of 2 engines, the Runtime Engine and the Dataflow Engine, a package consists of tasks (one of which is the all-important Dataflow Task) and containers. The development tool for SSIS is Business Intelligence Development Studio (BIDS).



MIOsoft

MIOsoft's MIOvantage software empowers its users to solve their data integration problems by providing extreme flexibility that maximizes its utility in virtually any situation. For businesses seeking a high-productivity data integration tool, MIOvantage includes essential capabilities: its graphical user interfaces allow technical and business teams to collaborate effectively, its beyond-match-and-merge technology allows unparalleled discovery of relationships within and between data, and its high-performance architecture can facilitate integrations at any scale and speed.



From the Company

"MIOsoft provides a complete array of data quality solutions through its MIOvantage software, which includes strong data integration capabilities. A key component of this is MIOvantage's flexible adapters, which can handle virtually any format and source system and are configured through a visual interface that maximizes understanding and efficiency.

Integrated data can be pushed immediately to a destination warehouse, or additional meaning can be extracted using MIOvantage's quality, analytics, and MDM capabilities. MIOvantage's high-performance architecture allows it to handle virtually any project scope—from tens to thousands of source systems—at any speed. Repeated integration management functions can be automated using included tools, freeing users for other tasks."

Key Features

Data Quality – Includes a wide array of essential data quality features, including visual parsing, profiling, and cleansing tools and an extensive transformation library.

Beyond Match and Merge – Enables intelligent decisions about when and how to match and merge data, delivering superior identity/entity resolution. Tools to discover relationships within and between data help maximize data's business value.

Flexibility – Connects to virtually any data source using visual adapters that make it easy to effectively prepare data from both standard formats and unique legacy or custom systems.

Bottom Line

MIOsoft's MIOvantage provides maximized capability and unparalleled flexibility to ensure that data integrations of all scales and types succeed. MIOvantage can connect to virtually any data source, including legacy and custom systems, using its configurable visual adapters.



Oracle

Oracle offers a unified Data Integration solution for deploying and managing data in SOA, Business Intelligence and data warehouse environments. Oracle's software combines data movement, synchronization, data quality, data management, and data services. The Oracle Data Integrator 12c is primarily for enterprise-scale businesses, and Oracle GoldenGate 12c is for streaming transactional data into Big Data systems.



From the Company

"Oracle Data Integrator is a comprehensive data integration platform that covers all data integration requirements: from high-volume, high-performance batch loads, to event-driven, trickle-feed integration processes, to SOA-enabled data services. It is fully integrated with Oracle Fusion Middleware, Oracle Database, and Exadata as well as out-of-box integration with ERPs, CRMs, B2B systems, flat files, XML data, LDAP, JDBC, and ODBC.

Oracle Data Integration delivers bulk or real-time data from heterogeneous systems; Log-based real-time change data capture (CDC) for minimizing impact on source systems; eliminating middle-tier transformation server for reduced TCO. Oracle's Data Integration solutions are optimized and certified for Oracle Exadata and can speed data loads to Exadata by 500%."

Key Features

Service-Oriented Architecture (SOA) – Provides support for high-volume, high-performance bulk data processing to an existing service-oriented architecture

Knowledge Modules – Implements data flow and defines templates for generating code across multiple systems involved in each integration process, thereby helping to capture and reuse IT expertise and reduce cost of ownership

Integration – Supports loading and transforming data by integrating with Hadoop, Hive, HDFS, and Oracle Big Data Appliance, and can stream data deployed in Apache Spark and Pig environments

Bottom Line

Instead of relying on a separate, conventional ETL transformation server, Oracle Data Integrator Enterprise Edition's E-LT architecture generates native code for disparate RDBMS engines (SQL, bulk loader scripts, for example). E-LT architecture extracts data from sources, loads it into a target, and transforms it using the database power.



