

The State of Enterprise Data Quality

As Data Volumes Grow So Does
the Need for Big Data Quality





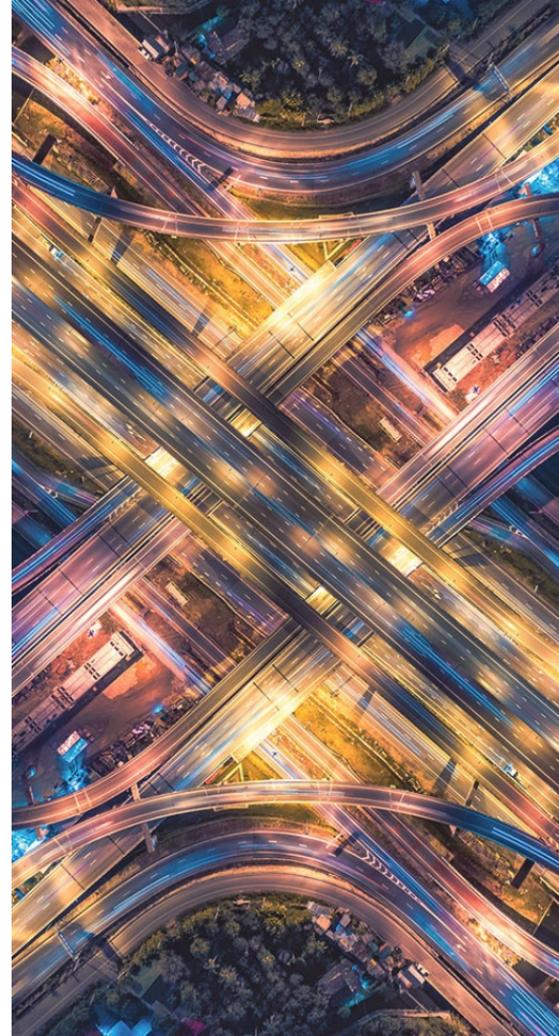
Survey Background

Syncsort's 2019 Enterprise Data Quality survey explores the challenges and opportunities for organizations looking to bring data quality across the enterprise as data volumes grow and new technologies emerge. Overall, Data Quality is growing in importance with 75% of respondents citing it as a high or growing priority in the next 12 months.

In the following report, we'll share highlights from the survey as well as a deeper look at the full results.

Respondent Profile

Syncsort polled 175 respondents, 69 percent of whom work for organizations with over 1,000 employees. Participants represented a range of industries, with the largest percentage coming from Financial Services (25%), as well as a range of positions, ranging from CDO to Data Analyst, with the majority in data-focused roles (29%).





Good Data Isn't Good Enough Anymore

There is a disconnect around understanding, confidence, and trust in the data and how it informs business decisions.

72 percent responded that the quality of the data used to run their business was good or better and 69 percent stated their leadership/c-suite trust data insights enough to inform business decisions on them. Yet, they also reported that only 14 percent of stakeholders had a very good understanding of the data and that less than 60 percent of the data was well understood by stakeholders.

More than 70 percent also reported that sub-optimal data quality negatively impacted business decisions, and almost half found that untrustworthy results or inaccurate insights from analytics were due to a lack of quality in the data fed into systems such as AI and machine learning.





Data Quality is a Top Challenge for Machine Learning

Poor data quality is enemy number one to the widespread, profitable use of machine learning. The phrase “garbage-in, garbage-out” has a multiplier effect with ML — first in the historical data used to train the predictive model and second in the new data used by that model to make future decisions.

With almost half reporting that untrustworthy results or inaccurate insights from analytics were due to a lack of quality in the data fed into systems such as AI and machine learning, it’s not surprising that “many sources of data” (69%) and “volume of data” (48%) are among the top 3 challenges companies face when ensuring high quality data.

3/4 of respondents also identified as having challenges profiling or applying data quality to large data sets.

A Wall Street Journal article revealed a recent report by Forrester Research Inc. found data quality a top challenge for AI projects and that *“companies pursuing such projects generally lack an expert understanding of what data is needed for machine-learning models and struggle with preparing data in a way that’s beneficial to those systems.”*

Full Survey Results



Understanding Data Across the Organization

How well do you (or other key stakeholders) understand the data that exists across your organization?



Defining “Good” Understanding of Data

If you answered, Very Good or Good, what percentage of your data is well understood by you/key stakeholders?



