

# Data Preparation

Enabling Self-Service and Support Across Business and IT

BEST  
PRACTICES  
REPORT



## A Note About This Research

December 2017

Ventana Research performed this research to determine attitudes toward and utilization of data preparation. This document is based on our research and analysis of information provided by organizations that we deemed qualified to participate in this benchmark research.

This research was designed to investigate data preparation systems, practices, needs and potential benefits. It is not intended for use outside of this context and does not imply that organizations are guaranteed success by relying on these results to improve data preparation. Moreover, gaining the most benefit from a data preparation system requires an assessment of your organization's unique needs to identify gaps and priorities for improvement.

The full report with detailed analysis is available for purchase. We can provide detailed insights on this benchmark research and advice on its relevance through the Ventana On-Demand research and advisory service. Assessment Services based on this benchmark research also are available.

We certify that Ventana Research wrote and edited this report independently, that the analysis contained herein is a faithful representation of our evaluation based on our experience with and knowledge of data preparation, and that the analysis and conclusions are entirely our own.

*Ventana Research*

This research is sponsored by



VENTANA RESEARCH

Bend, Oregon, USA

541-940-1010

[info@ventanaresearch.com](mailto:info@ventanaresearch.com)



*Data is essential to every aspect of business, and organizations that use it effectively are likely to gain advantages over competitors that do not. Information derived from this data is essential to address a variety of needs – to support analytics and decision-making, enable effective process improvements and optimize the customer experience. A variety of new factors are changing the data preparation process, including the growing importance of streaming data sources flowing into big data repositories and a resulting need to apply data science techniques to derive meaning from this data.*

*Data preparation typically involves a sequence of steps: identifying, locating and then accessing the data; aggregating data from different sources; and enriching, transforming and cleaning it to create a single uniform data set. Using data to accomplish organizational goals requires that it be prepared for use and to do this job of data preparation properly, businesses need flexible tools that enable them to enrich the context of data drawn from multiple sources and collaborate on its preparation as well as ensure security and consistency.*

*Ventana Research undertook this benchmark research to determine the attitudes, requirements and future plans of those who use data preparation technologies and to identify their best practices. We set out to examine both the commonalities and the qualities specific to major industry sectors and across sizes of organizations, exploring how organizations manage data preparation processes, issues they encounter and how their use of data preparation and related technology is evolving. Having analyzed the research, we offer the following set of best practices for those organizations considering ways to improve their data preparation processes and, as a result, business outcomes.*

## 10 Best Practice Recommendations

This benchmark research reveals significant new insights into the evolving nature and use of data preparation processes and systems. For organizations considering how to optimize the use of data preparation by employees, managers and executives and its value to the organization, we offer the following recommendations.

- 1. Establish a data preparation strategy.**  
Data preparation can be valuable to your organization; the research finds three-quarters of participants indicating that it has improved their activities or processes. Many organizations have already embraced data preparation, with nearly two-thirds (62%) reporting confidence in their ability to prepare data and slightly more than that (65%) reporting confidence in data quality. Further analysis indicates that data preparation has improved information quality and availability and has reduced manual processes.



However, more than one-third of organizations say they are not fully confident in their data preparation capabilities or the quality of the data produced by their data preparation processes. Your organization should identify and then target areas for improvement. Creating a data preparation strategy for business and IT can help your organization begin to realize these benefits as well.

## **2. Create clear goals for data preparation technology.**

The research shows that a large majority (88%) of organizations want to enable accessing and preparing data without the involvement of IT. However, fewer than half (42%) of the organizations that consider this important have accomplished this objective. Hesitance to employ self-service data preparation is most often due to security and governance concerns. However, differing attitudes toward the handling of data prior to its processing and integration are also an issue; more than half (51%) of participants in business functions are comfortable with line of business employees working with data that hasn't been processed by IT as opposed to only one-third (32%) of those with IT titles. These differences suggest that challenges go beyond technology and must be addressed with improvements in organizational processes. Consider the nature of the issues in your organization and look to make sure that any technology chosen is flexible, which will eliminate a problem cited by 37 percent of users, and that it doesn't require too many resources, cited by 36 percent of organizations. Work to clarify your organization's data preparation goals and regularly reevaluate your processes to ensure they are helping you accomplish them.

## **3. Utilize dedicated data preparation tools.**

The research reveals a generally even split between standalone data preparation tools and tools embedded within BI technologies. The latter are currently most often the primary tool, but half (47%) of all participating organizations are using a dedicated tool and more plan to do so in the future. Activities most often involved in preparing data include analysis, extracting and accessing data. More than half of organizations perform data quality activities as they prepare data. Dedicated data preparation tools can often provide more capabilities in these areas than embedded tools. Your organization should consider its specific needs and decide whether dedicated tools have a role in your data preparation processes.

