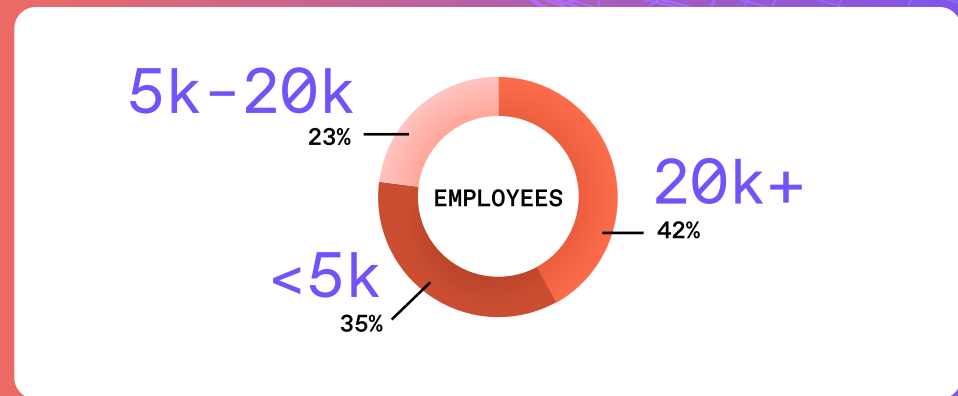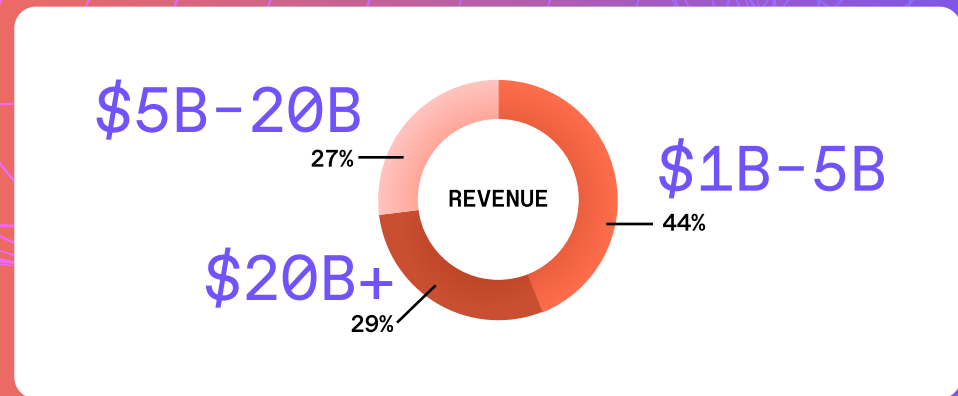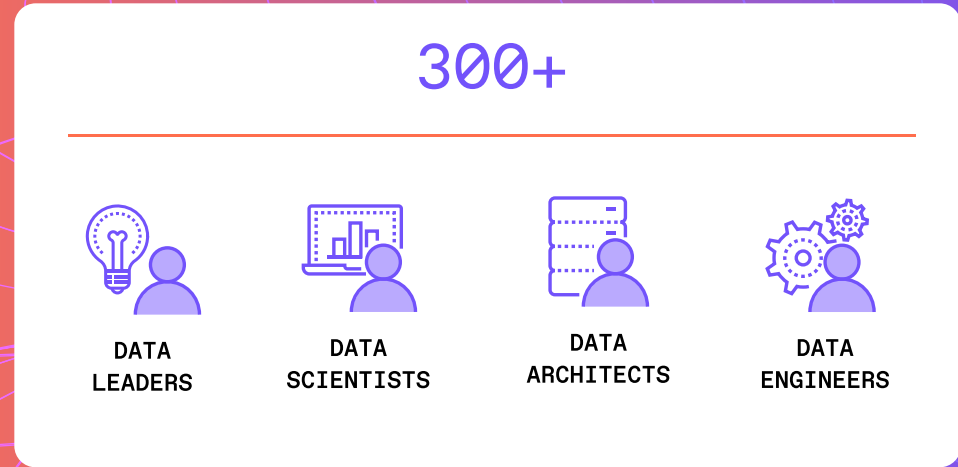Anomalo

State of Enterprise Data Quality 2024: Executive Brief

Executive Summary: Enterprise Data Quality Imperative in 2024

High-quality data is the new gold and it is imperative within the enterprise. In 2024, 100% of data leaders and practitioners are addressing data quality issues within their data warehouses and data lakes. Data quality is now imperative as 91% indicate that data quality is of greater importance than it was five years ago and 99% plan to invest in data quality within the next two years. Enterprise data teams' top data quality challenges include addressing an increase in data volume and unreliable external data sources within their complex data architectures. Also, 95% agree that data observability is not enough to determine the root cause of data quality issues and 87% indicate traditional rules-based approaches to data quality don't scale to meet today's needs. With the rise of GenAI, these challenges will compound exponentially. Anomalo commissioned Dimensional Research to survey over 300 data leaders and practitioners responsible for data quality to obtain insights into current trends and perspectives on data quality.

