Hakkōda

The State of Data and the High Cost of Data Sprawl

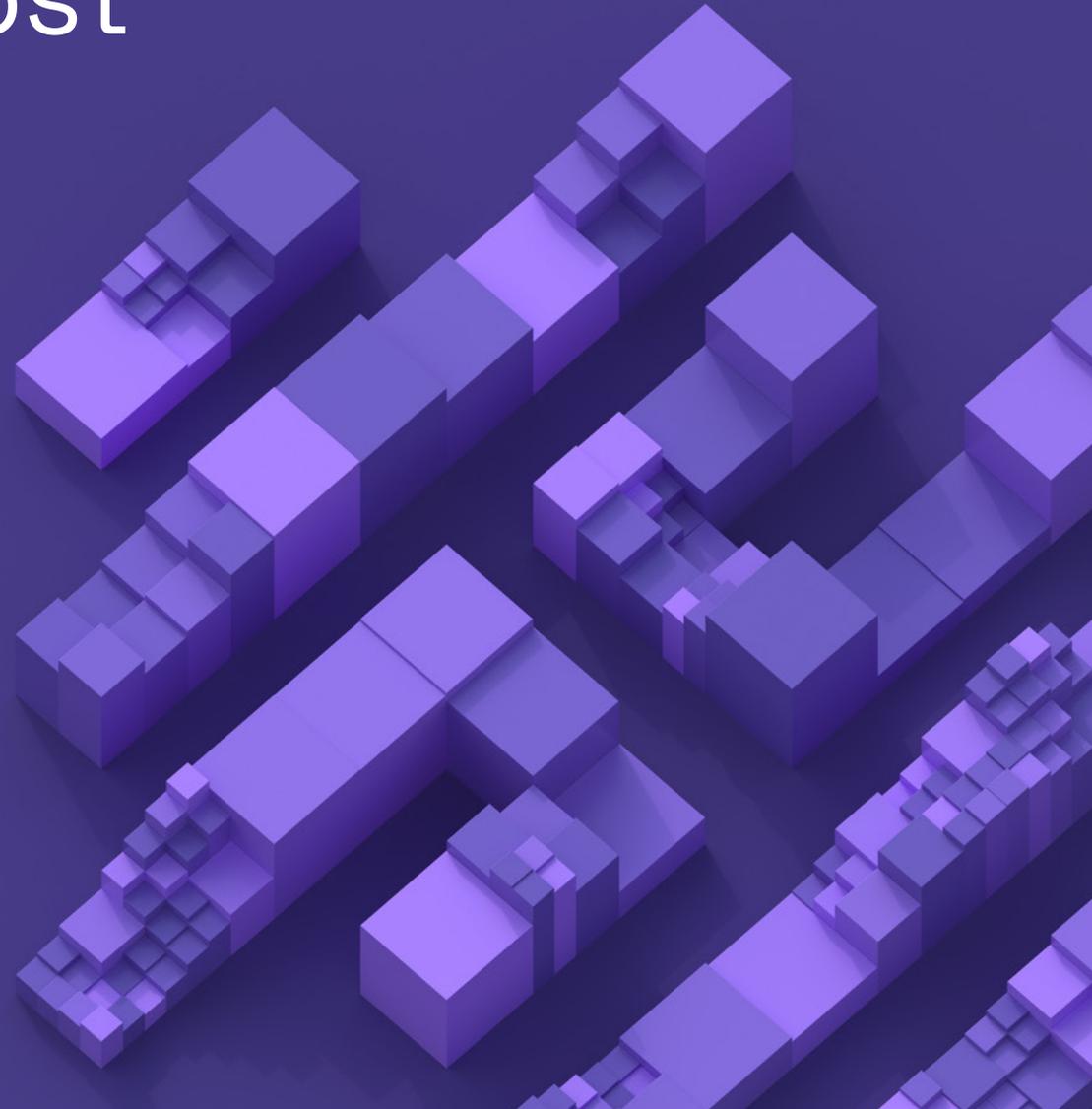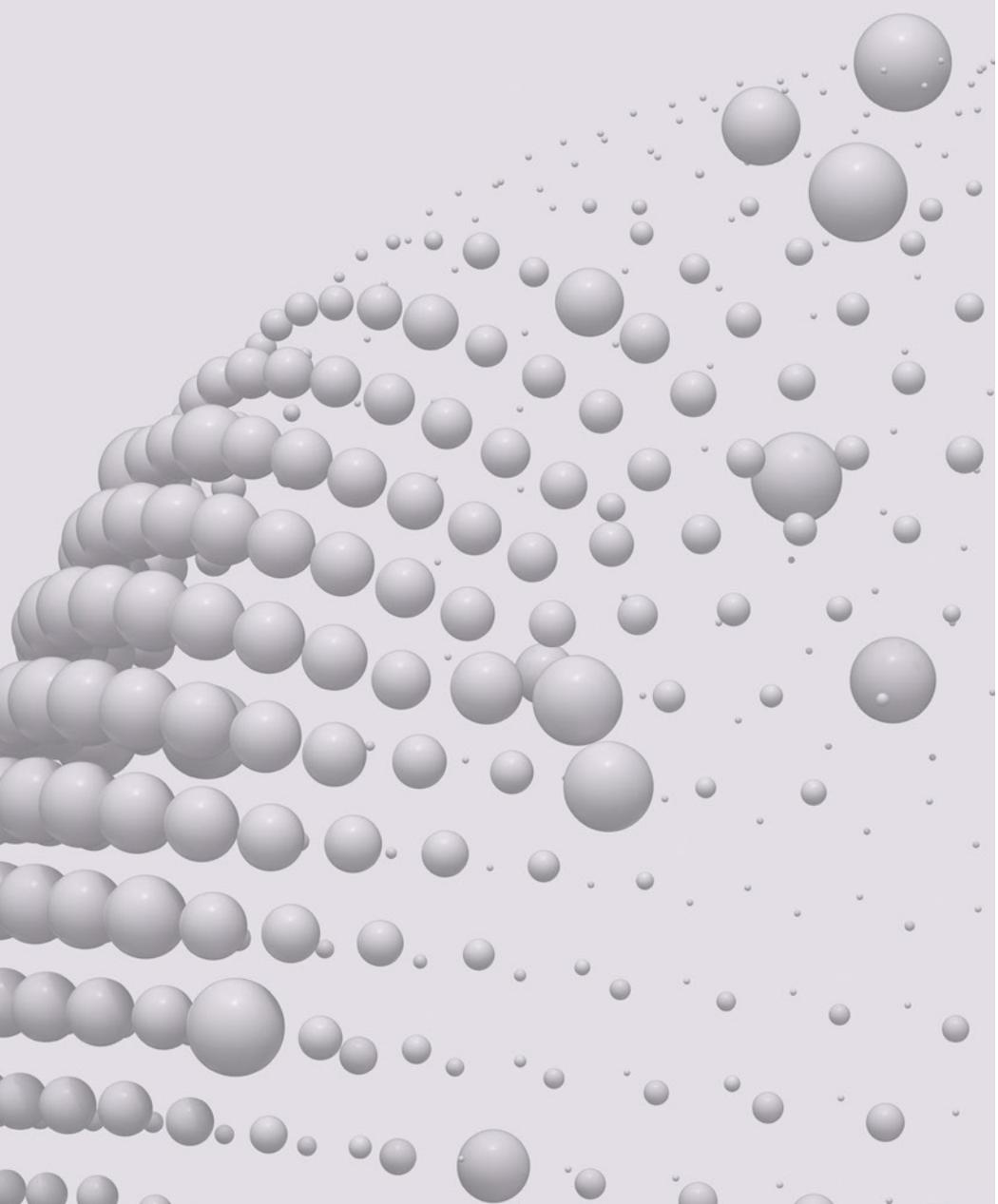**2022 Report**

