

# Data Drives Business

## Data Integration Considerations for ISVs and Data Providers

WHITE  
PAPER

 VENTANA RESEARCH

Sponsored by:

**Qlik** 



# Table of Contents

|  |          |
|--|----------|
| <b>Real-Time Data Drives Business Today .....</b>              | <b>3</b> |
| <b>Data Integration Is Not Easy .....</b>                      | <b>4</b> |
| <b>Advantages of a Modern Data Fabric for Integration.....</b> | <b>5</b> |
| <b>Illustrative Use Cases.....</b>                             | <b>6</b> |
| <b>Other Considerations.....</b>                               | <b>7</b> |
| <b>Next Steps .....</b>  | <b>8</b> |
| <b>About Ventana Research .....</b>                            | <b>9</b> |

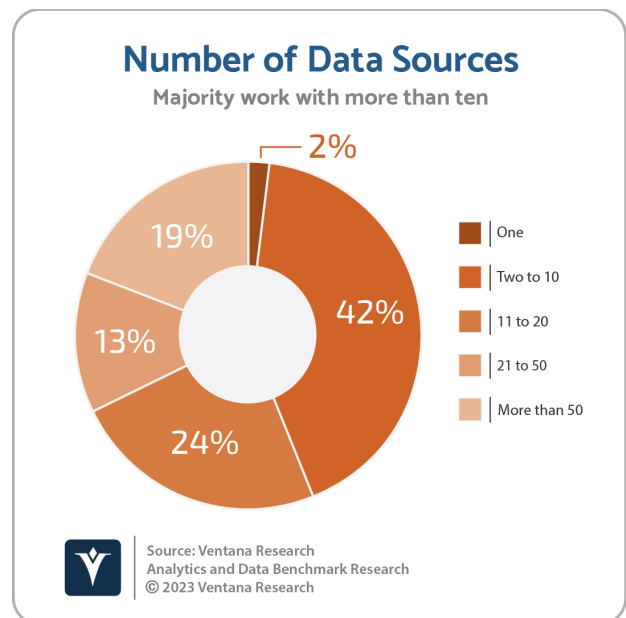


# Real-Time Data Drives Business Today

Data is an extremely valuable asset to almost every organization, and it informs nearly every decision an organization makes. It can be used to make better decisions at almost every level of an organization—and to make them more quickly. So, it is no surprise that nearly all participants in our research (91%) report that they have been able to use data to improve their activities and processes. Today, data is collected in more ways and from more devices than ever before. It can enable new methods of doing business and can even create new sources of revenue. In fact, the data itself can be a new source of revenue.

Independent software vendors (ISVs) and data providers understand this and they are designing products and services to help organizations step in and harness all this data-generated business energy. To maximize the opportunities, ISVs and data providers need to recognize that organizations use various types of data, including data from both internal and external sources. In fact, our research shows that the majority of organizations (56%) are working with 11 or more sources of data.

The most common types of collected data include transactional, financial, customer, IT systems, employee, call center, and supply chain. But there are other sources as well, many external to the organization. Nine in 10 organizations (90%) are working with at least one source of external data, which could mean location data, economic data, social media, market data, consumer demographics, government data, and weather data. To be useful, all of that must be integrated.



“Data integration” is the process of bringing together information from various sources across an organization to provide a complete, accurate, and real-time set of data that can support operational processes and decision-making. But nearly one-third of organizations (31%) report that it is hard to access their data sources, and more than two-thirds (69%) report that preparing their data is the activity where they spend the most time in their analytics processes. The process of data integration often places a burden on the operational systems upon which organizations rely.

At the same time, organizations also need to be able to integrate applications into their data processes. ISVs and data providers must bring data together with applications so it is easier for organizations to access and use the very data they provide.



## Data Integration Is Not Easy

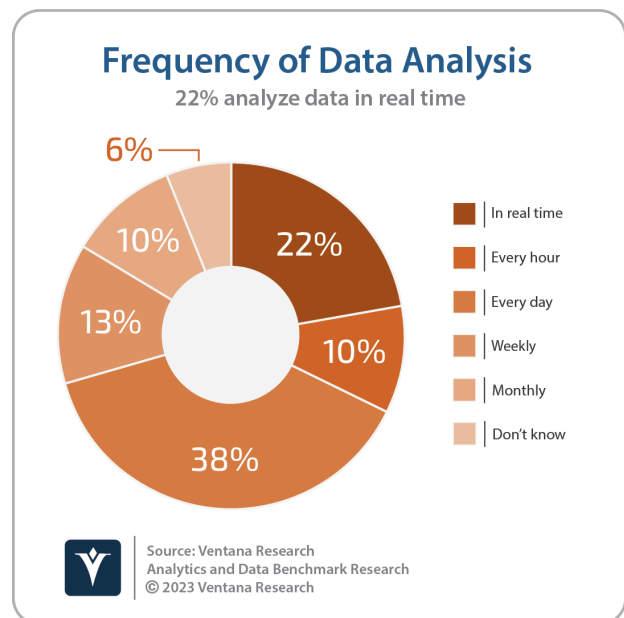
Simple linkages such as open database connectivity and Java database connectivity (ODBC/JDBC), or even custom-coded scripts, are not sufficient for data integration. While ODBC/JDBC can provide the necessary “plumbing” to access many different data sources, it offers little assistance to application developers in creating agile data pipelines. Simple connectivity also does nothing to assist with consolidating or transforming data to make it ready for analytics, for instance, in a star schema. Nor does simple connectivity provide any assistance in dealing with slowly changing dimensions which must be tracked for many types of artificial intelligence and machine learning (AI/ML) analyses.

