

# Data Preparation

Enabling Self-Service and Support Across Business and IT

WHITE  
PAPER



## 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)



# Table of Contents

<b>Executive Summary.....</b>	<b>4</b>
<b>Key Insights.....</b>	<b>8</b>
Organizations' data preparation performance varies widely.....	8
Organizations derive significant value from data preparation.....	9
Data preparation supports analytics and business intelligence.....	9
Data preparation utilizes frequent integration of multiple sources.....	10
Hope for self-service data preparation is not yet fulfilled.....	10
Cross-functional data preparation teams produce the best results.....	11
Data preparation requires usability and collaboration.....	11
Big data drives increased interest in data preparation.....	12
On-premises use cases dominate data preparation, but cloud computing is on the horizon.....	13
Organizations are reevaluating data preparation.....	13
<b>10 Best Practice Recommendations.....</b>	<b>15</b>
Establish a data preparation strategy.....	15
Create clear goals for data preparation technology.....	15
Utilize dedicated data preparation tools.....	16
Provide appropriate training for data preparation.....	16
Use cross-functional teams for data preparation.....	16
Enable frequent data integration.....	16
Select tools that provide usability and the necessary functionality.....	17
Support big data with your data preparation processes.....	17
Consider the value of cloud and hybrid deployments.....	17
Assess the shortcomings in your data preparation efforts.....	18
<b>About Ventana Research .....</b>	<b>18</b>
<b>Appendix: About This Benchmark Research.....</b>	<b>19</b>
Methodology.....	19
Qualification.....	19
Demographics.....	20
Company Size by Workforce .....	20
Company Size by Annual Revenue.....	21
Geographic Distribution.....	21
Industry .....	22
Job Title .....	22
Role by Functional Area.....	23



## Executive Summary

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; the most common uses are to support analytics and decision-making, enable effective process improvements and optimize the customer experience.

Ventana Research defines data preparation as 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

“

Businesses need flexible tools that enable them to enrich the context of data drawn from multiple sources and collaborate on its preparation.

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. Tools that provide these capabilities are referred to as data preparation tools. Users of these tools range from analysts to operations professionals in the lines of business to IT professionals.

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. These technical factors will likely increase the need for IT professionals to be involved in preparing data. Nonetheless, the trend toward

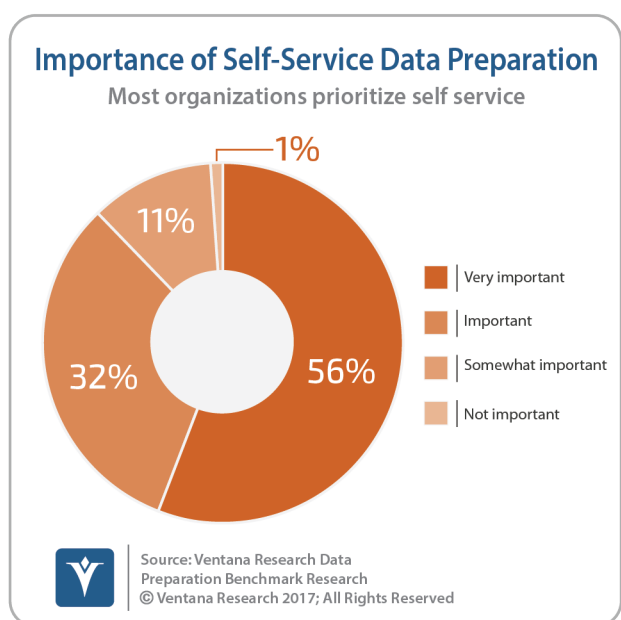
deploying tools that support self-service data preparation is growing. Self-service tools enable analysts to perform all or many of the data preparation tasks without the assistance of IT. Taken together, these two trends can lead to conflict for organizations that want to derive maximum business value from their data as quickly as possible while still maintaining the appropriate data governance, 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. The research explored how organizations manage data preparation processes, issues they encounter and how their use of data preparation and related technology is evolving.



Data preparation has unquestionably provided an opportunity for organizations to change the way they approach information management, but overall, organizations have not embraced these changes. Four years ago our Information Optimization Performance Index analysis found that more than half of organizations (52%) placed at the top two levels of our performance hierarchy compared to only 43 percent placing at those levels in this research. As in this research, the Information Optimization benchmark research also looked at the processes of collecting, preparing and deploying data throughout organizations. This decline suggests that many organizations need to improve their use of data preparation with a dedicated approach.