Table of Contents

Welcome to the World of Data Sprawl

Nearly everyone, in every organization, in nearly every function, uses data. As cliché as it is to say, data is the lifeblood of business.

Futurists imagine a data-driven utopia in which that lifeblood flows effortlessly throughout organizations–driving frictionless customer experiences, more informed real-time decisions, and innovative new offerings that elevate both brand and revenue.

What?

Data sprawl is the unplanned result of layer after layer of individual initiatives in which data is copied, exported, or sent to various applications.
It's the natural, but costly, by-product of organizations chasing opportunity.

But realists see the current state of data today–"pockets of greatness" in organizations held down by data sprawl. Sprawl that keeps data fragmented, hard to use and even harder to manage. Sprawl that keeps driving costs up exponentially, keeping the promised data-driven utopia just out of reach.

Data will compound, along with the sources of data and the platforms, apps and use cases for that data. The question is how to prevent sprawl.

To understand the cost, causes and business impact of data sprawl in organizations, Hakkoda commissioned Dimensional Research to survey 300+business and IT leaders responsible for data and analytics initiatives at mid-to-large size companies.

If you take away only three things from this ebook:

1. Data sprawl is expensive.

Costs are high across the data analytics lifecycle, with BI and reporting (54%) and data architecture (52%) costing the most. 1 in 4 respondents have more than 10 BI apps, 35% use five or more data warehouses, nearly a quarter rely on 10 or more service providers.

2. The costs of data sprawl are hard to pin down.

Half of the respondents know that manual processes (50%) and inefficient work (49%) are a problem, but struggle to capture those costs. Only 3% said they understand the full cost of their data analytics programs.

3. Sprawl and talent scarcity impacts innovation, but there are solutions.

Data sprawl is not only costly to organizations, it's a significant barrier to innovation. 94% of respondents reported barriers to innovation in their data programs, while 97% report problems finding talent. The data in this report offers insight into what's driving inefficiencies and preventing innovation, and investments companies can make to prevent sprawl and free up internal talent to focus on new initiatives.

Data matters. And since we're on subject, let's dig into it.

1 Data costs are high, the business impact is higher

As organizations have seen the power of data analytics, initiatives (and hopes) have soared. Leaders looking to make the transition to a digital and more globally distributed world need (and desperately want) data.

Collecting data is easy, but turning data into intelligence is hard, expensive, and not always entirely effortless. Which is why the cost of data is skyrocketing.
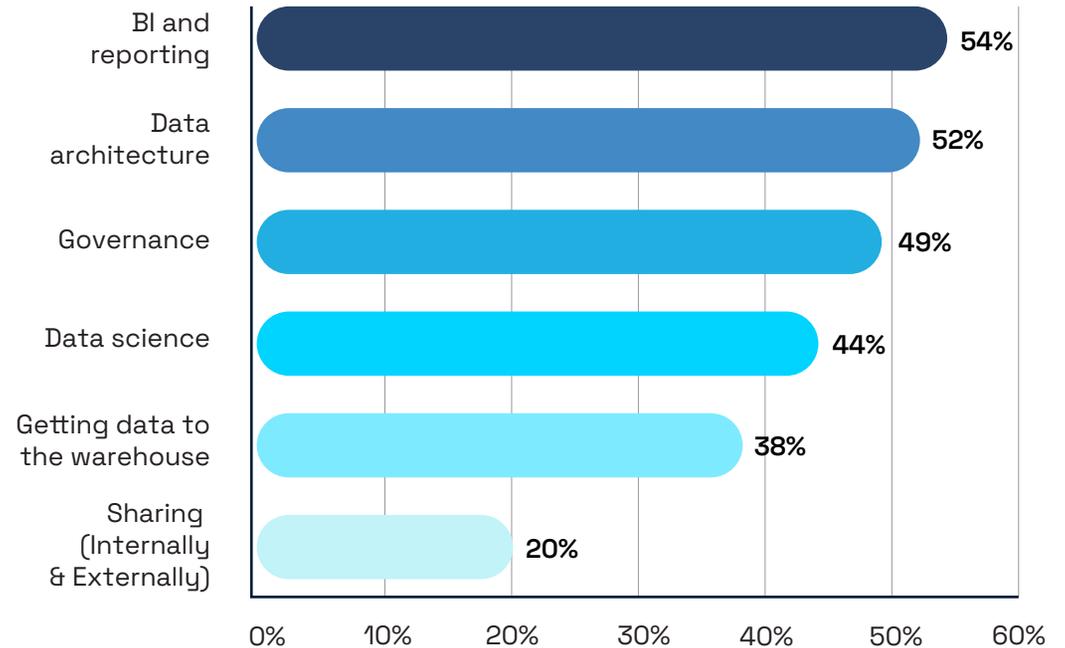
According to McKinsey, a mid-sized company with $5 billion of operating costs might spend $250 million on data management. And that's increasing every year.
A 2019 McKinsey survey showed spending on data-related costs was expected to increase between 47% and 80% each year between 2019 and 2021 depending on the industry. And that was before the pandemic hit.