Pentaho

Recently acquired by Hitachi, Pentaho is a software corporation based in Florida that offers an integration solution in two editions: Community and Enterprise. Pentaho's software follows an open-core business model, and provides support for Hadoop distributions from Cloudera, Hortonworks, and MapR. In addition, it provides plug-ins to NoSQL databases such as Cassandra and MongoDB, as well as connections to specialized data stores like Amazon Redshift and Splunk.



From the Company

"Founded in 2005, Pentaho tightly couples Data Integration with complete business analytics for Big Data, supporting Hadoop, NoSQL, and Analytic databases. Pentaho provides a full Big Data Analytics solution that supports the entire Big Data Analytics process from ETL and Data Integration to real-time analysis and Big Data visualization."

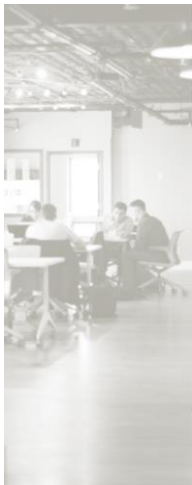
Pentaho Data Integration (PDI) enables organizations to extract data from complex and heterogeneous sources and create consistent, high-quality information for critical business applications. With a rich graphical user interface and a parallel processing engine, PDI offers high-performance ETL (extract, transform, and load) that covers all Data Integration needs."

Key Features

- **Visual Designer** – Offers visual tools to minimize coding necessity as well as an ETL tool for more efficient processing
- **Administration & Management** – Provides out-of-the-box capabilities to manage integration deployments, schedule regularly recurring workflows, and manage role-based security definitions from providers such as LDAP and Active Directory
- **Data Profiling & Quality** – Identify data that fails to comply with business standards and standardize, validate, deduplicate, and cleanse inconsistent or redundant data

Bottom Line

Pentaho offers a multi-threaded engine, data partitioning support, and clustered execution options for scaling out ETL jobs across a cluster of PDI servers through Hadoop. Also provides in-memory caching, dynamic look-ups, and parallel bulk-loaders.



Progress Software

Progress Software offers a DataDirect integration suite that allows organizations to transform both structured and unstructured data through unified data access. They offer Apache Hive ODBC and JDBC drivers to increase business performance. Furthermore, the company supports sources such as Amazon Redshift, Cloudera Impala, MongoDB and Cassandra. Their solution offers on-premise, cloud-based, and hybrid applications, and Panorama's DataDirect Cloud provides connectivity to on-premise and cloud data.



From the Company

"Progress Software supplies products for application development, SaaS enablement, cloud deployment, complex event processing (CEP), data connectivity and real-time decision management to solution partners and end users. Progress DataXtend is a Data Integration platform for service-oriented architecture that provides tools to help you ensure the free flow of information across your organization. Progress DataXtend Semantic Integrator (SI), the principal solution in this robust software suite, addresses the common information model challenges in SOA infrastructure and provides a way to ensure that data is not only correct, but that it is valid based on your business rules."

Key Features

Bulk Load – Manages 'single-driver' connectivity to enterprise databases and platforms, and deploys without application code changes or database vendor tools

Warehousing – Allows users to import data into an Oracle, DB2, Sybase, or SQL Server-based data warehouse without data latency issues

Prototyping – Offers assistance with building and testing SQL code in a web browser before deployment

Streaming – Allows customers to stream data in increments to optimize latency and facilitate a mobile experience for users

Bottom Line

Progress Software includes a complete graphical design environment for creating and managing Exchange Models. Furthermore, their DataXtend SI Engine, a runtime component of the DataXtend SI integration platform, converts and validates data before it is submitted to backend solutions.



SAP

SAP (System Analysis and Program Development) is a software corporation based in Germany that provides a suite with multiple Data Integration products to collect, manage, store, and retrieve data from multiple source: Data Services, Replication Server, Landscape Transformation Replication Server, Process Orchestration, Hana Cloud Integration, Enterprise Information Management (EIM), Agile Data Preparation, and PowerDesigner.



