Data Preparation

Enabling Self-Service and Support Across Business and IT

VENTANA RESEARCH

Prepared for

DATAWATCH

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

VENTANA RESEARCH
Bend, Oregon, USA
541-940-1010
info@ventanaresearch.com

Table of Contents

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 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.

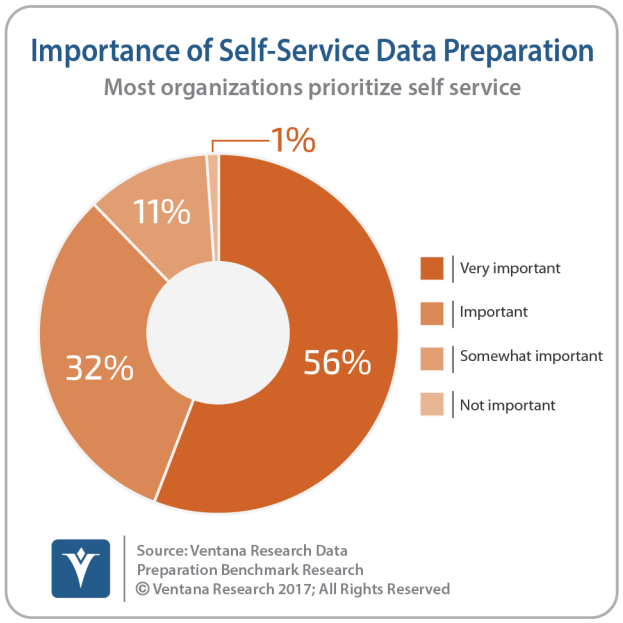
> Businesses need flexible tools that enable them to enrich the context of data drawn from multiple sources and collaborate on its preparation.

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.

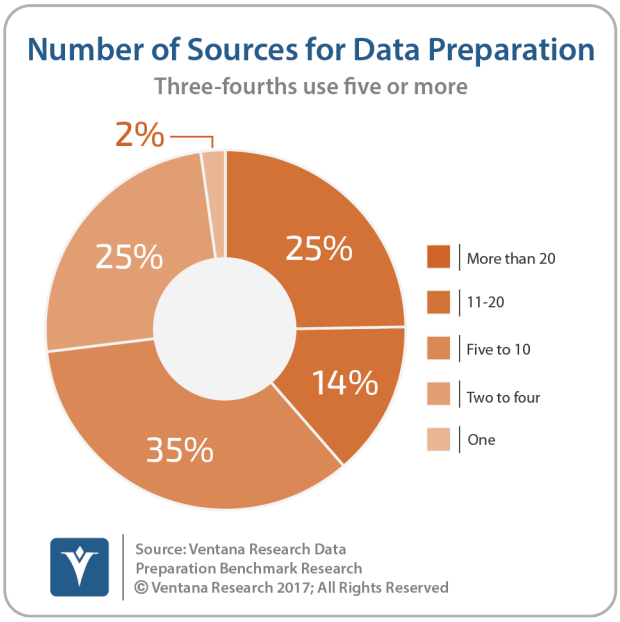
Two changes may be driving this decline: the growing complexity of data both in terms of volume and variety and a greater focus on enabling line-of-business users to work with data independent of their IT organizations. It's worth noting, though, that lackluster performance is not an indication of organizations'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

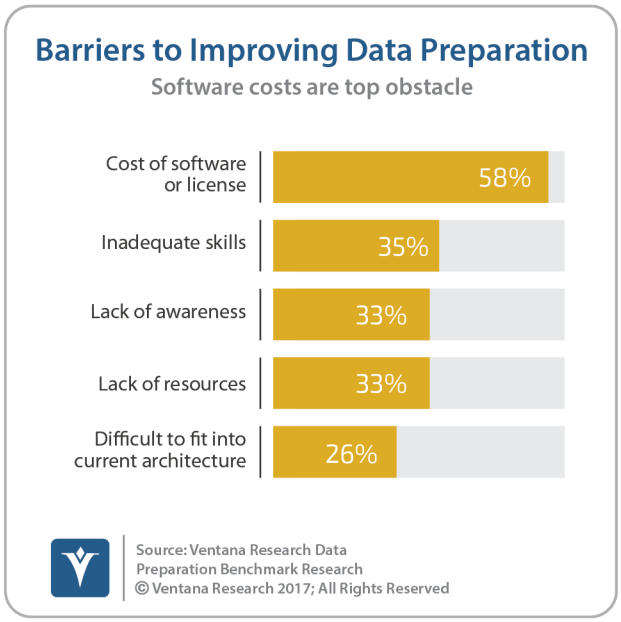
connect to databases and applications. These capabilities and others are delivering value, with three-quarters (76%) reporting their data preparation technologies have improved their organization's processes. They most often cited as benefits of having data preparation capabilities improved quality and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



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. 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 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.

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

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 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.

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

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 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.

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

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 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.

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

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.

Performance Index

In a competitive market environment, the organization that carries out its business operations most efficiently and accomplishes its business goals most effectively is likely to have a significant competitive advantage. Over time it will consistently be more profitable and more agile in responding to market shifts. It will attract the most talented people and, with all of these advantages, it will have the edge in gaining market share. These organizations – ones whose people, processes, stores of information and technology assets are deployed and used to maximum benefit – are the best performers in their market or industry. How they address their business challenges offers models for other organizations to learn from and emulate.

But sufficient information about those models typically isn't readily available. Businesses are not in the habit of sharing their differentiators with competitors. Often, reports of best practices passed along by technology vendors or IT analyst firms reflect the biases of those channels. What's needed is a trusted source of rigorous, unbiased and statistically reliable information about trends and best practices supplied by those who actually know whereof they speak.

The Ventana Research Performance Model

To satisfy this need, Ventana Research has developed the Ventana Research Performance Model™, a research-based evaluation of the performance of organizations in a given market or with respect to a business or IT process. Once Ventana Research has conducted benchmark research in an area, the Performance Model enables evaluation of the performance of a particular organization in comparison to the universe of organizations.

The Performance Model for a given business area is developed as part of the benchmark research methodology. It identifies in each of the four component areas of any business or IT operation – the people, processes, information and technology – the elements that are required to accomplish successfully the relevant business goals, then establishes for each the range of real-world approaches being used to do so. The analyst then builds into the research survey a series of questions that will ascertain where an organization stands on a performance hierarchy that spans from inefficient to efficient, ineffectual to effective, ad-hoc to automated, stand-alone to integrated – in short, how well it performs in that area.

The statistical performance analysis of a research data set begins by applying the performance model to establish the characteristics of each of the model's four levels of performance effectiveness – Tactical, Advanced, Strategic and Innovative – in each of the four component areas. When these individual analyses are complete, they are aggregated to yield an analysis of the distribution of overall performance by organizations in that area. We also

analyze performance across organizations by key demographic criteria: size by number of employees and annual revenue, geographic location and industry.

Since the analysis is based on answers to dozens of questions addressing the four component areas and uses correlations and crosstab analysis to elicit subtle but important patterns, the resulting analysis can be exceedingly detailed and insightful, yielding powerful guidance on initiatives that need to be undertaken to advance performance in lagging areas. It also provides a yardstick against which we can measure the relative performance of any organization's management and business processes.

In brief, the four levels of increasingly effective performance are:

Tactical: Organizations that are either unaware that they are lagging the norm or are complacent about their capabilities. These organizations embrace and often defend the status quo, and do not encourage thinking about innovation. To the extent that they perceive there to be a problem, they see it as isolated.

