

# Data Preparation

Enabling Self-Service and Support Across Business and IT

KEY  
INSIGHTS  
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

 DATAWATCH



VENTANA RESEARCH

Bend, Oregon, USA

541-940-1010

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



Ventana Research, the most authoritative and respected benchmark business technology research and advisory services firm, provides insight and expert guidance on mainstream and disruptive technologies through a unique set of research-based written, in-person and online offerings. The following **Key Insights** are drawn from our benchmark research and provided as a service to interested business employees, managers and executives. For more information on how to apply these findings and receive guidance on improving your workforce management capabilities, visit [www.ventanaresearch.com/ondemand/](http://www.ventanaresearch.com/ondemand/) and learn about our Ventana Research On-Demand services.

## Key Insights

This benchmark research yielded the following important general findings and key insights regarding data preparation and our previous benchmark research and experience in the data and information management markets. (We discuss performance levels in the Performance Index portion of the full research report; the actual questions asked in our survey and specifics of organization sizes are in appendices to the research report.)

### **Organizations' data preparation performance varies widely.**

Our Performance Index analysis finds more than half (56%) of organizations performing at the lower two levels of our four-step performance hierarchy. The analysis places one in five organizations at the highest Innovative level of performance, meaning they are able to use data preparation tools to innovate and compete effectively against others less adept at using this technology.



Data preparation has unquestionably changed the way organizations approach information management, but overall organizations have not kept up with these changes.

Data preparation has unquestionably changed the way organizations approach information management and support the operational and analytical needs, but overall organizations have not kept up with these changes. Only 43 percent of organizations placed at the top two levels of performance in this research. This suggests that many organizations still need to improve their use of data preparation. It's worth noting, though, that how well organizations perform is not an



indication of their interest in data preparation: 88 percent of participants said that data preparation is important to their organizations.

Analysis of the four dimensions into which we segment performance shows noticeably lower performance levels in two of the dimensions: In the People dimension two out of three (66%) organizations rank at the lowest two performance levels, which generally indicates a lack of familiarity with and understanding of data preparation. Reinforcing this, three of the four most-often cited barriers to making improvements to data preparation are inadequate skills in the organization (35%), lack of awareness (33%) and lack of resources (33%). The Process dimension also shows room for improvement with 56 percent at the lowest two levels. As new technologies such as data preparation emerge and evolve, organizations often struggle to develop the necessary skills and processes to take full advantage of the new capabilities.

### **Organizations derive significant value from data preparation.**

Three-quarters (76%) of organizations indicated that data preparation has improved their activities or processes, with an even greater percentage (90%) of line-of-business functions reporting such improvements. Participants said



Three-quarters of organizations indicated that data preparation has improved their activities or processes.

that data preparation has improved the quality of information, made information more available in a consistent manner and reduced manual processes. Since the preparation of data is essential to the analytics process and is often the most time-consuming part of it, these benefits carry through to analytics as well.

Many organizations have embraced data preparation. More than three in five (62%) said they are confident in their ability to do data preparation and nearly two-thirds (65%) said they are confident in the quality of their data. When it comes to technologies, however, more than half (56%) said their data preparation technology is adequate. This leaves room for improvement, with more than one-third of organizations not

completely confident in their data preparation ability, the quality of data or the adequacy of their technologies.

### **Data preparation supports analytics and business intelligence.**

The activities most often involved in preparing data include analysis (75%), extracting (64%), and accessing data (57%). Asked to identify the three data preparation tasks on which they spent the most significant amounts of time, research participants cited preparing data for analysis second-most-often.



[Q32] With such an emphasis on analysis, it is not surprising that organizations most often (41%) use their business intelligence tools as their primary data preparation tool. However, regardless of their choice of primary tool, nearly half (47%) of all participating organizations said they are using a dedicated tool specifically designed for data preparation. Another 36 percent said they plan to use such a tool in the future.

Data preparation must also support many data governance activities. More than half of organizations (54%) perform data quality activities as they prepare data. About one-third are managing metadata (34%) and securing (32%) and governing data (31%). One-quarter are auditing (28%) and profiling data (24%). So, while not as prevalent as analytics and business intelligence, governance activities are a key component of the data preparation process.