From the Company

"SAP is the market and technology leader in business management software, solutions, and services for improving your business process. The company's best-known software products are its enterprise resource planning application systems. SAP BusinessObjects Integration software provides direct connectivity to enterprise applications. This enables you to consolidate and transform vast amounts of data into reports, analyses, visualizations, and dashboards while helping ensure data security and compliance. With SAP BusinessObjects Integration, your organization can leverage a single Business Intelligence (BI) solution for better-informed, faster decision making and business performance optimization."

Key Features

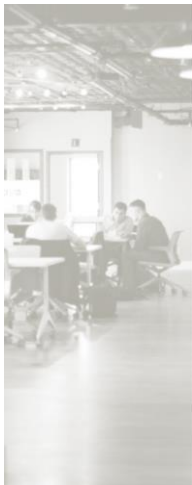
Functionality – Allows users to access and integrate data from any source and design an efficient integration process

Delivery – Supports operational and analytical initiatives, such as customer relationship management (CRM), enterprise resource planning, Business Intelligence, data warehousing, data migration, and Master Data Management

MemScale – Supports extreme transaction processing applications executed by large numbers of concurrent users for high-performance

Bottom Line

Data Integrator delivers ETL scalability with support for parallel processing, grid computing, real-time data movement, 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.



SAS

SAS Institute provides a solution with strong visual tools to build, implement, and manage data through any source, application, or platform. SAS enables the ability to collaborate within large enterprise projects share data throughout your organization. Their software also provides a common utility, Metadata Bridge Relationship Loader, which allows users to load third-party metadata into the SAS Relationship Service and analyze using the SAS Lineage Viewer.



From the Company

"SAS is a leader in business analytics software and services, and is the largest independent vendor in the Business Intelligence market. SAS Data Integration provides capabilities for enterprise data access and processing across systems and platforms; integrated data quality and an interactive, visual data integration development environment that enables collaboration and easy reusability, all with a single point of IT administration."

SAS Data Management Advanced enables organizations to extract, transform and load (ETL) data from across many different sources to create consistent, accurate information. A point-and-click process design desktop makes it easy to build logical workflows, quickly identify input and output data flows, and create business rules in metadata."

Key Features

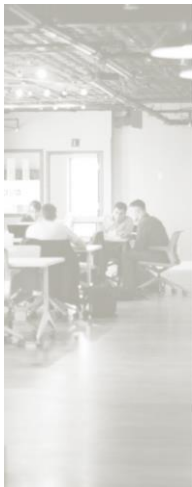
Productivity – Enables users to create data management rules and reuse them, providing a standard, repeatable method for integrating data

Collaboration – Provides the ability to update enterprise data, tweak processes, and analyze data results, as well as third-party metadata management/lineage visualization capabilities

Integration – Access to Hadoop via sources such as Impala or Pivotal HAWQ, allowing users to add Big Data to existing IT processes

Bottom Line

Organizations can take advantage of a grid-enabled, load-balanced, multithreaded parallel processing architecture that can quickly transform and move data between different platforms and systems. SAS also supports zero data movement by using SQL pass-through into popular database appliances, including Oracle, DB2, Teradata, Netezza, SQL Server, Aster, and Hadoop.



Stone Bond Technologies

Stone Bond Technologies' Enterprise Enabler® is a highly scalable, single platform for configuring, testing, deploying, and monitoring Data Integration for a point solution or throughout the enterprise. With Enterprise Enabler, combine data live from multiple sources and query the virtual data model for Data Virtualization, or use it to avoid staging data for ETL. For any type of data flow or on-demand consumption, the single powerful environment means up to 90% faster time to value for your business.