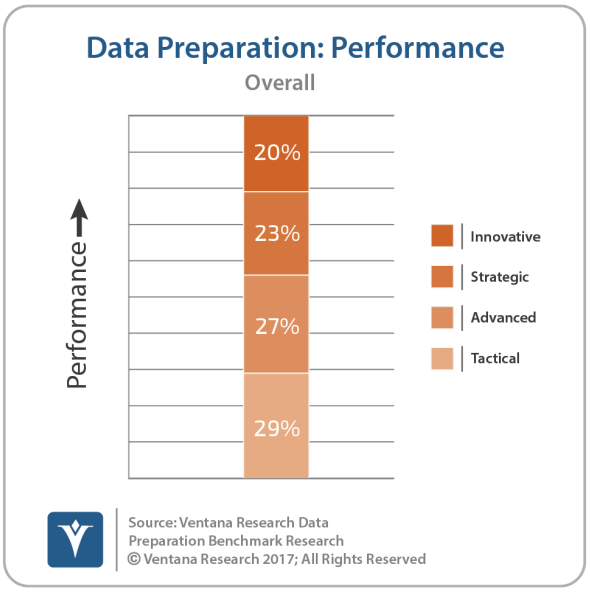
Advanced: Organizations that are somewhat dissatisfied with the status quo and somewhat open to the possibility of change. However, their employees are uncertain how to proceed. Some explore new processes or the acquisition of new technology but typically lack budget resources or executive support. They may suspect their individual problem has larger systemic implications, but they do not know how to pursue that suspicion.

Strategic: Organizations that are dissatisfied with the status quo. They have taken steps to address employee efficiency and performance. They consider process innovations and new technology that can help improve their decision-making and operational processes, but lacking key skills or grappling with organizational inertia or budget limitations may constrain the extent of their ability to change. They are open to examining larger systemic concerns.

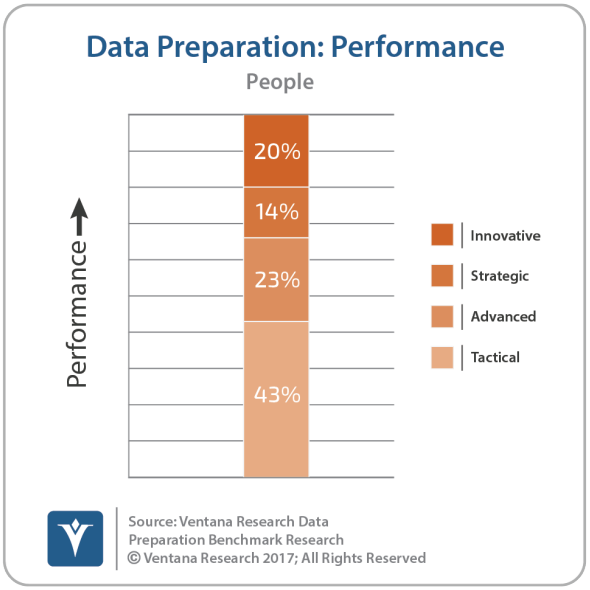
Innovative: Organizations that understand fully the value of performance improvement. They measure the performance of their people and processes, streamline processes and collect and make accessible all information that people need to do their jobs and make decisions. They understand that the operation of all systems is interrelated and so optimization must be managed across the enterprise.

Performance Index: Overall

Our overall performance analysis of data preparation practices finds that most organizations participating in this research have room for improvement: More than half (56%) of organizations occupy the two lowest

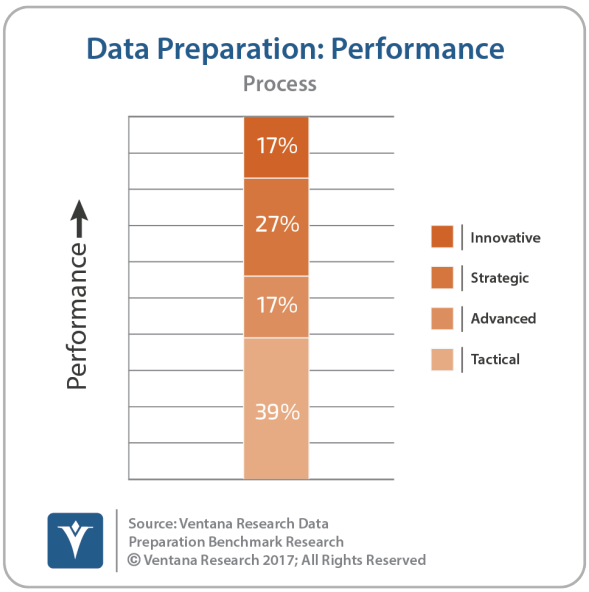


levels of our performance hierarchy, and only one in five organizations are at the top Innovative level. Further analysis among the four dimensions into which we segment this index reveals varying levels of performance, with the lowest performance in the Process and People dimensions. The highest percentage of Innovative organizations (21%) is in the Information dimension, but that figure is only four percentage points higher than the same figure for Process, the dimension with the fewest organizations at the Innovative level. Also, Process shows nearly two out of five (39%) organizations at the Tactical level. Performance is not much stronger in the People dimension, with 43 percent of organizations at the Tactical level and one in five at the Innovative level. Thus we conclude that nearly all organizations have at least some room for improvement in all phases of data preparation, and many have significant opportunities to better their performance.



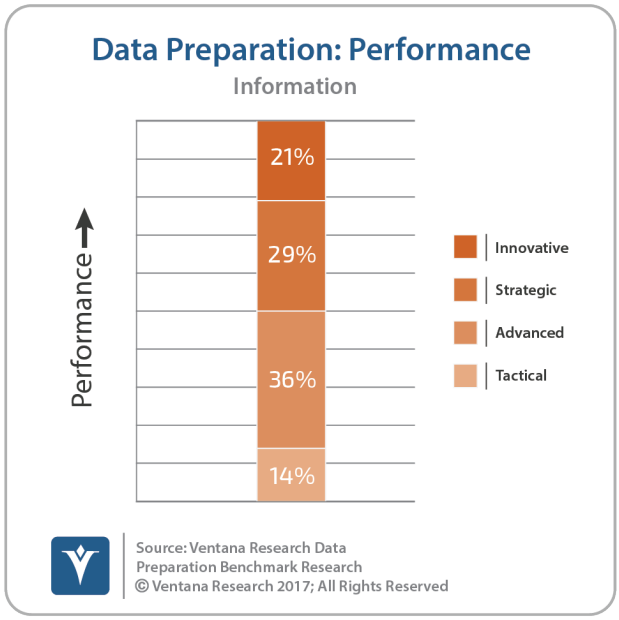
Performance Index: People

Our analysis reveals that many organizations struggle in the People dimension of data preparation. This is not surprising because data preparation can be complex and rapid advances in technology have led to a steep learning curve. Two out of three organizations rank at the lower two levels, with 43 percent occupying the lowest Tactical level. Drilling down into the results, many organizations reported that data preparation requires too many resources and too broad a skill set. Current training in technologies and techniques leaves room for improvement, with only 13 percent reporting that training is completely adequate; 17 percent said the same about training in data preparation concepts and techniques. The research also finds issues arising due to differing approaches to data preparation among business units, with 58 percent citing differences of either philosophy, access to information or budgets and priorities.



