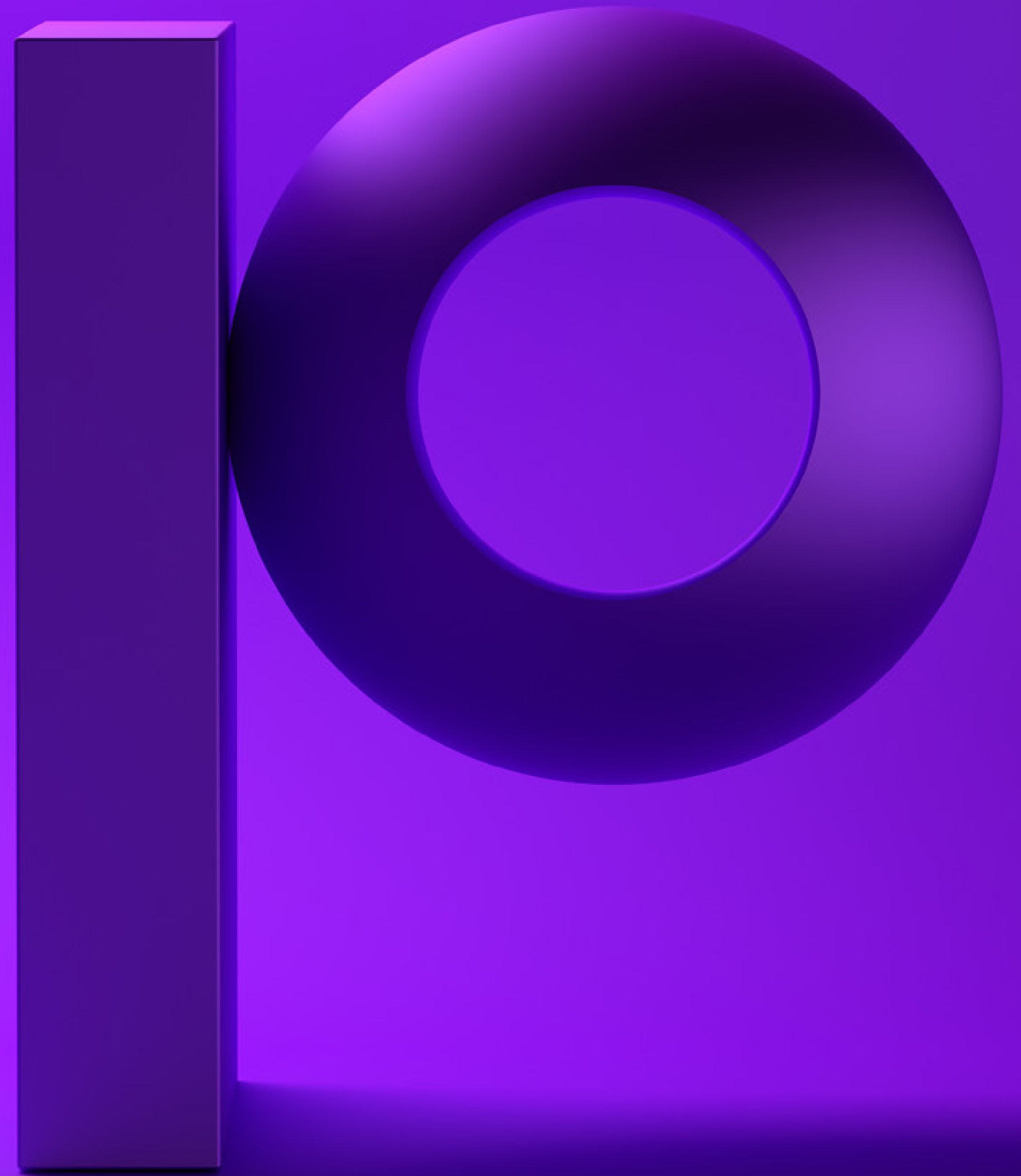precisely

Buyer's Guide and Checklist for Data Integration

Snapshot: The data integration landscape today

Gone are the days of sole batch ETL (extract, transform, and load) with the help of a few skilled developers fulfilling data integration requirements. A new dynamic and fluid model of data integration has taken its place – bringing data from across a business to users when and how they need it. Much of the changed approach is driven by a broader diversity of cloud data consumption models, as well as a surge in the number and types of applications demanding real-time data delivery.

Cloud delivery models and applications have become part of every organization's strategic business strategy, allowing them to expand capabilities, reduce costs, and drive digital transformation. However, bringing the right data from existing infrastructure to the cloud for business consumption can be an impossible task for many organizations. Cloud, and the benefits that it promises, requires a new way of thinking about data integration.

As a data integration leader in your organization, you can leverage moving to cloud as an opportunity to modernize existing integration approaches. To modernize, you cannot merely "copy" and "paste" existing data integration pipelines – traditional data pipelines are not adaptive enough. Instead, looking for tools that can build links between existing infrastructure and new cloud investments in an environment-agnostic, future-proof way is imperative. You need only look at what's transpired over the past ten years to understand why.

The rise of Hadoop and its broad array of on-premises offerings helped organizations to replace their traditional enterprise data warehouses. The early 2010s saw MapReduce fall to Spark and the introduction of an entirely new programming paradigm altogether. Shortly after that, a variety of managed Spark offerings from vendors like Databricks, AWS, Azure, and Google Cloud Platform promised improved agility and scalability at lower TCO. Successfully navigating these waves of revolution requires software that allows you to build integration frameworks that work seamlessly across multiple environments.

