DATA ENGINEERS: READY TO DELIVER MORE BUSINESS VALUE, YET MIRED IN MANUAL TASKS

A Global Survey of Data Professionals

March 2021

dimensional research

Sponsored by: Fivetran

EXECUTIVE SUMMARY

The research finds that…

While the top metric for data engineering success is data being used to make decisions, nearly half of those surveyed indicated key data has yet to be integrated -- diminishing capabilities for decision-making. Furthermore, 68% of data professionals shared that more insights could be extracted from existing data if they only had the time.

Nearly all participants (98%) revealed numerous challenges when building pipelines, which often take weeks to complete, require multiple tools and have heavy reliance on scripting. When pipelines break, participants report the business suffers, with decreases in efficiency, decision-making agility, dissatisfied customers and more. Yet 98% report their pipelines do break, and do so frequently, with half reporting it happens every month or even more often. The majority of companies indicated it takes longer than one business day to resolve pipeline issues, impacting the business and inhibiting data-driven decision-making.

Leadership is needed as professionals shared a lack of an overall data strategy, burdened by numerous stakeholders who often cannot clearly articulate what is needed. This lack of direction frequently leads to numerous ETL tools being used, but falls short of a comprehensive approach -- missing key capabilities such as data governance, version controlling, model dependency mapping, idempotent data replication (self-healing pipelines) and more. From this research, it is clear that data engineers deliver strong value to the business, as 79% reported they plan to hire more this year; but a strong data strategy and better tools could create more reliable data flow while improving data-driven decisions.

KEY FINDINGS

Data-Driven Decisions Enable the Business, but Many Orgs Are Failing

- Data available for decision-making leads all metrics for data engineering success
- 44% share that key data is not yet usable for decision-making
- 68% of data professionals shared that more insights could be extracted from existing data if they only had the time

Pipelines Are Slow to Build and Difficult to Maintain, Impacting the Business

- 98% face challenges building new pipelines
- 50% of companies require more than a business week to build a new data pipeline
- Company efficiency and agility suffer when pipelines break
- 98% reveal data pipelines break, with 51% stating it happens monthly or more frequently
- 55% report it takes longer than one business day to repair pipelines

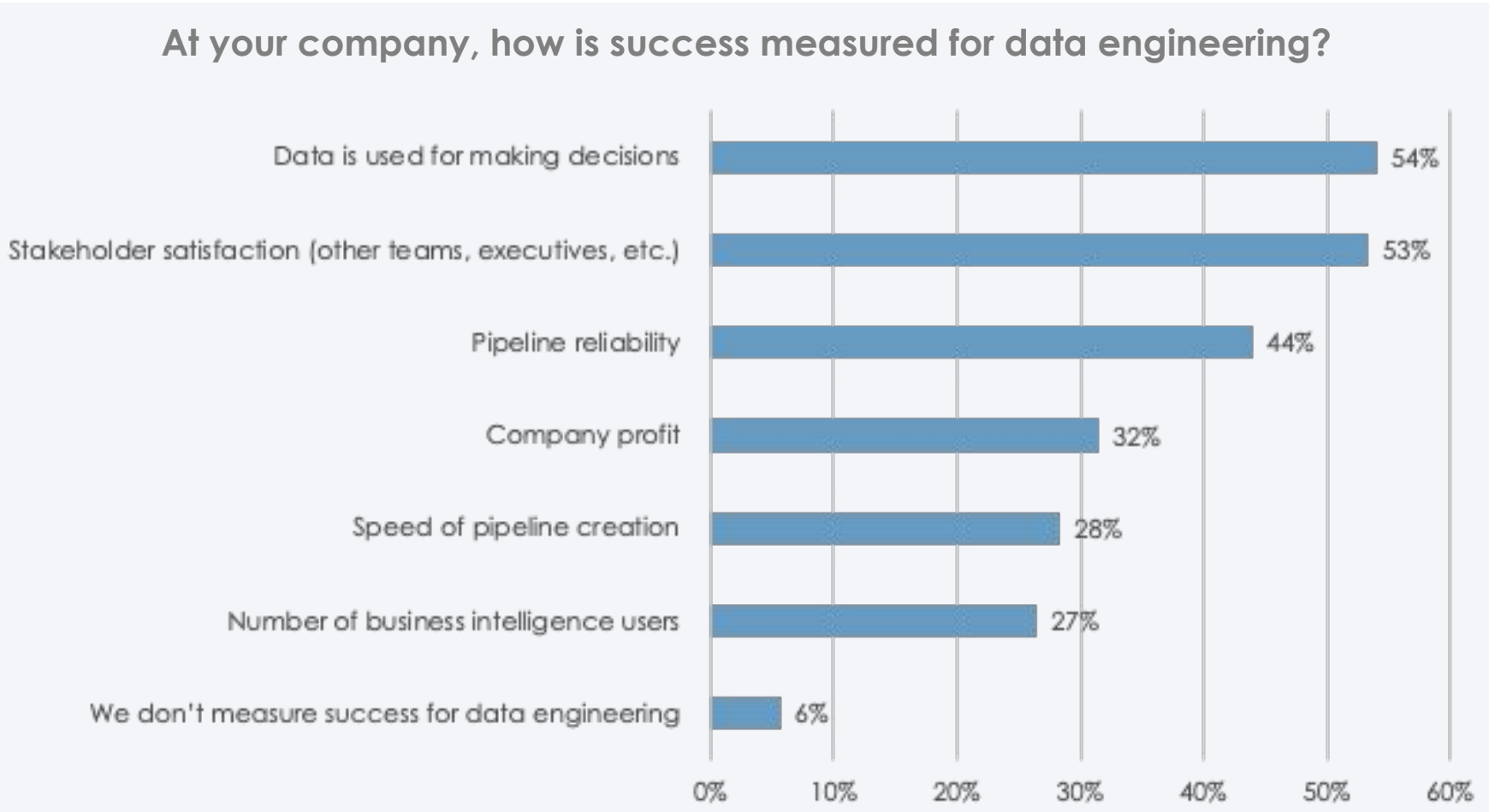
Data Engineering Needs Better Leadership

- Lack of strategy, numerous stakeholders and unreliable pipelines top the list of data challenges

DETAILED FINDINGS

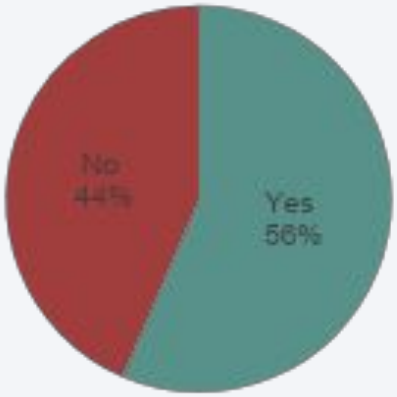
dimensional research

DATA AVAILABLE FOR DECISION-MAKING LEADS ALL METRICS FOR DATA ENGINEERING SUCCESS



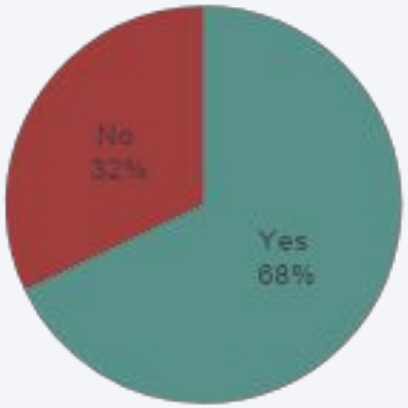
44% SHARE THAT KEY DATA IS NOT YET USABLE FOR DECISION-MAKING

At your company, are all valuable data sources currently integrated into pipelines?



