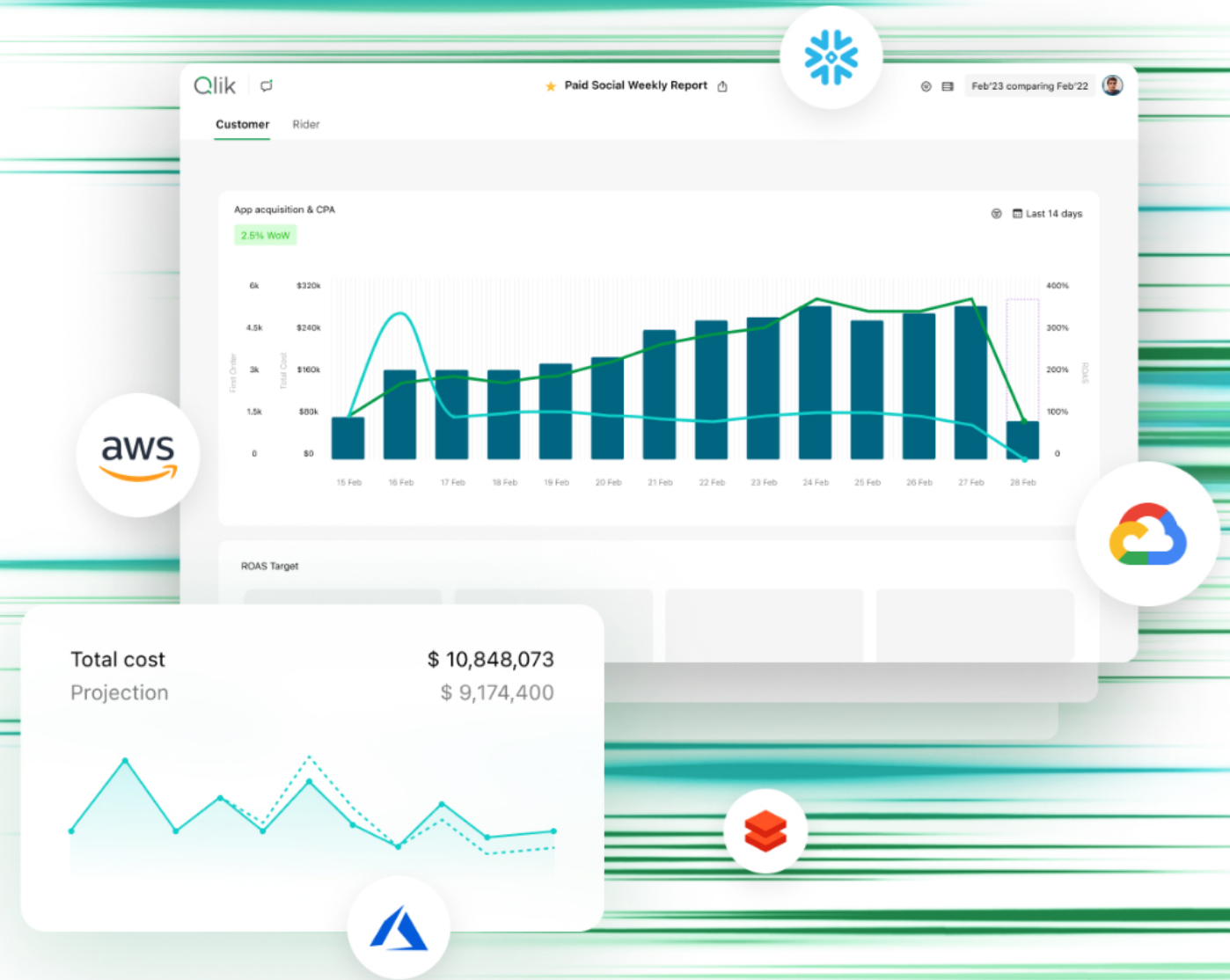




DATA INTEGRATION BUYER'S GUIDE

What To Look for in a Data Integration Solution



Meeting – and mastering – today’s data challenges

How to find the right data integration solution for your business

In business today, it’s vital to deliver clean, usable data for not only analytics but also AI, machine learning, and data science initiatives. But with more data and more complexity than ever before, that task can feel like a moving target:

- Data sources, varieties, and volumes are multiplying all the time
- Data needs to be secure and governed across the entire pipeline
- The underlying architecture has to have the flexibility to scale
- Costs, usability, and support must also be considered

The good news is, modern data technology can handle all of the above. With the right data integration solution, you’ll be well equipped to deliver.

