Choosing the Right Data Integration Solution for Your Needs

Introduction

Without data integration, accurate analytics are impossible to achieve. Imagine trying to make a decision based on incomplete data. The less information available, the more likely a decision leads to an undesirable outcome. Now, multiply this challenge—decisions will now involve millions of dollars, hundreds of data sources, and terabytes of data. In order to steer a business correctly, data integration needs to handle a heavy burden.

A Brief History of Data Integration

Data integration grew out of extract, transform, and load tools (ETL), which were designed simply: Take data from one source, transform that data until it’s in a form that another application can recognize, and then load it into that application. A basic example would be taking raw financial data from a bunch of receipts, putting them into a spreadsheet and loading it into an accounting program.

As time went on, enterprises began to both generate more data and expect more from it. In response, data integration tools have become both more user-friendly and more granular. Administrators can choose to migrate data either in transactions or in batches, they can transform and filter data during run time, and they can even set up conditions where the data integration platform handles errors automatically.

Businesses now face a bewildering array of choices and options around various types and features of data integration platforms. What platform is best for a particular business?

Furthermore, you may need pieces of one and not another. In this document, we run down the capabilities of three categories of enterprise integration solutions, their differences, and similarities, as well as a few potential vendors for each.

“Businesses now face a bewildering array of choices of data integration platforms. What platform is best for a particular business?”
For Automating Business Processes, Try Enterprise Application Integration

As of 2014, enterprises ran an average of 508 applications. That number has no doubt only increased since, but the salient point is this: These applications are not designed to communicate with one another. This is what enterprise application integration (EAI) is for.

In some EAI approaches, a single solution collects incoming data and pushes it out to relevant applications. This is known as a broker model. For example, if a salesperson closes a sale in the CRM, the EAI will push that information to accounts receivable to generate an invoice, payroll to generate a commission, and budget to bank that closed sale that quarter’s earnings.

The benefit of this approach is an automated workflow. Prior to EAI, the chain of events described above would involve a chain of emails or a “sneakernet”. At scale, this would translate to significant losses in terms of time and efficiency as workers manually transcribe and upload data. Therefore, an EAI solution can recapture a great deal of productivity.

When shopping for an EAI solution, it’s imperative that it connect to every data source in the organization. Without interoperability, companies are faced with the problem of data silos—pools of information that might play a useful role in analytics, but which are isolated from EAI. According to research from Aberdeen group, 74% of top performing companies have eliminated data silos altogether.

Another potential disadvantage of EAI is the broker model itself. Brokers will only scale to a certain extent. Once a large number of applications are involved, the broker becomes a bottleneck for data, and lag is introduced into the system. Depending on a company’s needs, it may be best to invest in a solution that makes use of enterprise service bus (ESB) architecture.

Application Integration Providers:

- Celigo
- Youredi
- Adeptia
- Software AG

“Without interoperability, companies are faced with data silos.”
For Flexible Data Visualization, Use Self-Service Data Preparation

Say that instead of automating a large series of tasks, a company wishes to analyze a large amount of data. Data analytics isn’t new, but its accessibility is. In the days of ETL, creating complicated analytics and data visualizations would require assistance from IT staff. By contrast, self-service data preparation is essentially what it says on the label—a way for business users to explore their data without needing assistance or specialized training.

The two large innovations in self-service data preparation revolve around the UI and what’s under the hood.

- By design, the UI for self-service data preparation is easy—most solutions are simply a variation on "drag and drop."
- Behind the scenes, machine learning algorithms are supported by detailed rulesets can normalize data quickly and almost automatically.
- The architecture of self-service data preparation solutions accesses data sets in RAM, which means that analytics can be conducted at the speed of thought.
- Users can explore large data sets and draw meaningful conclusions without much prior knowledge of data science.

