VIEWPOINT

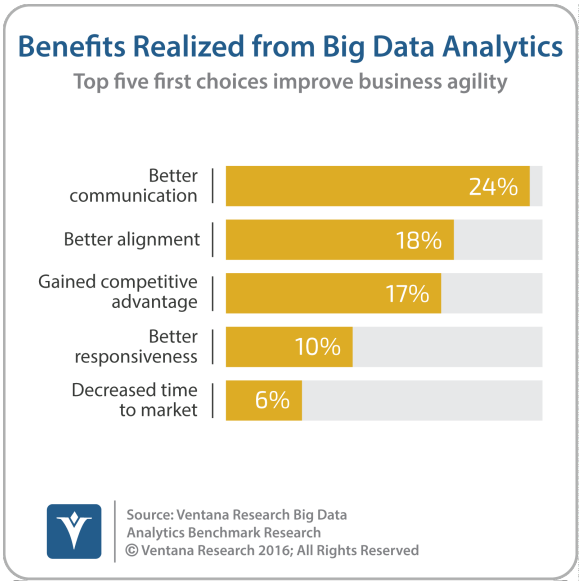
DAVID MENNINGER, SVP & RESEARCH DIRECTOR

VENTANA RESEARCH

Sponsored by
podium data

Make Your Data Lake a Data Marketplace

Data lakes have become all the rage, giving rise to discussions of the technologies involved but not often of their business value. A useful data lake should not be measured by the number of Apache projects it includes but rather by how well it enables an organization to operate more efficiently and more effectively, or to offer new products and services. If your data lake does not enable these improvements, you should rethink your strategy for using it. Our benchmark research on big data analytics shows that big data can deliver business benefits. Successful data lakes must automate delivery of these kinds of benefits to make data integration and data access simpler for business users. Nearly four in five (78%) organizations in our research said that automation of big data integration is important. Ultimately, your data lake should enable more finely grained analyses than could be done previously.



Automation and integration of data should accelerate access to and consumption of data and thus accelerate time to insight from use of analytics. Many organizations attempt to build the automation and integration with custom code or a variety of independent tools, which can be time-consuming and require specialized skills. Our research indicates that this choice of approach may be shortsighted: Fewer than half (47%) of organizations reported that they have the Hadoop-specific skills to support big data integration. As an alternative, an integrated platform approach can facilitate the process because the vendor has previously connected the required components and thus reduced the need for specialized skills.

Using an integrated platform also helps with the process of converting raw data into consumable data in a repeatable and efficient way. Since data preparation is often the most time-consuming part of the analytical process, the data lake should enable a new approach that provides analysts with self-service data preparation tools to transform raw data interactively into analytical data sets.

Once these preparation steps are defined, the platform should support automating the logic to create clean, consistent, high-quality data appropriate for broader consumption across the organization. Such an agile approach requires a well-integrated platform to ensure seamless handoffs and governance controls as data is matured from "raw" to "ready." Making information more available in a consistent manner across the enterprise is the most commonly reported benefit of big data integration, cited by 72 percent of participants.

The data lake should be designed to encourage broad usage. If it provides easy access to high-quality data it is much more likely to lead users to spread the word. The more people using the data the better, because each use of a properly managed data lake improves its quality and preparedness. Use generates value as users contribute more data which, in turn, gets cleansed and prepared for more analyses and other purposes. Frequent use also helps amortize the investment in the data lake.

Think of your data lake as providing "data as a service." Users want easy access to high-quality data from a variety of sources without having to know where it comes from. They just want it to be ready and available. IT can add significant value by efficiently handling the infrastructure tasks and services to manage the data. Ideally these data services can be extended to allow users to bring their own data, too. As the provider and steward of the data, IT must ensure that it is secure while providing self-service capabilities. Concern about security is the most common obstacle to integrating big data: More than half (56%) of organizations cited this as an issue.

A data lake with a two-way exchange in which users both consume and contribute data creates a data marketplace. As with other marketplaces, participation in the market (here, use of the data) is a measure of success. Users are attracted because markets are transparent and efficient. This transparency in the data along with its quality and consistency create trust in the data and encourage more usage. A healthy data marketplace provides data on demand, ready to incorporate into analytics and business processes that will help your organization succeed.

David Menninger is SVP and director of data and analytics research for Ventana Research. David is responsible for the firm's research direction in data, information and analytics technologies, including big data, business intelligence, predictive analytics, information management, operational intelligence and IoT. For more insights by David Menninger, see https://davidmenninger.ventanaresearch.com.