Two changes may be driving this decline: the growing complexity of data both in terms of volume and variety and a greater focus on enabling line-of-business users to work with data independent of their IT organizations. It's worth noting, though, that lackluster performance is not an indication of organizations' interest in data preparation:



88 percent of participants said that self-service data preparation is important to their organizations. Those organizations that didn't consider it important cite security, governance or risk issues as their main concerns. Despite a high level of interest in providing self-service data preparation, the reality is that organizations have not succeeded in deploying these capabilities.

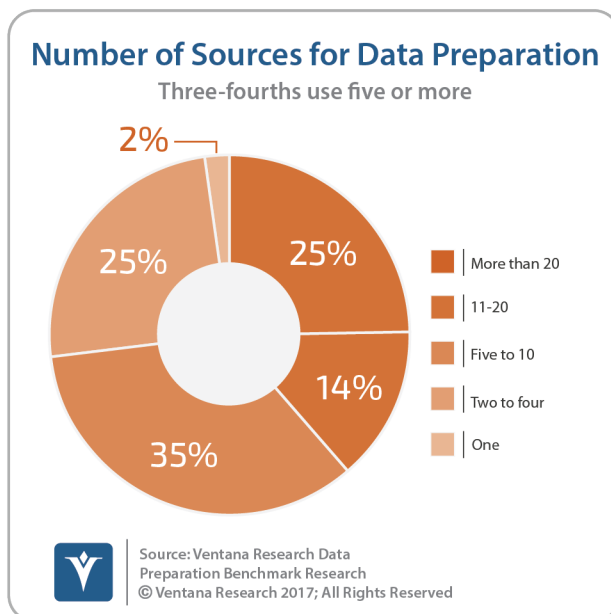
Organizations take different approaches to data preparation. Nearly half (47%) of organizations participating in the research use a dedicated product for self-service data

preparation. However, this is not typically their primary tool. The largest portion (41%) of organizations use analytics or BI tools as their primary tool for data preparation. Overall, two-thirds (67%) of organizations are satisfied or somewhat satisfied with their current technology, suggesting there is some room for improvement. Even though dedicated data preparation tools may not be the primary tool, organizations using a dedicated tool report satisfaction at higher rates (87%) than those that do not have a dedicated tool available (50%).

Regardless of the approach they use, organizations want their data preparation tasks readily available for reuse and they need to be able to join disparate data sources during data transformation, the most commonly reported critical data preparation capabilities. Users emphasized reusability and IT personnel emphasized joining data. In terms of system-level capabilities, organizations most often want to be able to process large volumes of data and



connect to databases and applications. These capabilities and others are delivering value, with three-quarters (76%) reporting their data preparation technologies have improved their organization's processes. They most often



cited as benefits of having data preparation capabilities improved quality and consistency of information, meeting the organization's analytic needs more easily and reducing or eliminating manual processes.

Ironically, while data preparation is helping organizations meet their analytic needs, three out of four participants reported that analysis is the activity most often required for data preparation. Other top requirements include extracting data, accessing data and data quality, each cited by more than half of the participants. Data preparation

typically involves data sources including accounting or financial management systems, data warehouses and operational data stores. Cloud computing business applications were most often cited as an important external data source for data preparation. Three-quarters (73%) of all participants are working with five or more data sources in their data preparation activities. Organizations are also often working with big data sources; about half (46%) use their current data preparation technology to work with big data sources.

Data preparation tools are meeting organizations' needs in some cases but the research suggests plenty of room for improvement. Just more than half (56%) consider their data preparation technologies completely or mostly adequate. A slightly higher percentage (62%) report confidence in their organization's ability to prepare data. However, fewer than half (44%) are comfortable allowing users to work with data not prepared by IT. Furthermore, many users complain that their data preparation technology is not flexible or adaptable when change is needed and IT's top complaint is that it requires too many resources. This difference points to a broader disconnect between business units and IT: They do not always see eye to eye on data preparation issues. Nearly half (45%) of participants report that the top issue between the two groups is their differing view on access to data, with business units preferring an expansive approach and IT preferring a controlled approach.