From the Company

"Since 2002 Stone Bond Technologies has been providing agile integration with Enterprise Enabler, on premise or in the cloud. With the current technology focus on Big Data and Data Virtualization, many customers are using our technology to prepare the data as virtual data models for use as a Logical Data Warehouse, as building blocks for Master Data Management projects or Enterprise Service Layers. Stone Bond's team ensures the best value for our customers with hands-on support, training, implementation, and mentorship"

Key Features

AppComm Technology - Efficient and intelligent configurable endpoint access technology – with Enterprise Enabler, connect anything to anything

Design Studio - Exceptionally easy to use graphical interface for configuring and testing integrations

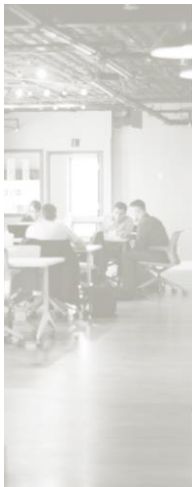
Active Integration Interface (AII) - Powerful native transformation engine that operates across endpoints live, without staging

Process Designer - Defines composite applications, data workflows, and caching scenarios

Integration Integrity Manager (IIM) - Monitors for change in schemas, determines impact, issue notifications and traces definitive lineage

Bottom Line

Enterprise Enabler is a one of a kind data virtualization platform that unifies information silos for visibility across multiple data sources all in real-time. Users can analyze business data as it is occurring to maximize the use of assets, minimize costs, and improve/refine business processes. Enterprise Enabler incorporates Data Virtualization, ETL, EAI, SOA, and MDS in a single platform and is implemented in up to 90% less time than other data integration solutions, all while eliminating the need of costly data storage and warehouses.



Software AG

Software AG, a software company based in Germany, provides a solution called Terracotta In-Memory Data Fabric Platform, and allows users to store data within its datasets to ensure accessibility. Their software scales to TBs of memory to accelerate application performance and can predict, identify, and inform of potential risks to allow for timely mitigation and problem solving, which allows data-driven organizations to handle risks more aptly and be more equipped for such events in the future. Software AG's Terracotta suite includes BigMemory, Universal Messaging, Enterprise Ehcache, Enterprise Suite, Quartz Scheduler, and Web Sessions.



From the Company

"Founded in 1969, Software AG brands, including ARIS, webMethods, Adabas, Natural, CentraSite, Terracotta and IDS Scheer Consulting, represent a portfolio encompassing process strategy, design, integration and control; SOA-based integration and data management and efficient management of Big Data.

WebMethods ConnecX is software that integrates enterprise data with existing and new applications by providing real-time, SQL-based access to relational and non-relational data sources. Data can be accessed as if it were stored in one relational database even though it may come from different applications, web services and other data sources."

Key Features

In-Memory Data Management – Provides a platform for low (and predictable) latency at any scale, and can develop real-time insights and intelligent action based on instant access to business data

Connectivity – Enables users to connect to Hadoop, social, enterprise, web, mobile, and other sources from a single, scalable connectivity solution

Detection – Protect company assets with real-time fraud detection and data monitoring to reduce time-to-detection, thereby minimizing losses and maintaining regulatory reporting

Bottom Line

Software AG's WebMethods' OneData is a Master Data Management platform that provides users with five data out-of-the-box solutions and simplifies integration. WebMethods provides a single view of data from non-relational and relational databases, including: Adabas, VSAM, C-ISAM, DISAM, RMS, DataFlex, Powerflex, Oracle, DB2, SQL Server, RDB, and Sybase.



Syncsort, Inc.

Based out of New Jersey, Syncsort provides a Data Integration software solution that facilitates Hadoop, Microsoft Windows, UNIX, Linux, and mainframe systems. Their DMX software can extract, transform, and load data using little CPU utilization. Syncsort's Smart ETL Optimizer tool provides Data Integration that works with the most efficient algorithms based on an organization's data structures and system, and from there logs what it encounters at run-time.



