

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 get data to users fast. But with more data and more complexity than ever before, that task can feel like a moving target:

- Data sources, varieties, and volume are multiplying all the time
- Actionable data is needed not only for analytics but also for data science, AI, and machine learning initiatives
- Data needs to be secure and governed across the entire pipeline
- The underlying architecture has to have the flexibility to scale
- And you’ve got to consider costs, usability, and support, too

The good news is, modern data technology can handle all of the above. And 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.

“Increasingly, companies must come to recognize and appreciate that data is a business asset that flows through an organization. Data cuts across traditional organizational boundaries, often without clear ownership. The fluidity of data compounds the complexity of managing this asset in a way that consistently delivers business value.”¹

HARVARD BUSINESS REVIEW

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
- Scale more easily when demands increase
- Replace ETL with self-service capabilities
- Analyze more data structures as well as unstructured data
- Make it faster and easier to develop new applications
- Support more advanced analytics
- Better leverage AI, machine learning, IoT, and decision automation
- 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-premise:

- 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 streaming, data warehouse automation, and data lake creation.



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.



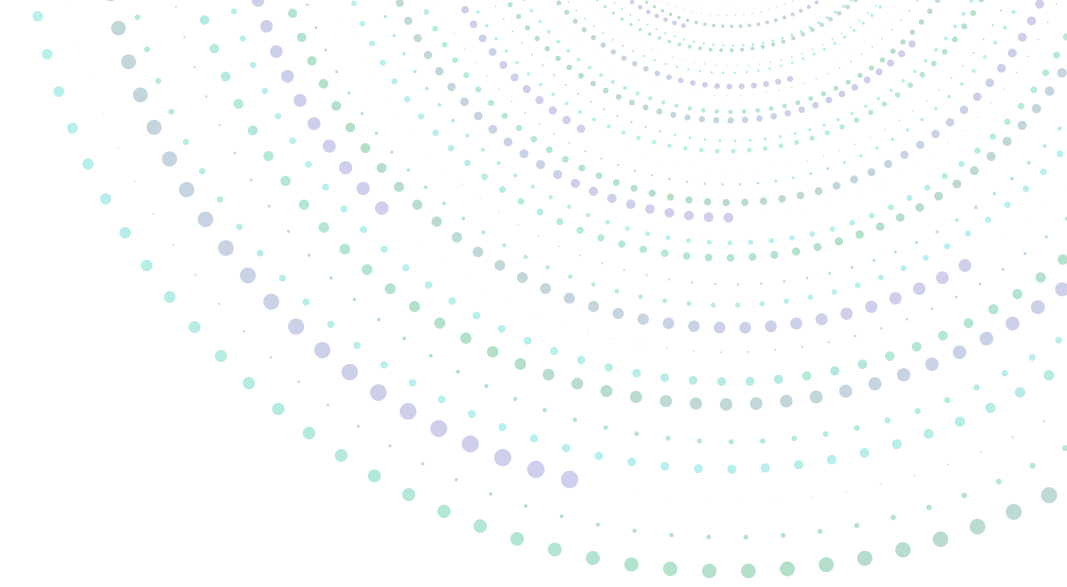
SUPPLIER SERVICES

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



CHANGING THE GAME

Discover the three recent developments that 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

- Publishing production database transactions in multiple topic streams to multiple data lake/data warehouse targets
- Making database transactions available in real time for major streaming systems, including the ability to provision one producer to many streaming consumers
- Integrating data from mainframe/legacy or other sources with microservices-based development
- Auto-mapping heterogeneous sources to a single consumption format

Data warehouse automation

- Facilitating the migration and operationalization from traditional EDWs to cloud data warehouses (e.g., Snowflake, AWS Redshift, and Azure Synapse)
- Streamlining the process of implementing, managing, updating, documenting, and cataloging data warehouses and marts
- Rapidly propagating source or model changes through the data warehouse environment
- If necessary, quickly adapting to data warehouse technologies of the future

