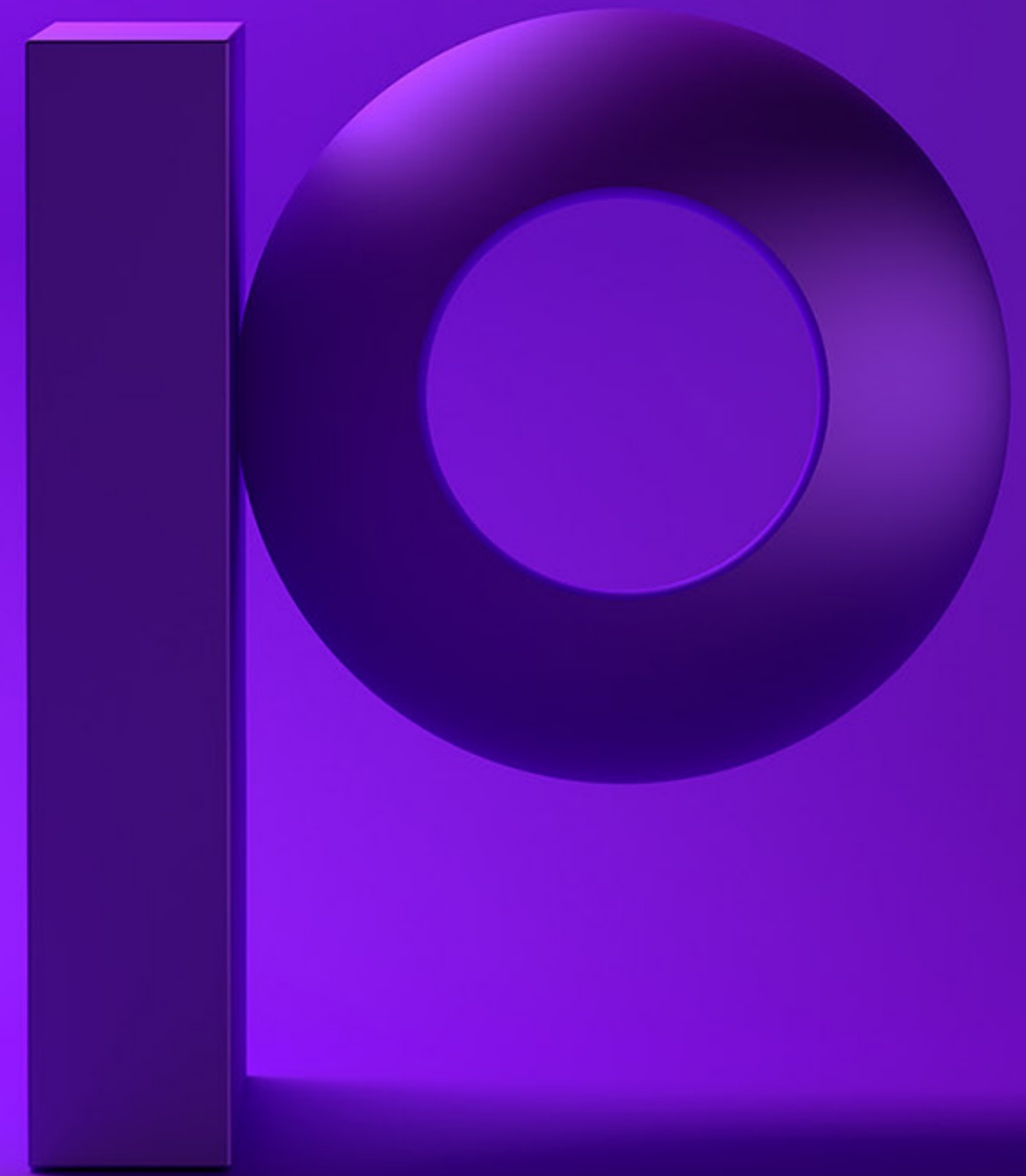


precisely

Buyer's Guide and Checklist for Data Integration

2026 Edition —Agentic AI Readiness



Introduction

Data integration is no longer just infrastructure — it's the foundation of AI readiness.

For years, buyers judged data integration platforms mainly by one question: Which vendor could best move workloads and data from on-premises systems to the cloud? That question still matters, but it's no longer the one that matters most. Today, enterprise data leaders are asking something different, and there's much more at stake.

Can my data integration strategy support autonomous AI system operating at enterprise scale?

As a data integration leader, you're navigating a fast-moving landscape, and the choices you make now will determine how ready your organization is for what comes next.