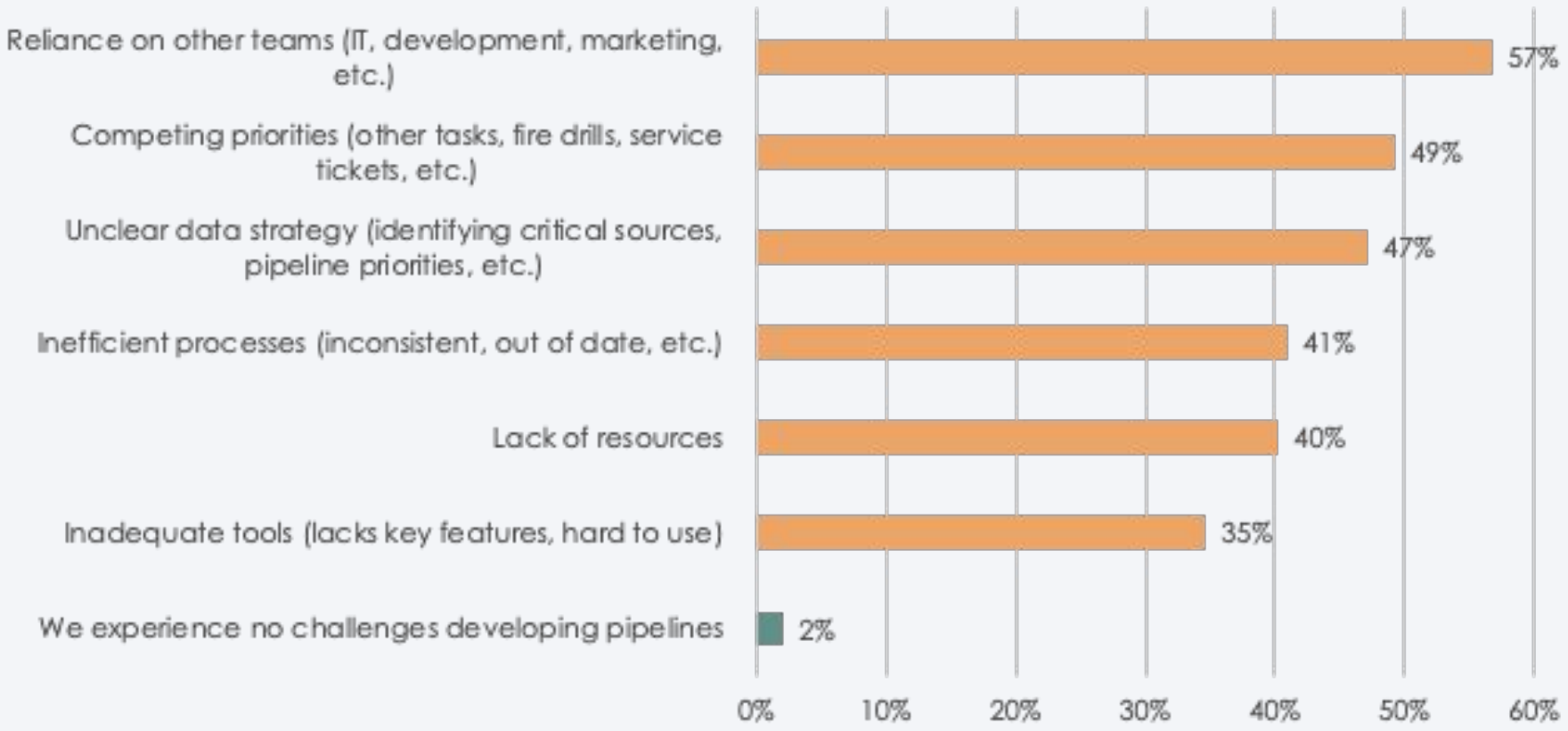
68% STATE THAT ADDITIONAL BUSINESS INSIGHTS CAN BE EXTRACTED FROM EXISTING DATA

In your opinion, is too little time spent focusing on creating new business insights from existing data?



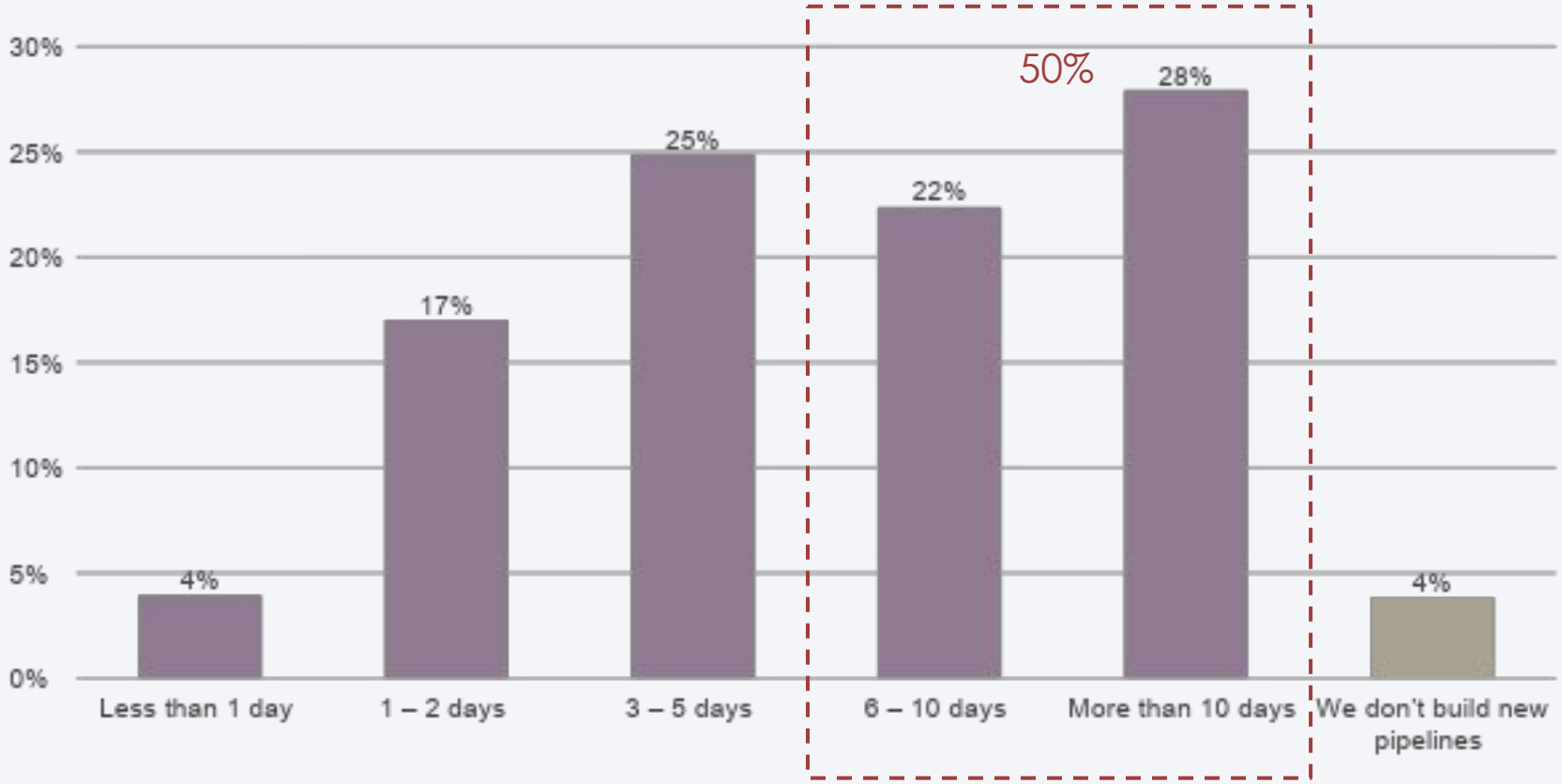
98% FACE CHALLENGES BUILDING NEW PIPELINES

What challenges do you experience when developing new pipelines from data sources?



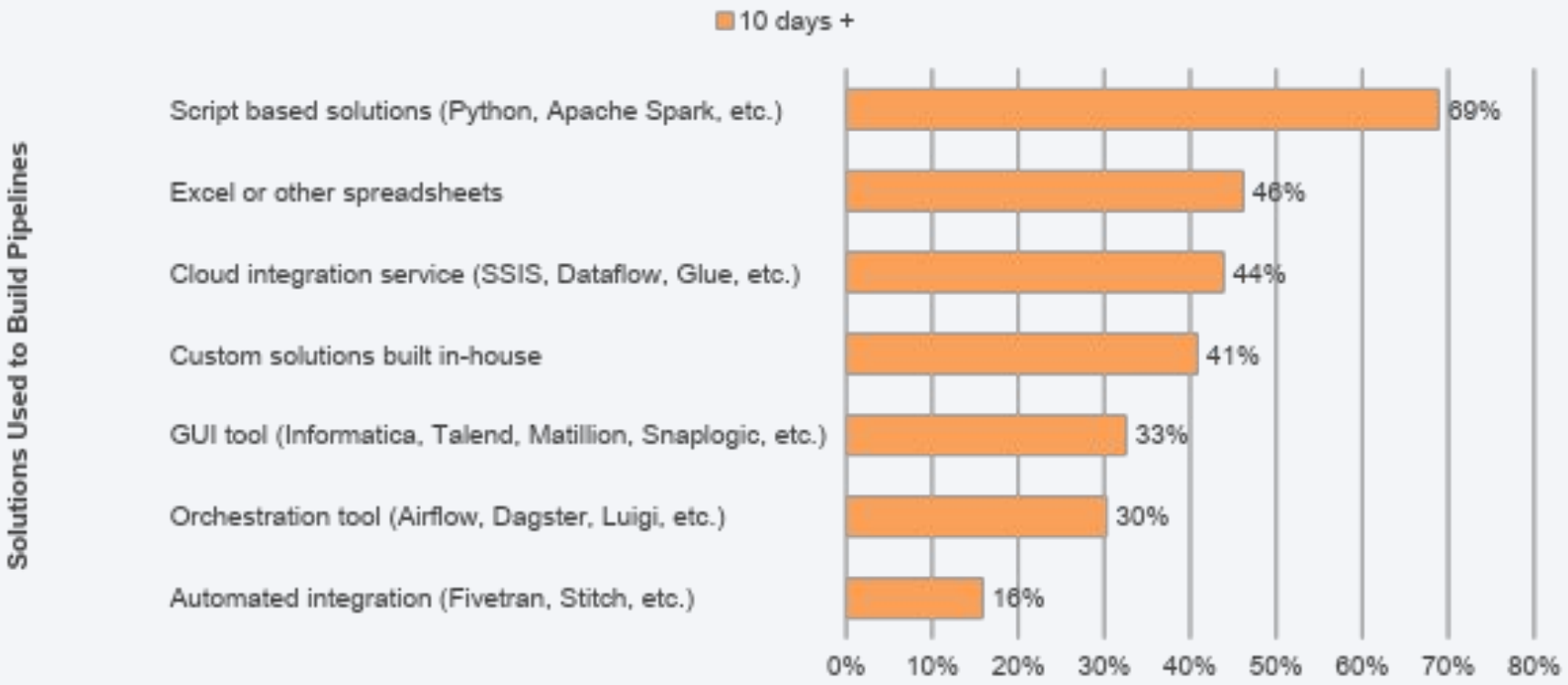
50% OF COMPANIES REQUIRE MORE THAN A BUSINESS WEEK TO BUILD A NEW DATA PIPELINE

How long does it typically take to create a new data pipeline from a data source?