This flexibility can sometimes be its own enemy. Inexperienced or overenthusiastic users can sometimes misuse the product in a way that draws erroneous conclusions, or slows down the application itself (for example, by trying to connect too many data sources to a single analysis platform). On the other hand, some solutions may be too simple to satisfy professional data scientists. When choosing a self-service data preparation solution, organizations should tailor their choice to the level of expertise available.

"By design, the UI for self-service data preparation is a variation on "drag and drop."

Self-Service Data Prep Providers:

- Denodo
- Datawatch
- Trifacta
- Pentaho
For Businesses with Hybrid Clouds, use Integration Platforms as a Service (iPaaS)

By its very nature, the iPaaS model incorporates both EAI and self-service data preparation products. These services operate in the cloud, and can connect data sources in the cloud to applications that are on-premise and vice versa.

Most businesses—85% according to RightScale—operate some form of hybrid cloud, with a heavy emphasis on the public cloud. In fact, most business don’t go with a single public cloud provider. The average is 1.8 public clouds per business.

Data integration across clouds is its own problem. Bottlenecks in integration could previously be solved by adding more storage and compute resources into the mix, but in the cloud, interoperability becomes a problem. Taking an application like a CRM and moving it into the cloud means that it’s much more difficult to connect its data to apps that are normally hosted on premise. Almost 20% of businesses are concerned about integration in the cloud, and it’s become enough of a problem that 40% of businesses have moved at least some data back on-premise.

The iPaaS market has emerged to take care of those concerns. Improvements in architecture, implementation, and standards allow these services to quickly process data between separate clouds, and between private clouds and legacy on-premise apps.

“This market is extremely new, having just attained prominence in 2015, and in some ways it’s still finding its feet. Solutions are generally more focused on quickly connecting systems, as opposed to providing large amounts of granular options. The market is also new enough that it’s difficult to tell which vendor offerings will scale best with increasing data requirements.

iPaaS Providers:

- HVR
- Jitterbit
- MuleSoft
- Scribe Software
What about ETL?

Formerly the market standard for moving data from source to target, ETL (Extract, Transform and Load) tools increasingly are increasingly taxed by the expansion in unstructured data volumes that come from non-traditional sources. For some, ETL may be the most direct way to unify enterprise data assets, but for others, it might create a bottleneck that contributes to latency. Thankfully, integration tools branch out in many different directions to offer functionality that meets a wide variety of analysis styles.

Organizations in specific verticals may choose to store all of their data in a single repository (Data Lake) and run self-service data preparation on it there. Still, too, some use cases may warrant federating data so that a virtual copy can provide a unified view of all enterprise data. In other scenarios, fine-tuning the relationship between applications so that they may communicate with one another is most important.

Vendor Comparison Map

Aimed at simplifying the time-consuming vendor selection and evaluation process, the Venn diagram below offers an at-a-glance reference guide for 20 Data Integration players, the solutions they provide, and the markets they address.

Vendors on the outside of the circles tend to offer highly-specialized point solutions in those markets, while vendors towards the center offer more comprehensive platforms, which address multiple needs but may not do so with the same depth and granularity as point solution providers.

Many vendors included in this map offer various solutions addressing each of the markets concerned, and we have tried to reflect this in their positioning.

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.
Vendor List

**Actian**

Actian provides Data Management tools that enable automated action across various business operations. With support for Big Data, Actian DataConnect integrates with a wide range of data volumes and sources on-premises, in the cloud, or in hybrid environments. The tool’s visual interface was designed to allow for integrating disparate data in real-time or via the batch process. The company’s Integration Platform as a Service (iPaaS) solution Actian DataCloud features enterprise-grade functionality that business users can handle.

**Adeptia**

Adeptia’s self-service software assists IT teams with management and security, and provides a hybrid offering of both cloud and on-premise integration. Adeptia’s software platform is entirely web-based and offers three distinct flavors of integration: data, application and B2B, that enables the creation of simple app-to-app connections with little technical knowledge. The vendor offers variations of its software for specific users, covering everyone from line-of-business folks to IT developers and EDI specialists.
Attivio

Attivio 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 data from any location or format. Attivio’s flagship solution also provides a predictive analytics module to more efficiently analyze, search, and manipulate data. Attivio 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.