Data lake creation

- Ingesting, automating, and cataloging metadata at enterprise scale to establish a data lake
- Integrating a data lake with streaming architectures
- Supporting multi-zone data lake methodology with full historical data store
- Rapidly propagating source or model changes through the data lake environment
- If necessary, quickly adapting to data lake technologies of the future



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
- Log-based change data capture with zero footprint (to minimize source impact)
- Rapid, agentless deployment
- Ability to easily change sources or targets as data requirements evolve
- Continuous, real-time data replication
- Ability to automate many processes
- An intuitive GUI to configure database publication
- Immediate availability of production data for analytics and microservices

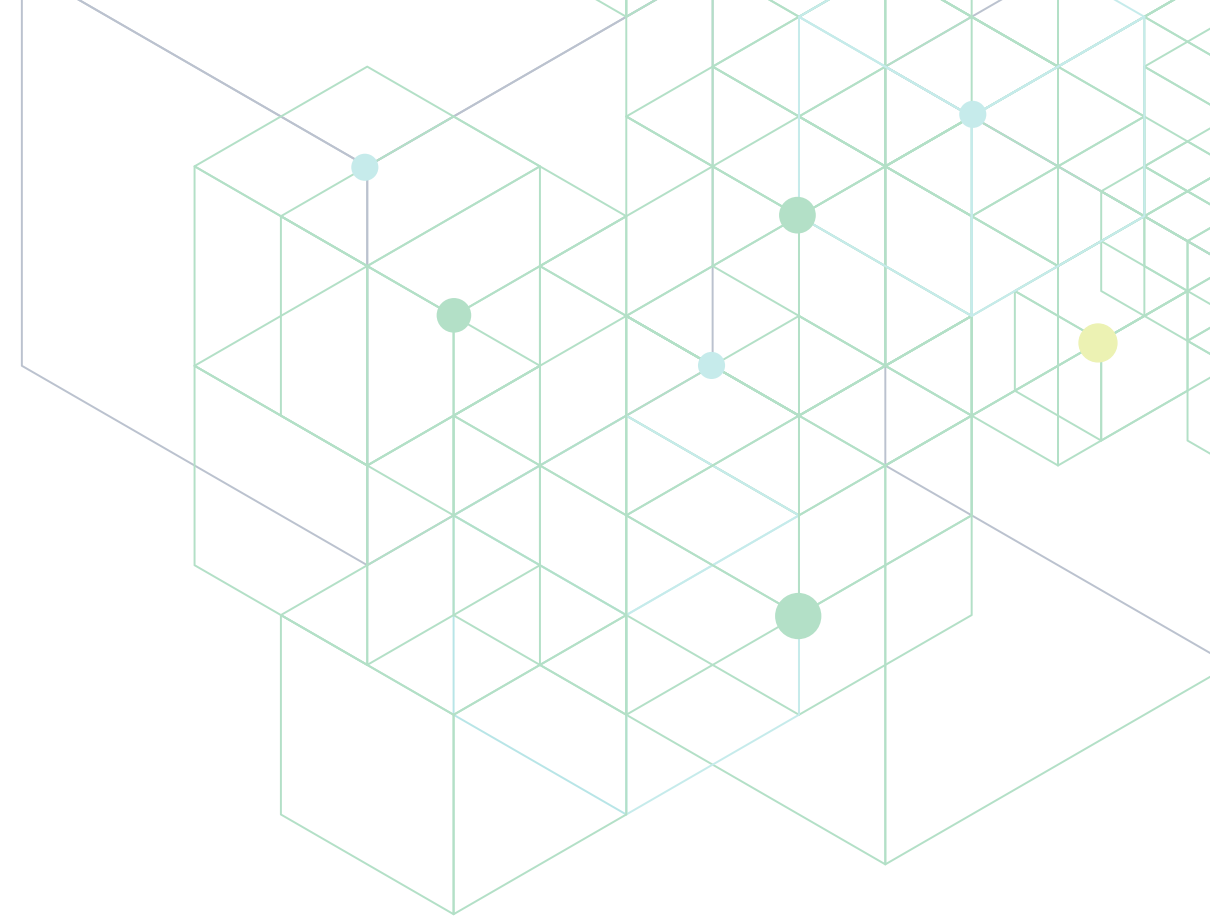




CORE CAPABILITIES

Data transformation

- Reusable transformation templates to automate data set manipulation
- 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, 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)
- Conforming table structures according to rules, such as “Add _____ to column names,” “Insert namespace,” and “Add calculated columns”
- Conforming 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
- Automatic conforming of multiple entities to standard data models or approaches, e.g., Data Vault 2.0, Dimensional Modeling (star schema), MS Common Data Model, HL7, and Swift
- ACL-based transformations enabling different scopes, including personal, departmental, tenant/environment, or company





CORE CAPABILITIES

Catalog and lineage

- A secure data marketplace with a single catalog enabling users to easily find, consume, and share data marts
- Accessibility to all required data sources, whether they're on-prem or in the cloud
- A complete catalog of metadata associated with each data source
- Preserved lineage so users can understand the origin, evolution, and meaning of sets
- Analysis of unique data fields' impact on direct or indirect downstream datasets and analytic applications
- A global mechanism for offering governed data sources for analysis
- Automatic validation, profiling, and quality-checking of data
- Self-service selection of new data sets by users
- Sufficient governance so that users can access only authorized data
- Selective protection of personally identifiable information 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





SOLUTION MUST-HAVES

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:

Data quality

- Categorization of clean vs. dirty records
- Quarantining of data that doesn't meet quality standards
- Automatic cleaning of dirty records according to organization-specific rules
- Use of data-quality rules to trim, merge, and calculate
- Dumping of problem data into error marts