Performance Index: Process

Our analysis finds the weakest dimensional performance in the Process dimension of data preparation: only 17 percent of organizations are at the Innovative level, and nearly two out of five are at the Strategic level. Our analysis suggests the reasons for this include subdued confidence in processing large quantities of data at high velocities – only 13 percent of participants said they were very confident in their organization's ability to process high-velocity data and 17 percent said they were very confident in their organization's processing of large volumes. However, there are signs of improvement. Nearly nine in 10 organizations (88%) said self-service is important in data preparation, and most (64%) said they were evaluating or planned to evaluate the way they process data. Cost is a frequently-cited barrier, as is a lack of awareness of the topic, but those organizations that have improved data preparation capabilities report improved efficiency and savings.

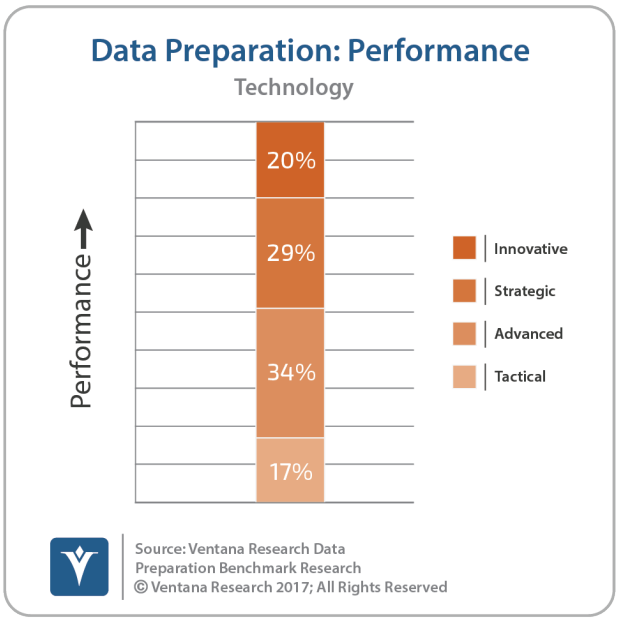


Performance Index: Information

Organizations perform best in the Information dimension of this analysis, with half at the top two levels of performance and only 14 percent at the lowest Tactical level. Most organizations do reasonably well at amassing and using data, and not many do poorly at managing it. Nearly two-thirds (64%) of participants reported confidence in their organization's ability to process data and a similar percentage (65%) expressed confidence in the quality of data. The speed with which data is integrated has become table stakes, with nearly one in five organizations reporting real-time integration capabilities. Nearly half (48%) integrate data every day and 18 percent perform this activity on a weekly or monthly basis.

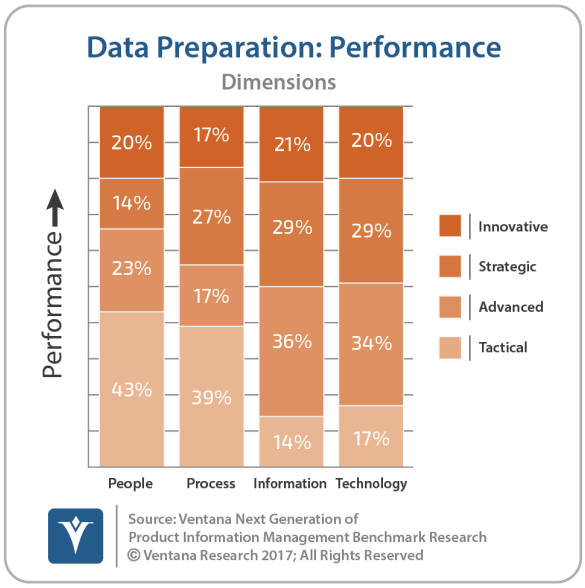
Performance Index: Technology

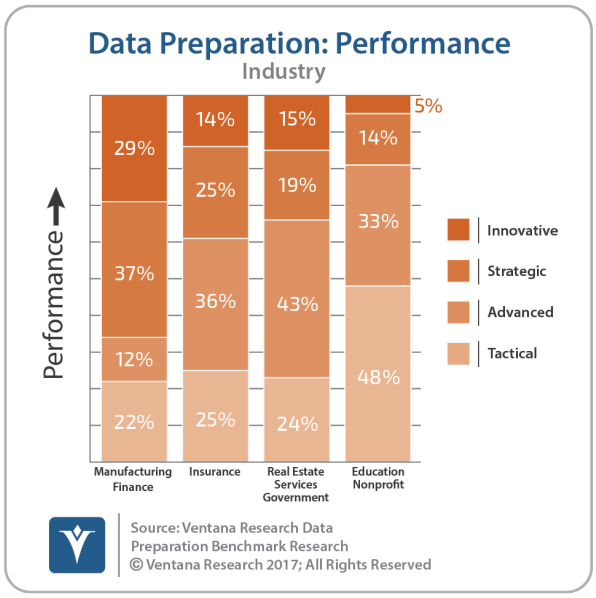
In the Technology dimension our research finds a distribution of participating organizations similar to that in the Information dimension. One in five organizations are at the highest Innovative level, with slightly fewer (17%) at the Tactical level. More than two-thirds (68%) of participants reported satisfaction with data preparation technology and 57 percent said the technology adequately meets the organization's needs. Nearly half of organizations use a dedicated product for data preparation, and another 14 percent said they plan to use such a product within a year. Furthermore, more than four out of five organizations say that data preparation technology has improved business activities either significantly or slightly. When selecting technology for data preparation, organizations prioritize collaboration, cited as important by 83 percent of organizations, and mobile access, cited as important by 44 percent.



Performance Index: Key Dimensions

Our analysis shows that organizations participating in this research have room for improvement across the four dimensions, particularly in the People and Process dimensions, where more than half of organizations occupy the bottom two tiers. In all dimensions but Process, though, most organizations are at the Advanced and Strategic levels. Overall, organizations are able to handle their data preparation with varying degrees of competence. Those that need to advance can do so by improving training and collaboration among departments in order to align differing strategies and priorities. Technology designed specifically for data preparation can significantly bolster these efforts.



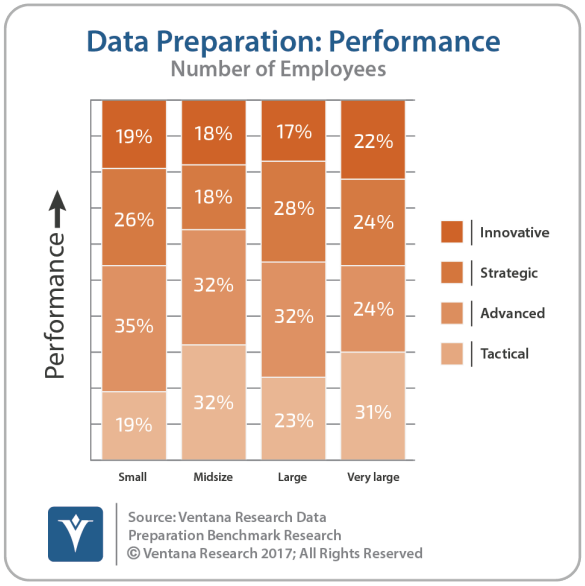


Performance Index: Industry Sector

In analyzing the research findings by industry sector, we find that manufacturing organizations perform best in data preparation; two-thirds of organizations are at the two highest levels. This of course is a sector that depends on data preparation to maintain adequate inventories and up-to-date shipment schedules. Organizations in the Finance and Services sectors didn't perform as well and reveal similar results, indicating in each case room for improvement. Most organizations in those industries are at the middle Advanced and Strategic levels, with one in four at the Tactical level and fewer than one in five at the Innovative level. Nearly half (48%) of Government, Education and Nonprofits are at the Tactical level, revealing a basic competence in data preparation.