How can you find the data integration solution that will perform best for your organization? This guide reveals the criteria you’ll need – including high-level considerations, use cases, and capabilities at the product and platform levels.

¹ Randy Bean, Harvard Business Review, “Why becoming a data-driven organization is so hard,” February 24, 2022, <https://hbr.org/2022/02/why-becoming-a-data-driven-organization-is-so-hard>.

² McKinsey Digital, “The data dividend: Fueling generative AI”, September 15, 2023, <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-data-dividend-fueling-generative-ai>



Increasingly, companies must come to recognize and appreciate that data is a business asset that flows through an organization.”¹

HARVARD BUSINESS REVIEW



If your data isn’t ready for generative AI, your business isn’t ready for generative AI”.²

MCKINSEY & COMPANY

Before you begin

Before you evaluate features and capabilities, get clear on the basics: what you're trying to do and how much it will actually cost.

Your goals

Prioritize your list based on needs.

- Modernize legacy applications
- Deliver real-time data from existing systems
- Provide a strong foundation for generative AI initiatives with quality, governed data
- Deliver trusted contextual data to LLMs
- Provide quality data to Machine Learning and Prompt Engineers
- Scale more easily when demands increase
- Replace coding and highly technical tasks with self-service capabilities
- Support more advanced analytics
- Reduce infrastructure costs

Total cost of ownership

Look at the whole picture, not just the initial price.

If SaaS:

- Annual or ongoing subscription costs
- User training and enablement
- Implementation and ongoing support costs

If on-premises:

- Software subscription and maintenance costs, including third-party products and underlying technologies
- Hardware costs, including networking, storage, and servers for development and maintenance
- IT resources and overhead costs
- Implementation and ongoing support costs
- User training and enablement

What to consider

From use cases to user communities, it's important to look at every aspect of a data integration platform.



USE CASES

See a list of the most common use cases within data integration.



CORE CAPABILITIES

Make sure you're ticking every box when it comes to functionality, down to the all-important details.



SOLUTION MUST-HAVES

Check out a list of all the non-negotiable features you'll need at the solution level.



USABILITY AND PRICING

Find out what to look for in terms of ease of use and clarity in pricing.



SUPPLIER SERVICES

Learn what to consider in a solution provider, including service, support, and the user community.



CHANGING THE GAME

Discover how you can revolutionize data delivery.



USE CASES

Consider current and future needs

Look for a data integration solution that grows as your needs change, accommodating all your use cases within a unified, governed framework. Common use cases include:

Data streaming

- Replicating data from multiple sources like databases, data warehouses, and SaaS-based applications to targets like a cloud data platform
- Replicating data between databases for consistent and up-to-date data integration across systems and applications on a one-to-one, one-to-many (and vice versa), or even many-to-many basis
- Integrating data from SAP/mainframe/legacy or other sources with microservices-based development
- Replicating data between on-premises, SaaS applications, and cloud databases for consistent and up-to-date target data in the cloud
- Integrating with event streaming platforms, such as Apache Kafka or Amazon Kinesis, and synchronizing data between databases and other data targets

Data warehouse / lake modernization

- Facilitating the migration and operationalization from traditional relational databases and EDWs to cloud data warehouses (e.g., Snowflake, AWS Redshift, and Azure Synapse), lakehouses (e.g. Databricks, Apache Iceberg), and platforms (e.g. Microsoft Fabric)
- Streamlining the process of implementing, managing, and updating data warehouses, data lakehouses, and data marts
- Rapidly propagating source or model changes through the data warehouse/lakehouse environment
- If necessary, quickly adapting to data management technologies of the future

Data quality and governance

- Profiling your data to help you discover data quality issues, risks, and overall trends
- Standardizing formats and matching records to ensure you have the data ready for analytics and AI initiatives
- Establishing a metadata-powered catalog to enforce consistency and foster collaboration
- Enriching and augmenting data so that users can get a more complete understanding of the information they're consuming
- Defining the models and rules to validate data and resolve data errors



CORE CAPABILITIES

Dig into the details

Take a close look at what each solution offers, and connect capabilities with the outcomes they enable. Here are key options to consider.