Celigo

Celigo’s integrator.io platform is an Integration Platform-as-a-Service (iPaaS) tool that provides business users with a set of wizards to build data flows between systems without the help of IT. Non-technical users utilize a data loader to upload data to cloud applications via SmartConnectors. Requiring no coding, SmartConnectors are prebuilt integrations between NetSuite, Salesforce and many other applications. Celigo’s integration platform is offered as a subscription service, and can be purchased in standard, premium or enterprise.

ClearStory Data

ClearStory Data engineers data tools that enable business users to discover, analyze and consume data from different sources. ClearStory Data provides a connection to disparate sources for data preparation and blending for any data source, API-based application or Hadoop. Smart Data Discovery provides data values, patterns and correlations within a single window or across blended data so users can zone in on critical data to arrive at the most meaningful conclusion. The solution includes a visual data intelligence engine and StoryBoards, a feature that gives users the ability to combine insights from different analysis into one.

Datawatch

Datawatch Monarch works by extracting data into analytics-ready rows and columns. Users can connect to all major relational and Big Data Sources, as well as 30 additional sources pre-defined via the platform. Datawatch enables users to bypass manual data preparation steps with manual processes by utilizing workflows built within the solution. Monarch allows users to extract data from multi-structured sources and web pages then combine, join, and blend it via joint analysis and fuzzy matching. Monarch features native support for many of the most popular visualization and advanced analytics tools on the market.
**Dell Boomi**

The Boomi AtomSphere Integration Platform-as-a-Service (iPaaS) supports a variety of integration processes between cloud platforms, SaaS applications, and on-premises systems. It utilizes a visual interface to configure application integrations that features drag-and-drop and data mapping. The runtime engine, Boomi Atom, allows integrations to be deployed whenever needed via the Boomi Atom Cloud, public or private cloud, or on-premises. Regardless of deployment option, integration processes are managed directly through the Boomi platform.

**Denodo**

Denodo offers high-performance Data Integration and abstraction across a wide range of Big Data, enterprise, cloud, unstructured data sources and real-time data services. Denodo also enables access to unified business data for agile BI, analytics, and single-view applications. With the Denodo Platform, it is possible to leave all of the source data exactly where it is, stored across a myriad of heterogeneous systems, and to establish a virtual, "logical data lake" for accessing all of the data.

**HVR**

HVR provides real-time data replication for Business Intelligence, Big Data, and hybrid cloud. HVR can solve for a variety of Data Integration use cases including data migrations, Data Lake consolidation, geographic replication, database replication, and cloud integrations. HVR uses data capture between data sources as well as the most commonly used file systems. Beyond generic performance optimizations, HVR offers faster log-based capture from SQL server, improved loads into Teradata and Amazon Redshift, and support for log-based Change Data Capture on open source PostgreSQL.

"HVR provides real-time data integration in a single unified environment."

**IBM**

IBM’s InfoSphere Information Server helps to transform data and deliver it to any system. IBM provides several options for data delivery: ETL (extract, transform, load), virtual (federated), and incremental (data replication). InfoSphere also decreases deployment time using a graphical interface, and allows integration with DMBS, Big Data sources, messaging queues, ERP, and other packaged applications.
Informatica

Informatica’s Data Integration 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 increases productivity and help to ensure that open source platform innovations can be adopted without sacrificing maintainability or reusability. Informatica also provides Data Management and Integration Platform-as-a-Service (iPaaS) functionality that solve a wide variety of enterprise use cases from cloud data warehousing to business process analytics.

Information Builders

With tools for real-time integration, data movement and delivery, and data federation, Information Builders enables organizations to manage both structured and unstructured information. The company’s iWay tools accelerate the deployment of integration projects such as ETL and Big Data processing. Information Builders enables WebFOCUS users to issue distributed queries that correlate and manipulate data from disparate relational databases, packaged applications, structured data files and legacy databases.