From the Company

"Syncsort Incorporated is a software company specializing in high speed sorting products, as well as Data Integration and backup software and services, for Windows, Unix, Linux, and mainframe systems. DMExpress Syncsort's flagship Data Integration product is the fastest version with high-performance compression technology and high-performance join algorithms. All of the components required for Data Integration acceleration are included in this comprehensive edition. DMExpress also contains new capabilities, such as Metadata Interchange, allowing you to easily import jobs from other platforms, such as Informatica and IBM DataStage, to accelerate deployment."

Key Features

DMExpress – Deploys as part of the Hadoop Cluster, and delivers what is required to collect, prepare, blend, transform, and distribute important data

Ironcluster – Allows users to provision a full-featured Data Integration environment for Amazon EC2 and Amazon EMR, and can connect to RDBMS, mainframe, HDFS, Salesforce and Redshift

FirePort – Runs on all major UNIX systems, and transforms mainframe any type of data to UNIX format

Bottom Line

DMExpress eliminates data latency by reducing CPU time, elapsed time, and disk I/O activity while utilizing minimal resources on commodity hardware. It also shortens deployment and development timeframes by leveraging a graphical, drag-and-drop development environment. DMExpress utilizes proprietary sorting algorithms, I/O optimization, parallel processing, and dynamic environmental monitoring techniques.



Talend

Talend offers an enterprise-class integration solution that allows users to natively connect databases, flat files, and cloud-based applications. The company provides graphical drag-and-drop tools, test creation, and code generation in numerous languages. Talend Open Studio, a free service, can manage and monitor projects and allow teams to leverage a shared repository and versioning tools. Their other offering, Talend Open Studio for Data Integration is subscription based and specifically facilitates enterprise-scale projects, such as team collaboration tools, industrial-scale deployment, and load balancing.



From the Company

"Founded in 2005, Talend is an open-source software vendor that provides Data Integration, data management, enterprise application integration and Big Data software and solutions. Talend's data integration products provide an extensible, highly-performing open-source set of tools to access, transform and integrate data from any business system in real-time or batch to meet both operational and analytical Data Integration needs. With 450+ connectors, it integrates almost any data source. The broad range of use cases addressed include: massive scale integration (Big Data/ NoSQL), ETL for Business Intelligence and data warehousing, data synchronization, data migration, data sharing, and data services."

Key Features

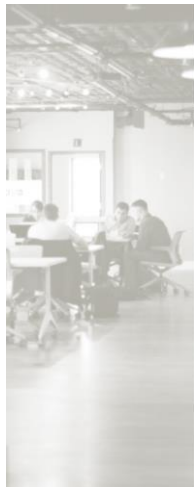
Data Quality – Provides, cleanses, anonymizes, and masks business data while monitoring quality over time. Can also reduce costs, increase sales, and improve performance

Development – Features 900+ connectors to natively connect databases, flat files, and cloud-based applications. Enterprise BMP assists with modeling and evaluating processes

Continuous Delivery – Helps organizations manage and monitor projects to lessen software defects and increase time-to-market

Bottom Line

Talend's flagship scalable tool allows users to integrate, cleanse, mask and profile vital business data. Talend connects natively to databases, packaged applications (ERP, CRM, etc.), SaaS and cloud applications, mainframes, files, Web services, data warehouses, data marts, and OLAP applications.



About Solutions Review:

Solutions Review is a collection of technology news sites that aggregates, curates and creates the best content within leading technology categories. Solutions Review mission is connect buyers of enterprise technology with the best solution sellers.

Over the past three years, Solutions Review has launched ten tech Buyer's Guide sites in categories ranging from Cybersecurity to Wireless 802.11ac as well as Mobility Management and Business Intelligence, Data Analytics, Data Integration and Cloud Platforms.

Information for this report was gathered via a meta-analysis of available online materials and reports, conversations with vendor representatives, and examinations of product demonstrations and free trials. Solutions Review does not endorse any vendor, product or service depicted in this publication and does not advise technology users to base their vendor selection entirely on this research. Solutions Review disclaims all warranties, expressed or implied, regarding this research, including any warranties of merchantability or fitness for a particular purpose.