Simple connectivity does little to help organizations transform the data to ensure its standardization or quality. Data from various sources often contains inconsistencies, for instance in customer reference numbers or product codes. Accurate analyses require that these inconsistencies be resolved as the data is integrated. Similarly, data quality is an issue that must be addressed as the data is integrated. Our research shows these two issues of data quality and consistency are the second most common time sinks in the analytics process.

Nor does simple database connectivity help organizations effectively integrate data from files, applications or application programming interfaces (APIs). With the proliferation of cloud-based applications, many of which only provide API access, ODBC/JDBC connectivity may not be an option. And many organizations still need to process flat files of data, as our research shows that these types of files are the second most common source of data for analytics.

Data integration is not a one-time activity, either. It requires the establishment of data pipelines that regularly collect and consolidate updated data. A greater infrastructure is needed around these pipelines to ensure that they run properly and to completion. ISVs and data providers that rely only on simple connectors must create and maintain this extra infrastructure themselves.

Those data pipelines also need to be agile enough to support a variety of styles of integration. Batch updates are still useful for bulk transfers of data, but other more frequent styles of updating are needed as well. Our research shows that nearly one-quarter of organizations (22%) need to analyze data in real time. Since the most common sources of information are transactional and operational applications, it is important to create pipelines that can access this data as it is generated. Incremental updates and change data capture (CDC) technology can solve this problem and these are becoming competitive necessities.





Real-time requirements are even more demanding when we consider event data, where nearly one-half (47%) of organizations process it within seconds. Then, as applications and organizational requirements change, the data pipelines must reflect those changes. Therefore, the tools used to support such a wide variety of ever-changing sources need to be open enough to be easily incorporated into a wide variety of processes.

But if ISVs and data providers focus their energies on maintaining data pipelines, it distracts resources from the core business. Creating data pipeline infrastructure that is highly performant and efficient requires years of engineering. Simple bulk movement of entire data sets is slow and inefficient, even though it may be necessary for initial data transfers. Subsequent data transfers, however, should use a CDC approach. In this way, only the changes themselves need to be processed, creating much smaller data transfers and much faster processes.

Data infrastructures also typically rely on replication schemes to continuously copy and move data around where and as necessary. Such replication requires scheduling and monitoring, which, again, would require additional development and maintenance on the part of ISVs and data providers.

## Advantages of a Modern Data Fabric for Integration

A modern data fabric is based on a cloud-native architecture and includes orchestration and automation capabilities that enhance the design and execution of data pipelines that

“

**Deployment flexibility is important and a modern data fabric must support cloud, multi-cloud, on-premises and hybrid configurations.**

consolidate information from across the organization. Deployment flexibility is important and a modern data fabric must support cloud, multi-cloud, on-premises and hybrid configurations. Containerization is a key component of cloud-native architectures, allowing the parallelization and elasticity that is demanded by today's applications. As more processing resources are needed, additional containers can be instantiated, and the workload can then be split among the various containers.

Orchestration and access via APIs also allows remote invocation of data pipelines common to modern architectures and cloud deployments, while at the same time allowing coordination and synchronization of various related application processes, even when

they are distributed across different cloud applications and services. These APIs need to span all aspects from provisioning to core functionality for orchestration to be effective. Automation of these orchestration tasks can enhance many aspects of data pipelines to make



them both more efficient and more agile. Automated data mapping, automated meta data creation and management, schema evolution, automated data mart creation, and data warehouse and data lake automation can quickly and efficiently create analytics-ready data. When combined with orchestration, automation can also provide “reverse integration” to update data in source systems when necessary and appropriate.

Modern data integration platforms employ AI/ML to streamline and improve data processing. AI/ML can be used to automatically detect anomalies in data pipelines, such as whether the pipelines suddenly processed an unusually small number of records. Such an anomaly could indicate a problem somewhere else in the pipeline. AI/ML can also be used to automatically deal with errors in pipelines and routine changes, such as those in the sources or targets. AI/ML can also determine the optimal execution of pipelines, including the number of instances to create or where different portions of the pipeline should be processed. AI/ML can be used to enrich data with predictions, scoring or classifications that help support more accurate decision-making. We assert that by 2026, more than two-thirds of all data processes will use AI/ML to accelerate the realization of value from the data.

**Data**  
*Market Assertion*

By 2026, more than two-thirds of all data processes will use artificial intelligence and machine learning to accelerate the realization of value from the data.