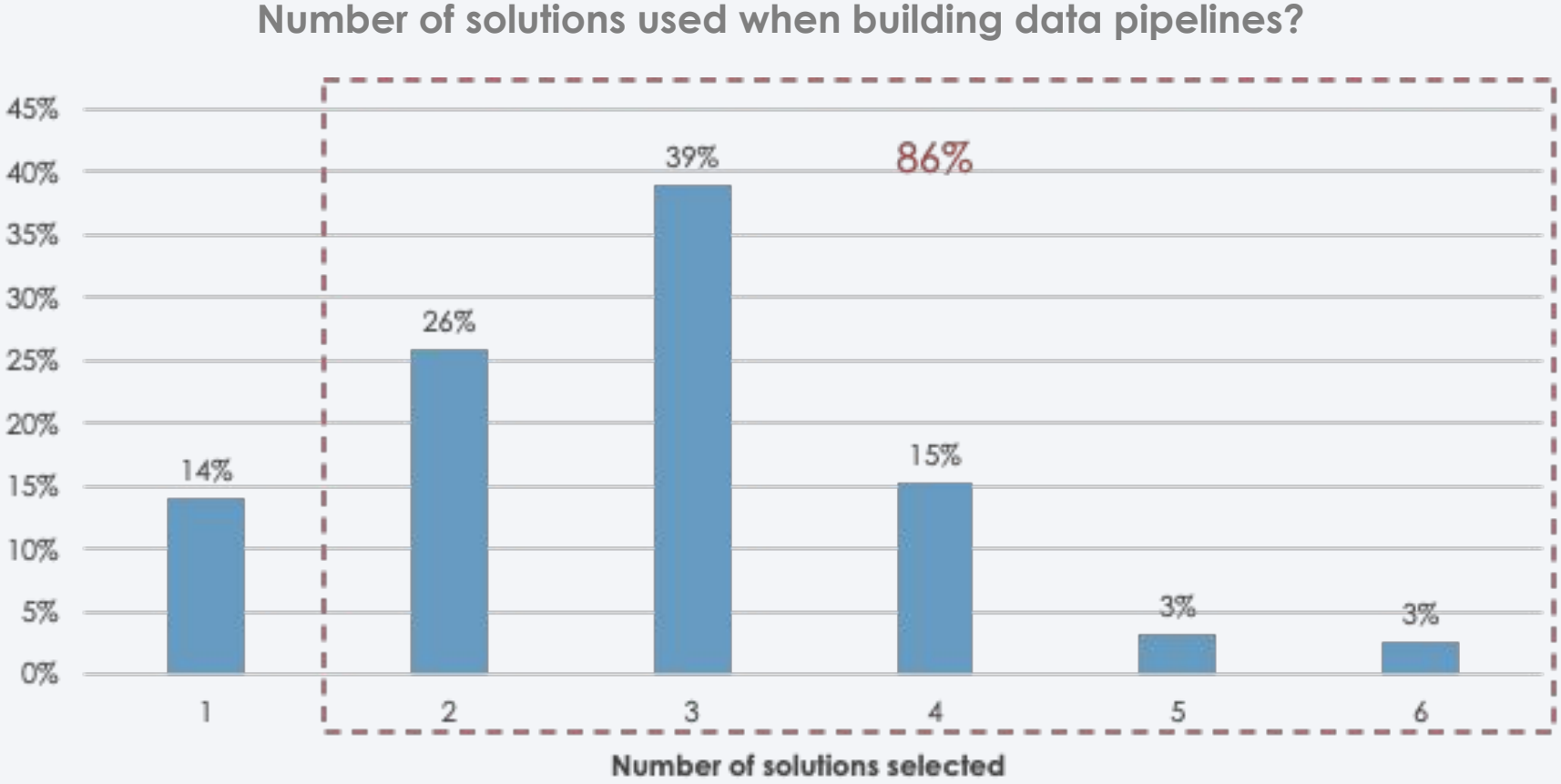


SCRIPT-BASED TOOLS AND SPREADSHEETS RESULT IN LONGER PIPELINE CREATION TIMELINES

Those that required 10 or more days to create pipelines compared to tools used

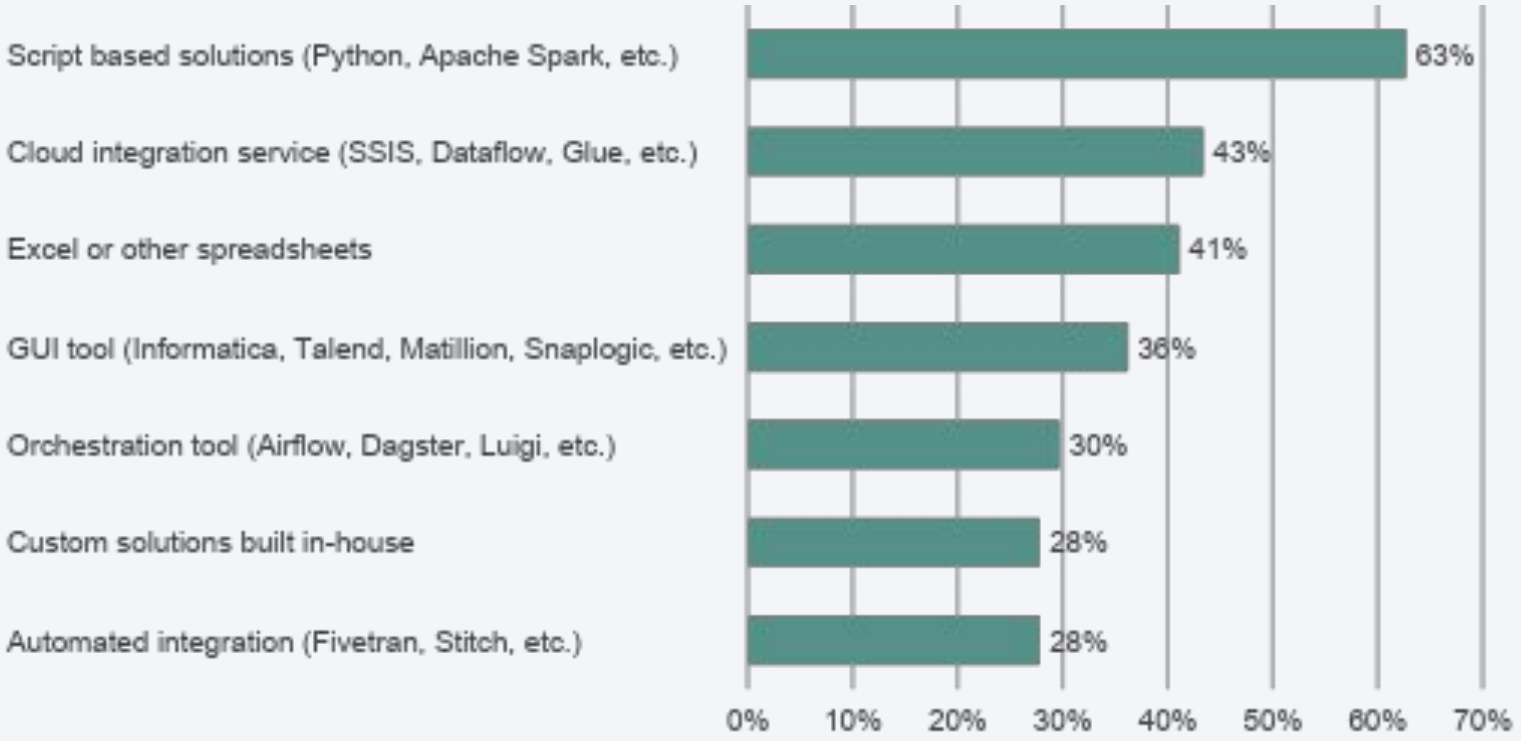