### Data preparation utilizes frequent integration of multiple sources.

Organizations are often processing large volumes of data from multiple sources. More than half (53%) said that processing large volumes of data and providing connectors to databases and applications (51%) are important system



Nearly three-fourths of organizations integrate data daily while an additional 17 percent integrate data in real time.

capabilities. One-fourth of organizations access more than 20 data sources, but the largest group (35%) works with five to 10 data sources. Size of organization seems to be related to number of data sources, with 36 percent of very large organizations accessing more than 20 data sources and only 10 percent just two to four, while 16 percent of small organizations access more than 20 sources and 45 percent two to four. Working with these data sources is time-consuming; in the data preparation process, research participants reported spending the most significant amounts of time connecting to data sources for access and integration.

Data preparation must be done frequently. The research suggests that daily data integration is table stakes now: Nearly three-fourths (72%) of organizations integrate data daily while an additional 17 percent integrate data in real time. Those organizations that are integrating data in real time reported higher levels of confidence and satisfaction, and said they are more comfortable allowing business users to access data without the assistance of IT. Scheduling data preparation jobs is a necessity for frequent data integration, and nearly half (48%) of companies report that this is an important system capability.



## Hope for self-service data preparation is not yet fulfilled.

A substantial majority (88%) of organizations report that self-service data preparation – accessing and preparing data for analysis without the involvement of IT – is important to their organization. The research finds that hesitance to employ self-service data preparation is most often due to security and governance concerns. Despite this stated importance of self-service, fewer than half (42%) of organizations are comfortable allowing business users to work with data that has not been integrated or prepared for them by IT. Here, though, the views of those with business functions differ significantly from those in IT; half of business users (51%) said they are comfortable while only one-third (32%) of IT reported they are comfortable. These differences suggest challenges that go beyond technology and that likely must be addressed with improvements in organizational processes.

That isn't to say data preparation technology can't get better. Only one-third of organizations reported they are satisfied (31%) with the technology and an additional one-third (36%) are somewhat satisfied with it. The research suggests areas for improvement: More than one-third (37%) complained their technology is not flexible enough (37%) and requires too many resources (36%). And while users of tools designed specifically for data preparation are less likely than the users of other categories of tools for data preparation to

complain that they are hard to maintain or too slow, they are more likely than others to complain that they are inflexible and require too many resources.



The mix of skills needed to prepare data successfully reinforces the notion that cross-functional teams would perform best.

### Cross-functional data preparation teams produce the best results.

Data preparation spans the IT and line-of-business functions in organizations. Business intelligence and data warehousing teams within IT are the group most likely (28%) to design and deploy data preparation tasks. Combined with centralized IT and line-of-business IT functions, IT leads data preparation 46 percent of the time. Line-of-business analysts, data scientists and line-of-business operations lead the process 36 percent of the

time. However, the 17 percent of organizations that use cross-functions teams with shared responsibility feel best about their results. They report the highest levels of satisfaction with their data preparation technology and their ability to support big data and those organizations are more comfortable allowing business users to work with data without the assistance of IT.



Only 17 percent of organizations said that no data preparation issues arise between IT and line-of-business functions. The top issue, cited by 45 percent, is disagreement over expansive vs. controlled access to data. The mix of skills needed to prepare data successfully reinforces the notion that cross-functional teams would perform best. More than three-fourths (77%) of organizations identified analytic skills and two-thirds (62%) cited business skills as necessary for successful data preparation. More than one-third (35%) said that big data technology and programming skills are necessary as well. This cross section of skills is hard to find in a single group, which may explain why the cross-functional teams tend to perform better.

### Data preparation requires usability and collaboration.

As we often see in our research, usability followed by functionality ranks as the most important technology or vendor consideration influencing purchases of data preparation systems. These priorities make sense given the importance of self-service. Looking at specific data capabilities, nearly half of participants (48%) said they want to manage tasks in a repository for reuse and an equal



Collaboration capabilities can help foster and support the cross-functional line-of-business and IT participation that produces the best results.

number said they want to join disparate data sources during transformation. More than four in 10 (44%) said they want to provide real-time processing to further their data preparation efforts and 41 percent want to design graphical workflows of steps to process data.