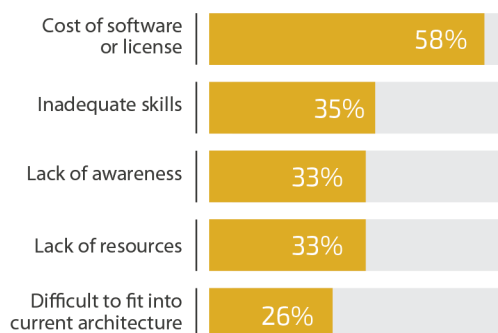
Many organizations (45%) expect to be reevaluating the way they assess and select data preparation technology in the next 12 to 18 months. When





### Barriers to Improving Data Preparation

Software costs are top obstacle



Source: Ventana Research Data  
Preparation Benchmark Research  
© Ventana Research 2017; All Rights Reserved

considering technology options and vendors, organizations rated usability and functionality the most important evaluation criteria. However, cost is a barrier for these organizations: Nearly six in 10 organizations (58%) cite it, making it far and away the most often selected 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 are slightly more popular at 42 percent. These purchase preferences correspond with the primary uses of data preparation: analysis, extracting and accessing data. As organizations evaluate data preparation processes, they should consider primary-use cases to determine the types of tools that would be most valuable to their organization.



## 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.

“

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

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 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 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,

“

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

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 capa-

“

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

bilities. 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.

### **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

“

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

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 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 impor-

“

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

tant. 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

“

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.

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.

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.





## 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 that can help your organization begin 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 an 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

## Methodology

Ventana Research conducted this benchmark research on the web from January through July 2017. We solicited survey participation via email, our website and social media invitations. Email invitations were also sent by our media partners and by vendor sponsors.

We presented this explanation of the topic to participants prior to their entry into the survey:

Using data effectively requires first that it be prepared for use. This typically involves a sequence of steps: accessing the data, perhaps through search; aggregating it; and enriching, transforming and cleaning data from different sources to create a single, uniform data set. To do the job of data preparation properly, organizations need flexible tools that enable them to enrich the context of data drawn from multiple sources, collaborate on its preparation and govern the process of preparation to ensure consistency and security. Users of these tools range from analysts to operations professionals in the lines of business to IT professionals.

The following promotion incited participants to complete the survey:

What's In It For You? Upon completion of the research, all qualified participants will receive a report on the findings of this benchmark research to support their organization's efforts, along with a \$25 Amazon.com gift certificate. In addition, all qualified participants will be entered into a drawing to win one of 25 benchmark research reports and a 30-minute consultation, a package valued at US\$1,495 or €1,232. Thank you for your participation!

## Qualification

We designed the research to assess the use of and plans for data preparation across organizations and industries. Qualification to participate was presented to participants as follows:

The survey for this benchmark research is designed for individuals who participate in data- and analytics-related processes. Solution providers, software vendors, consultants, media and systems integrators may participate in the survey, but they are not eligible for incentives and their input will be used only if they meet the qualifications. Incentives are provided to qualified participants in the research and also are conditional on provision of accurate and verifiable contact information including company name and company email address that can be used for fulfillment of incentives.



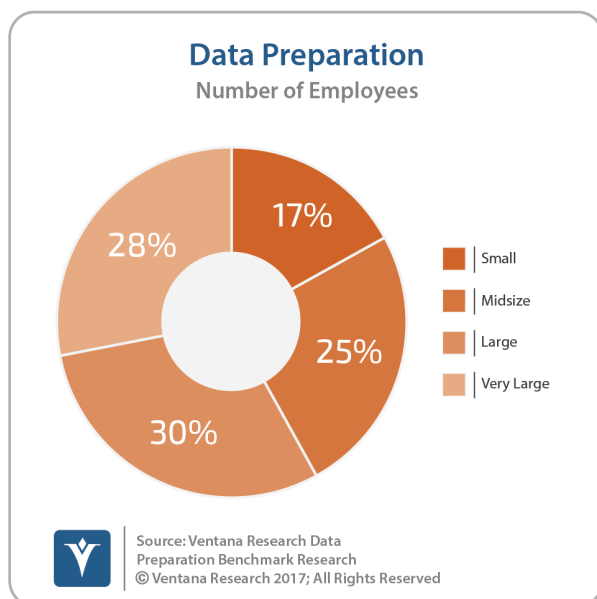
Further qualification evaluation of respondents was conducted as part of the research methodology and quality assurance processes. It entailed screening out responses from companies that are too small, questionnaires that were not materially complete, or those where the submission is from an inappropriate submitter or appears to be spurious.

## Demographics

We designed the survey used for this research to be answered by executives and managers across a broad range of roles and titles working in organizations. We deemed 179 of those who clicked through to this survey to be qualified to have their answers analyzed in this research. In this report, the term “participants” refers to that group, and the charts in this section characterize various aspects of their demographics and qualifications.