- Transformation of data via business rules and, once defined, enabling the rules to be broadened to other pipelines
- Application of business standards across all data (e.g., making sure that the same data filters are used against all sources)



SOLUTION MUST-HAVES

Data security

- User authentication
- Detailed access controls that can be enabled by role or individual user
- Table-based or column-based access control
- Data-field obfuscation to mask sensitive information
- Encryption of data at motion and at rest (i.e., during the transport and storage stages)
- Automatic allowance of embedded governance with rules and policies to influence discovery early and often
- Injection of data-quality improvements based on the proper context for individuals and workflows

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-prem, SaaS, CSP, and VPC)
- Cloud, on-prem, 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



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 Data Integration Platform®. You can read key findings and get a framework for evaluating the potential savings in your own organization.

[The Total Economic Impact™ of the Qlik Data Integration Platform →](#)



CHANGING THE GAME

Modern advances in 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**

Log-based change data capture technology automatically and securely replicates changes to data as they occur.

✔ **Automated analytics data pipelines**

Scalable data-to-analytics pipelines transform data with push-down SQL.

✔ **Complete data governance for explainable analytics**

Integrated metadata catalogs give users an easy, intuitive way to access the data they need as soon as it's available.

At Qlik, we lead the industry in all of the above. By automating data streaming, refinement, cataloging, and publishing, we make it possible to deliver analytics-ready data in real time throughout your organization. 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 Data Integration Products](#) →

[Qlik Cloud® Data Integration Solutions](#) →

Ready to see what Qlik can do for your business?

[Get in Touch](#)

ABOUT QLIK

Qlik's vision is a data-literate world, where everyone can use data and analytics to improve decision-making and solve their most challenging problems. Qlik offers real-time data integration and analytics solutions, powered by Qlik Cloud, to close the gaps between data, insights and action. By transforming data into Active Intelligence, businesses can drive better decisions, improve revenue and profitability, and optimize customer relationships. Qlik serves more than 38,000 active customers in over 100 countries.



qlik.com

© 2022 QlikTech International AB. All rights reserved. All company and/or product names may be trade names, trademarks, and/or registered trademarks of the respective owners with which they are associated.

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>.