Read this eBook to discover 10 key features to help you choose a vendor that offers both software and an approach that can grow and change with your organization, with recommendations on:

- **From cloud migration to agentic AI readiness** — what it means to have integrated, governed, and enriched data as a prerequisite for autonomous AI
- **From fragmented data pipelines to continuous, automated data delivery** — including real-time replication, cloud-native ETL/ELT, and low-code transformation
- **From point-in-time compliance to ongoing data governance** — as a runtime requirement for AI systems

The Data Integration Landscape Today

Precisely's 2026 State of Data Integrity and AI Readiness report, produced with Drexel University's LeBow College of Business, found that organizations are significantly overestimating their AI readiness, with data integrity gaps standing out as the primary bottleneck. Closing that gap, not simply adopting more AI tools, is what will separate organizations that scale agentic AI from those that stall.



1. The agentic AI trigger

Just a few years ago, the challenge was feeding AI models relevant data. Now, the challenge is feeding autonomous AI agents real-time, governed, trustworthy data — continuously. Agents don't wait for a pipeline refresh cycle; they need data that is always-on and always-trusted. Analyst research has consistently found that most agentic AI pilots fail not because of the AI itself, but because the underlying data isn't clean, current, or governable enough. When agents act on bad data, the cost isn't a bad dashboard, it's a bad decision made autonomously, at machine speed, before anyone notices.

2. The pipeline fragmentation crisis

Organizations managing data across on-premises systems, multiple clouds, and SaaS applications commonly run five to ten or more disconnected ETL tools. The result is a governance blind spot: no single team can see, let alone certify, how data actually moves across the enterprise. Buyers should ask a simple but pointed question: How many tools does it take to move, transform, and trust my data?

3. The rise of low-code, cloud-native ETL

The shift to cloud-native ETL/ELT means data teams no longer need specialized developer resources to build and manage pipelines. Modern platforms offer visual, drag-and-drop pipeline builders that dramatically reduce time to value. Buyers should evaluate whether vendors have embraced this model or are still relying on legacy ETL paradigms. ETL powered by Matillion, for example, brings embedded low-code ETL natively into the Data Integrity Suite — one illustration of what this shift looks like in practice.

4. Legacy data is still mission-critical

Despite cloud-first narratives, mainframe, IBM i, SAP, and Oracle systems still power the most critical enterprise transactions. Any modern integration strategy that ignores these sources is incomplete. The best vendors bridge legacy and cloud natively without compromise or separate tooling.

5. The MCP moment

Model Context Protocol (MCP) is rapidly becoming the standard by which AI agents access and act on enterprise data. Buyers should ask vendors whether their platforms expose data via MCP, enabling AI agents to access trusted data programmatically without custom engineering overhead. Vendors who don't support MCP will increasingly require custom integration work for every new agent framework that comes to market; a cost that compounds as the agent ecosystem grows, and one that widens the gap between MCP-native platforms and everyone else.



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

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

- ❑ **Strategic direction** — Today, having an AI strategy isn't enough. The important question is: What is the vendor's roadmap for agentic AI readiness, and how does data integration fit within it? Vendors should articulate how their integration offering connects to broader AI and analytics outcomes, not just data movement.
- ❑ **Partnership** — A vendor that partners with the cloud platforms and AI tools you already use, such as Snowflake, Databricks, Confluent, AWS, and Azure, is more valuable than one that requires you to adapt your stack to theirs. Ask about certified, production-tested integrations, not just claimed compatibility — and ask how the vendor plans to maintain and expand those partnerships as the ecosystem continues to evolve

- ❑ **Flexibility with existing investments** — With so many emerging AI and data platforms, flexibility is more critical than ever. Ask vendors explicitly: Can I switch my AI model, cloud target, or analytics platform without rearchitecting my integration layer?
- ❑ **Industry contributions** — Vendors contributing to open standards such as MCP, Apache Kafka, Databricks Delta Lake, and Apache Iceberg are better positioned to keep pace with a rapidly evolving landscape. Make sure you're working with a leader, not a follower.
- ❑ **Demos and references** — Ask specifically for references in your industry and for your specific use case. Ask vendors to demonstrate integration across legacy and cloud sources, not just cloud-to-cloud scenarios.
- ❑ **Training and support** — Today, data democratization means business analysts and data stewards need to work within integration platforms alongside developers. Ask whether the platform offers AI-assisted interfaces that lower the skill floor.
- ❑ **AI governance and explainability** — As autonomous agents use data integration pipelines to act on business decisions, buyers need to understand how vendors enable governance inside the integration layer. Ask how the platform ensures that data moving through pipelines is tagged, audited, and traceable for compliance purposes.
- ❑ **Agentic readiness** — Ask vendors specifically whether their platform is designed to serve AI agents, not just human users. Can your data integration solution expose pipelines, schemas, and data products via APIs or MCP so that agents can access and act on trusted data without manual handoffs?