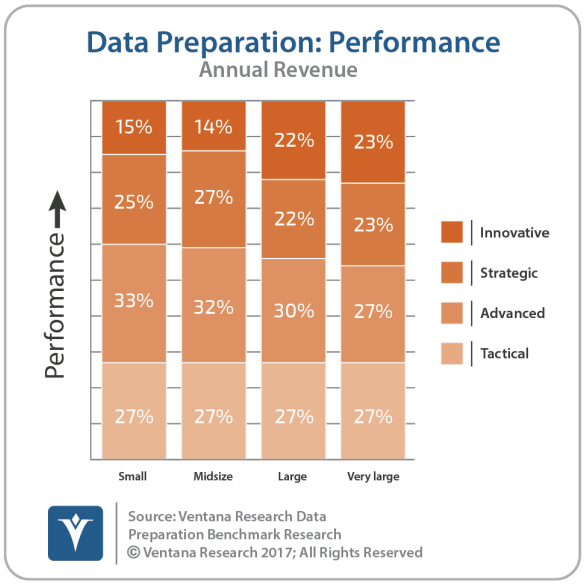
Performance Index by Size: Number of Employees

We assessed the performance of organizations in relation to their size as determined by how many employees they have. (In our definition, small organizations have fewer than 100 employees; midsize organizations have 100 to 999 employees; large organizations have 1,000 to 9,999 employees; and very large organizations have 10,000 or more.) We find slightly better performance at the high end of the size spectrum, but overall, most organizations occupy the lowest two levels. Performance is distributed somewhat evenly for large and very large companies, but midsize companies have nearly two-thirds (64%) at the Tactical and Advanced levels. Small companies performed better, with nearly half (45%) at the Innovative or Strategic levels.



Performance Index by Size: Revenue

We also assess the performance of organizations in relation to their size as determined by their annual revenue. (In our definition, small organizations have revenue of up to US$100 million; midsize have revenue between US$100 million and US$500 million; large organizations have revenue between US$500 million and US$10 billion; and very large organizations have revenue of more than US$10 billion.) Measured by revenue, the advantage of large and very large companies becomes clearer, with 46 percent of very large companies and 44 percent of large companies at the Strategic or Innovative levels. Only 14 percent of midsize and 15 percent of small organizations are at the Innovative level, with distribution otherwise relatively even. Similar to when measured by number of employees, most organizations occupy the Advanced and Strategic levels.



Ventana Research's Recommendations to Datawatch

This benchmark research was designed to explore and understand actions, intentions, perceptions and trends. It examines in detail to what degree organizations are implementing data preparation technology and what issues they could successfully address if they had advanced capabilities. We base the following recommendations on the findings of this research as well as our knowledge of data preparation and the market for such software. The perspectives offered here are drawn from the work we have done in this area in the past 15 years.

1. Help your customers establish a data preparation strategy.

Data preparation can be valuable to organizations; the research finds three-quarters of participants indicating that it has improved their activities or processes. Drilling down into results, 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, with more than one-third of organizations not fully confident in data preparation, data quality or related technologies, your organization should target areas for improvement. Helping your customers and prospects create a data preparation strategy can position your organization to provide the software that realize these benefits.

2. Help organizations understand the full set of business and IT requirements for data preparation.

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, we see differing views regarding working with data prior to its integration; more than half (51%) of those in business functions say are comfortable with business users working with data that hasn't been processed by IT while only one-third (32%) of those in IT say that. These differences suggest that challenges go beyond technology and must be addressed with improvements in data preparation processes. Consider the nature of the issues in organizations and aim to provide technology that is flexible (an issue cited by 37 percent of users) and that doesn't require too many resources, cited by 36 percent of organizations.

3. Encourage organizations to assess their need for dedicated tools.

The research reveals a generally even split between the use of standalone data preparation tools and tools embedded within BI systems. Capabilities embedded within BI tools 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. Emphasize how your tools provide these capabilities and their importance to the data preparation process. Your sales processes should help your customers consider their organization's specific needs and decide whether dedicated tools have a role in their data preparation processes.

4. Provide appropriate training on 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 organizations have not kept up with these changes. As you develop a data preparation training curriculum, pay particular attention to preparing big data and web- and cloud-based data as the research finds training in these topics is the least adequate.

5. Encourage organizations to use cross-functional teams.

Data preparation involves a balance between functions within line of business 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 organizations to use such an approach; it is not the approach most used today. Business intelligence and data warehousing teams within IT often are the groups most likely to design and deploy data preparation tasks as are analyst teams in line of business. Unfortunately but perhaps not unsurprisingly, issues between IT and line-of-business functions are common, the most common being 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 customers with the best results.

6. Show the value of 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 an 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. Processing large volumes of data and providing connectors to databases and applications are important system capabilities. Help organizations see the value of data preparation processes that enable frequent data integration.

7. Prioritize usability and functionality in your presentation to prospects and customers.

The two evaluation criteria considered most important by participants are usability and functionality, priorities that make sense in the context of 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 evaluation criteria can support the cross-functional teams requirements and provide the best results.

8. Show value of big data with data preparation processes.

The research shows many organizations (46%) working with big data, and those organizations that have been working with big data 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 easier and faster way to process this data. Ensure your prospects and customers see the potential of utilizing big data as it becomes a higher priority over next 12 to 18 months.

9. Show the future of data preparation in the cloud and the continued value of on-premises.

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

10. Capture the opportunity as organizations change their data preparation processes.

Almost half (45%) of the organizations participating in this research said they are planning to change their data preparation processes over the next 12 to 18 months. As your customers assess their data preparation activities, consider how your organization and its products can help them through the process. Nearly six in 10 organizations cite cost as a barrier issue; other barriers include inadequate skills, limited awareness and lack of resources. As you develop data preparation technology, consider your customers' primary-use cases and the specific requirements of 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.

Appendix 1: 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 incented 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.

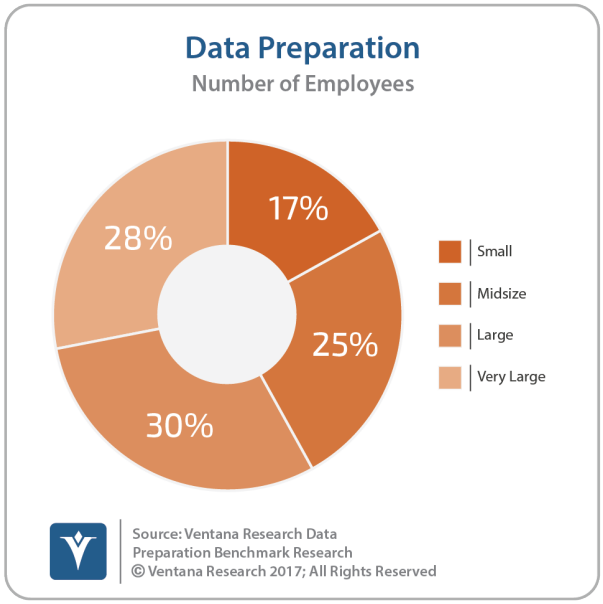
Further qualification evaluation of respondents was conducted as part of the research methodology and quality assurance processes. It entailed screening out responses from companies that are too small, questionnaires that were not materially complete, or those where the submission is from an inappropriate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