What's driving those costs?

According to IT and Line of Business (LOB) data leaders we surveyed, there isn't one major driver. Costs are significant across the lifecycle, led by Business Intelligence (BI) and reporting processes (54%) and data architecture (52%). However, governance (49%) and data science (44%) were not far behind.

Costs can be high across the data analytics lifecycle

What stage of your company's data analytics efforts costs the most (pick up to 3)?

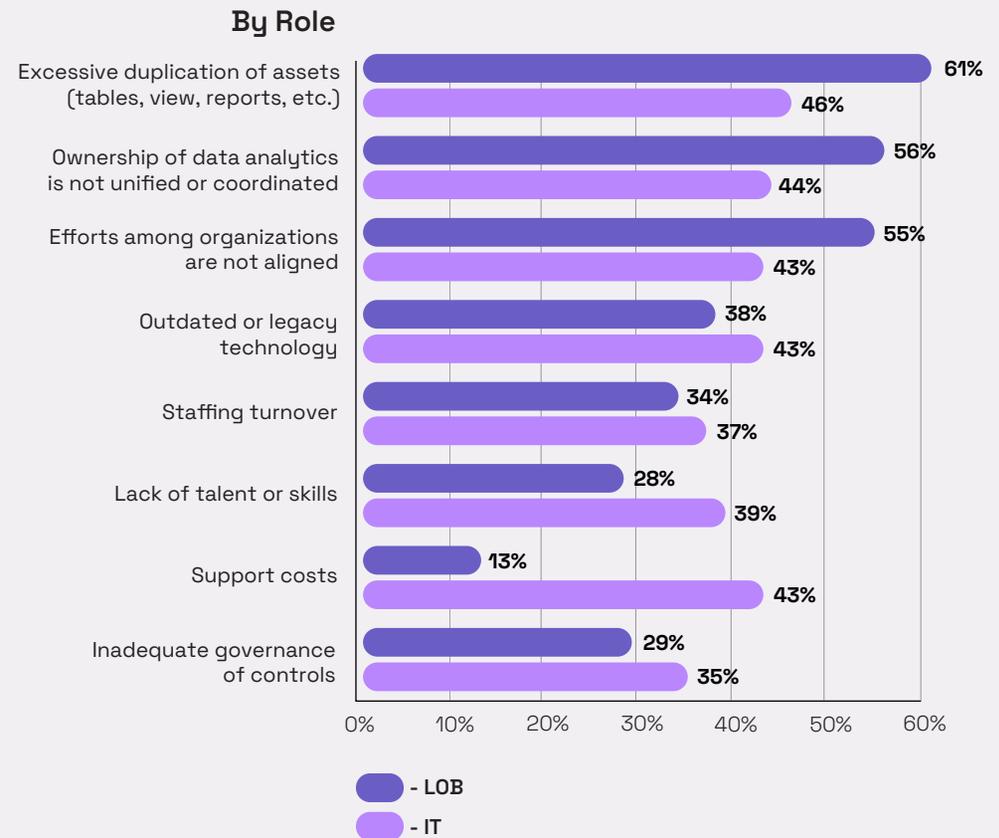| Category	Percentage
BI and reporting	54%
Data architecture	52%
Governance	49%
Data science	44%
Getting data to the warehouse	38%
Sharing (Internally & Externally)	20%

Cost inefficiencies range from duplication and lack of alignment to outdated technology and lack of talent.

Business leaders (61%) and IT (46%) both see that excessive duplication of assets drives costs, but IT is more keenly aware of how the underlying lack of talent or skills (39%) and support costs (43%) exacerbates the problem. More about that in Key Finding 3.

Business and IT leaders both see that excessive duplication of assets drive up costs

LOB leaders are more concerned with duplication and alignment; IT more with support costs and talent

What factors within your company's data analytics program cause inefficiencies?

By Role

| Factor	LOB	IT
Excessive duplication of assets (tables, view, reports, etc.)	61%	46%
Ownership of data analytics is not unified or coordinated	56%	44%
Efforts among organizations are not aligned	55%	43%
Outdated or legacy technology	38%	43%
Staffing turnover	34%	37%
Lack of talent or skills	28%	39%
Support costs	13%	43%
Inadequate governance of controls	29%	35%

- LOB
- IT

2 Data sprawl makes it difficult to capture and manage costs

In modern business, long-term efficiencies are almost always sacrificed on the altar of near-term imperatives. But the need to move fast has created a bit of a mess inside organizations.

In fact, 6% of data leaders characterized the organization and efficiency of their data program as "a dumpster fire."

97% of organizations have multiple BI or reporting applications; 1 in 4 have more than 10

How many different BI or reporting applications has your company purchased or deployed?