Greater than 70%



70% - 50%



50% - 30%



30% - 10%



10% or less

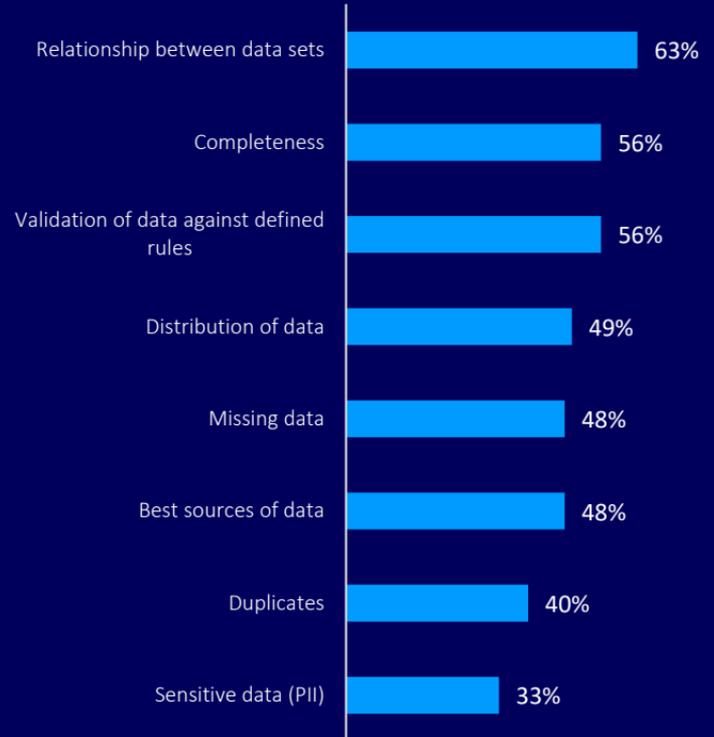


Data Attributes Lacking Visibility

Of those who responded that had partial, minimal or very little understanding of their data, the top three attributes respondents lacked visibility into were:

- Relationship between data sets
- Completeness of data
- Validation of data against defined rules

If you answered partial, minimal or very little understanding, what attributes of your data do you lack visibility into?



Note that this was a select all that apply question so responses will not total to 100%.

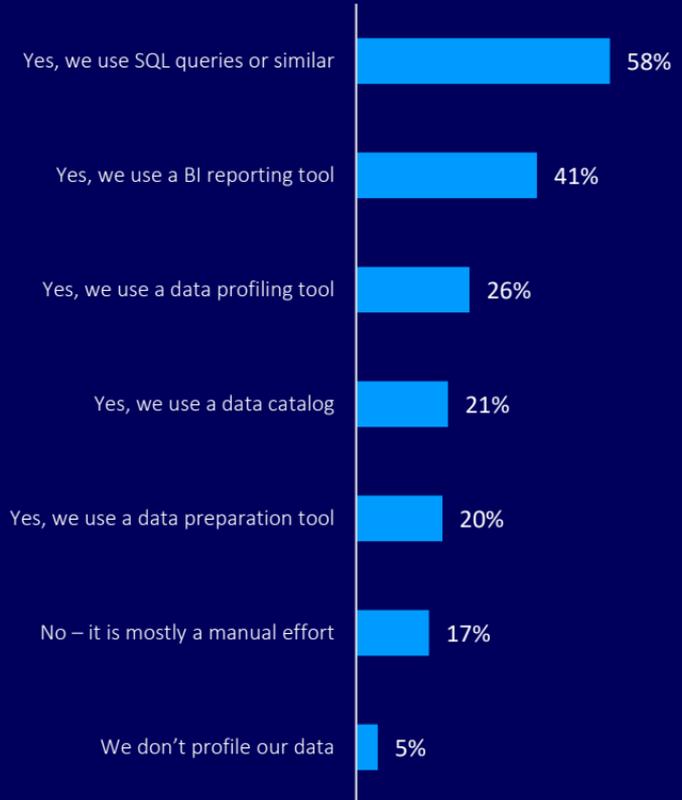
Use of Data Profiling Tools

Less than 50% of respondents take advantage of a data profiling tool or data catalog where insight may be centrally provided for broad access.

Instead, respondents rely on other methods to try to gain understanding of data, with more than 50% of respondents using SQL queries or similar and over 40% using a BI tool.