Questions to ask vendors:

- What is your roadmap for agentic AI readiness, and how does data integration fit within it?
- Can I switch my AI model, cloud target, or analytics platform without rearchitecting my integration layer?
- How is data moving through your pipelines tagged, audited, and made traceable for compliance purposes?

A Note on the Competitive Landscape

The vendor landscape has shifted quickly and significantly — including several major consolidation events that buyers should factor into any evaluation.

The data integration market is consolidating around two models:

- **Specialized, best-of-breed platforms:**
Purpose-built tools, cloud-native ETL, automated connectors, and transformation frameworks that do one thing well and integrate broadly across stacks. This approach can offer speed and flexibility for teams with focused, modern use cases. Worth noting: this space is itself consolidating, with major players merging to offer more complete ingestion-to-transformation pipelines — narrowing the gap between best-of-breed and unified platform approaches.
- **Unified data management platforms:**
Platforms that combine integration, quality, governance, enrichment, and AI readiness in a single, interoperable environment. This approach reduces tool sprawl and simplifies governance — important for organizations with complex hybrid environments or stringent compliance requirements. As consolidation continues, buyers should evaluate not just current capabilities but long-term vendor independence: when a data management platform becomes part of a larger ecosystem, roadmap priorities, pricing structures, and support for non-native integrations can shift in ways that affect customers over time.

Neither model is inherently superior. The right choice depends on your organization's complexity, governance requirements, and AI ambitions. Key questions to ask yourself:

- How many integration tools does my organization currently run? (If more than three or four, fragmentation costs may be significant.)
- Do I have legacy sources such as mainframe, IBM i, SAP, or Oracle that require specialized connectors and deep platform expertise?
- Are my AI initiatives likely to require governed, auditable data pipelines — or is exploratory analytics the primary near-term use case?
- How much do I value flexibility in AI model and cloud platform choices versus the simplicity of a tighter, single-vendor experience?
- What is the vendor's ownership structure, and how might that affect their roadmap and pricing over your contract horizon?

10 Key Features to Look For

Successful data integration solutions should be able to 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, including support for autonomous AI.

- ❑ **Wide support for enterprise-grade sources and targets** — from VSAM and COBOL copybooks to JSON and Kafka, and from Databricks, Snowflake, Azure Synapse, and Google BigQuery to lakehouse formats like Delta Lake and Apache Iceberg. Critically, don't lose sight of legacy sources: mainframe (z/OS), IBM i, SAP, and Oracle. Today, AI readiness requires connecting all of these, not just the modern ones. An integration strategy that only works cloud-to-cloud leaves critical enterprise data behind.
- ❑ **Quickly and easily add sources or targets** — Look for low-code and no-code pipeline creation. Today's business users and data stewards are increasingly involved in pipeline configuration alongside technical teams, so vendors should offer visual, drag-and-drop pipeline builders rather than just developer-facing APIs. Ask vendors: Who can build a new pipeline on your platform, and how long will it take?
- ❑ **Tech stack integration** — Look for certified integrations across the AI and analytics ecosystem, not just AWS, Databricks, Snowflake, and Confluent, but AI platforms (AWS Bedrock, Azure OpenAI, Google Vertex AI), orchestration layers (Apache Airflow, dbt), and the AI agent frameworks your organization is beginning to adopt. Keep in mind that the ecosystem is shifting — Fivetran and dbt Labs completed their merger in June 2026, so evaluate whether integrations you rely on today will remain independently supported as consolidation continues. Ask specifically: Does the vendor support MCP? MCP is rapidly becoming the standard for AI agents to access enterprise data, and vendors who expose integration capabilities via MCP give agents direct, governed access to pipelines without custom engineering, a differentiator that matters more every quarter.
- ❑ **Ease of deployment** — An experienced vendor should offer solutions that enable quick deployment, but deployment speed alone isn't the full picture. A platform that's fast to deploy but requires months of data cleanup before it can feed an AI initiative hasn't solved your actual problem. Ask: How long does it typically take to go from deployment to the production of AI-ready data outputs?
- ❑ **Design once, deploy anywhere** — This principle is more relevant than ever with hybrid and multi-cloud environments as the norm. Look for unified pipeline design across hybrid and multi-cloud environments, including on-premises legacy systems and ask whether "design once" extends to AI model targets. Can the same pipeline serve a Snowflake analytics workflow and an AI agent running on a separate model without redesign? That architectural flexibility will determine how quickly you can pivot as AI technology evolves.
- ❑ **Exceptional performance and scalability** — Look for software that scales to accommodate growing volumes of data and users, as well as unpredictable peak usage. Ask about pushdown transformation execution, where transformations run within the cloud data platform itself (Snowflake, Databricks) rather than in the integration layer, maximizing performance and minimizing data movement costs.
- ❑ **Resiliency and backup** — Think of this as data pipeline continuity, not just systems resiliency. In an agentic AI environment, a broken pipeline doesn't just create a reporting gap; it can cause AI agents to act on stale or incomplete data, with real operational consequences. Ask: How does your platform detect, alert on, and recover from pipeline failures, and is any of that recovery automated? The most advanced platforms today are beginning to use AI agents themselves to monitor and self-heal pipelines — a significant operational advantage worth asking about. Precisely's Data Integration Agent, for example, automates monitoring of data pipelines and can detect and flag inconsistencies.