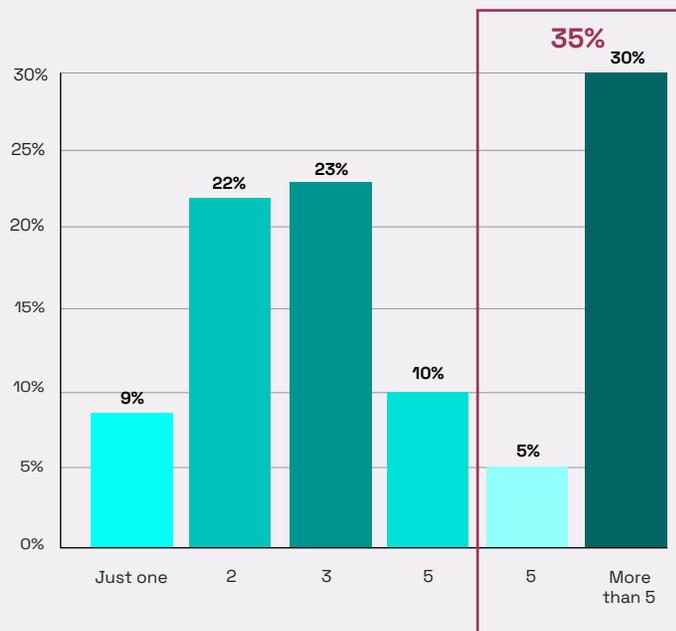| Category	Percentage
Just one	3%
2-3	24%
4-5	32%
6-10	17%
More than 10	24%
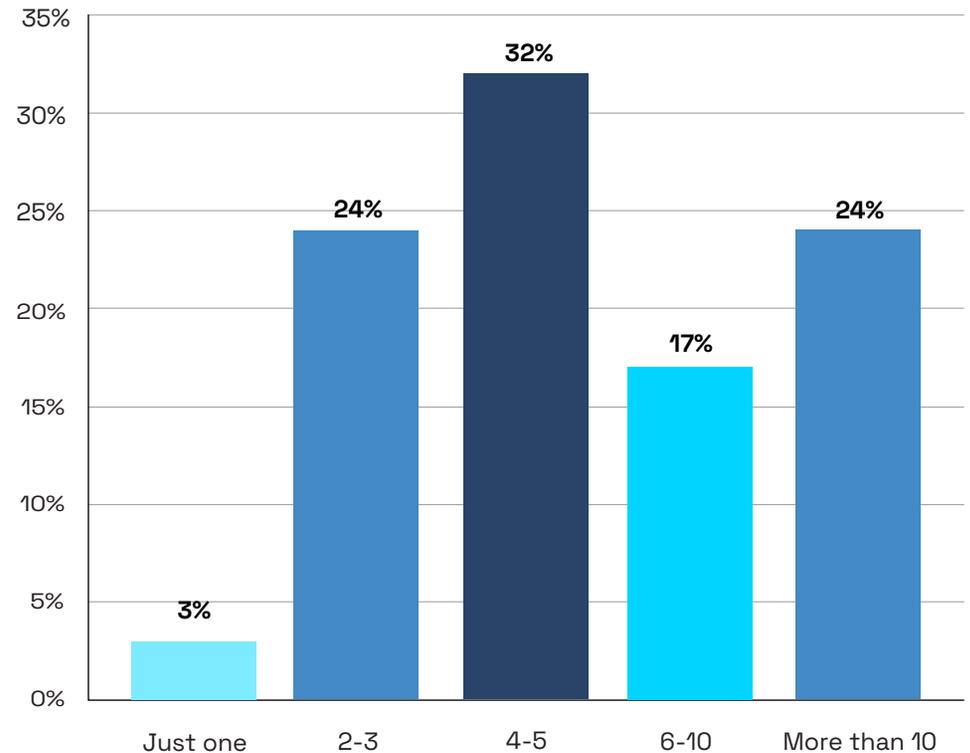
After decades of data initiatives, most companies have an ever-growing tangle of BI apps, data warehouses, and service providers. 97% of data leaders have multiple BI/reporting apps–and one in four data leaders reported having more than 10.

More than a third of those surveyed (35%) use five or more data warehouses, and nearly a quarter (23%) work with more than 10 service providers for analytics.

This isn't surprising. It's a lot easier to add new technology that delivers incremental value on what the company had before than it is to rip out or re-architect what is in place. However, these investments pile up, creating data sludge rather than real return.

To solve the problems that this sludge is creating, organizations add people (employees, contractors, outsourced labor) and manual workarounds, which only adds to the complexity.

More than 1 in 3 (35%) use five or more different data warehouses

How many different data warehouses (including lakes and marts) does your company use currently for analytics or decisioning?

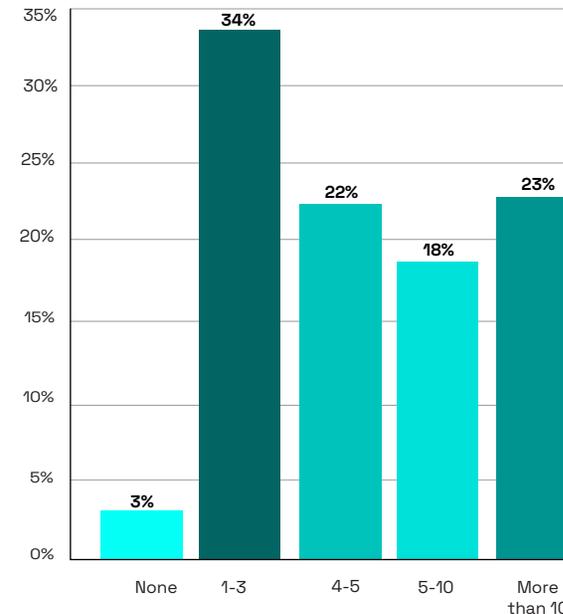## 97% of organizations work with service providers for analytics programs

How many different service providers (contractors, consulting partners, managed service providers, etc.) do you work with to support data analytics programs across your company?

More than three in four executives surveyed (77%) report large numbers of staffers doing basic reporting and analytics, including a quarter (25%) that had more than 500 data analysts in the organization.

If each data analyst earns between $80,000 and $100,000 a year then those organizations are spending as much as $50 million dollars per year just trying to keep up with data sprawl.

But that's just the beginning.

