The importance of self-service data preparation and pipelines to lower the barriers to analytic insight

The 451 Take

Self-service data preparation software has emerged in recent years to become a critical enabler of delivering ana-lytic insight. Self-service data preparation enables enterprises to reduce the time spent on data prep by enabling business analysts and data scientists to prepare their own data for analysis rather than relying on IT and data man-agement staff to prepare data for them.

Putting data preparation capabilities in the hands of these business analysts and data scientists should reduce the time it takes for users to discover, integrate, cleanse and enrich data to make it suitable for analysis. Reducing this preparation overhead should also enable users to expand the scope of their analysis – accessing more datasets and greater volumes of data. In many cases, capabilities for automation and advanced governance also prove useful in this area, as does collaborative functionality that enables analysts to access trusted datasets and pre-queried results from their colleagues.
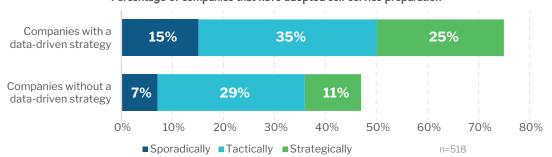
Enterprise-wide data preparation initiatives have the potential to unite multiple constituents across the company, in-cluding data engineers, analysts and data scientists, by providing a common platform for collaboration that increas-es the productivity of all stakeholders. Enterprises that adopt self-service data preparation software can accelerate data-driven decision-making by business and data analysts and support the democratization of access to data by decision-makers while freeing up time for data professionals to enable them to focus on higher-value initiatives.

Data from 451 Research surveys illustrate that adoption of self-service data preparation goes hand in hand with data-driven decision-making. A recent 451 Research survey indicates that 67% of organizations have adopted self-service data preparation. Of those, 15% of respondents are embracing self-service data prep strategically across the organization. However, the results also show that adoption is significantly higher among enterprises that are tak-ing active steps to be more 'data driven,' and that those companies are also more likely to be embracing self-service data-preparation strategically.

Three-quarters of organizations with a formal strategy to become more data-driven have adopted self-service data prep (25% strategically). In comparison, just 47% of companies without any strategy to become more data-driven have adopted self-service data prep (11% strategically).

Companies with More Strategies to Be Data-Driven are Greater Adopters of Self-Service Data Prep
Source: 451 Research's Voice of the Enterprise: Digital Transformation, Q4 2018

Percentage of companies that have adopted self-service preparation

	Sporadically	Tactically	Strategically
Companies with a data-driven strategy	15%	35%	25%
Companies without a data-driven strategy	7%	29%	11%

n=518

WWW.451RESEARCH.COM

Business Impact

SELF-SERVICE DATA PREPARATION SOFTWARE ENABLES ENTERPRISES TO REDUCE THE BURDEN ON IT to prepare data for end users, particularly business and data analysts, but also potentially senior decision-makers, which can accelerate project delivery through the avoidance of IT backlogs.

REDUCING THE TIME TAKEN FOR USERS TO DISCOVER, INTEGRATE, CLEANSE AND ENRICH DATA to make it suitable for analysis better enables them to create new analytics on demand in an agile man-ner and makes it possible for them to expand the scope of their analysis – accessing more datasets and greater volumes of data.

COLLABORATION AMONG USERS CAN EXPEDITE THE TIME TO VALUE, enabling analysts to access trusted datasets and pre-queried results.

THE COMBINATION OF SELF-SERVICE DATA PREPARATION AND DATA ENGINEERING FUNCTIONALITY on a common, collaborative platform enables enterprises to streamline both ends of the data preparation pipeline.

SELF-SERVICE DATA PREPARATION HAS A KEY ROLE TO PLAY in ensuring that enterprises gener-ate value from their investments in data lake environments, enabling integration, transformation, discovery and operationalization of data from multiple data sources for multiple purposes.

Looking Ahead

The early adoption of self-service approaches was largely driven by semi-autonomous functional departments that were able to select analytics and preparation technologies that gave data ana-lysts and data scientists the freedom to prepare and analyze data. There are clear potential ben-efits to empowering data analysts and data scientists to prepare and analyze their own datasets. But this needs to be balanced with the requirements of data stewards and IT to maintain control over who has access to what data, and for what purpose, driven by privacy, security, regulatory and data quality concerns. This is especially relevant given the rise of multi-function data lakes, designed to provide a unified data platform serving multiple applications. 451 Research believes that the combination of data governance and self-service data preparation is key to delivering a functional data lake.

We have also observed that many enterprises are beginning to alter their view of data governance from something that is put in place to limit access to data to something that is put in place to en-able access to the right data by the right people. While self-service data preparation offerings were initially targeted specifically at data analysts and data scientists, these products are evolving to meet the needs of data stewards and IT professionals to manage data governance without getting tied down in actually preparing the data for analysis.

Datameer provides hybrid-cloud analytics data management products that help organizations have successful, lower risk cloud journeys. Datameer X is an enterprise-grade data preparation, data pipeline and dataops platform that helps organizations gain greater business value from their data lakes, reduce analytics cycles by up to 90% and lower their data engineering costs and faciliates easy last-mile self-service data preparation. Datameer brings together all your personas on a common platform – data engineers, data analysts, data scientists – with capabilities to match their unique needs while also allowing collaboration in a secure and well-governed manner. To learn more, please visit www.datameer.com.