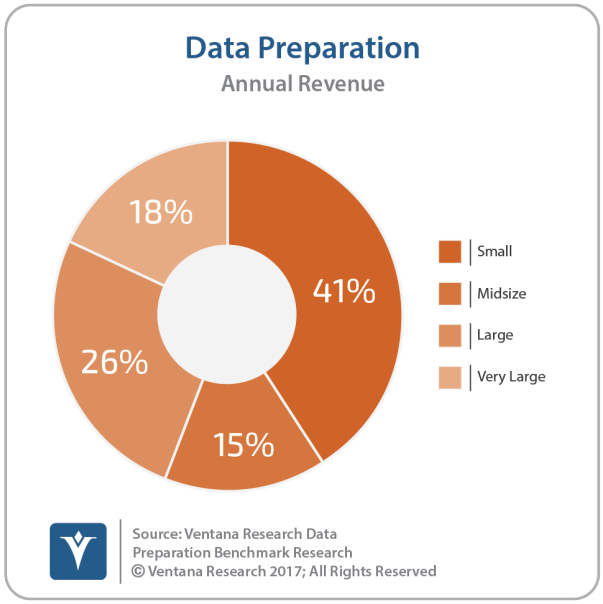
Company Size by Workforce

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.



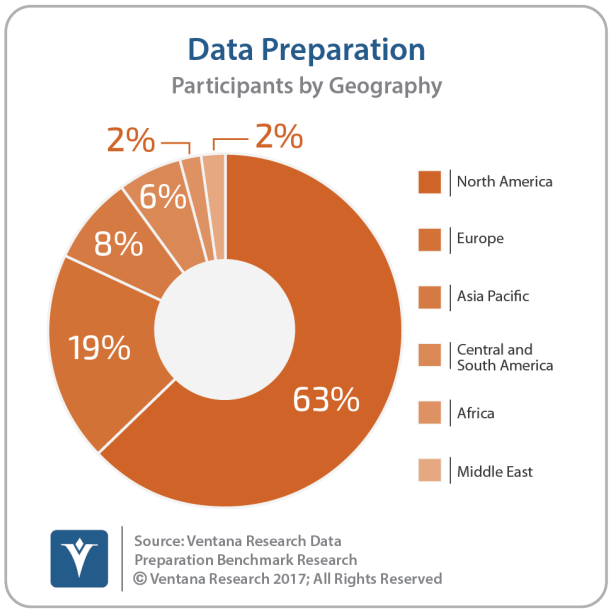
Company Size by Annual Revenue

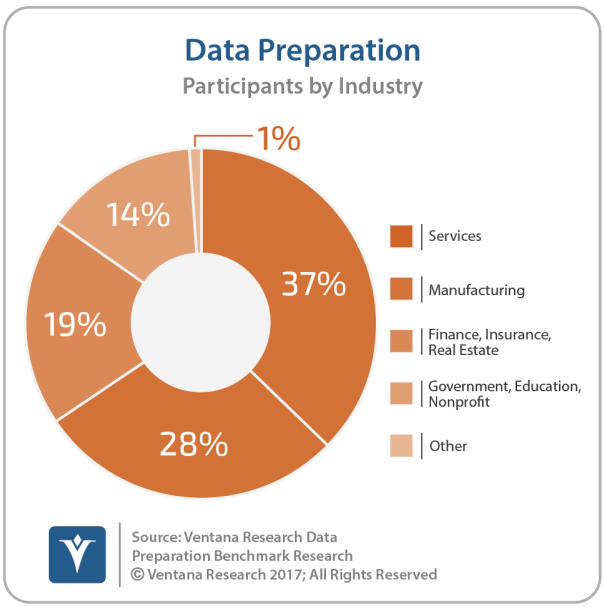
When we measured size by annual revenue, the distribution of categories shifted downward between the two largest and two smallest divisions. By this measure, 10 percent fewer are very large companies (having revenue of more than US$10 billion), 4 percent fewer are large companies (having revenue from 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

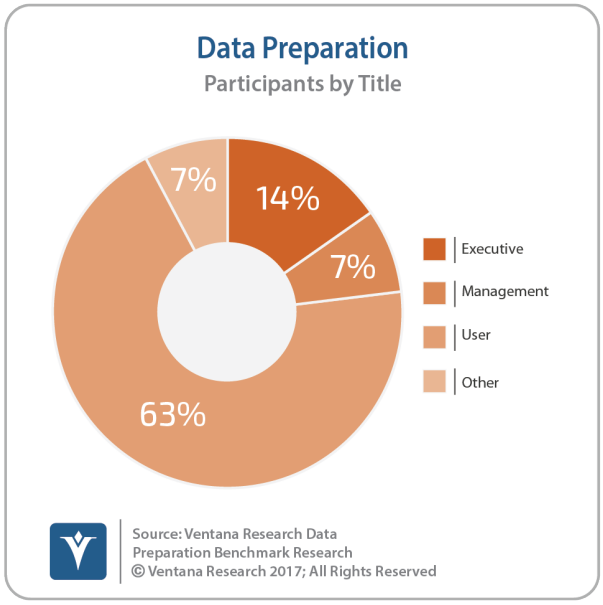
A majority (63%) of the participants were from companies located or headquartered in North America. Those based in Europe accounted for 19 percent and Asia Pacific for 8 percent. Of the remainder, 8 percent were from Central and South America and 2 percent were from Africa and the Middle East, respectively. 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.





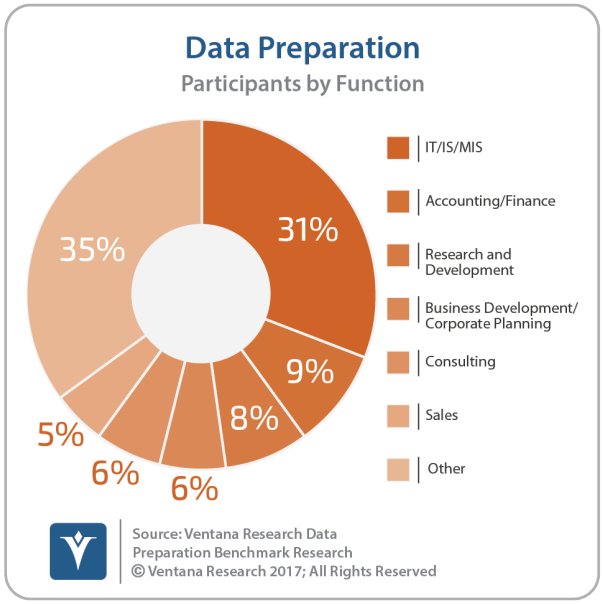
Industry

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.