In addition to the hard costs of hiring (and retaining) people and adding licenses, there's a long tail of often-unmeasured costs that come with
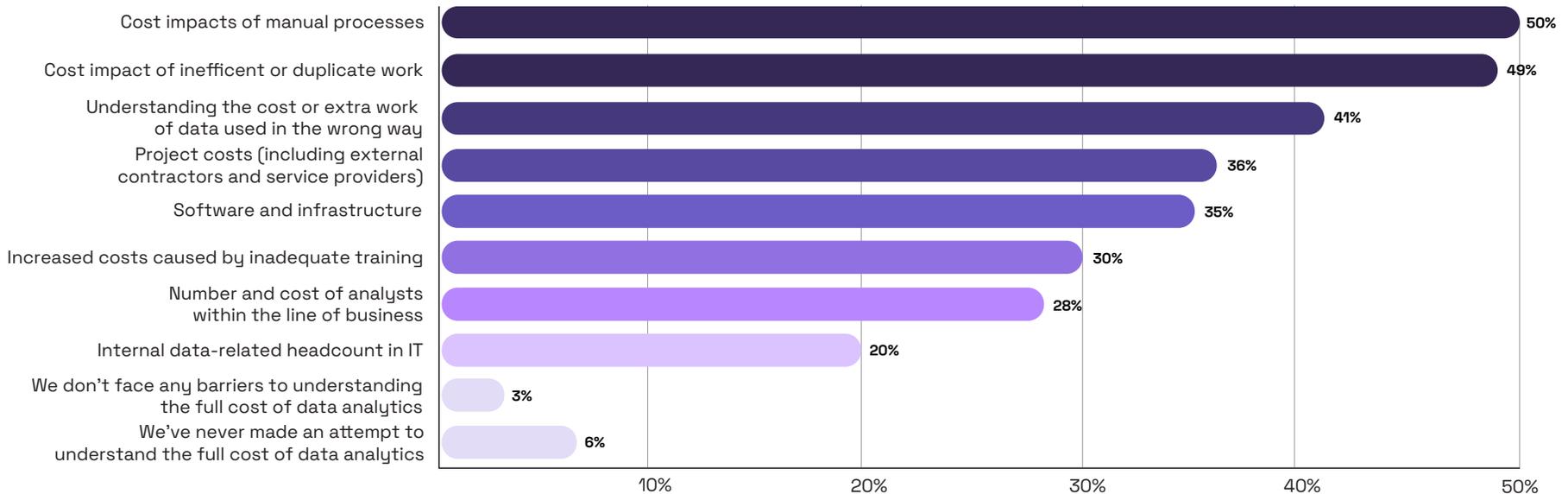
these manual processes, inefficiencies, duplicate work, poor training, and delayed work. Yet despite the steep investment, Gartner still estimates that every year, poor data quality costs organizations an additional average $12.9 million.

91% of data leaders report that it's hard to capture the costs of their data analytics program. Half (50%) struggle to capture the costs of manual processes. More than a third (36%) say they struggle to capture the project costs of adding external contractors and service providers. A full 6% don't even bother.

When data spawl is the norm, so is the need to manage around it.

The hard and soft costs of data sprawl

What types of costs does your company struggle to capture in order to fully understand
the cost of data analytics programs?

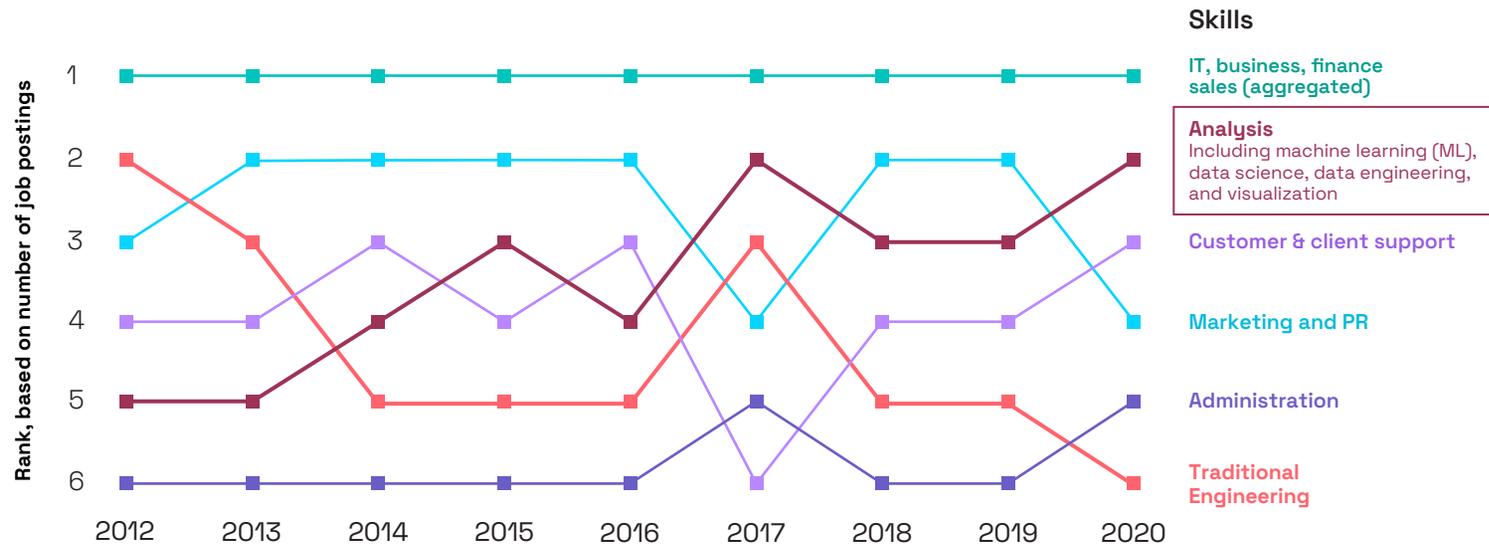| Category	%
Cost impacts of manual processes	50%
Cost impact of inefficent or duplicate work	49%
Understanding the cost or extra work of data used in the wrong way	41%
Project costs (including external contractors and service providers)	36%
Software and infrastructure	35%
Increased costs caused by inadequate training	30%
Number and cost of analysts within the line of business	28%
Internal data-related headcount in IT	20%
We don't face any barriers to understanding the full cost of data analytics	3%
We've never made an attempt to understand the full cost of data analytics	6%

3

Talent scarcity is adding costs, and preventing innovation

We've established that one of the ways organizations manage around the inefficiencies of data sprawl is by hiring people.