The rise of the cloud data warehouse has brought with it promises of unlimited scalability, unrivaled user concurrency, zero administrative overhead, and improved data sharing across the organization. Delivering on these promises requires data pipelines to provide access to information around the business, including legacy systems such as mainframe and IBM i. As a result, investing in data integration software that can natively connect to both cloud and legacy sources is imperative.

That said, data integration nirvana isn't just connecting data from legacy platforms to the cloud, or providing a centralized source of all data – it's also about delivering data in real-time. As the speed and volume of transactions accelerate to support your business, data is continuously generated, accessed, processed, and stored across your entire IT landscape – much of it containing latent, unmined insights. To realize the full value of your all your data, it must be provided to users, applications, and data consumers downstream, when and where it's needed. To do this, your data integration tool must include real-time replication capabilities. The software should be able to communicate changes to data, regardless of the data source or target, without negatively impacting the performance of the source transactional systems via intrusive triggers.

Whether or not you already have a set of data integration tools, what follows is a checklist designed to help you evaluate data integration vendors and software that will help you meet the challenges of the new data integration paradigm, while complementing and optimizing both current and future investments.

Learn more about Connect

Buying the right data integration solution

A checklist for choosing the best vendor and software for your organization's needs

Checklist for client-vendor partnership

This checklist has been created to help you to analyze data integration vendors and their offerings. It should help you to determine the software solution that is right for your organization. Regardless of your choice, the purpose of this check list is to build the best client-vendor relationship that help you meet your goals.

Where to start

When working with any vendor, you should know exactly what you are, and are not, getting. As you begin to consider different vendors, make sure to evaluate the following.

- **Strategic direction** – Ask where the vendor is headed. Are they taking steps to ensure they are innovating with the times? Understand how they address cloud, advanced analytics and machine learning integration requirements. Find out about their strategy and approach to delivering the latest features and capabilities to customers.

- **Partnership** – A vendor should not only supply software but be a partner in your data integration transformation. Determine the willingness of vendor to partner with you to understand your unique business or technical needs.

- **Flexibility with existing investments** – Don't be shy. Ask lots of questions about how the vendor can integrate into your existing environments and technology stack. Discuss your specific environment and what you need to get done.

- **Industry contributions** – Make sure you're working with a leader, not a follower. Ask about vendor contributions to the wider technology landscape, as they are more likely to provide deeper integration and support for the rapidly evolving tech ecosystem.

- **Demos and references** – Ask for demos and customer references. Talk to industry experts; get their recommendations.

- **Training and support** – From innovative technologies to new employees, there's always a need for training. Make sure you have access to the continued support you need.

10 key features to look for

Data integration solutions can remain in place for years. The following checklist can help you choose
a vendor that offers both software and an approach that can grow and change with your organization.

☐ **Wide support for enterprise-grade sources and targets** – VSAM. COBOL copybooks. JSON. Cloud
data warehouses. Kafka. The list goes on. A vendor should have technology to support a wide range
of enterprise data sources and targets for all your ETL and CDC needs.

☐ **Quickly and easily add sources or targets** – The software should let you quickly respond to new
requirements with ease. Users should be able to add sources and targets at the click of a button.
Ask about the process and amount of time required to add new or existing sources and targets.

☐ **Tech stack integration** – Can the vendor easily integrate their products into your existing technology
stack? Ask the vendor if they have certified and proven solutions with technology leaders (ex. Cloudera,
Databricks, Snowflake, and more).

☐ **Ease of deployment** – An experienced vendor should offer solutions that enable quick deployment.
Make sure you understand what to expect around both deployment and support. The deployment
should in no way slow down your operations. Always check references.

☐ **Design once, deploy anywhere approach** – Look for vendors with a record of providing customers
with a path to new platforms without bringing down users or service interruptions. Ask how much work
is involved when moving from one platform to another. The software should support a variety of platforms
and enable movement between platforms with no manual intervention.

☐ **Exceptional performance and scalability** – Look for software that scales to accommodate growing
volumes of data, users, and unpredictable peak usage demands. Ask vendors how they deliver both
predictable performance and scalability.

☐ **Resiliency and backup** – Data integration software should help you focus on your business, not your
systems. Ask how vendor solutions provide data delivery resiliency and back up in the event of a service
disruption. Understand if manual coding or tuning is necessary for bringing systems back online.

☐ **Proper security and governance** – Ask how data is protected when moving through the data pipeline.
Ask about the data governance controls in place to ensure compliance throughout the data pipeline.
A vendor should be able to present proof of secure data transfer and have integrations with top security
software to provide data protection.

☐ **Future proof investments** – You shouldn't need to continually throw resources at complex deployments
(ex. integrations across cloud, hybrid, or on-premises environments). The software you choose should take
a development approach that helps insulate you from future technology requirements and disruption
without coding, tuning, or redevelopment work.

☐ **Fast time to value** – Implementation of data integration software should not increase tech stack bloat.
Deployments should require no specialized skills, be resource-efficient, and targeted to your use case.
Dig into the ease of use of a vendor offering to ensure you are getting the fastest time to value possible.

precisely

Precisely is a global leader in data integrity, providing accuracy and consistency in data for 12,000 customers in more than 100 countries, including 90 percent of the Fortune 100. Precisely enables companies to integrate, verify, locate, and enrich their data to power better business decisions. To learn more, visit www.precisely.com.

www.precisely.com