**Matt Aslett**  
VP & Research Director

VENTANA RESEARCH

Modern data integration platforms must also incorporate all appropriate capabilities for data governance. Data sovereignty issues may require that data pipelines be executed only within certain geographies. Compliance with internal or regulatory policies may require single sign-on or the use of additional credentials to appropriately track and govern data access and use. Therefore, a platform with built-in capabilities for governance can help identify personally identifiable information and other sensitive or regulated data. But implementing any of these modern data integration platform requirements can impose a significant burden on ISVs and data providers.

## Illustrative Use Cases

### Smart Factories

There are many use cases where a data integration platform including industrial “internet of things” (IIoT) software is employed to support smart-factory operations. Smart-factory IIoT software requires the collection and processing of large amounts of data that is generated by devices scattered across the manufacturing process. Plus, historical data is needed for analysis but integrating that bulk of data can bog down the production systems. So, offloading



historical data using a CDC approach minimizes the burden on the production system, while at the same time making large amounts of data available for analysis, which helps optimize factory operations. A data integration platform also provides the openness to support other databases as necessary.

## Insurance Industry

Insurance technology data providers can use data integration to help their customers be more competitive by providing access to up-to-date information that enables online quotes. Data is the key to the accurate pricing of insurance liabilities, and many of the sources and



**Insurance technology data providers can use data integration to help their customers be more competitive by providing access to up-to-date information that enables online quotes.**

targets exist in the cloud, but they require support for a variety of endpoints. By using CDC-based replication, however, both claims and market data can be collected, consolidated, and distributed within minutes. As a result, millions of quotes can be generated each day where each incorporates real-time analysis of vast volumes of data.

## Other Applications

Data integration can be the key to many other ISVs and data providers. Mobile application providers can integrate location data with transaction data to provide broader market data on consumer behavior. Talent management ISVs can integrate data relating to internal performance and compensation with external market data to improve employee acquisition and retention. Foreclosure data can be collected, consolidated, and distributed to support loan origination and servicing operations. Vendor data can be collected and provided to improve procurement

processes augmenting supplier performance analyses with risk, diversity, sustainability and credit scores. And regardless of the vertical industry or line-of-business function, faster access to more data generally produces better results.

## Other Considerations

Once data is integrated, it can provide the basis for a broad range of analytics. By supporting these analytics, ISVs and data providers can extend the value of their capabilities and therefore increase their revenue opportunities. A platform that supports the entire life cycle of data—including the lineage of that data through the processes—makes it easier to find and understand the data needed for particular operations and analyses. Collecting and sharing metadata in a data catalog provides better understanding and access to the data while also improving data governance. Our research shows that organizations that have embraced data



catalogs are three times more likely to be satisfied with their analytics and have achieved greater rates of self-service analytics.

Choosing a data integration platform that also supports analytics will make it easier for organizations to capture this revenue. In fact, our research shows that reports and dashboards are the most common types of analytics used by more than 80% of organizations. However, when considering analytics providers, look at those that support other newer techniques as well, such as AI/ML and natural language processing, which are projected to be required by 80% of organizations in the future.

Organizations need to use data to help drive actions. Data can help them understand what has happened and why, but organizations ultimately need to process what they have learned and then take action. In many situations, however, there is simply no time to review data to determine what course of action to take. ISVs and data providers can help their customers derive more value from data by using real-time information to trigger the appropriate actions.

ISVs and data providers are using technology to add value to business processes. While all business processes typically require data, data integration itself is merely a means to the end. If the process is not done properly, it can detract from the overall approach, so it requires careful design and development. Organizations should ideally spend their time on core competencies, not on developing data integration technology. By using a full-featured, purpose-built data integration platform, they can ensure that the data needed by ISVs and data providers is robust and available in a timely manner.

## Next Steps

- Explore all available data sources that can boost the value of your services.
- Consider platforms that go beyond simple connections to data sources and that minimize the amount of development and maintenance work required.
- To maximize performance and minimize the impact on production systems, create repeatable and agile pipelines that operate efficiently.
- Look for platforms with significant automation capabilities to maximize productivity and responsiveness.
- Ensure that your architecture provides a modern, cloud-native approach.





## About Ventana Research

Ventana Research is the most authoritative and respected benchmark business technology research and advisory services firm. We provide insight and expert guidance on mainstream and disruptive technologies through a unique set of research-based offerings including benchmark research and technology evaluation assessments, education workshops and our research and advisory services, Ventana On-Demand. Our unparalleled understanding of the role of technology in optimizing business processes and performance and our best practices guidance are rooted in our rigorous research-based benchmarking of people, processes, information and technology across business and IT functions in every industry. This benchmark research plus our market coverage and in-depth knowledge of hundreds of technology providers means we can deliver education and expertise to our clients to increase the value they derive from technology investments while reducing time, cost and risk.

Ventana Research provides the most comprehensive analyst and research coverage in the industry; business and IT professionals worldwide are members of our community and benefit from Ventana Research's insights, as do highly regarded media and association partners around the globe. Our views and analyses are distributed daily through blogs and social media channels including [Twitter](#), [Facebook](#) and [LinkedIn](#).

To learn how Ventana Research advances the maturity of organizations' use of information and technology through benchmark research, education and advisory services, visit [www.ventanaresearch.com](http://www.ventanaresearch.com).