But what happens when you need five times more people than your talent acquisition department can find? That may happen sooner than you imagine.

According to Deloitte, yet again in 2020, the number of jobs posted by tech companies for analysis skills like Machine Learning (ML), data science, data engineering and visualization, was higher than traditional engineering skills, QuantHub's latest report shows a shortage of 250,000 data scientists, with three times as many job postings as job searches.

The U.S. Tech University is looking for more 'analytical' and less 'engineering' talent

Skills

IT, business, finance sales (aggregated)

Analysis
Including machine learning (ML), data science, data engineering, and visualization

Customer & client support

Marketing and PR

Administration

Traditional Engineering

Y-axis: Rank, based on number of job postings

X-axis: 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020

The results from Hakkoda's survey confirm what the industry is seeing. Only a lucky 3% reported they had no problem finding talent.
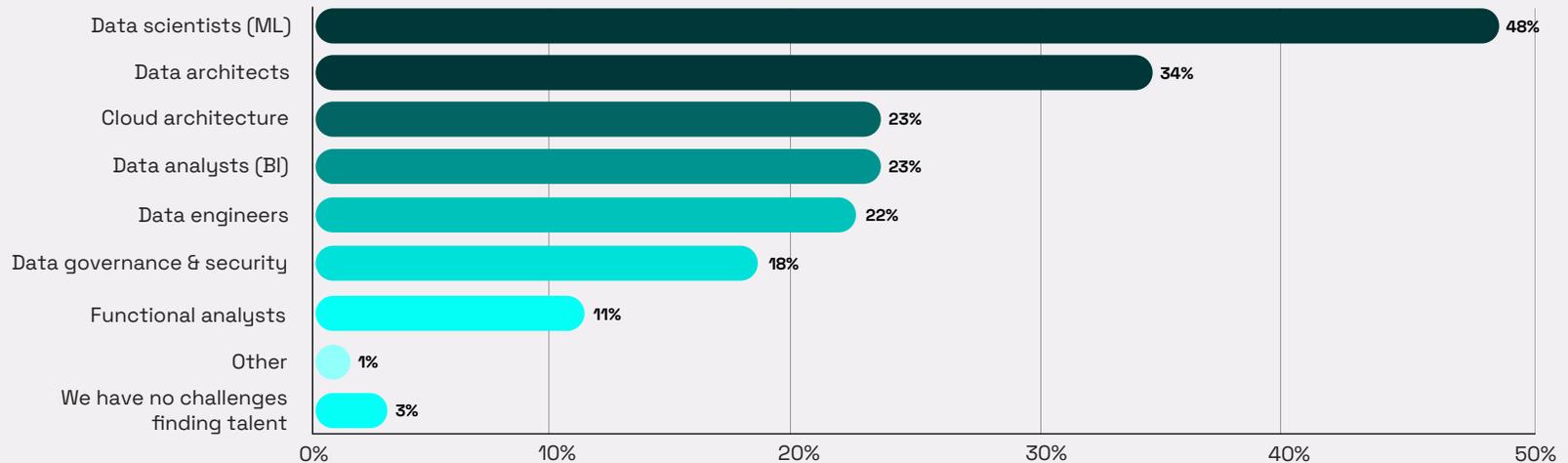
Organizations struggled most with finding Machine Learning-experienced data scientists (48%) and data architects (34%). Experts in cloud architecture (23%) and data analysts and engineers (23%) were a little easier to find… but not by much.

It's interesting to note that when asked which roles provided the most business value, data analysts (41%) and data architects (35%) topped the list. Data scientists were a close third.

When asked which roles provided the most business value, data analysts (41%) and data architects (35%) topped the list.

Data scientists and architects are the hardest roles to find

In which areas does your company have the MOST difficulty finding resources and talent for your data analytics program?

Role	Percentage
Data scientists (ML)	48%
Data architects	34%
Cloud architecture	23%
Data analysts (BI)	23%
Data engineers	22%
Data governance & security	18%
Functional analysts	11%
Other	1%
We have no challenges finding talent	3%

Finding the right talent will always be a challenge.

After all, there is not an infinite pool of good data analytics employees in the world, never mind great ones. Imagine how much more productive those great ones could be if organizations addressed the duplication of assets and manual inefficiencies.
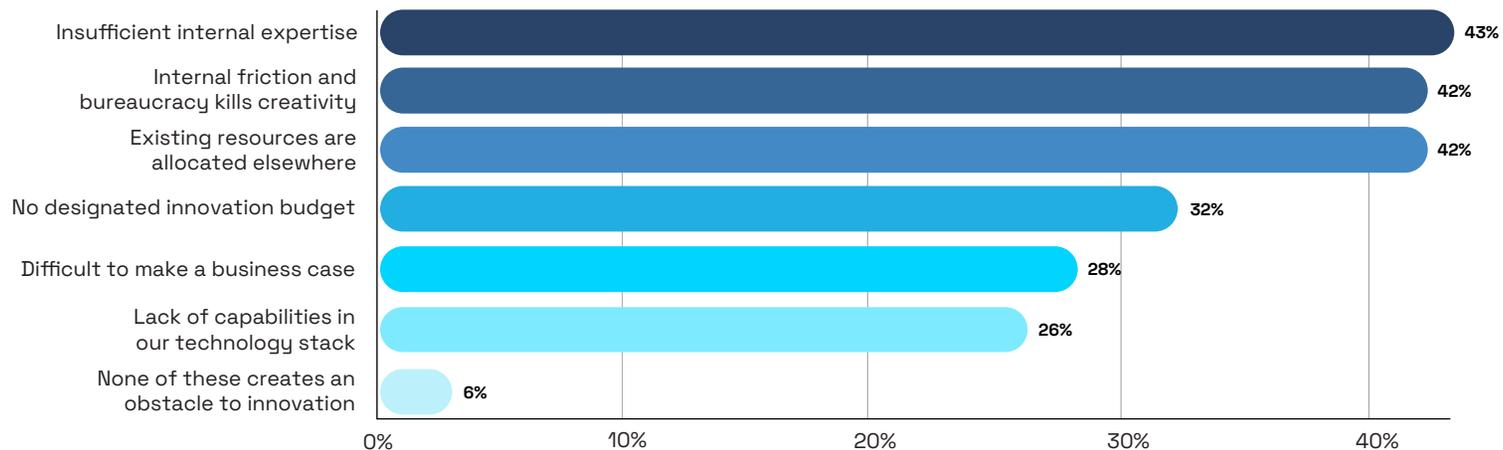
Freeing or finding great talent could also make a big impact on innovation.