Only 17% are profiling data manually.

Do you use tools for profiling data? (select all that apply)



Note that this was a select all that apply question so responses will not total to 100%.

Profiling Large Data Sets

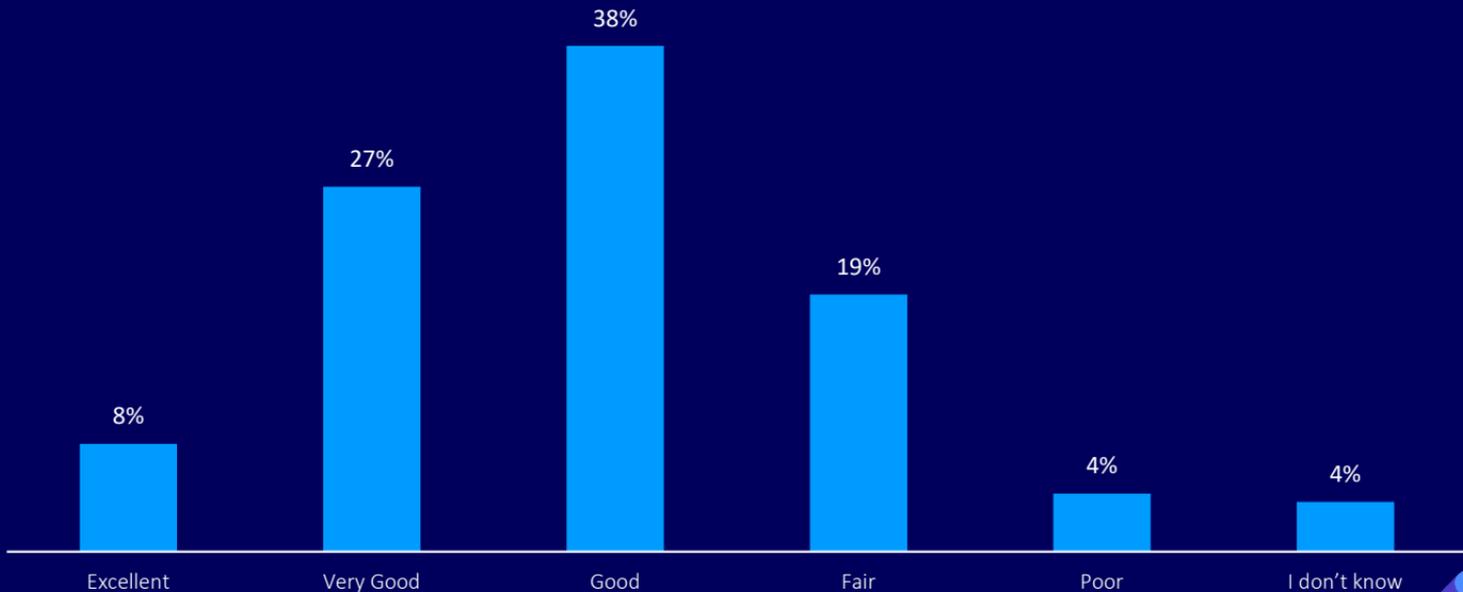
3/4 of respondents identified as having challenges profiling or applying data quality to large data sets.

Do you have challenges profiling and applying data quality to large data sets?