### Company Size by Workforce

We require participants to indicate the size of their entire company. Our



research repeatedly shows that size of organization, measured in this instance by employees, is a useful means of segmenting companies because it correlates with the complexity of processes, communications and organizational structure as well as the complexity of the IT infrastructure. In this research, participants represented a broad range of organization sizes in similar numbers: 28 percent work in very large companies (having 10,000 or more employees), 30 percent work in large companies (with 1,000 to 9,999 employees), 25 percent work in midsize companies (with 100 to 999 employees), and 17 percent work in small companies (with fewer than 100

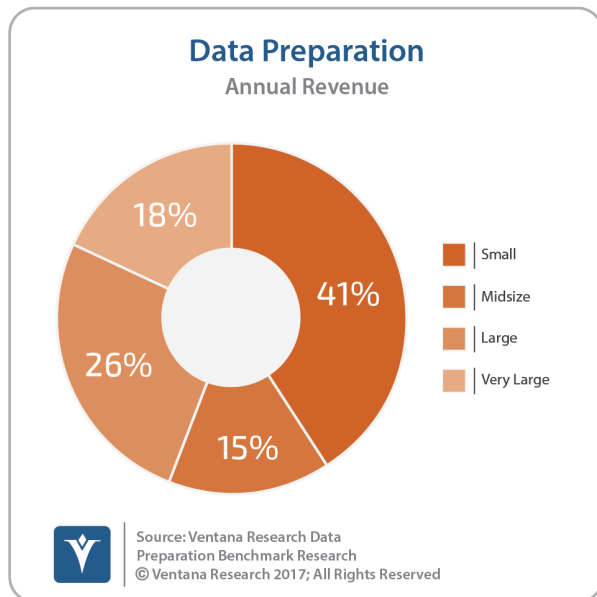
employees). This distribution is consistent with prior benchmark research and our research objectives and provides a suitably large sample from each size category.



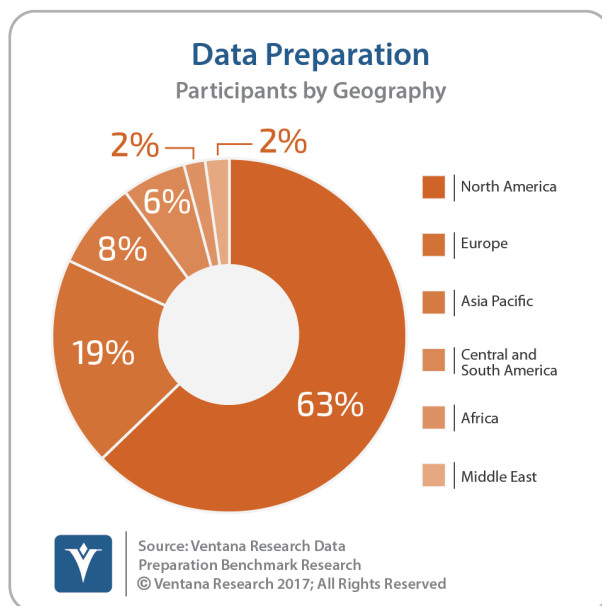


## Company Size by Annual Revenue

When we measured size by annual revenue, the distribution of categories

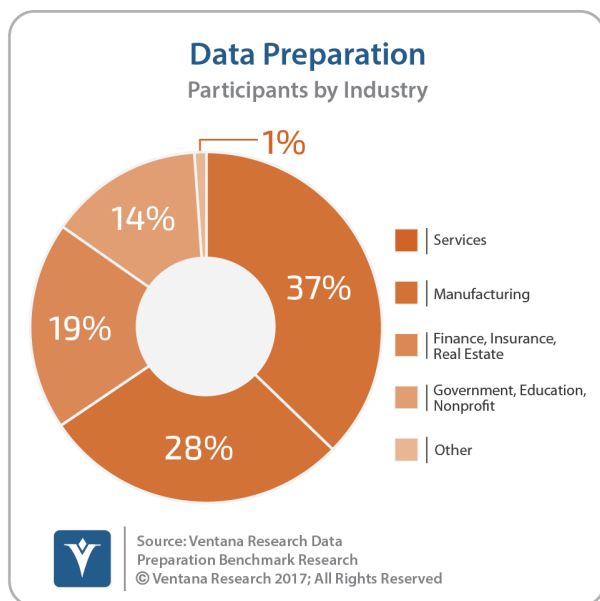


shifted downward between the two largest and two smallest divisions. By this measure, 10 percent fewer are very large companies (having revenue of more than US\$10 billion), 4 percent fewer are large companies (having revenue from US\$500 million to US\$10 billion), and 10 percent fewer are midsize companies (having revenue from US\$100 to US\$500 million), but 24 percent more are small companies (with revenue of less than US\$100 million). This sort of redistribution is typical in our research when we measure by revenue instead of head count.