86% USE MULTIPLE SOLUTIONS TO BUILD NEW DATA PIPELINES

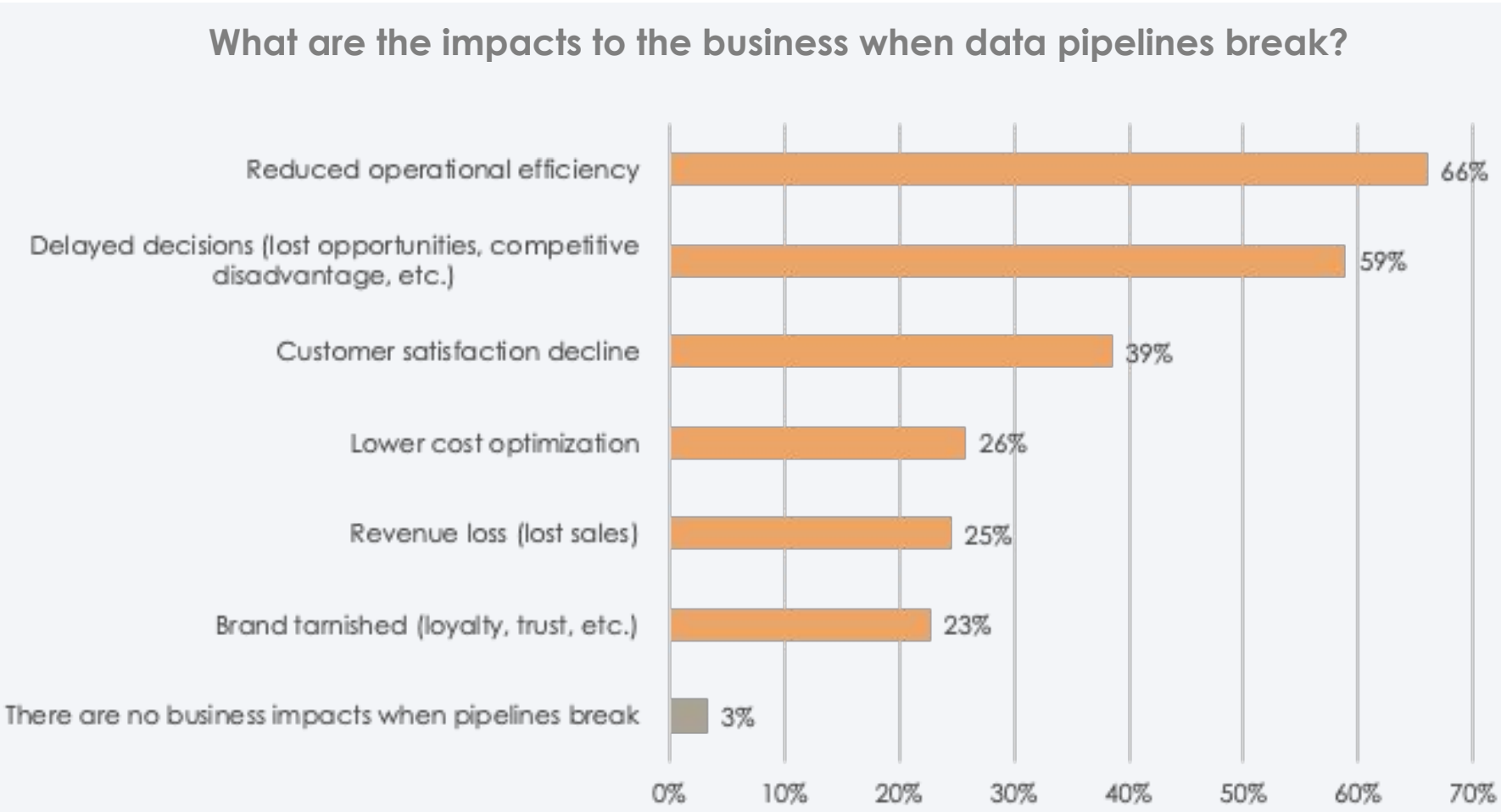


63% STILL RELY ON MANUAL SCRIPTING

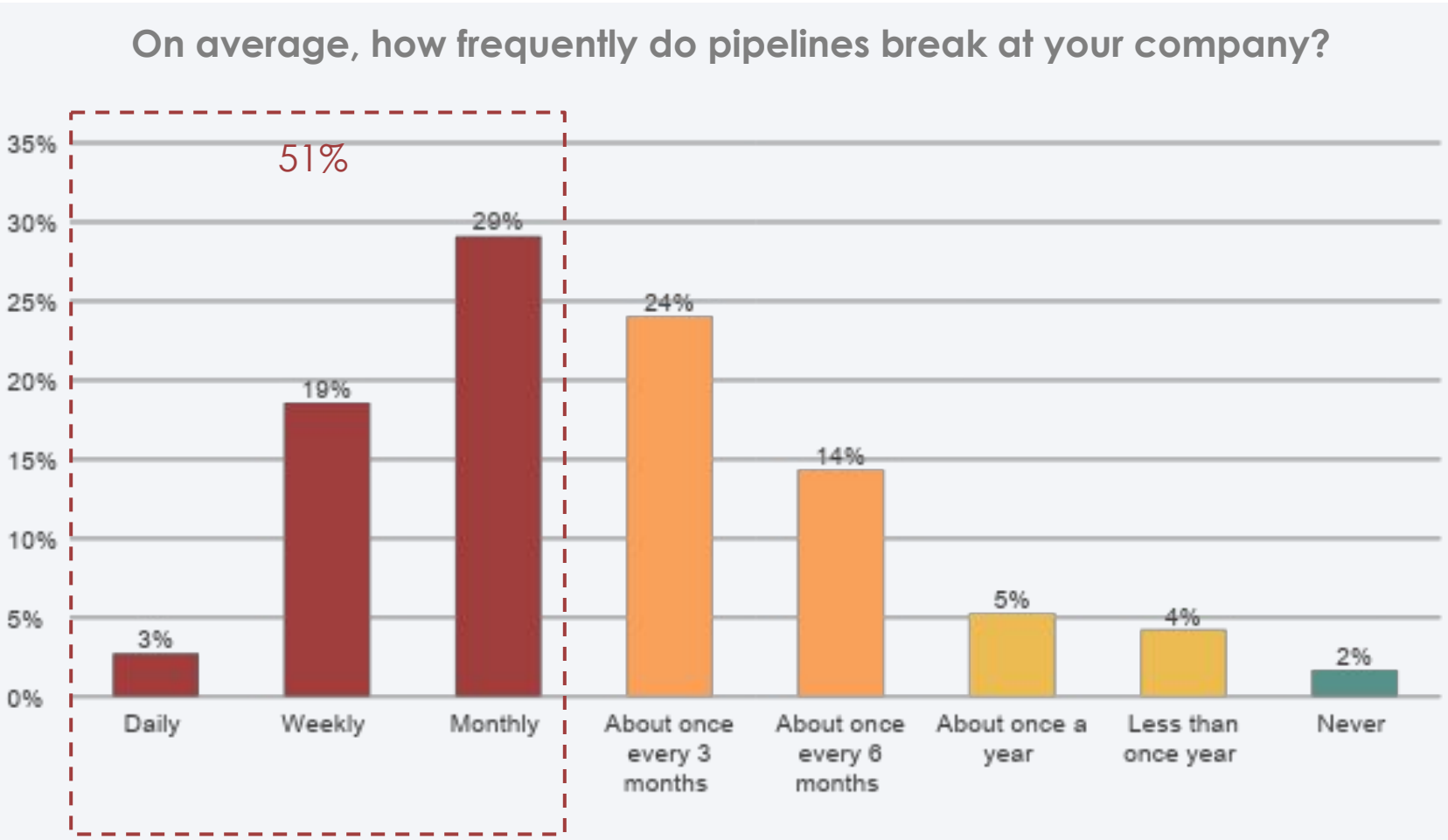
Which of the following solutions do you use when building data pipelines?