The vast majority (94%) of respondents reported some barriers to innovation in their data programs. When asked what is preventing organizations from innovating or creating new offerings, internal bureaucracy and inability to tap into existing resources were a close second to lack of talent.

Perhaps this is why so many organizations work with so many service providers.

A lack of technology isn't preventing innovation; Internal expertise, bureaucracy and a lack of prioritization is

Which of the following makes it difficult for your company to innovate and create new offerings or experiment using your data?

Insufficient internal expertise — 43%
Internal friction and bureaucracy kills creativity — 42%
Existing resources are allocated elsewhere — 42%
No designated innovation budget — 32%
Difficult to make a business case — 28%
Lack of capabilities in our technology stack — 26%
None of these creates an obstacle to innovation — 6%

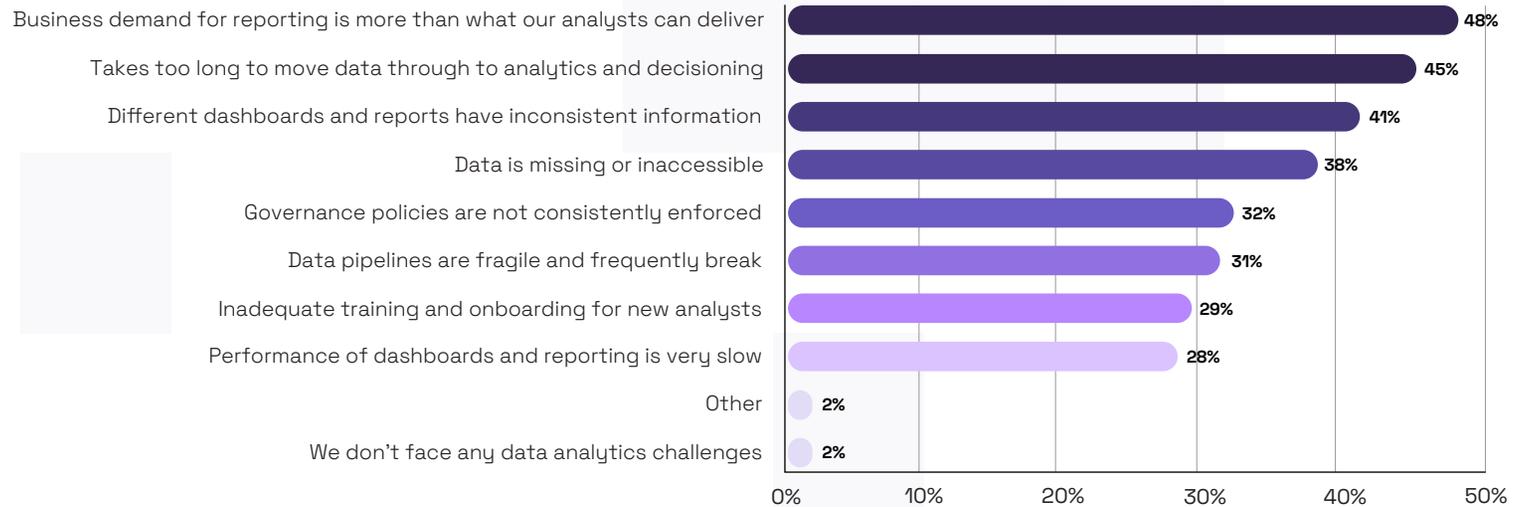# 4 Addressing sprawl and inefficiencies can solve many data challenges

Nearly all IT and LOB executives surveyed (98%) report some challenges with their current approach to data analytics, ranging from overwhelming business demand (48%) to delays in moving data through to decisioning (45%), to trying to decipher different dashboards showing inconsistent data (41%). Data sprawl is the direct cause of all of these challenges.

As GI Joe once said, knowing is half the battle.

Nearly all IT and LOB executives surveyed (98%) report some challenges with their current approach to data analytics.

Organizations cite the inability to keep pace with business demand their biggest data challenge

What challenges does your company face with your current approach to data analytics?

