Selecting a Platform for Cloud Integration

Five Considerations for the Evaluation Process

Research Perspective
Sponsored by snapLogic
Table of Contents

Challenge and Opportunity 3

Five Key Evaluation Criteria 4
- Easy access to an increasing variety of cloud data sources 4
- Design that handles cloud-based data sources 4
- Ease of use to encourage adoption 5
- Data quality services to help build trust 5
- Scalability for increasing data volumes and numbers of users 5

Consider How Your Organization Works 6

About Ventana Research 7
Challenge and Opportunity

The advent of cloud computing and the proliferation of business applications that operate in that context have introduced new challenges for business and IT. Integrating cloud-based data with data inside the enterprise is very important to 44 percent of organizations as shown by our business technology innovation benchmark research. Yet according to our research on data in the cloud, 56 percent of organizations export data to spreadsheets and text files for this purpose. These outdated approaches cannot provide either secure access or data that is guaranteed to be accurate. Nor are established extract, transform and load (ETL) or data integration tools an optimal choice as they were not designed to connect to cloud-based applications and data sources.

Cloud-based data integration differs significantly from traditional on-premises data integration. Typically the latter has involved large projects that change infrequently with the projects generally focused on a few relatively large data feeds that have been optimized for batch processing. In contrast, cloud-based data integration often involves many data sources, with new sources being added frequently. This integration challenge is likely to intensify; our research finds that half of organizations are planning to integrate data from cloud computing into enterprise systems and 44 percent need cloud-to-cloud integration. Moreover, these cloud sources may change at any time since they are not under the control of the enterprises using them.

Organizations need to understand that cloud-based data sources are different even beyond where they reside. Far more often than on-premises systems, they include unstructured data – social media conversations, for example – in addition to structured data that can easily be organized into rows and columns. Processing data from these systems may be done in batch mode, but it also can be event-based or based on continuous streaming of information. Our research finds that one-fourth or more organizations need to improve access to 10 types of cloud-based business data.
To cope with this new and evolving situation, many organizations are considering tools that also are based in the cloud and are specifically designed to integrate these kinds and volumes of data. We offer the following criteria to help decide how to select the right data integration cloud platform and tools for an organization’s needs.

Five Key Evaluation Criteria

As part of the process of examining data integration platforms in the cloud, we recommend looking for each of the following capabilities:

**Easy access to an increasing variety of cloud data sources**

Our benchmark research has found that on average enterprises need to access and analyze from five to more than a dozen data sources, and the popularity of the cloud is expanding this number. The research also shows that organizations plan to adopt more cloud-based services to support business processes. A data integration platform should provide access to a variety of data sources and make it easier to incorporate both existing and new ones into an organization’s information management architecture without significant added investment or cost. And it should be flexible enough to accommodate new data sources as new applications are used in the cloud and business requirements demand integration of the data.

**Design that handles cloud-based data sources**

To integrate data for both cloud-based and on-premises data sources in a consistent way, organizations should base their processes on the Web standards common to these applications. Representational State Transfer (REST) has become the standard architecture for Web interfaces that provide stateless interactions. This mode of interaction, where each request is independent of any other request, is important for Web-based applications since the request may be routed to any of a number of servers. RESTful interfaces generally work with HTTP, XML and other Web standards. Using this standards-based approach provides access to nearly any cloud-based application since virtually all already support browser-based interfaces.
Ease of use to encourage adoption
Organizations need to find ways to get data to more users more quickly. Currently more than two-thirds (69%) of users spend more time preparing data for analysis than actually analyzing it. And almost half (44%) of organizations take six or more days to deliver metrics and key performance indicators to those who need them. One way to speed up this process is to make it easier for users to access the data. Point-and-click graphical user interfaces that automatically map data elements based on field names and contents can guide users through tasks. Tools that eliminate coding make modifications easier. And while code generators, an alternative approach, may appear to offer some of these benefits, they do not deliver optimal ease of use since they require technical resources to package and deploy the generated code – which, of course, must then be maintained.

Data quality services to help build trust
Data quality is paramount; it is the most important concern associated with integrating cloud-based data for 62 percent of organizations. Our research also shows that users trust cloud-based data sources much less than on-premises data sources. Ensuring data quality can help increase not only the accuracy of data but also users’ trust in it. Therefore, we advise selecting a product that includes or provides close partnering with data quality services such as address verification and deduplication. Another way to increase trust is to reduce reliance on exports to error-prone spreadsheets or text files, which is still the most common technology used for integrating cloud-based data.

Scalability for increasing data volumes and numbers of users
As data volumes continue to expand, integration technologies must be able to handle increasing and elastic workloads while still providing timely access to information. Almost half of organizations (45%) require cloud integration on a daily basis. To make the best use of each individual server – to scale “up” – multithreading applications can take advantage of multiple CPU cores and efficient in-memory processing. Other optimizations that can be chosen include bulk loaders and parallel readers. Scaling “out” spreads the processing across multiple servers (a “cluster”). More processing power is
available as nodes are added. If clustering is done well, performance scales linearly in proportion to the number of nodes. However, clusters also require load-balancing to distribute workloads and prevent bottlenecks on an individual server. For maximum efficiency and cost-effectiveness, look for data integration tools that support a range of upward scalability but also can scale down as needed.

Consider How Your Organization Works

Cloud-based applications are changing the way organizations work; 55 percent are using them, according to our business technology innovation research. Because fewer in-house IT resources are involved in managing cloud-based applications, end-user organizations have more choices as they select applications to meet their business needs. We advise you to anticipate a proliferation of cloud-based applications and thus an increase in the number of off-premises data sources your organization needs to manage and integrate. Find a platform that fits your needs today, but as part of your search plan for growth tomorrow by anticipating more data, new data sources and changes to the existing sources. The research finds this is important to more than three-quarters (76%) of organizations; it should be important to your organization, too. Use tools that can put high-quality data from any source into the hands of your users easily and quickly to maximize the value of your organization’s data.
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, LinkedIn and Google+.

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.