COMPANY EFFICIENCY AND AGILITY SUFFER WHEN PIPELINES BREAK

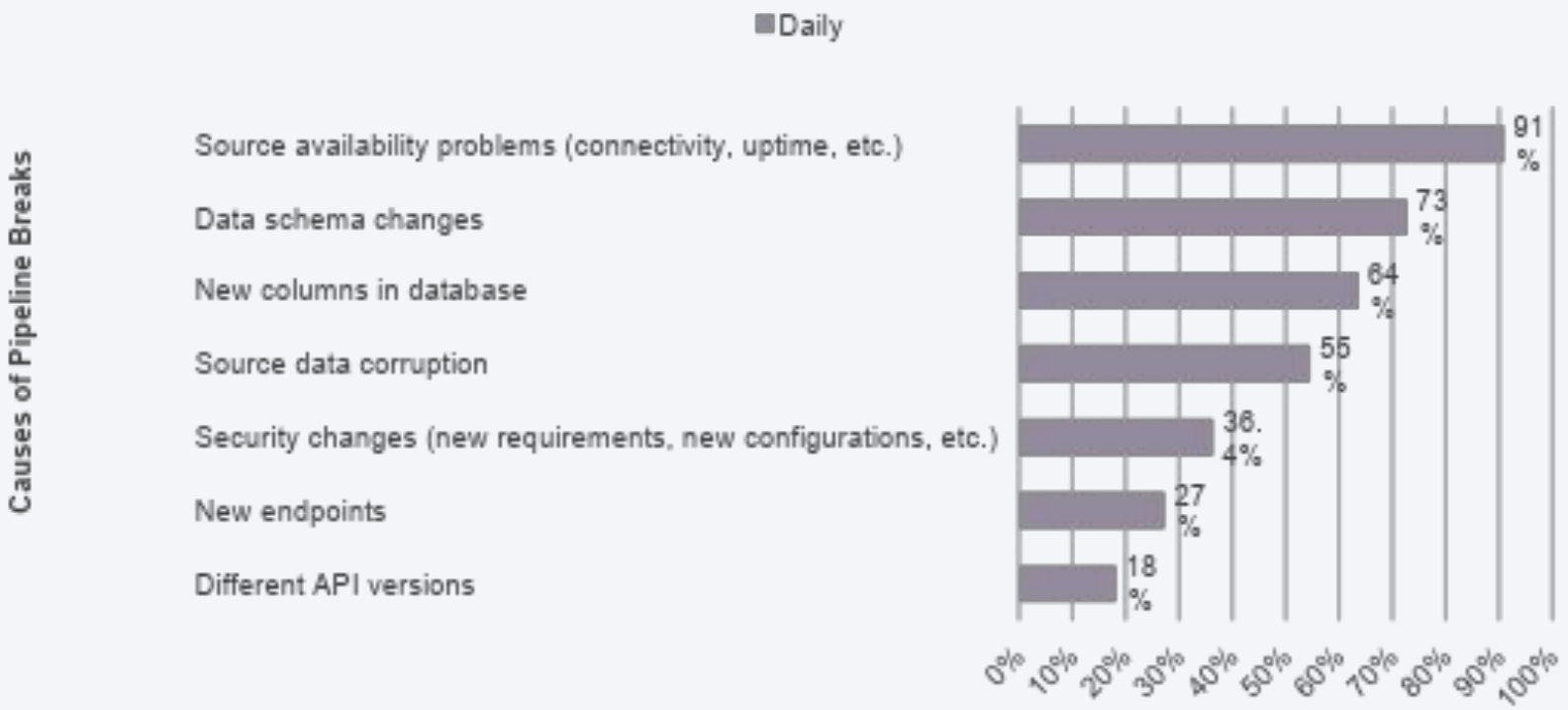


98% REVEAL DATA PIPELINES BREAK, WITH 51% REPORTING IT HAPPENS MONTHLY OR MORE FREQUENTLY

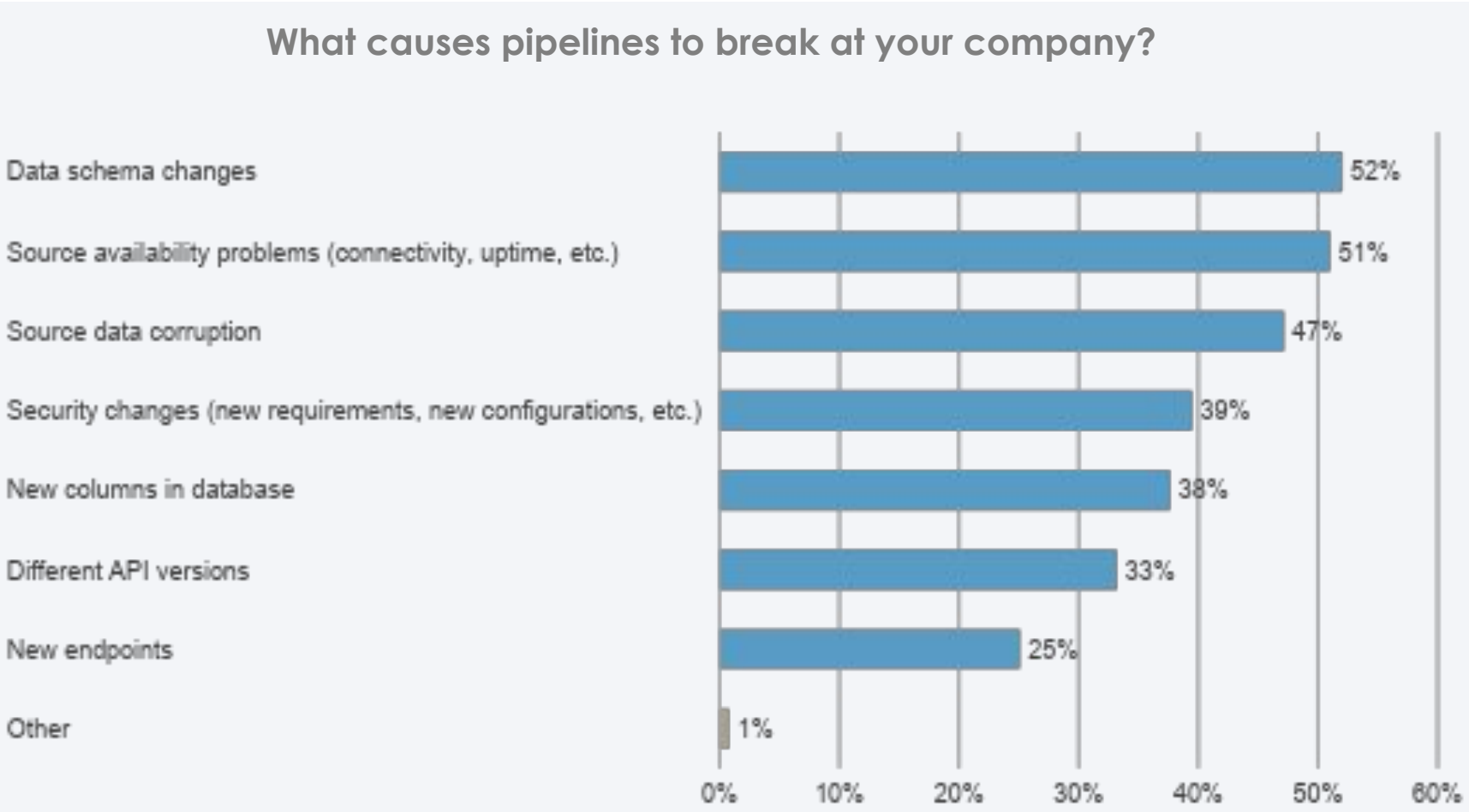


SOURCE AVAILABILITY AND DATA SCHEMA CHANGES TOP REASONS WHY PIPELINES BREAK DAILY

Those that have pipelines break daily mapped to what causes them

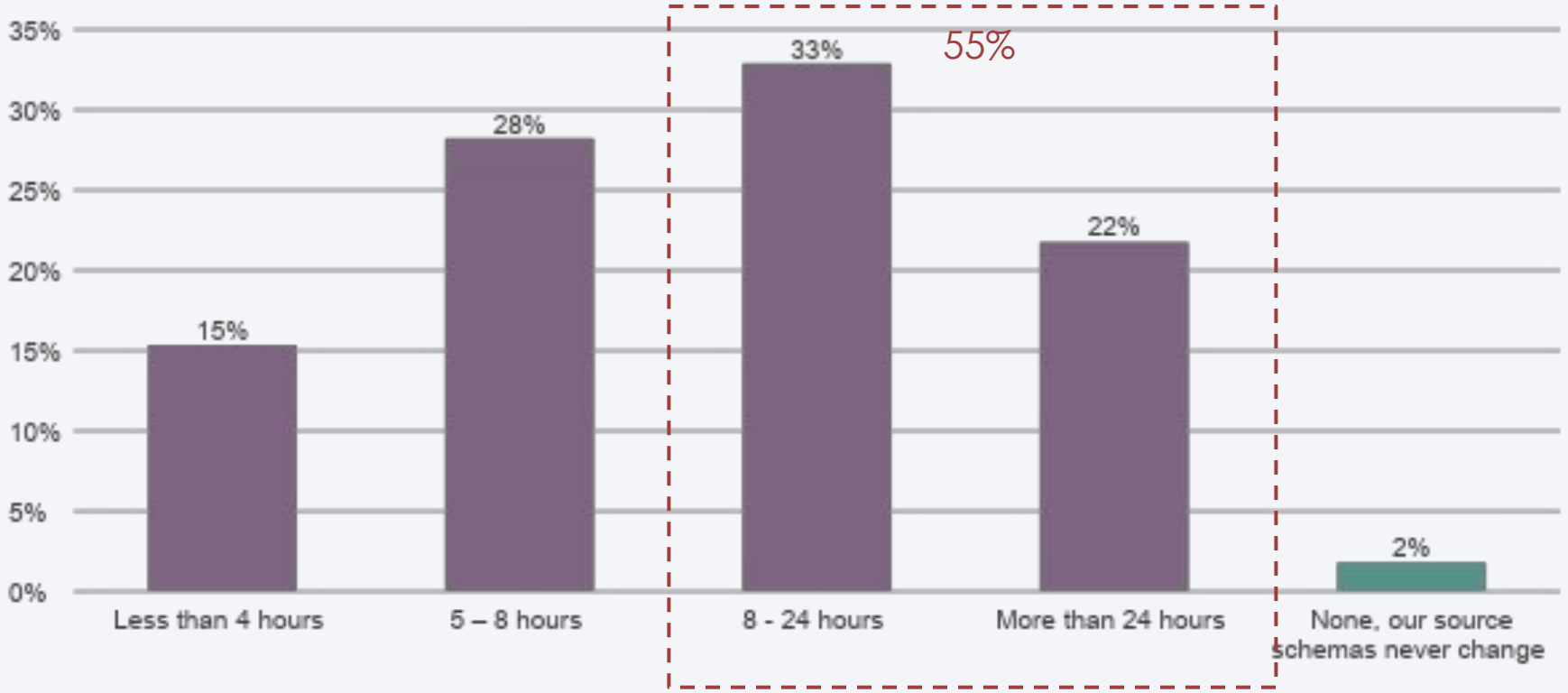