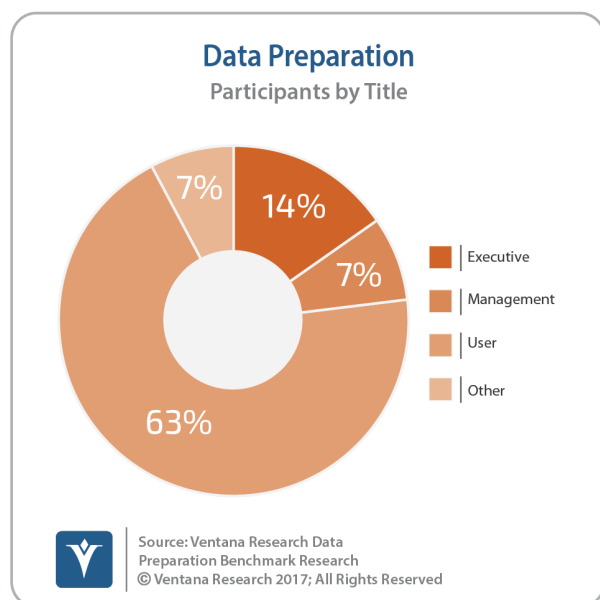
## Geographic Distribution

A majority (63%) of the participants were from companies located or headquartered in North America. Those based in Europe accounted for 19 percent and Asia Pacific for 8 percent. Of the remainder, 8 percent were from Central and South America and 2 percent were from Africa and the Middle East, respectively. This result was in keeping with our expectations at the start of this investigation, since organizations participating in our research most often are headquartered in North America. However, many of these are global organizations operating worldwide.



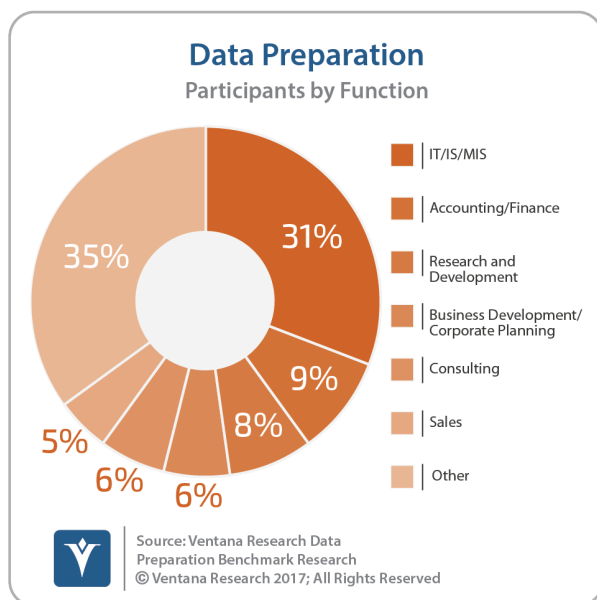
## Industry

The companies of the participants in this benchmark research represented a broad range of industries, which we have categorized into four general categories as shown in the adjacent chart. Companies in services accounted for 37 percent and those in manufacturing accounted for 28 percent. Those in finance, insurance and real estate accounted for 19 percent. Government, education and nonprofits accounted for 14 percent and a miscellaneous other category for the balance.



## Job Title

We asked participants to choose from among 12 titles the one that best describes theirs. We sorted these responses into four categories: executives, management, users and others. Nearly two-thirds identified themselves as having titles that we categorize as users, a grouping that includes analyst (23%), senior manager or manager (22%), director (10%) and staff (8%). One in seven are executives, and 7 percent are management, by which we mean vice presidents. Others, in this case consultants, accounted for the balance.



## Role by Functional Area

We asked participants to identify their functional area of responsibility as well. This enabled us to identify differences between participants who have differing roles in the organization. Predictably, nearly one-third of the participants identified themselves as being in the IT/IS/MIS function; 9 percent work in accounting; 8 percent in research and development and 6 percent in business development. Six percent of participants work in consulting and 5 percent in sales; the remaining 35 percent comprise the Other category.