Appendix 2: Questions

1. Which of the following best describes your title or role in your company?

CEO, President
COO or Head of Operations
CIO or Head of Information Technology
CFO or Head of Finance
Other CxO
Executive (EVP, SVP)
VP
Director
Senior Manager
Manager
Analyst (Business, Financial, etc.)
Staff
Consultant
Professor/Teacher
Student
Other (Please specify)

2. Which of the following best describes your functional area of responsibility?

Accounting/Finance
Administration
Business Development/Corporate Planning
Call or Contact Center
Compliance/Internal Audit
Consulting
Customer Support
Data Center/Networking
Distribution and Fulfillment
Education
Executive/Management
Field Service
Human Resources
IT/IS/MIS
Legal
Manufacturing
Marketing
Operations
Product Management/Marketing
Product Planning
Public Relations
Purchasing
Research and Development
Sales
Supply Chain
Training
Transportation
Warehousing
Other (Please specify)

3. How many people work in your entire company?

100,000 or more

50,000 to 99,999
25,000 to 49,999
10,000 to 24,999
5,000 to 9,999
2,500 to 4,999
1,000 to 2,499
500 to 999
250 to 499
100 to 249
1 to 99

4. What is your level of expertise in data preparation?

Business expert – I have a deep understanding of how to prepare data for business analyses but not necessarily of the technologies available.
Technology expert – I have a deep understanding of data preparation technologies but not necessarily of the business use and impact.
Knowledgeable – I have a good understanding of data preparation but am not an expert.
Novice - I have done a limited amount of data preparation.
None – I have not done data preparation.

5. How important is self-service data preparation – that is, accessing and preparing data for analysis without involvement of IT – to your organization?

Very important
Important
Somewhat important
Not important

6. How important is self-service data preparation – that is, accessing and preparing data for analysis without involvement of IT – to your organization?

Very important
Important
Somewhat important
Not important

7. Does your organization use or plan to use a dedicated product for self-service data preparation (rather than capabilities embedded in products such as data integration or business intelligence tools)?

Yes, we have had and used one for more than one year.
Yes, we began to use one within the last 12 months.
Not yet, but we will begin to use one within 12 months.
Not yet; we intend to begin to use one but do not know when.
No, we do not use or intend to use one.

8. What technology does your organization primarily use for data preparation?

Analytics or business intelligence tools
Custom programming
Database or data warehouse tools
Data integration tools
Data preparation tool dedicated to this task
Microsoft Office (Excel)

9. How satisfied are you with your organization's current data preparation technology?

Satisfied

Somewhat satisfied
Neither satisfied nor not satisfied
Not satisfied
Don't know

10. What are your major complaints about the current data preparation technology? (Select all that apply.)

Too slow
Too expensive
Too unreliable
Hard to use
Hard to build
Hard to maintain
Not flexible or adaptable to change
Requires too many resources
Hard to access our data sources
Technology does not have sufficient capabilities

11. To prepare data in your enterprise, which of the following data activities are required? (Select all that apply.)

Accessing
Analysis
Archiving
Auditing
Blending
Extracting
Governing
Loading
Masking
Mastering
Messaging
Metadata management
Migrating
Profiling
Quality
Real-time loading
Replicating (copying)
Securing
Synchronizing
Transforming
Virtualizing

12. Does your organization use or plan to use big data technology for storage of data? ("Big data technology" encompasses an RDBMS, a data warehouse appliance, a specialized DBMS, Hadoop, an in-memory database or another non-SQL approach to manage very large sets of data.)

Yes, we have had and used one for more than one year.
Yes, we began to use one within the last 12 months.
Not yet, but we will begin to use one within 12 months.
Not yet; we intend to begin to use one but do not know when.
No, we do not use or intend to use one.

13. Which of the following big data technologies does your organization use or plan to use?

In use now
Plan to use within 12 months
Plan to use in 12 to 24 months
Still evaluating
No plans to use

RDBMS (e.g., IBM DB2, Microsoft SQLServer, MySQL, Oracle) on standard hardware
Data warehouse appliance (e.g., Netezza, Exadata, Teradata)
Specialized DBMS (e.g., Amazon Redshift, Aster Data, HP Vertica, Infobright, SAP/SybaseIQ, Snowflake)
Flat files
Hadoop
In-memory databases (e.g., MemSQL, Oracle TimesTen, SAP HANA, VoltDB)
NoSQL approaches (e.g., Cassandra, CouchDB, MongoDB, Redis, Riak)
Other

14. How confident are you in your organization's ability to prepare for use the variety of data you'll need to gain value for your business?

Very confident
Confident
Somewhat confident
Not confident

15. Which of the following sources of information are important for your organization's data preparation activities? (Select all that apply.)

Accounting or financial management systems
Big data
Business process or IT outsourcing
Call or contact center events
Cloud computing or software as a service applications and systems
Content (documents, Web pages)
Customer data sources
Data warehouse or operational data store
Employee or workforce applications
External sources (e.g., data from U.S. census or marketing services company)
Financial market data
Internet of Things device data
IT systems management alerts and events
Machine data (including log files and system information)
Network and router traffic
Online commerce or website
Supply chain systems
Social media
Transactional data from applications (e.g., CRM, ERP, PLM, SCM)
Other (Please specify):

16. Which of the following external data sources are important for your organization's data preparation efforts? (Select all that apply.)

Cloud computing business applications (e.g., Salesforce.com)
Consumer demographic sources (e.g.,Acxiom, Experian)
Economic data sources
Government demographics sources including U.S. Census
Internet information sources (e.g., Google, wikis)
Location sources (e.g., TomTom)

Market sources (e.g., Nielsen, SymphonyIRI)
Social media (e.g., Facebook, LinkedIn, Twitter)
Weather sources
Web applications
None of the above

17. How confident are you in the quality of the data being generated by your organization's data preparation efforts?

Very confident
Confident
Somewhat confident
Not confident

18. How adequate are your organization's data preparation technologies?

Completely adequate
Mostly adequate
Somewhat adequate
Inadequate
Don't know

19. Which of the following patterns of data movement occur or are planned to occur as part of your organization's data preparation processes?

In use now
Plan to use within 12 months
Plan to use in 12 to 24 months
Still evaluating
No plans to use

On-premises to on-premises
On-premises to cloud (SaaS or hosted)
Cloud (SaaS or hosted) to on-premises
Cloud (SaaS or hosted) to cloud (SaaS or hosted)

20. Which of the following data capabilities are critical for your organization's data preparation efforts? (Select all that apply.)

Deploy data preparation jobs as data services
Design graphical workflow of steps to process data
Design transformations
Develop and manage metadata that can be shared across analytic systems
Enable calls to external programs or systems
Enable user-defined expressions and objects
Establish rules for processing and routing data
Join disparate data sources during transformation
Manage data preparation tasks in repository for reuse
Perform data sampling as part of preparation tasks
Provide real-time processing of data
Provide self-documenting tasks
Set notifications of exceptions

21. Which of the following system-level capabilities are important for your organization's data preparation efforts? (Select all that apply.)