❑ **Proper security and governance** — This is the item that has changed the most. Today data governance isn't just a compliance checkbox, it's a prerequisite for safe AI deployment. Look for:

- **Data lineage and auditability** — Can the platform track data from the source to the AI model to the business decision? Regulators and internal risk teams increasingly require this end-to-end visibility.
- **Role-based access control for machine actors** — As AI agents access data programmatically, governance must extend beyond human users. Ask how vendors manage access permissions for automated systems and agents.
- **EU AI Act implications** — Fully applicable as of August 2026, the EU AI Act requires organizations to demonstrate that data used in AI systems is governed, traceable, and compliant. Your integration layer is the first place where this must be enforced. Ask vendors specifically what capabilities they offer to support EU AI Act compliance.
- **Shadow data risks** — Ungoverned pipelines feeding AI systems represent one of the fastest-growing enterprise risk categories. Ask vendors how they surface and govern data flows that may have developed outside official IT processes.

Questions to ask vendors:

- How do you support compliance with the EU AI Act's data governance and traceability requirements?
- How do you manage access permissions for AI agents and other automated systems, not just human users?
- How do you help us surface and govern shadow data flows that develop outside official IT processes?

❑ **Future-proof investments** — In the past "future-proof" meant avoiding rework when migrating between platforms. Today, it means being agentic-ready — building an integration architecture that can serve not just today's analytics needs but tomorrow's autonomous AI systems. The most forward-looking organizations aren't asking "can this platform replicate data?" anymore, they're asking "can this platform produce data that AI agents can trust, access, and act on autonomously?"

Look for:

- A **data product model** — packaged, governed, reusable data assets that can be shared across teams and consumed by AI agents without rebuilding pipelines.
- Data exposed via **APIs and MCP**, enabling AI agents to access integration capabilities programmatically.
- A roadmap oriented around **reducing manual effort** for routine data tasks through intelligent automation, not just adding AI features on top of legacy architecture.

❑ **Fast time to value** — Implementing data integration software shouldn't increase tech stack bloat. Deployments should require no specialized skills, be resource-efficient, and be targeted to your use case. Look for AI-assisted onboarding and configuration. Leading platforms now include AI assistants that help users configure pipelines, write transformation logic, and troubleshoot issues in natural language, dramatically reducing the time from deployment to productivity. Ask: Does your platform include AI-assisted setup and natural language interfaces for non-technical users? This is an area where the gap between vendors is widening quickly. Precisely's Gio AI Assistant, for example, enables both technical and business users to manage data integration tasks through conversational interaction.



As a global leader in data integrity, Precisely ensures that your data is accurate, consistent, and contextual. Our portfolio, featuring the Precisely Data Integrity Suite, brings together software, data, and data strategy consulting services. This unique combination enables organizations to move to Agentic-Ready Data, the highest-quality of data that is integrated, governed, and enriched for AI, automation, and analytics initiatives at enterprise scale. Over 12,000 organizations in more than 100 countries, including 95 of the Fortune 100, trust Precisely to support some of the world's most complex, regulated, and mission-critical data environments.

Learn more at www.precisely.com.

©2026 Precisely. All other company and product names used herein may be the trademarks of their respective companies.