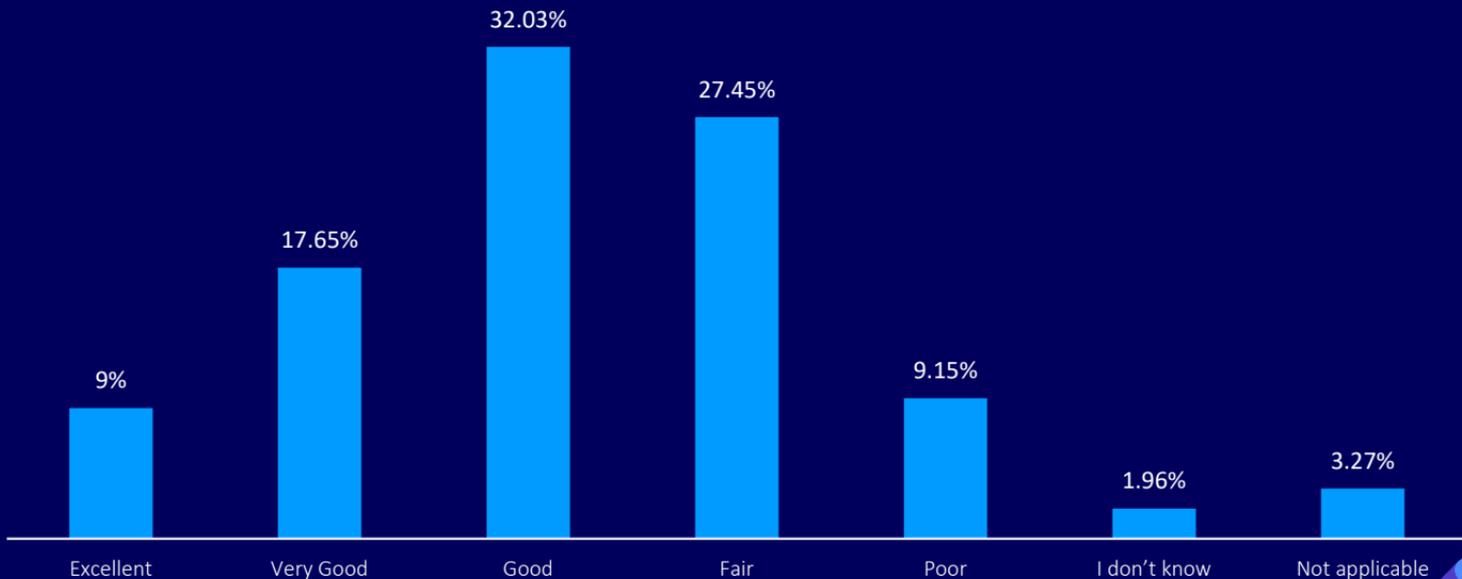
How would you rate the quality of the data used to run your business?

Only 8% of respondents reported having excellent data quality.



How would you rate your organization's ability to get a single view of customer?

More than 30% of respondents lack ability to get a single view of the customer.

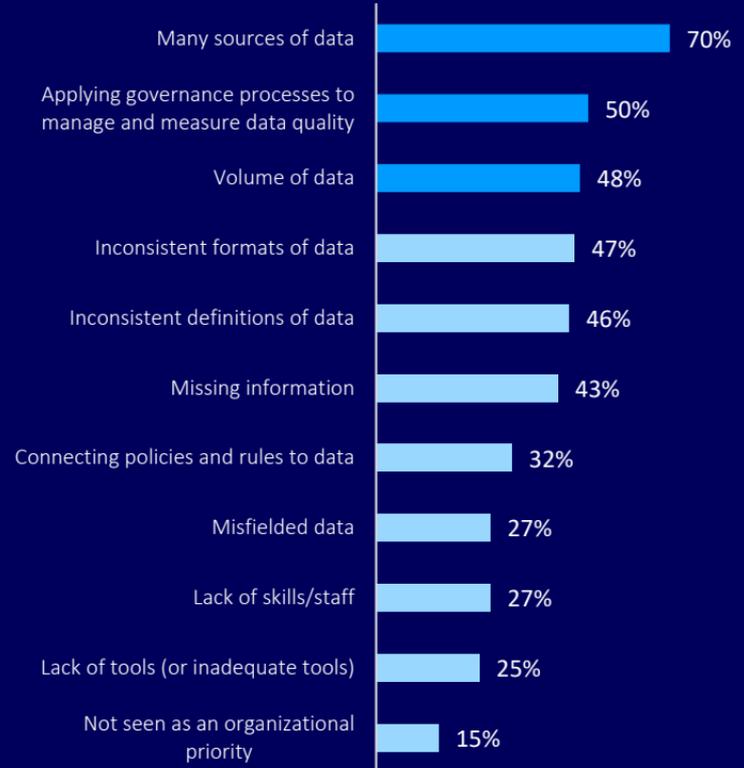


Challenges To Ensuring Data Quality

Many sources of data (70%) and volume of data (48%) are among the top 3 challenges companies face when ensuring high quality data.

Applying governance processes to manage and measure data quality is second with 50%.