Collaboration and mobile capabilities can also make data preparation more usable and functional. One-fourth of participants reported the task in which they spend the most significant amount of time in their data preparation work is collaborating with others. The research also shows that more than four in five (83%) participants consider collaboration around data preparation tasks important. Collaboration capabilities can help

foster and support the cross-functional line-of-business and IT participation that produces the best results. While not as important as collaboration, mobile capabilities can also help make data-preparation tasks more accessible and usable. Nearly half (45%) of organizations said they consider mobile access important.

### Big data drives increased interest in data preparation.

The research finds that nearly half (46%) of organizations are using their data preparation technologies to work with big data and more than half (53%) indicate that it is important to their organization to process large volumes of



data. As organizations spend more time working with big data, their appreciation for data preparation increases. Those who have been using big data for more than a year report (72%) that self-service data preparation without the involvement of IT is very important more often than those using it for a shorter period of time or not using it. They also are most likely to report they are satisfied with their data preparation technology (48%).

Working with big data can be challenging because of the size and complexity of the data sets. One-third (35%) of organizations reported that big data technology skills are necessary to prepare data successfully. Data preparation tools can provide an easier and faster way to process this data. Those organizations using big data for more than a year are least likely to complain that their data preparation technology is too slow. Overall, accessing big data technologies is one of the least-cited barriers (12%) to making improvements to data preparation. Big data is also influencing changes, as more than one-fourth of organizations (27%) said they will consider utilizing big data as they assess and select data preparation technology in the next 12 to 18 months.

### **On-premises use cases dominate data preparation, but cloud computing is on the horizon.**

On-premises to on-premises processes dominate the data preparation landscape with nearly two-thirds (64%) of organizations processing data in this manner. Among European participants, on-premises processing is even more prevalent (71%). Approximately one-fourth (27%) of participants are moving data from on-premises to the cloud or vice versa, but only 15 percent are performing data preparation processes that operate entirely within the cloud. Over the next 12 to 24 months the highest priority for use is on-premises to cloud followed by cloud to cloud (13%) providing insight to the future of data preparation.



Almost half of research participants said they plan to change the way they assess and select data preparation technology in the next 12 to 18 months.

These patterns are consistent with our prior research, which suggests there are no unusual requirements for data preparation that inhibit the adoption of cloud-based technologies. These patterns also match the way organizations prefer to deploy data preparation software with nearly six in 10

(57%) preferring on-premises deployments. Nearly two in five (38%) have no preference or prefer the cloud, reinforcing the notion that data preparation software can be deployed in the cloud.



## **Organizations are reevaluating data preparation.**

Data preparation technology has advanced significantly in the last few years and organizations are reevaluating their approach to these processes. Almost half (45%) of research participants said they plan to change the way they assess and select data preparation technology in next 12 to 18 months. Most often (53%) these changes are driven by a business improvement initiative, which provides the appropriate rationale for such an investment but cost is a barrier: Nearly six in 10 organizations (58%) cite it as a barrier issue, followed by inadequate skills (35%), limited awareness (33%) and lack of resources (33%). On the other hand, issues such as latency, big data and scalability are least likely to be barriers, suggesting the obstacles are organizational rather than technical.

When organizations decide to purchase data preparation technologies, they most often prefer to acquire these capabilities from a business intelligence vendor. The research finds that two-thirds (68%) said they would purchase from BI vendor, whereas half as many (35%) indicated they would purchase from a specialized vendor. Data integration vendors appears slightly more popular at 42 percent. These purchase preferences correspond with the primary uses of data preparation for analysis, extracting and accessing data. As organizations evaluate data preparation processes, they should consider primary-use cases and the needed requirements for the roles who need these types of tools that would be most valuable to their 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.)

**These Key Insights are drawn from the full Ventana Research Benchmark Research report. The full report is available for purchase, payable by check or credit card. Advice and focused guidance based on this benchmark research can be purchased through our Ventana On-Demand service. For more information about the full Benchmark Research report or assessment of your organization using our Performance Index methodology, please contact us at [sales@ventanaresearch.com](mailto:sales@ventanaresearch.com).**