

Data Preparation

Enabling Self-Service and Support Across Business and IT

RESEARCH
REPORT
EXECUTIVE
SUMMARY



A Note About This Research

December 2017

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

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

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

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

Ventana Research

This research is sponsored by

 DATAWATCH



VENTANA RESEARCH

Bend, Oregon, USA

541-940-1010

info@ventanaresearch.com



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

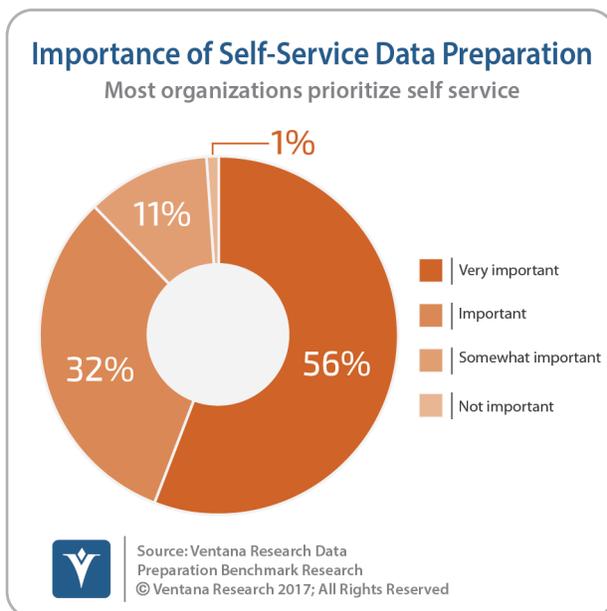
deploying tools that support self-service data preparation is growing. Self-service tools enable analysts to perform all or many of the data preparation tasks without the assistance of IT. Taken together, these two trends can lead to conflict for organizations that want to derive maximum business value 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.

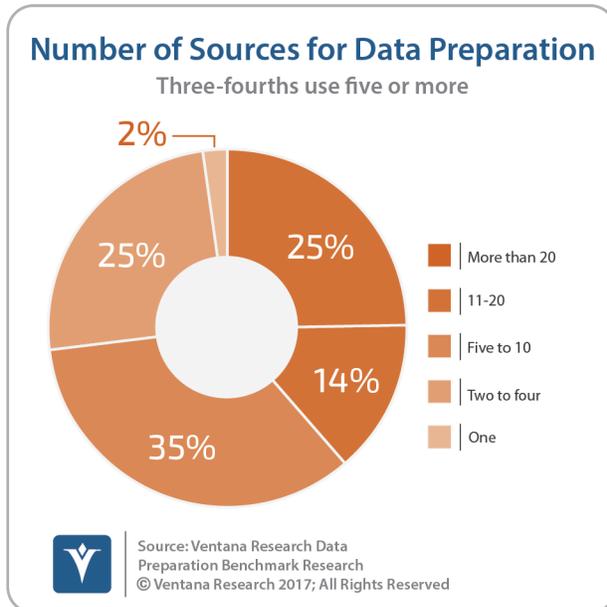
Organizations take different approaches to data preparation. Nearly half (47%) of organizations participating in the research use a dedicated product for self-service data

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



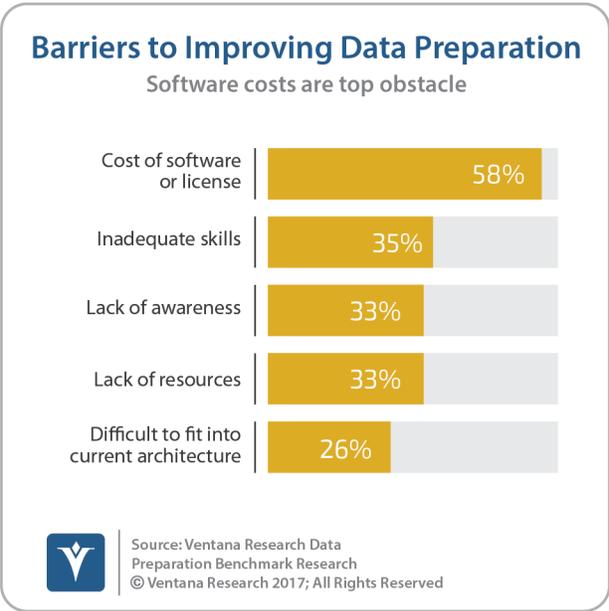
cited as benefits of having data preparation capabilities improved quality and consistency of information, meeting the organization's analytic needs more easily and reducing or eliminating manual processes.

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



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.



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: About This Benchmark Research

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

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