Data movement

- Uni-, bi-, and multi-directional movement of data across endpoints
- Batch, micro-batch, and continuous delivery of data
- Ability to easily change sources or targets as data requirements evolve
- Log-based change data capture (CDC) with zero footprint (to minimize source impact)
- Support for popular data streaming services such as Apache Kafka, Confluent, Amazon Kinesis, Azure Event Hub, and Google Pub/Sub
- Rapid, agentless deployment
- Automated, continuous, real-time data replication
- Ability to automate many processes via APIs
- A low code/no code approach to configuring data pipelines
- Immediate availability of production data for analytics and microservices
- Auto-mapping heterogeneous sources to a single consumption format





CORE CAPABILITIES

Data transformation

- Create reusable transformations and control how they are generated. Materialization options to control cost and performance
- Automatic SQL generation that conforms data to standardized formats
- Fit-for-purposes transformations, including source-to-target mappings, rule-based row-level transformations, star schema, column standardization, and custom SQL
- Support for popular cloud data warehouses like Azure Synapse, Google BigQuery, AWS Redshift, and Snowflake
- Refinement and merging of data into analytics-ready structures
- Data transformation to common formats such as AVRO, Parquet, CSV, and QVD
- Lightweight transformations performed on the fly (e.g., filtering by date) as well as global transformations (e.g., consistency with one date format)
- Standardize metadata, fill gaps, and generate keys to easily transform values, narrow datasets, mask information, validate, and uncover relationships effortlessly
- Conform values to rules, such as “Match schema types,” “Set missing value to null,” “Uppercase,” “Split,” and “Character replace”
- Advanced data pipeline creation – such as Type 2 transformation – based on policy rules, active metadata, and reusable patterns
- Handle diverse data sources with agility and flexibility by mixing and matching data for various use cases regardless of the data ingestion tool or process.
- Automatic data warehouse and mart creation — customize marts, optimize incremental processing, and deduce data relationships
- Write custom SQL code to describe data relationships and manipulate data. Utilize Common Table Expressions (CTEs) or use aliases and parameters to select and transform data
- ACL-based transformations that enable different scopes, including personal, departmental, tenant/environment, or company



CORE CAPABILITIES

Data governance

- A data governance framework that can set accountabilities and delegate authority to sufficiently manage data privacy and data access
- A complete catalog of metadata associated with each data source
- Preserved lineage so users can understand the origin, evolution, and meaning of datasets
- Analysis of unique data fields' impact on direct or indirect downstream datasets and analytic applications
- Single and multi-faceted search so one can optionally specify different parameters to perform an advanced search
- A business glossary to bridge the gap between technical metadata and business terminology, and align the understanding of business terms across your entire organization
- Sufficient governance so that users can access only authorized data
- Data classification to selectively protect sensitive data such as personally identifiable information (PII) and, if necessary, the ability to obscure it from general viewing
- The ability to interact with multiple BI tools
- The ability to manage other related assets, such as analytic applications, pre-defined workflows, and collaborative notes





CORE CAPABILITIES

Data quality

- Data profiling to assess your data for completeness, accuracy, timeliness, and consistency
- Data observability, which is not only focused on completeness, accuracy, uniqueness, timeliness, and anomalies; but also data volume changes in data pipelines, data schema changes and lineage, and data availability
- Robust data cleansing capabilities, including the ability to remove duplicates, correct errors, and standardize data
- Data protection to safeguard the privacy, availability, and integrity of sensitive data
- Data enrichment to add additional data elements to the dataset that will provide more comprehensive information for analysis and decision making
- Data validation by applying business-centric rules to ensure that the data conforms to various standards and is within the acceptable business parameters
- Data standardization to ensure that the solution can enforce established standards for data format, structure, and content



Evaluate the solution at a system level

After you check the boxes on individual features, think about what you'll need from the solution as a whole.

To lay a solid foundation for success, look for the following:

AI Readiness

- Data pipeline productivity features like natural language to SQL functionality
- Data transformation features that create datasets for machine learning (ML) projects and Large Language Model (LLM) initiatives
- Support for Vector Databases such as Elasticsearch, Neo4J, Pinecone, PostgreSQL, Redis, and Snowflake

DevOps

- Utilize a broad collection of APIs and an event-driven architecture to develop new integrations and applications
- The ability to use advanced data mapping transformation tools and industry standards such as HL7 and EDI

- Native support for CI/CD deployment and containerized micro-services
- Process workflow automation capabilities