NUMEROUS ISSUES CREATE PIPELINE FAILURES



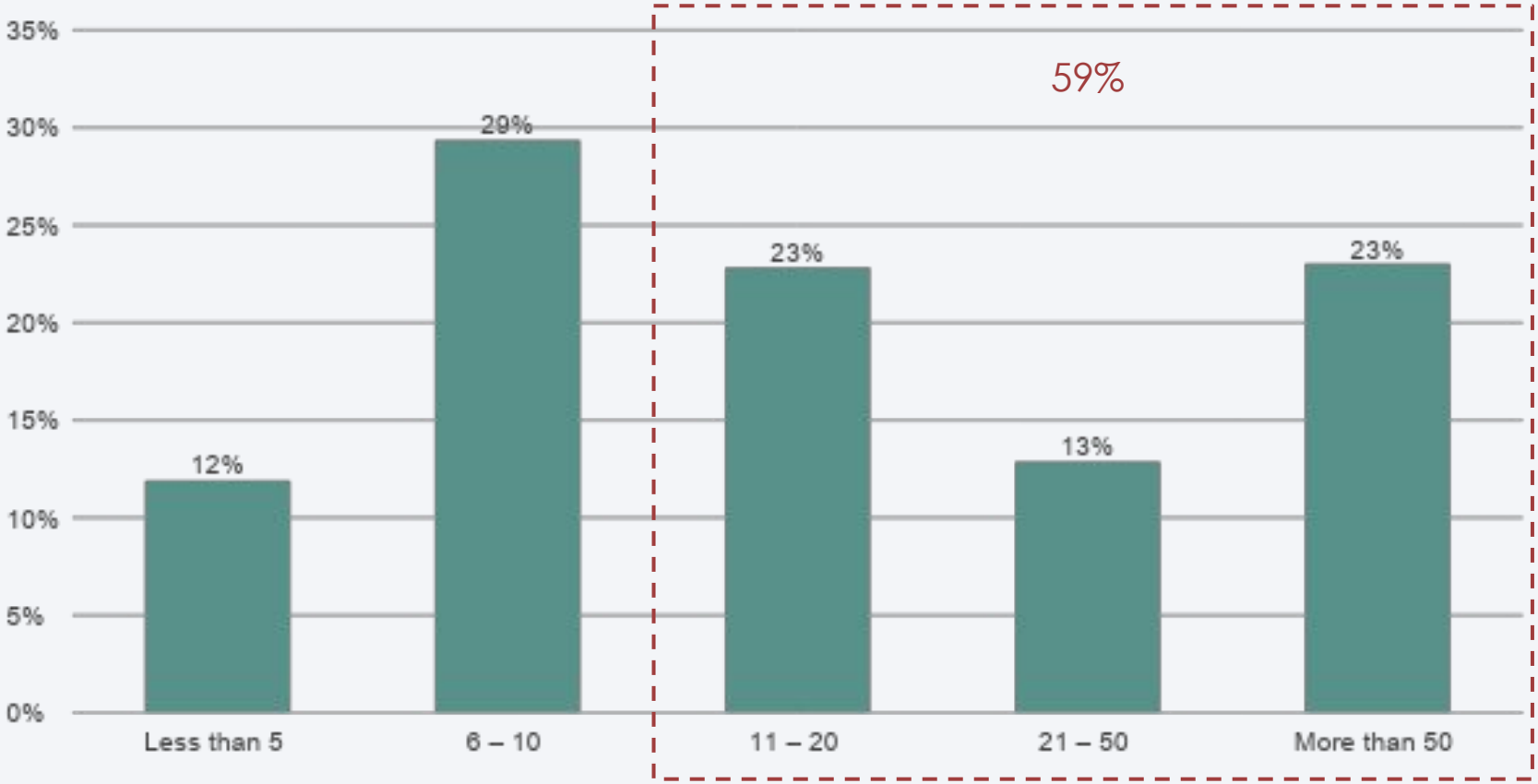
55% REPORT IT TAKES LONGER THAN ONE BUSINESS DAY TO REPAIR PIPELINES

How long does it typically take to repair a pipeline due to source schema changes (API versions, new endpoints, new columns in database, etc.)?

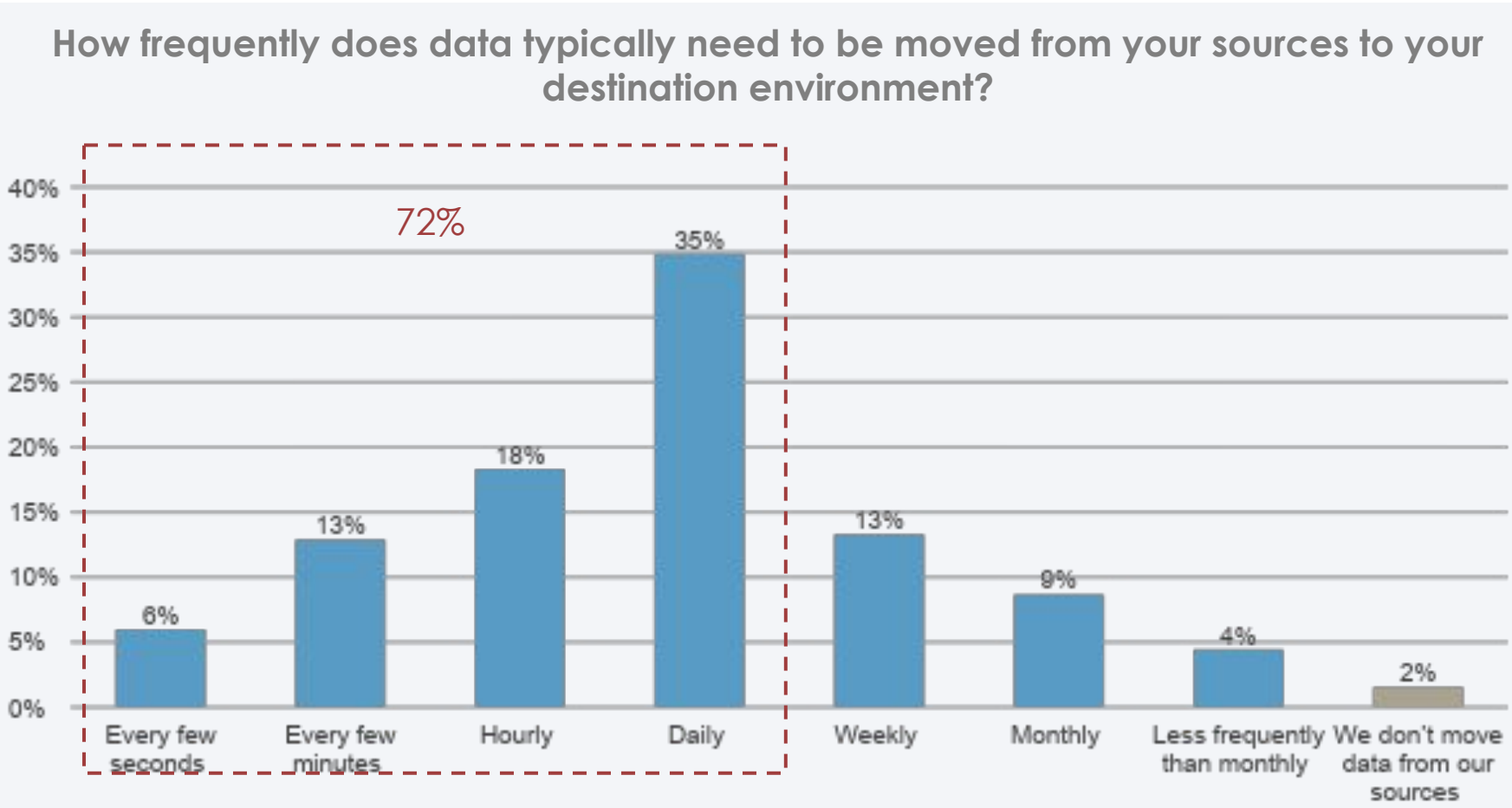


59% OF COMPANIES USE 11 OR MORE DATA SOURCES

Approximately, how many data sources are currently used at your company?