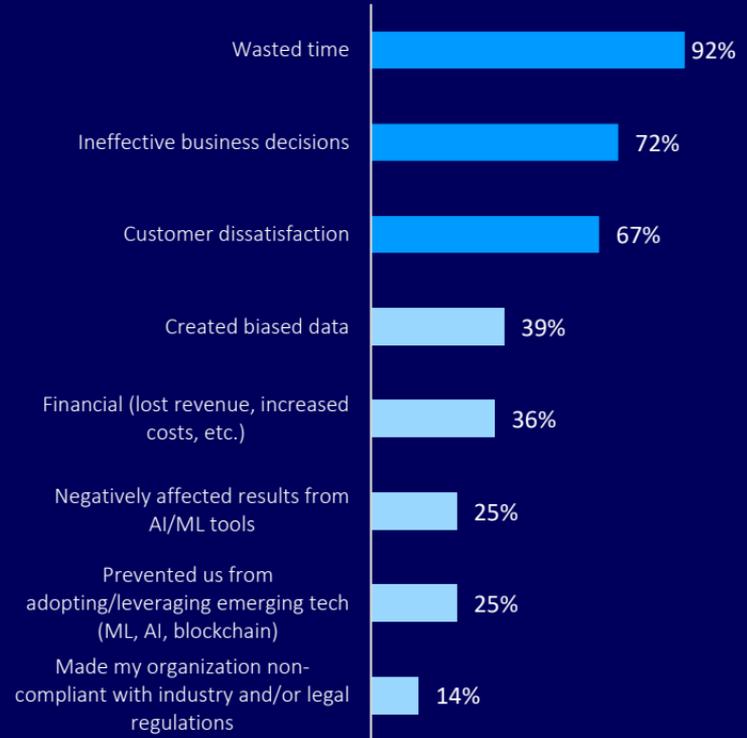
What are the greatest challenges you face when ensuring high data quality?



Consequences of Poor Data Quality

Those who reported Fair or Poor data quality cited Wasted Time as the number one result (92%), followed by Ineffective Business Decisions (72%) and Customer Dissatisfaction (67%).

If you answered the previous question a Fair or Poor, as a result of sub-optimal data quality, your organization has experienced the following repercussions:

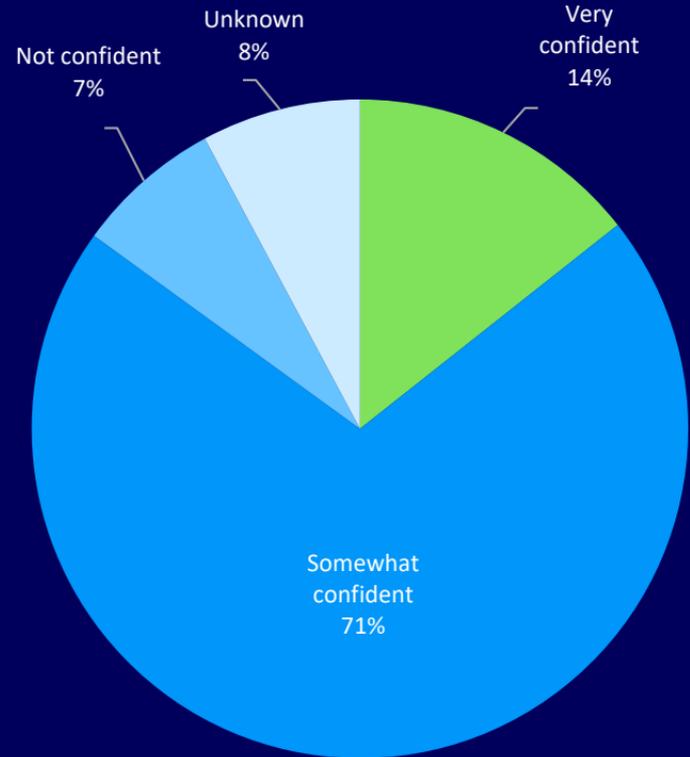


Note that this was a select all that apply question so responses will not total to 100%.

Confidence in Data Sent to Analytics Platforms

70% of respondents are “Somewhat confident” in the data their organization sends to analytics and data visualization applications.

How confident are you in the data your organization sends into analytics and data visualization applications?