Change data capture in processing of data
Deploy as cloud computing service (SaaS)
Deploy through browser

Enable end-to-end processes (linked tasks or workflows)
Enable error handling
Enable mobile monitoring of preparation tasks
Enable multiuser support of preparation tasks
Enable systems management and monitoring
Make test or sample data available during authoring
Process large volumes of data
Provide connectors to databases and applications
Provide development and testing environment
Provide graphical authoring environment for preparation tasks
Provide sandbox environment to experiment with data
Schedule data preparation jobs to run repeatedly
Secure access to data for preparation and use
Share data for preparation and processing
Utilize in-database operations to prepare data

22. Which of the following types of analytics require data preparation in your organization? (Select all that apply.)

Data discovery
Data science, including predictive analytics or advanced statistics
Event discovery
Exploration of data
Forecasting, planning or what-if analysis
Information discovery (content or text in documents)
Presentation of information in a dashboard or portal
Query and reporting
Right-time and real-time analytics
Search for information (data or content)
Spreadsheet access
Visual discovery

23. Do you use your organization's existing data preparation technology to work with big data sources?

Yes
No
Don’t know

24. How satisfied are you with your organization's current data preparation technology for supporting big data?

Satisfied
Somewhat satisfied
Neither satisfied nor not satisfied
Not satisfied
Don't know

25. Why are you not satisfied with the data preparation technology? (Select all that apply.)

Current infrastructure not fast enough
Data not accessible for integration or preparation
Difficult to fit into current architecture
Existing technologies too complicated to integrate
Hard to build and maintain
Low scalability or reliability of data processes
Not adaptable or flexible to change

Not enough skilled resources
Not reliable enough
Technology inadequate
Too expensive
Too slow

26. How many data sources does your organization need to access in its data preparation efforts?

1
2-4
5-10
11-20
More than 20

27. How confident are you in your organization's ability to process large volumes (the overall quantity) of data?

Very confident
Confident
Somewhat confident
Not confident

28. How confident are you in your organization's ability to process data arriving at a high velocity?

Very confident
Confident
Somewhat confident
Not confident

29. In which of the following organizational areas or business processes is your organization using data preparation capabilities? (Select all that apply.)

Accounting/Finance
Administration
Business Development/Corporate Planning
Call or Contact Center
Compliance/Internal Audit
Consulting
Customer Support
Data Center/Networking
Distribution and Fulfillment
Education
Executive/Management
Field Service
Human Resources
IT/IS/MIS
Legal
Manufacturing
Marketing
Operations
Product Management/Marketing
Product Planning
Public Relations
Purchasing
Research and Development
Sales

Supply Chain
Training
Transportation
Warehousing
Other (Please specify)

30. To what extent has use of data preparation technology impacted your organization's activities and processes?

Improved them significantly
Improved them slightly
Had no impact
Reduced them slightly
Reduced them significantly
Don't know

31. Which of the following are the five most important benefits your organization has realized as a result of having data preparation capabilities? (Please rank them in order.)

Improved quality of information across enterprise
Made information more available in a consistent manner across enterprise
Met analytic needs of business more easily
Reduced costs for license fees
Reduced costs for implementation time and fees
Reduced effort and staff required
Reduced or eliminated manual processes
Reduced time required for data collection and preparation
Retained and analyzed more data
Utilized computing resources more efficiently

32. In working with data preparation processes as part of your job, on which of the following tasks do you spend the most significant amounts of time? (Please rank the top three.)

Collaborating with others
Connecting to data sources for access and integration
Deploying data preparation tasks into production systems
Designing data preparation tasks
Experimenting with data in a sandbox environment
Preparing data for analysis
Reviewing data for quality and consistency issues
Waiting for approval to get secure access to systems
Waiting for data and information
Waiting for resources to be available to design integration tasks

33. How comfortable is your organization in allowing business users to work with data that has not been integrated or prepared for them by IT?

Very comfortable
Comfortable
Somewhat comfortable
Not comfortable
Don’t know

34. Which of the following areas of your organization are realizing the benefits and value of investment in data preparation the most? (Select all that apply.)

Customer Service

Field Service
Finance
Human Resources
IT
Manufacturing or Supply Chain
Marketing
Operations
Sales
Other (Please specify)

35. Does your organization use or plan to use cloud computing products (SaaS or hosted) for data preparation?

Currently use
Plan to use
Do not plan to use
Don’t know

36. Why doesn't your organization plan to use cloud computing for data preparation? (Select all that apply.)

Regulatory reasons
Security reasons
Too difficult to integrate and transport data
Not enough skills or resources
Loss of control of technology
Concerns about performance

37. How often does your organization have to integrate data?

Real time
Every hour
Every day
Weekly
Monthly
Yearly
Don't know

38. How adequate is the training your organization provides in each of the following areas of data preparation?

Completely adequate
Mostly adequate
Somewhat adequate
Inadequate
Don’t know

Training on application of data preparation to business problems
Training on data preparation concepts and techniques
Training on data preparation technology
Training on data preparation methods specific to big data
Training on data preparation methods for text and unstructured data
Training on data preparation methods for Web- and cloud-based application data
Other

39. Which of the following general skill sets do people in your organization need to prepare data successfully? (Select all that apply.)

Analytic skills

Big data technology skills
Business skills (domain knowledge)
Communication skills
Data skills
Excel skills
Programming skills
SQL skills
Vendor-specific technology skills
Other (Please specify)

40. Which of the following data preparation issues arise between business units and IT in your organization? (Select all that apply.)

Cost focus vs. revenue focus
Different charters or direction set by the top of the organization
Different intellectual approaches (creative vs. logical)
Differing views on access to data (expansive vs. controlled)
Entrenchment of budgets and priorities
No issues, the two groups get along well

41. How important is collaboration among people in your organization on tasks related to data preparation?

Very important
Important
Somewhat important
Not important

42. How important is mobile access for data preparation tasks?

Very important
Important
Somewhat important
Not important

43. Which of the following organizations is primarily responsible for designing and deploying data preparation tasks in your organization?

Cross-functional business team with no particular organization leading
Data scientists or data miners
Line of business analysts
Line of business operations or nonanalyst roles
IT organization responsible for BI and data warehouse
Centralized IT team
IT function within line of business

44. Which of the following data preparation-related vendors' products does your organization use today or plan to use in the next 12 to 18 months? (Select all that apply.)

In use now
Plan to use within 12 to 24 months

Actian
Alpine Data
Alteryx
AtScale
Attunity
Datameer

Datawatch
Dell Boomi
IBM
Informatica
Information Builders
Jitterbit
Lavastorm
Microsoft
MicroStrategy
Oracle
Paxata
Pentaho
Pitney Bowes
Qlik
SAP
SAS
SnapLogic
Syncsort
Tableau
Talend
Tamr
Tibco
Trifacta
Zoomdata
Other

45. How important is each of the following technology and vendor considerations in influencing the purchase of data preparation technology?

Very important
Important
Somewhat important
Not important

Usability – Software meets business needs
Manageability – Administration and security
Reliability – Architectural fit, performance and scalability
Functionality – Capabilities of software for business
Adaptability – Customization, development and integration
Validation of vendor – References, viability and commitment
TCO/ROI – Software meets cost and benefit requirements

46. From which of the following types of vendors would your organization consider purchasing data preparation technologies? (Select all that apply.)

We would purchase data preparation from a business intelligence or analytics vendor.
We would purchase data preparation from a data integration vendor.
We would purchase data preparation from a database vendor.
We would purchase data preparation from a specialized vendor.

47. Are you satisfied with the results of the group that is identifying and selecting your organization's data preparation technology?