## **4. Provide appropriate training for data preparation.**

Fewer than 40 percent of participants said they considered their training in data preparation technology and techniques completely or mostly adequate. Data preparation has substantially changed the way organizations approach information management, but our analysis indicates that overall organizations have not kept up with these changes. As you develop a data preparation training curriculum, pay particular attention to training on



handling big data and preparing web- and cloud-based data as the research finds that training in these topics is the least adequate.

## **5. Use cross-functional teams for data preparation.**

Data preparation involves a balance between line of business functions and IT. The 17 percent of organizations that use cross-functional teams reported the highest levels of satisfaction with their data preparation technology and their ability to support big data and are most comfortable allowing business users to work with data without the assistance of IT. Encourage your organization to adopt a cross-function approach; it is not the approach most organizations use. Business intelligence and data warehousing teams within IT are the group most likely (28%) to design and deploy data preparation tasks. Unfortunately but perhaps not unsurprisingly, issues between IT and line-of-business functions are common, most often involving disagreement over expansive vs. controlled access to data; only 17 percent reported no issues between these groups. Nevertheless, the mix of skills needed to successfully prepare data means that a cross-functional team would likely provide your organization with the best results.

## **6. Enable frequent data integration.**

Those organizations that are integrating data in real time report higher levels of confidence and satisfaction and are more comfortable allowing business users to access data without the assistance of IT. Working with data sources is time-consuming, particularly if your organization is accessing a large number of sources. One-fourth of organizations access more than 20 data sources, but the largest group (35%) works with five to 10. If you are like most organizations, you know that processing large volumes of data and providing connectors to databases and applications are important system capabilities. Create data preparation processes and technology that enable frequent data integration.

## **7. Select tools that provide usability and the necessary functionality.**

Participants evaluating data preparation technology considered usability and functionality as the two most important evaluation criteria, priorities that make sense for self-service. Since data preparation processes span both line of business and IT, it is important that your organization address the needs of both of these groups. Collaboration and mobile capabilities can also make data preparation more usable and functional; more than four in five (83%) participants cited collaboration around data preparation tasks as important. These capabilities can support the cross-functional teams that provide the best results.



## **8. Support big data with your data preparation processes.**

The research shows many organizations (46%) work with big data, and those organizations that have been doing so the longest most often report that self-service data preparation without the involvement of IT is very important. They also are most likely to report satisfaction with data preparation technology. Working with big data brings challenges because of the size and complexity of data sets, but data preparation tools can provide an easy and fast way to process this data. Consider utilizing big data as you assess and select data preparation technology in the next 12 to 18 months.

## **9. Consider the value of cloud and hybrid deployments.**

Consider both current and future needs as you evaluate your requirements for on-premises and cloud deployments. This research finds organizations undertaking a mix of cloud and on-premises data preparation activities. Although there are currently more on-premises data preparation processes (64%), it is important that your organization assess the potential benefits of cloud-based and hybrid processes. Adopting cloud-based technologies can often lead to faster deployments and reduced IT budgets.

## **10. Assess the shortcomings in your data preparation efforts.**

Almost half (45%) of the organizations in this research said they are planning to change their data preparation processes over the next 12 to 18 months. As you evaluate your data preparation activities, consider the adequacy not only of technologies but also of your people, the types of information involved and the processes you put in place. Nearly six in 10 organizations cite cost as a barrier to improving their data preparation processes; other barriers include inadequate skills, limited awareness and lack of resources. As you evaluate data preparation processes, consider primary-use cases and the specific people and technology requirements of your organization.





## 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).



## Appendix: About This Benchmark Research

Ventana Research designed this benchmark research to assess the use of and plans for data preparation. We conducted this research on the web from January 2017 through July 2017. Applying our standard methodology and quality assurance criteria, we identified 179 qualified participants. They represent a range of organization sizes: 28 percent work in very large companies (having 10,000 or more employees), 30 percent in large companies (with 1,000 to 9,999 employees), 25 percent in midsize companies (with 100 to 999 employees), and 17 percent in small companies (with fewer than 100 employees). A majority (63%) of the participants are from companies located or headquartered in North America, although many of these are global organizations operating worldwide. Among industry categories, companies that provide services accounted for 37 percent, those in manufacturing for 28 percent and those in finance, insurance and real estate for 19 percent. Government, education and nonprofits accounted for the remaining 14 percent. Categorized by their job title, one in seven are executives, and 7 percent are management, and nearly two-thirds are what we categorize as users. Predictably, nearly one-third of the participants identified themselves as being in the IT/IS/MIS function. (More demographic details about the participants are available in the full research report.)