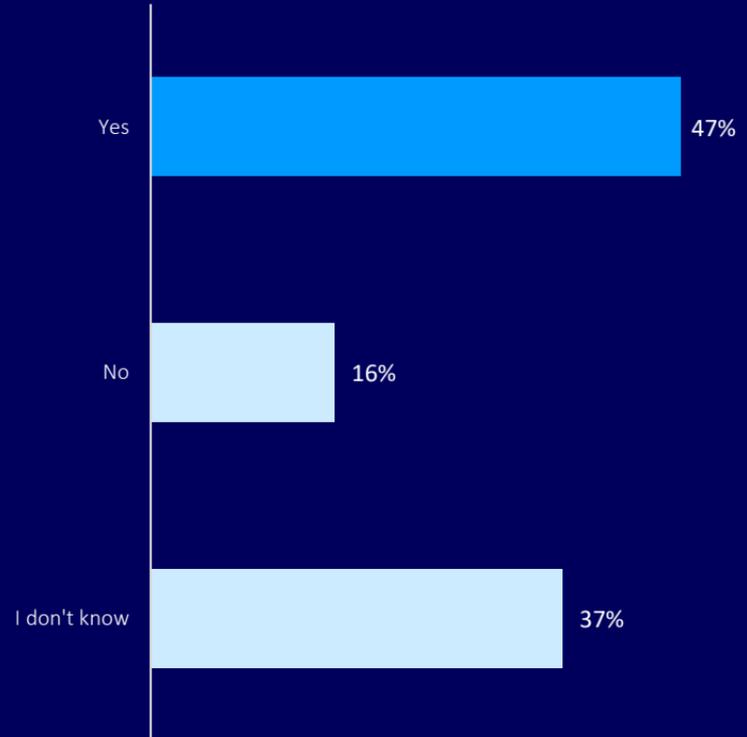


Poor Data Quality Leads to Inaccurate Data Insights

47% of respondents had untrustworthy or inaccurate insights from analytics due to lack of quality.

Only 16% are confident they aren't feeding bad data into AI and ML applications.

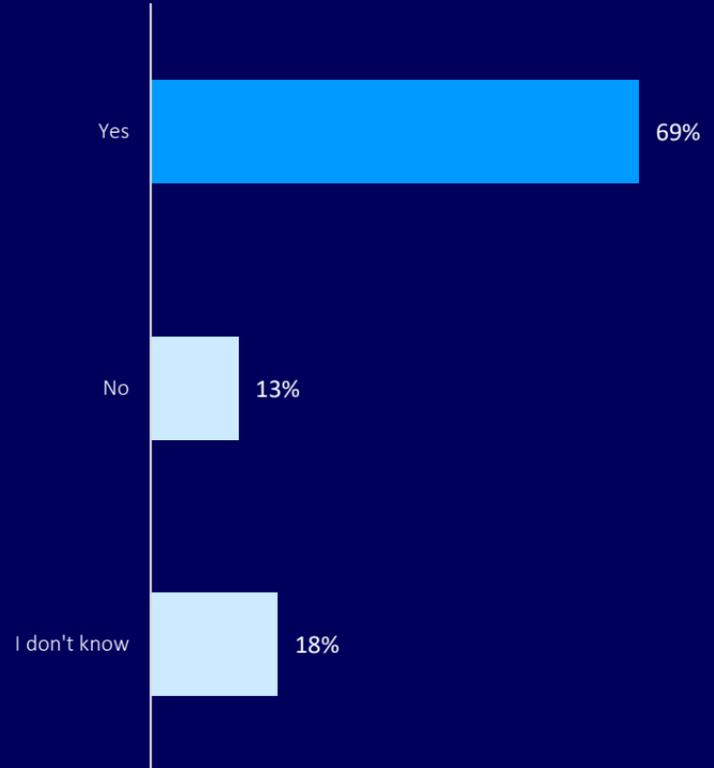
Have you ever had un-trustworthy results or inaccurate insights from analytics due to a lack of quality in the data being fed into those systems (such as AI, ML)?



Leadership Trust in Data Insights

Although confidence in data sent to analytics systems is lukewarm and almost half reported they'd had untrustworthy results from analytic platforms, nearly 70 percent of respondents still state that their leadership trusts the insights enough to inform business decisions.

Does your leadership/c-suite trust your data insights enough to inform business decisions on them?