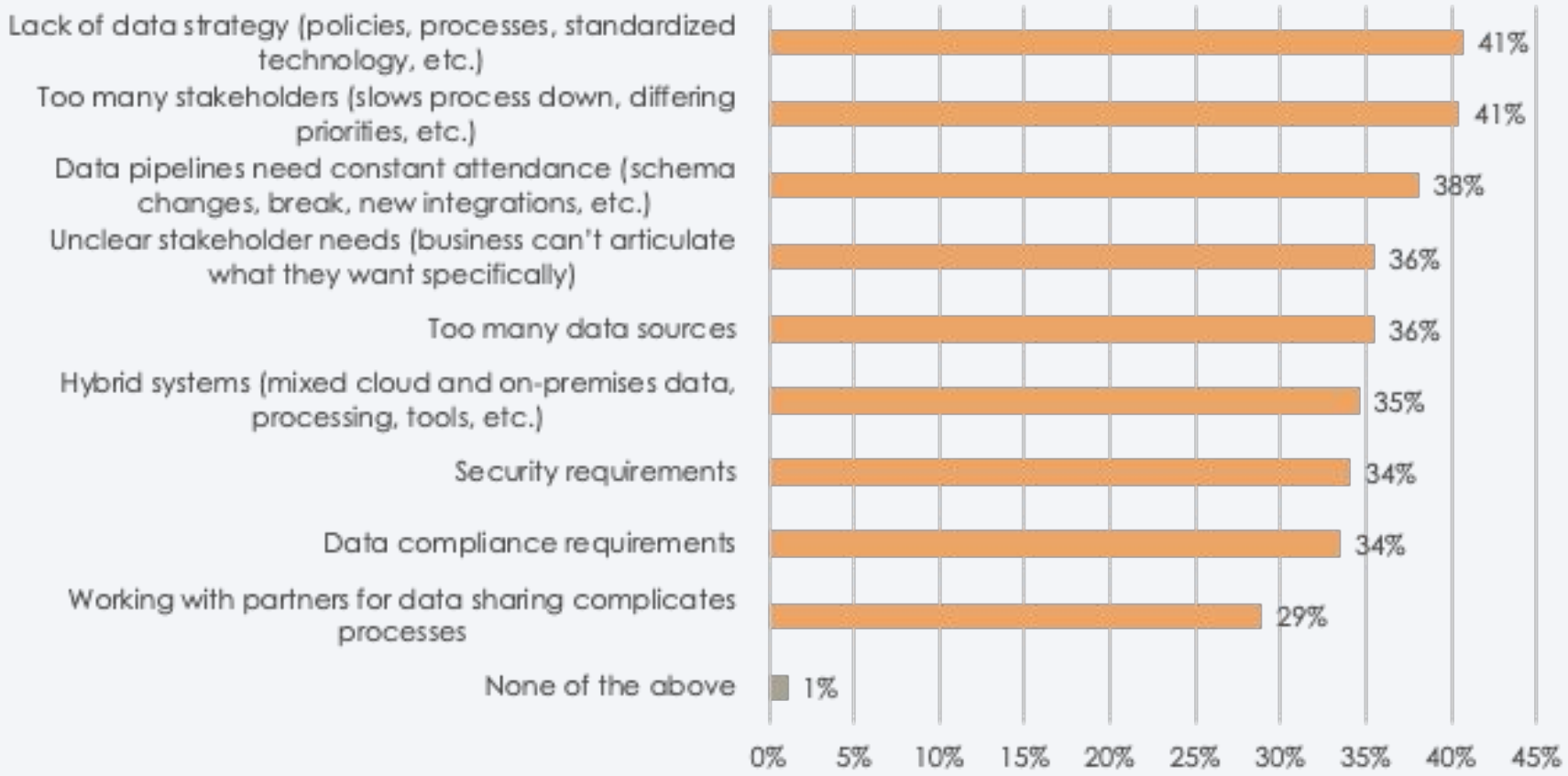


72% MOVE SOURCE DATA DAILY OR MORE FREQUENTLY

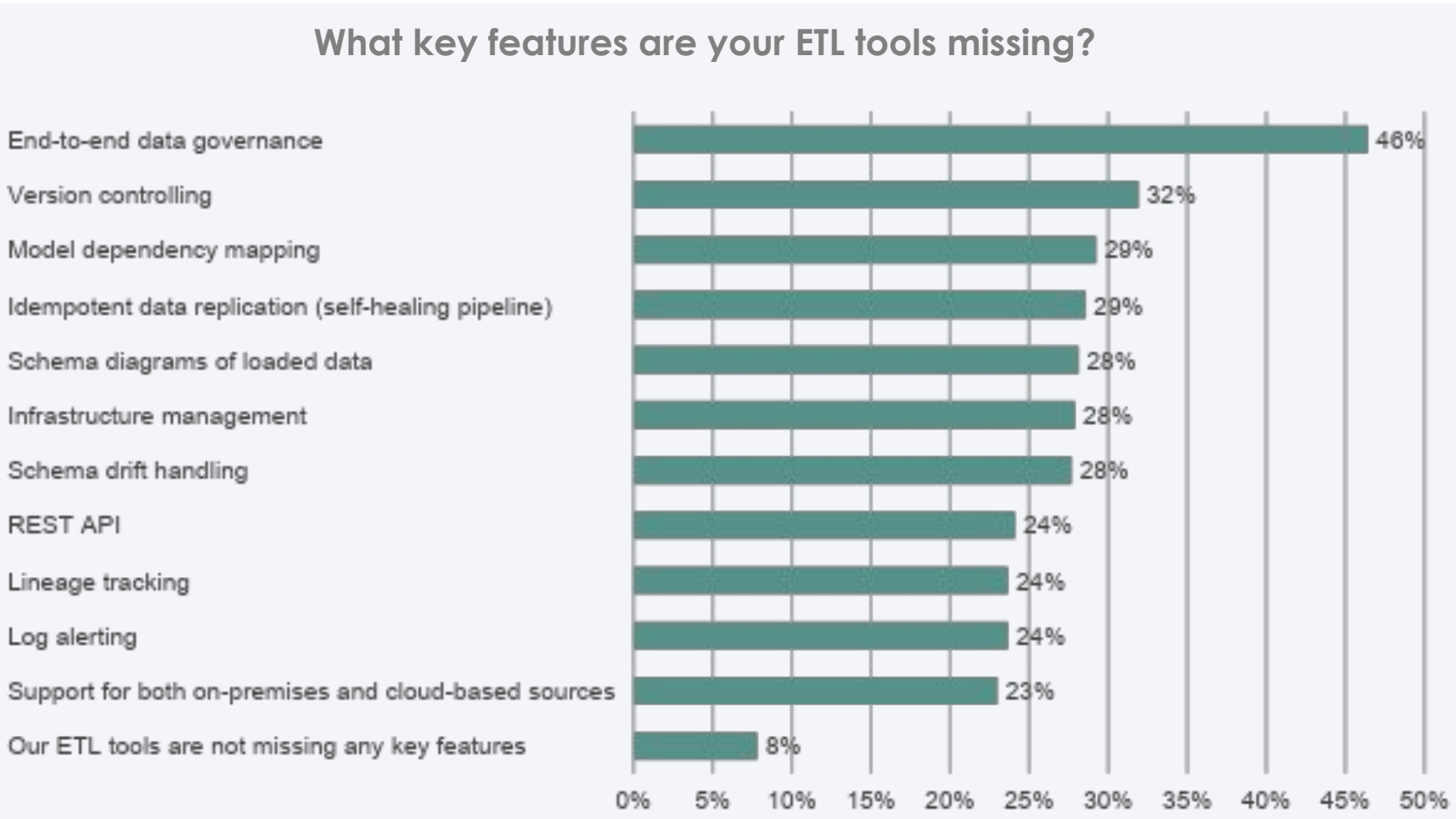


LACK OF STRATEGY, NUMEROUS STAKEHOLDERS, UNRELIABLE PIPELINES TOP THE LIST OF DATA CHALLENGES

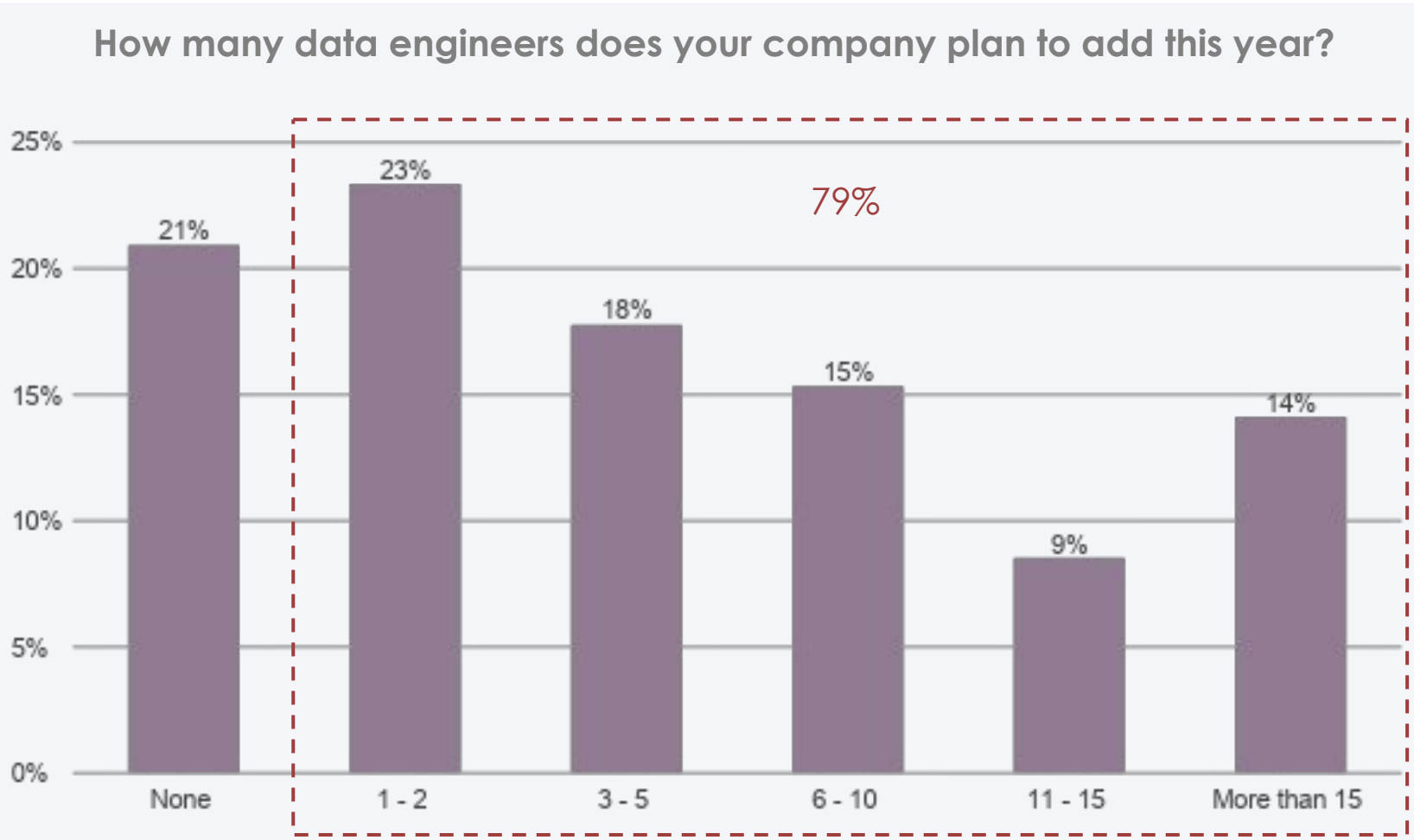
Which of the following data-related challenges does your organization experience?