Jitterbit

Jitterbit’s graphical “no-coding” approach simplifies the configuration and management of otherwise complex integration projects. 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

The Liaison ALLOY Platform provides unified Data Integration and management capabilities as managed services that can connect any two application endpoints no matter where they reside. The solution stores data in a Big Data repository that provides on-demand self-service access to clean data. The ALLOY Platform is an alternative to traditional ESB or iPaaS in that it allows organizations to gradually migrate legacy integration operations while maintaining security and compliance.

“Uses a Big Data repository that provides on-demand self-service access to clean data.”
**MuleSoft**

MuleSoft is a B2B application delivery network that connects data, applications and devices with APIs. It enables organizations to improve their applications through integration while also providing API connectivity to a wide variety of on-premises and cloud-based applications and systems. The California-based vendor provides both iPaaS and traditional integration tools. Anypoint Platform is MuleSoft’s integration application and API suite. It comprises a toolset that enables organizations to create, integrate, model, build, and deploy services, APIs and applications.

**Pentaho**

Pentaho Data Integration prepares and blends data via a graphical ETL designer that features drag-and-drop connectivity to any data source. The platform features an expansive library of pre-built components to access, prepare and blend data from relational sources, Big Data stores, and enterprise applications. Pentaho accelerates design and deployment of analytics by utilizing a visual data experience from anywhere in the pipeline so users can take advantage of self-service data preparation. Pentaho’s multi-threaded integration engine is architected to scale up and out, including deployment to cluster and cloud environments.

“Accelerates design and deployment of analytics by utilizing a visual data experience.”

**Scribe Software**

Scribe offers no-coding data tools that can be used to integrate virtually any application, data source or Software as a Service platform. Scribe Online is the company’s cloud-based Integration Platform as a Service (iPaaS) tool that provides application integration via a graphical interface for developers and business analysts. Scribe Insight connects business systems with a suite of on-demand adapters for CRM and ERP applications to streamline integration projects for users that prefer to work on-premises.

**Software AG**

Software AG webMethods Integration Platform is a pre-integrated suite that features an Enterprise Service Bus for integration of systems, services, processes, business partners and data to improve business performance. In addition to application, cloud and B2B integration, the platform manages file transfers, master data management and mobile applications.
**Tamr**

The Tamr Enterprise Data Preparation Platform is comprised of three core capabilities that allow cataloging, connection and consumption to unify and process disparate data. The tool catalogs metadata associated with internal and external sources in a platform-neutral location so users can discover data and match it against business requirements. Tamr identifies data quality issues across all sources regardless of size or format so users can standardize values and fix blank or missing data.

**Talend**

Talend’s 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. 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.

**Trifacta**

Trifacta presents automated visual representations of data based on its content in an interactive format that enables users to select specific elements of the profile to prompt transformation. The data visualization provides exploration of the data in a granular way that makes it easy to spot outliers and invalid data. Trifacta Data Wrangler was designed for non-technical users and provides support for open source and vendor-specific security, metadata management and governance frameworks.

“Trifacta Data Wrangler was designed for non-technical users.”

**Youredi**

Youredi connects to all IT systems and can be hosted in an organization’s private data center or in the cloud. Youredi Integration Platform as a Service (iPaaS) provides adapters that simplify connecting to on-premise and public applications. The platform also makes it possible to combine traditional batch-based transfers with modern event-based protocols including large-volume IoT scenarios. This effectively enables businesses to combine legacy data with modern interfaces for real-time integration.
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’s mission is to connect buyers of enterprise technology with the best solution sellers.

Over the past three years, Solutions Review has launched 14 tech Buyer’s Guide sites in categories ranging from Cybersecurity to Mobility Management, Business Intelligence, Data Integration, Cloud Platforms and Content Management.

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.