Administration

- A centralized control panel for efficiently managing data replication and transformation processes
- The ability to design, execute, monitor, and analyze integration tasks across multiple environments
- Seamless integration with larger IT systems and practices to simplify enterprise-wide management tasks and boost security and compliance
- Management of high-scale data consolidation across servers and end points to help meet/exceed service-level agreements

Flexible deployment options

- Portable data flow design across multiple infrastructures (on-premises, SaaS, CSP, and VPC)
- Cloud, on-premises, and hybrid deployment options
- The ability to bring together data from multiple clouds
- Independent data storage (with no lock-in)

Scalability

- Deployment options that protect data location and local governance needs without impacting performance or scale
- Elastic scalability to meet the peak demands of data, users, and complex use cases



USABILITY AND PRICING

Look for ease of use and straightforward pricing

A solution that plays well with users will increase adoption, and pricing transparency is good for business.

Adoption and usability

- Simple, intuitive, and consistent user experience across product areas
- A smooth, progressive increase in UX complexity as users need and want it
- The right balance between self-service (offering a full range of capabilities) and IT administration and oversight

Lines of business

- Defined solutions/offerings for business function areas (sales, finance, IT, etc.)
- App templates and starter apps
- Industry and functional experts with deep domain experience

Pricing and packaging

- Simple pricing and packaging
- Subscription-based pricing options
- Easy-to-understand upgrade options
- Pricing metrics based on actual system usage (ex. consumption or capacity)



Consider the vendor

A strong data integration provider will offer services and community well beyond the product, and industry analysts like Gartner can help you situate the vendor in the marketplace.

Services, training, and support

- Consulting services spanning requirements, development, and deployment
- A wide variety of training and enablement, on demand and in person
- 24/7 proactive worldwide support
- Partner and ecosystem support

Reputation in the industry

- Respect from major industry analysts
- A broad base of satisfied and loyal customers
- A known commitment to customer success, including ongoing engagement
- A clear roadmap for planned features and innovations

User community and ecosystem

- Active community of users for knowledge sharing
- Robust ecosystem of partners focusing on specific industries and use cases
- Online marketplaces for partner solutions and offerings

Hear from the experts

The Gartner® Magic Quadrant™ for Data Integration Tools provides a view of the entire data integration landscape, enabling you to quickly compare tools and see how they align to Gartner's criteria for completeness of vision and ability to execute.

[Gartner Magic Quadrant for Data Integration Tools](#)

Qlik® commissioned Forrester Consulting to run a Total Economic Impact™ study of the potential ROI of deploying the Qlik Talend Data Integration Solution. You can read key findings and get a framework for evaluating the potential savings in your own organization.

[The Total Economic Impact™ of Qlik Talend](#)



A trusted foundation for data delivery

Every business has unique needs around data integration, and by clarifying yours, you'll be equipped to make the best decision in a solution. But no matter how you've set up your infrastructure or what industry you're in, you can benefit from three modern capabilities that have revolutionized data delivery — dramatically increasing speed, accuracy, and usability:

✓ **Real-time data movement**

Deliver large volumes of real-time analytics- and AI-ready data into streaming and cloud platforms, data warehouses, and data lakehouses.

✓ **Automated analytics data pipelines**

Deploy scalable data-to-analytics pipelines that transform data with pushdown SQL.

✓ **Complete data quality and governance for explainable analytics and trusted, generative AI results**

Discover, remediate, and share clean data to build trust and drive usage of your AI and analytics initiatives.

At Qlik, we lead the industry in all of the above. By automating data streaming, refinement, quality, and governance, we make it possible to deliver business-ready data in real time to power your AI, ML, and analytics solutions. As a result, you can close the gaps between data, insights, and action — enabling your teams to respond in the moment with data-informed business moves.

Curious to learn more?

Qlik Talend™ Data Integration and Quality



Ready to see what Qlik can do for your business?

Get in Touch



About Qlik

Qlik transforms complex data landscapes into actionable insights, driving strategic business outcomes. Serving over 40,000 global customers, our portfolio leverages advanced, enterprise-grade AI/ML and pervasive data quality. We excel in data integration and governance, offering comprehensive solutions that work with diverse data sources. Intuitive and real-time analytics from Qlik uncover hidden patterns, empowering teams to address complex challenges and seize new opportunities. Our AI/ML tools, both practical and scalable, lead to better decisions, faster. As strategic partners, our platform-agnostic technology and expertise make our customers more competitive.

qlik.com