Yes
No
Don't know

48. Which of the following are major complaints? (Select all that apply.)
They take too long.
They do not communicate their strategy.
Their selection does not make it simple for the end user to access data.
They are not reliable in delivering the promised data preparation technology.
They do not engage with users and discuss data preparation technology opportunities and challenges.
Other (Please specify)

49. Who is or will be responsible for managing your organization's use of data preparation technology?
Line of business analysts
IT organization
Across lines of business and IT
Consultants
Don't know

50. In which of the following ways does your organization prefer to deploy data preparation software?
On premises (installed in organization)
On demand (multiple tenants, software as a service)
Hosted by supplier (managed and installed at third-party provider)
No preference

51. Is your organization planning to change the way it assesses and selects data preparation technology in the next 12 to 18 months?
Yes, we do this continuously
Yes
No
Don't know

52. Which of the following reasons are driving the change? (Select all that apply.)
Business improvement initiative
Business unit complaints about efficiency of business processes
Compliance with new regulations
Errors and mistakes impacting business
Improvement in business-to-business efforts
Improvement in quality of business processes
Legacy modernization efforts
Mergers and acquisitions create need to advance technology efforts
Methodology implementation (e.g., Six Sigma, balanced scorecard, lean supply chain)
New initiative to provide applications and information on mobile technologies
Operational efficiency and cost savings initiative
Part of a wider analytics and BI initiative
Part of wider big data efforts
Part of wider business-to-customer experience or satisfaction improvement efforts
Sales and revenue generation improvement
Simplify access to information and technology using cloud computing
Visibility into business and IT operations

53. Which of the following parties primarily fund data preparation initiatives?
General IT budget
Line of business IT budget

Shared service funded by business
General business budget
Don't know

54. Which of the following are barriers to making improvements to data preparation? (Select all that apply.)

Cost of software or license
Current infrastructure not fast enough
Difficult to fit data preparation into current architecture
Hard to access big data technologies
Hard to define ROI and build business case
Inadequate skills in the organization
Lack of adequate software capabilities to handle big data
Lack of awareness
Lack of resources to take on improvement
Latency of data to process for business
Low priority
Low scalability or reliability of data processes
No executive support
Not enough requests from our customers or users

55. What was your company's total revenue last year?

$50 billion or more
$25 billion to $49.99 billion
$10 billion to $24.99 billion
$5 billion to $9.99 billion
$1 billion to $4.99 billion
$500 million to $999 million
$250 million to $499 million
$100 million to $249 million
$50 million to $99 million
Less than $50 million

56. Which of the following best describes the industry in which your company competes?

Advertising or Public Relations
Aerospace and Defense
Agriculture
Apparel
Automotive
Banking
Business Services
Chemicals
Communications
Construction
Consulting
Consumer Products
Education
Electronics
Energy
Engineering
Entertainment
Environmental

Finance
Food and Beverage
Government
Government - Federal Civilian
Government - Federal Military
Government - Local or City
Government - State or Provincial
Healthcare
Hospitality
Insurance
Life Sciences
Machinery
Manufacturing
Media
Medical
Mining
Not for Profit
Publishing
Real Estate
Recreation
Retail
Security Products and Services
Shipping
Technology
Telecommunications
Transportation
Utilities
Wholesale
Other (Please specify)

57. In what country are you located?

United States
Afghanistan
Albania
Algeria
American Samoa
Andorra
Angola
Anguilla
Antarctica
Antigua and Barbuda
Argentina
Armenia
Aruba
Australia
Austria
Azerbaijan
Bahamas
Bahrain
Bangladesh
Barbados
Belarus
Belgium
Belize

Benin
Bermuda
Bhutan
Bolivia
Bosnia
Botswana
Bouvet Island
Brazil
British Indian Ocean Territory
Brunei Darussalam
Bulgaria
Burkina Faso
Burundi
Cambodia
Cameroon
Canada
Cape Verde
Cayman Islands
Central African Republic
Chad
Chile
China
Christmas Island
Cocos (Keeling) Islands
Colombia
Comoros
Congo
Congo, The Democratic Republic of The
Cook Islands
Costa Rica
Cote D'Ivoire
Croatia
Cuba
Cyprus
Czech Republic
Denmark
Djibouti
Dominica
Dominican Republic
East Timor
Ecuador
Egypt
El Salvador
Equatorial Guinea
Eritrea
Estonia
Ethiopia
Falkland Islands (Malvinas)
Faroe Islands
Fiji
Finland
France
French Guiana
French Polynesia

French Southern Territories
Gabon
Gambia
Georgia
Germany
Ghana
Gibraltar
Greece
Greenland
Grenada
Guadaloupe
Guam
Guatemala
Guinea
Guinea Equatoriale
Guinea-Bissau
Guyana
Haiti
Heart Island and Mcdonald Islands
Holy See (Vatican City State)
Honduras
Hong Kong
Hungary
Iceland
India
Indonesia
Iran, Islamic Republic of
Iraq
Ireland
Israel
Italy
Jamaica
Japan
Jordan
Kazakhstan
Kenya
Kiribati
Korea
Korea, Democratic People Republic of
Kuwait
Kyrgyzstan
Lao People's Democratic Republic
Latvia
Lebanon
Lesotho
Liberia
Libyan Arab Jamahiriya
Liechtenstein
Lithuania
Luxembourg
Macao
Macedonia
Madagascar
Malawi

Malaysia
Maldives
Mali
Malta
Marshall Islands
Martinique
Mauritania
Mauritius
Mayotte
Mexico
Micronesia, Federated States
Moldova
Monaco
Mongolia
Montserrat
Morocco
Mozambique
Myanmar
Namibia
Nauru
Nepal
Netherlands
Netherlands Antilles
New Caledonia
New Zealand
Nicaragua
Niger
Nigeria
Niue
Norfolk Island
Northern Mariana Islands
Norway
Oman
Pakistan
Palau
Palestinian Territory
Panama
Papua New Guinea
Paraguay
Peru
Philippines
Pitcairn
Poland
Portugal
Puerto Rico
Qatar
Reunion
Romania
Russian Federation
Rwanda
Saint Helena
Saint Kitts and Nevis
Saint Lucia
Saint Pierre and Miquelon

Saint Vincent and The Grenadines
Samoa
San Marino
Sao Tome and Principe
Saudi Arabia
Senegal
Serbia
Seychelles
Sierra Leone
Singapore
Slovakia
Slovenia
Solomon Islands
Somalia
South Africa
South Georgia and the South Sandwich Islands
Spain
Sri Lanka
St. Martin
Sudan
Suriname
Svalbard and Jan Mayen
Swaziland
Sweden
Switzerland
Syrian Arab Republic
Taiwan, Province of China
Tajikistan
Tanzania, United Republic of
Tasmania
Tchad
Thailand
Togo
Tokelau
Tonga
Trinidad and Tobago
Tunisia
Turkey
Turkmenistan
Turks and Caicos Islands
Tuvalu
Uganda
Ukraine
United Arab Emirates
United Kingdom
Uruguay
US Minor Outlying Islands
Uzbekistan
Vanuatu
Venezuela
Vietnam
Virgin Islands, British
Virgin Islands, US
Wallis and Futuna

Western Sahara
Yemen
Yugoslavia
Zambia
Zimbabwe

58. Are you willing to take part in a brief, confidential interview on this topic with a Ventana Research analyst?
Yes
No

59. Are you willing to receive additional information on this topic from Ventana Research and/or the sponsors of our research?
Yes
No

60. Do you currently have an approved budget for an initiative in this area?
Yes
No
Don't know

61. When will you purchase information technology for this area?
Immediately
Next 3 to 6 months
Next 7 to 12 months
More than 12 months from now

62. If you would like to qualify for participation incentives and access to the results of this benchmark research, please provide the following required information:
(Name, business email address, title and company are required to receive any incentives and copy of the report.)
First name
Last name
Business email address
Job title
Company
Phone
Address1
Address2
City
State or province
Postal code
Country