300+

DATA LEADERS   DATA SCIENTISTS   DATA ARCHITECTS   DATA ENGINEERS

REVENUE

$5B–20B 27%
$1B–5B 44%
$20B+ 29%

EMPLOYEES

5k–20k 23%
20k+ 42%
<5k 35%

Key Highlights

High Data Quality Imperative in Enterprise

100%

Have data quality issues in their data warehouses or data lakes

9 in 10 report that data is more important to business leadership now than it was 5 years ago

99%

Will invest in data quality in the next 2 years

Low Data Quality and Business Impact

95%

Have had a data quality issue with a direct impact upon business outcomes

91%

Say stakeholders won't make data-driven decisions if they don't trust their data

94%

Report data issues impact trust among business executives

Key Highlights

Traditional Data Observability Solutions Not Sufficient
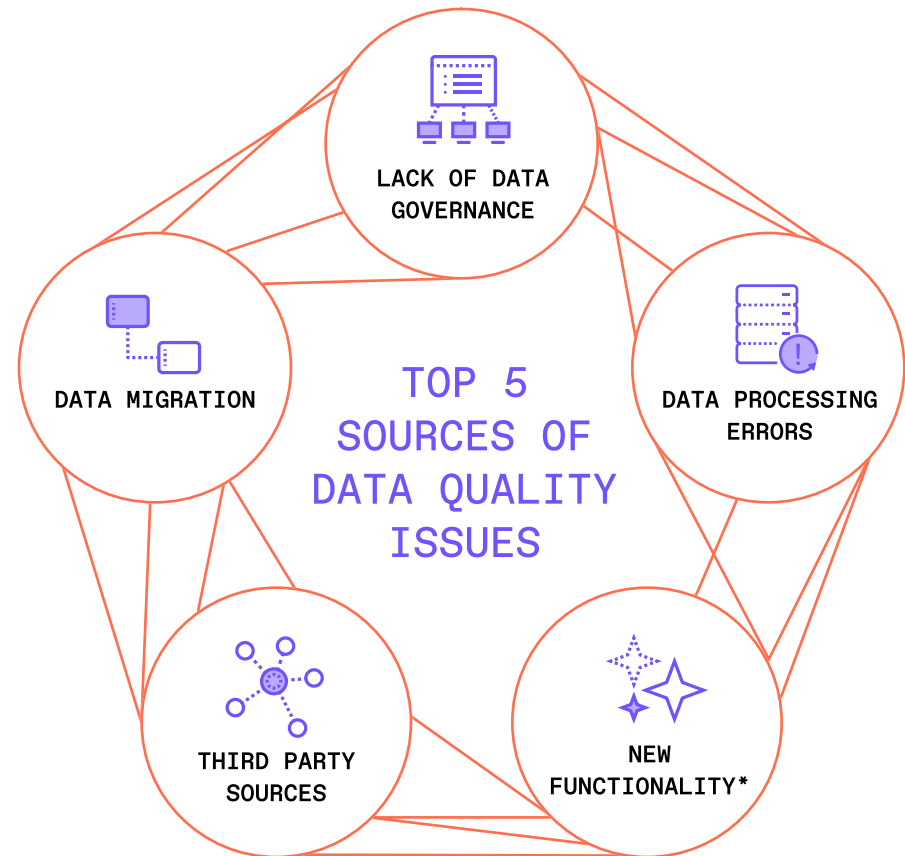
95%

Agree that data observability is not enough to determine the root cause of data quality issues

87%

Say that traditional rules-based approaches to data quality don't scale to meet today's needs

99%

Report challenges delivering data quality, top challenges: complex data architecture, an increase in data volume, and external data sources

TOP 5 SOURCES OF DATA QUALITY ISSUES

- LACK OF DATA GOVERNANCE
- DATA PROCESSING ERRORS
- NEW FUNCTIONALITY*
- THIRD PARTY SOURCES
- DATA MIGRATION

*e.g. features, upgrades, bug fixes, etc.

Data Leaders: Key Data Quality Challenges

Enterprise data leaders are in the unique position of being responsible for both technology and business outcomes related to the scaling of data quality. Quality challenges abound in the data flow and within the data itself. Traditional data observability is necessary but not sufficient, as it monitors the metadata about the tables. Yet, data observability does not address the quality issues within the data itself.