DATA GOVERNANCE TOPS LONG LIST OF KEY ETL FEATURES MISSING



79% OF COMPANIES PLAN TO HIRE DATA ENGINEERS THIS YEAR



FOR MORE INFORMATION…

About Dimensional Research

Dimensional Research® provides practical market research for technology companies. We partner with our clients to deliver actionable information that reduces risks, increases customer satisfaction, and grows the business. Our researchers are experts in the applications, devices, and infrastructure used by modern businesses and their customers.

For more information, visit www.dimensionalresearch.com.

About Fivetran

Fivetran, the leader in automated data integration, delivers ready-to-use connectors that automatically adapt as schemas and APIs change, ensuring consistent, reliable access to data. Fivetran improves the accuracy of data-driven decisions by continuously synchronizing data from source applications to any destination, allowing analysts to work with the freshest possible data. To accelerate analytics, Fivetran automates in-warehouse transformations and programmatically manages ready-to-query schemas. Fivetran is headquartered in Oakland, California, with offices around the globe.

For more information, visit www.fivetran.com.

METHODOLOGY AND PARTICIPANTS

dimensional research

GOALS AND METHODOLOGY

Research Goal

The primary research goal was to understand the direct business value data engineering enables to support data-driven decision making. The research also focused on the role of the data engineer, their tasks, data pipeline creation and maintenance, tools used, and challenges.

Methodology

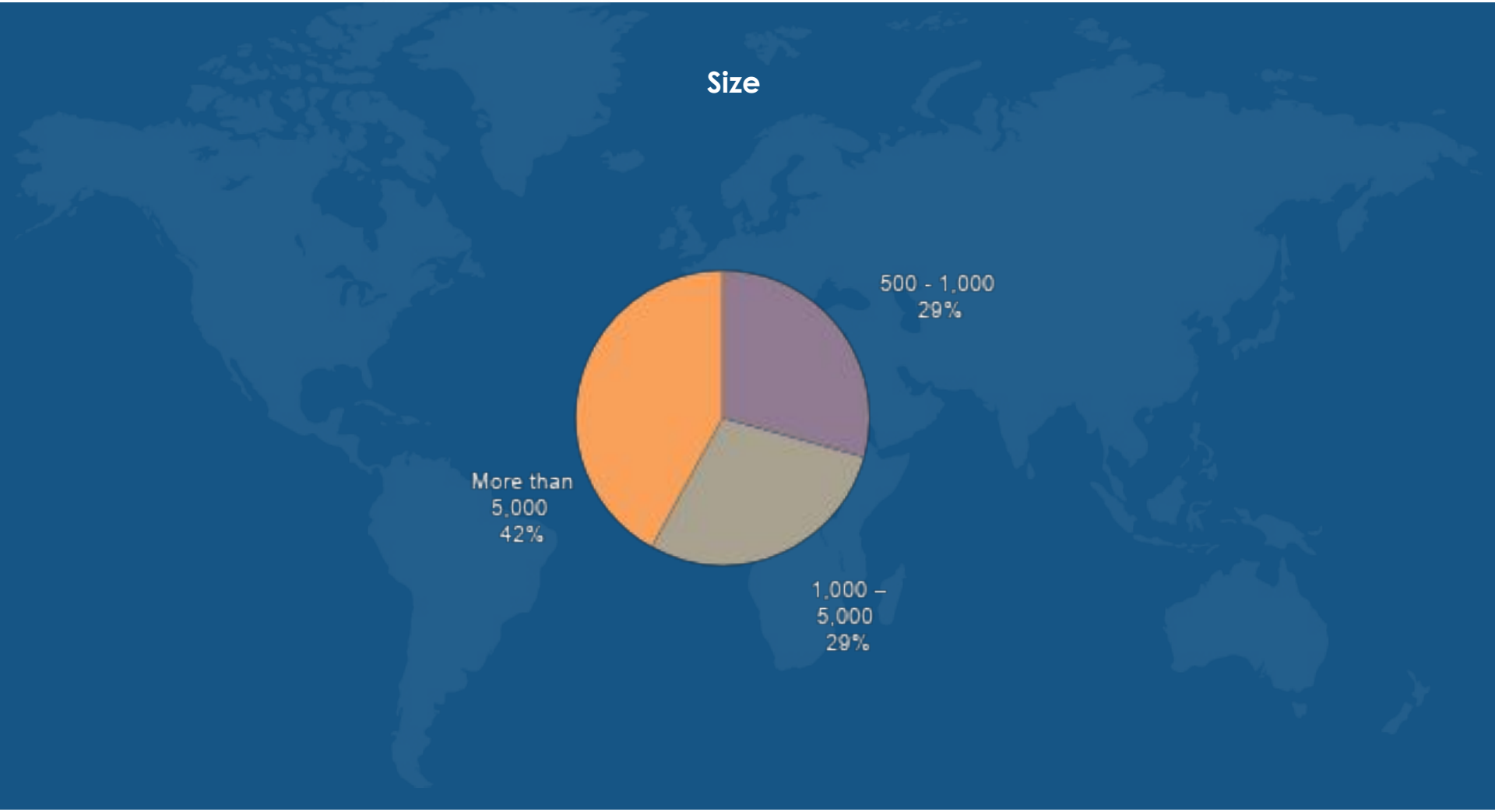
Data professionals at medium-sized to enterprise companies representing all seniority levels were invited to participate in a survey on their company's data engineering processes, objectives, expected business value and specific tool use.

The survey was administered electronically, and participants were offered a token compensation for their participation.

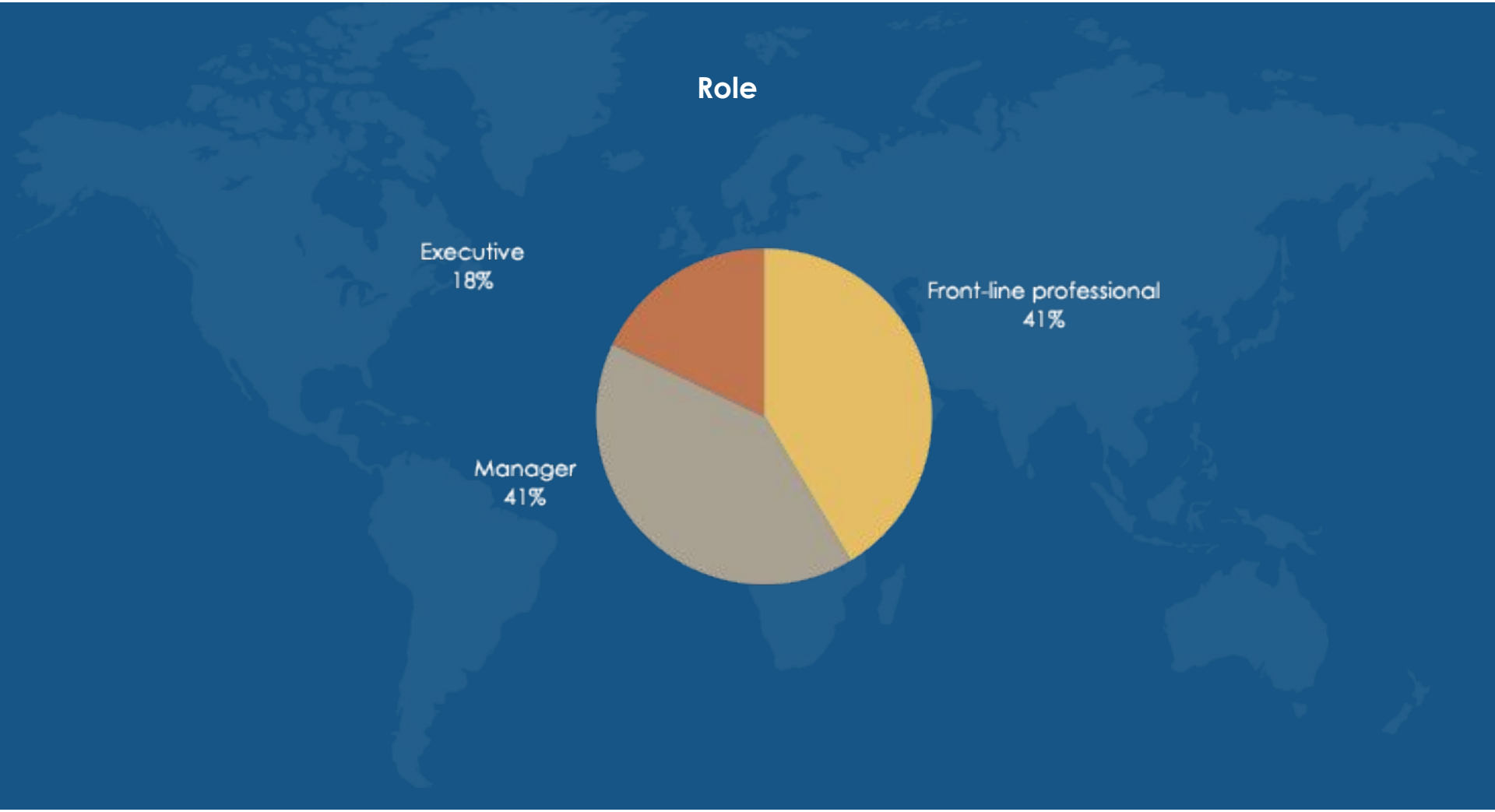
Participants

A total of **543 qualified participants** completed the survey. All participants were data professionals. Participants represented 5 continents.

COMPANIES REPRESENTED



INDIVIDUALS REPRESENTED



FOR MORE INFORMATION:

Check out a free demo!

dimensional research