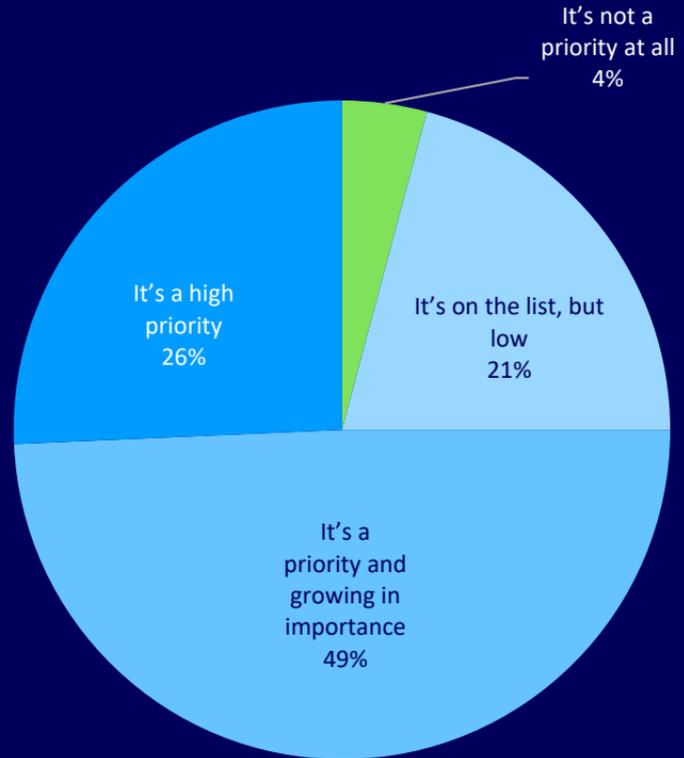


Data Quality is Growing in Priority

Although levels of confidence and trust in data appears mixed, 75% of respondents cite data quality as a high or growing priority.

Only 4% feel data quality is not a priority.

How high of a priority is Data Quality for your company in the next 12 months?





Data Quality in the Cloud

73%

Leveraging cloud computing
for strategic workloads.

48%

Have partial to no
understanding of the data
that exists in the cloud.

22%

Rate the quality of their
data in the cloud as Fair or
Poor.





Data Quality in Big Data

55%

have a data lake or enterprise data hub leveraging distributed computing platforms like Hadoop or Spark

26%

do not have a process for applying data quality to the data in the data lake or enterprise data hub

19%

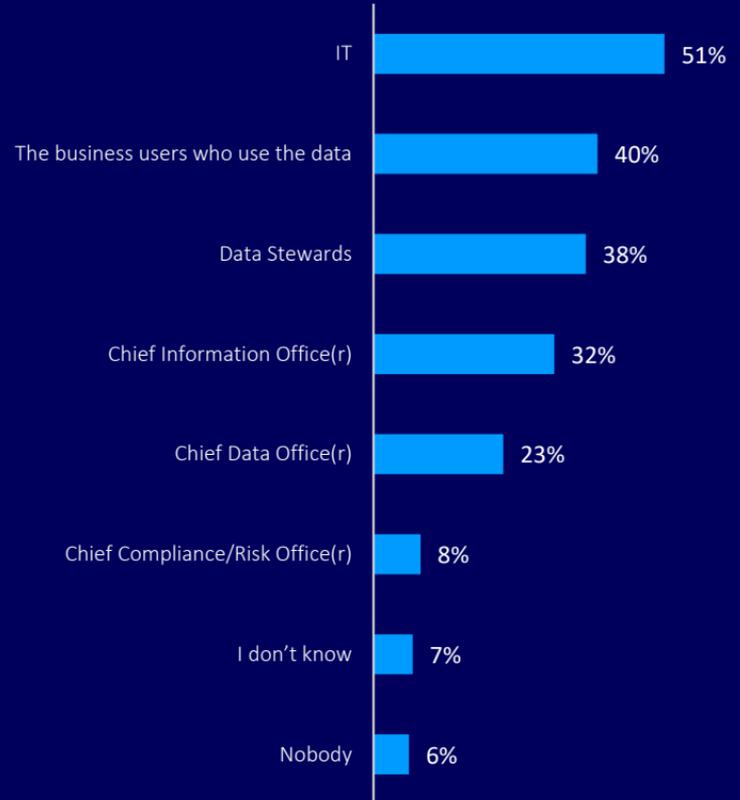
Rate the quality of their data in the data lake or enterprise data hub as Fair or Poor, while 32% rate their data as “Good”



Responsibility for Data Quality

51% reported that IT is responsible for Data Quality, while business users and data stewards play a critical role.

Who is responsible for Data Quality at your company?



Note that this was a select all that apply question so responses will not total to 100%.



[Learn more at syncsort.com](https://www.syncsort.com)