Delivering data quality is a known challenge within enterprise companies, and 95% of those surveyed have had a data quality issue with a direct impact on business outcomes. Also, 94% report that data issues impact trust amongst business executives, and 91% indicate that stakeholders won't make data-driven decisions if they don't trust their data.
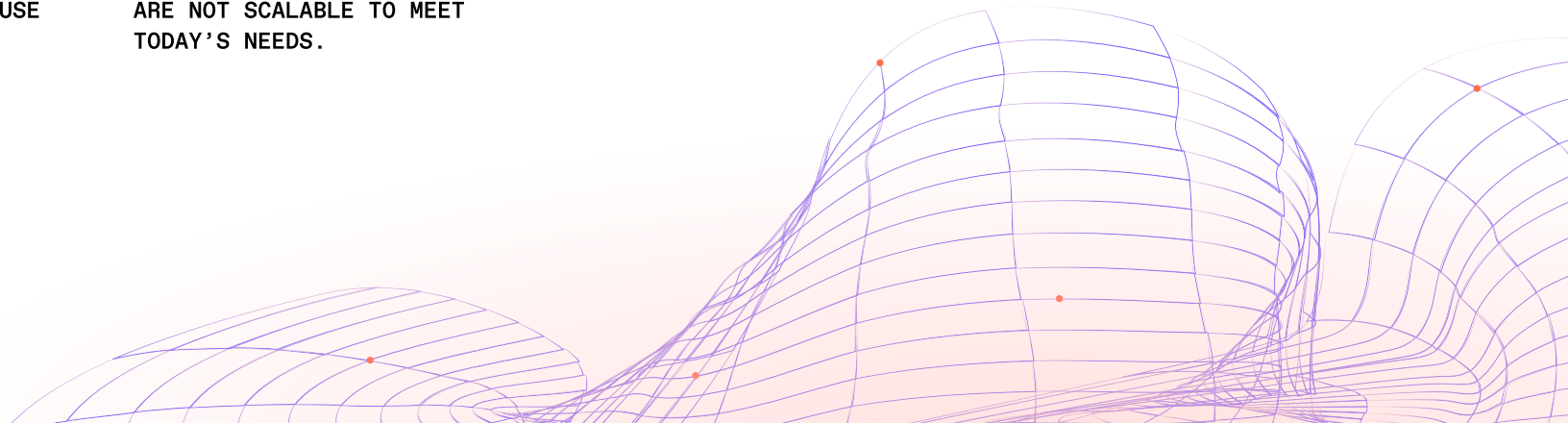
As enterprise data team leaders are responsible for a wide breadth of technology and business outcomes related to data quality, any solution they consider must enable multiple data teams to detect the root cause of quality issues within the data itself, alert appropriate stakeholders, and provide personalized insights to help cross-functional data teams towards resolving the data issues.

95% AGREE THAT DATA OBSERVABILITY IS NOT ENOUGH TO DETERMINE THE ROOT CAUSE OF DATA QUALITY ISSUES.

87% AGREE THAT TRADITIONAL RULES-BASED APPROACHES ARE NOT SCALABLE TO MEET TODAY'S NEEDS.

Next Steps: Evaluating Data Quality Solutions

As companies grow and look to scale their data quality solution, traditional data observability, open-source, and home grown solutions are potentially limiting. Oftentimes, the limitations may include not having the capability to support or serve multiple data teams and help them towards effective root cause detection and issue resolution at scale.

When evaluating comprehensive data quality solutions, enterprise companies benefit from being able to easily automate root cause detection, alert the right people or teams about the data quality issues at the right time, and empower multiple data teams to collaborate together towards resolution. Enterprise companies also benefit from ensuring their data quality solutions have the appropriate technical and security compliance (i.e., ISO 27001, SOC 2, HIPAA, GDPR).

For additional information to support the next steps, review Anomalo's *The Definitive Guide to Data Quality Monitoring Vendor Selection*.

About Anomalo

Anomalo helps enterprises build confidence in the data they use to make decisions and build products. Enterprises can simply connect Anomalo's complete data quality platform to their data warehouse or lakehouse and begin monitoring their data in less than 5 minutes, all with minimal configuration and without a single line of code. Then, they can automatically detect and understand the root cause of data issues before anyone else. Anomalo is backed by SignalFire, Norwest Venture Partners, Foundation Capital, Two Sigma Ventures, Village Global, First Round Capital, and Databricks Ventures. For more information, visit anomalo.com or follow at linkedin.com/company/anomalo/.

STUDY METHODOLOGY

Anomalo commissioned the independent firm Dimensional Research to capture the state of enterprise data quality through a quantitative data survey in May 2024. Participants included 300+ data leaders and practitioners with responsibility for data quality at enterprise companies with at least $1B USD in revenue.