Challenge	%
Business demand for reporting is more than what our analysts can deliver	48%
Takes too long to move data through to analytics and decisioning	45%
Different dashboards and reports have inconsistent information	41%
Data is missing or inaccessible	38%
Governance policies are not consistently enforced	32%
Data pipelines are fragile and frequently break	31%
Inadequate training and onboarding for new analysts	29%
Performance of dashboards and reporting is very slow	28%
Other	2%
We don't face any data analytics challenges	2%

The good news is that there are solutions out there. Nearly 40% of data leaders said that additional investment in automation could provide a big impact on the business. 35% pointed to more efficient BI and reporting, and modernizing architecture and platforms.
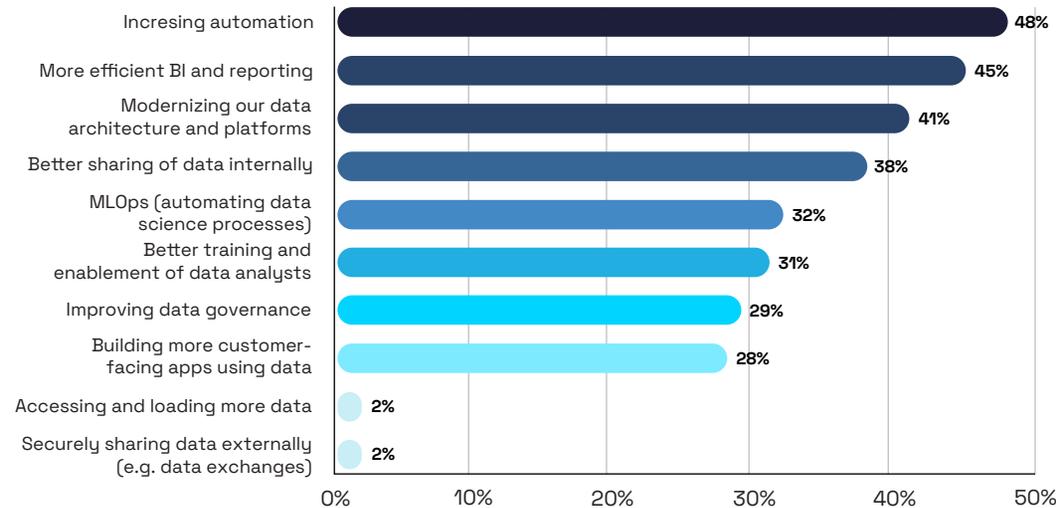
While lower on the list, a quarter of IT and LOB data leaders pointed to better trainig for analysts and automating data science processes through ML. The future depends on the ability of organizations to make the rapidly multiplying firehoses of data work for them instead of overwhelming them.

The good news?
There's an answer – and it works.

Nearly 40% of data leaders said that additional investment in automation could provide a big impact on the business.

Investments in increasing automation would have the most business impact

Where in the data value chain would your company get the
most business impact from increased investment?

Category	Value
Increasing automation	48%
More efficient BI and reporting	45%
Modernizing our data architecture and platforms	41%
Better sharing of data internally	38%
MLOps (automating data science processes)	32%
Better training and enablement of data analysts	31%
Improving data governance	29%
Building more customer-facing apps using data	28%
Accessing and loading more data	2%
Securely sharing data externally (e.g. data exchanges)	2%

0%   10%   20%   30%   40%   50%

How Hakkoda can help?

Many data leaders (and vendors) use data sprawl as an argument for tighter controls. Sure—security, standarization andgovernance are vital.

But tighter controls are not the way to unlock the value of data. Data derives its value from utility.

Companies won't find data enlightenment by making the false choice between control or speed. Modern business demands control and speed.

This is why more than 4,900 organizations have shifted to Snowflake's Data Cloud. Snowflake's underlying architecture and use of cloud technology, along with its ability to bring apps to data rather than data to apps, is a proven, powerful antidote to sprawl. As pure-play Snowflake partner, with expertise across the data value chain, Hakkoda helps organizations address the problem of data sprawl, without choosing between control and speed.

Here are a few ways we can help with the findings in this report:

Manage costs and sprawl:

Our data and cloud architects assess the true cost of your analytic programs and evolve your architecture, governance, and pipelines with efficient modern data practices that stick.

Manage through talent scarcity:

Our subscription-based, Scalable Teams model provides data leaders with a way to fill in the gaps or provide capacity when resources or skills are scarce. Our Snowflake Concierge service provides teams with analyst onboarding, governance, enablement, and training to make data organizations infinitely more productive.

Add automation:

Once we help you get your data house in order, we set out to automate processes for data ingestion, transformation and validation to increase speed and accuracy of data moving to your analytics. Further, our MLOps capabilities navigate the shift from human to machine decisioning.

Bring back innovation:

Our no-code, enterprise-grade applications help organizations start creating again with their data. With Hakkoda customers earn Innovation Credits, which they can use to experiment and test your craziest ideas... on us.

To learn more about how Hakkoda can help you reduce data sprawl and position your company for a smarter future, go to:
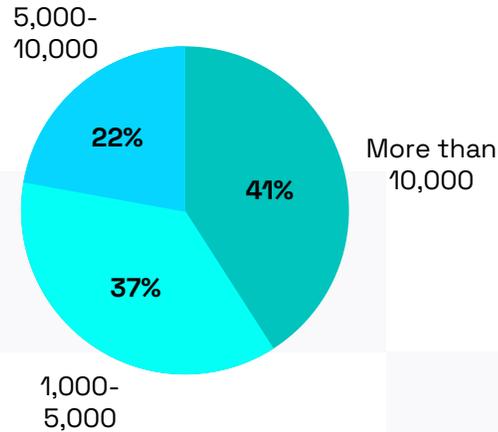
https://hakkoda.io
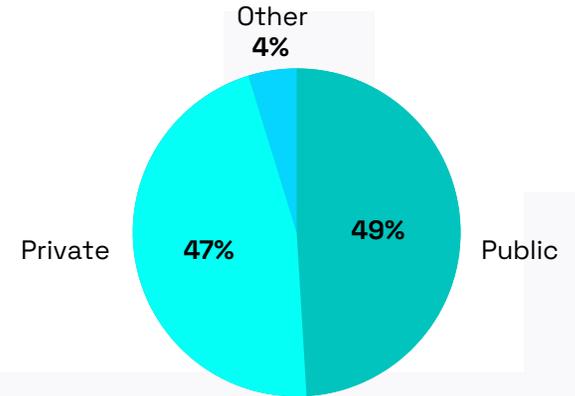
Methodology & Demographics

Hakkoda commissioned a third-party research firm, Dimensional Research, to survey 312 IT and Line-of-Business leaders responsible for data and analytics initiatives about the current state of analytics environments at mid-to-large size companies.

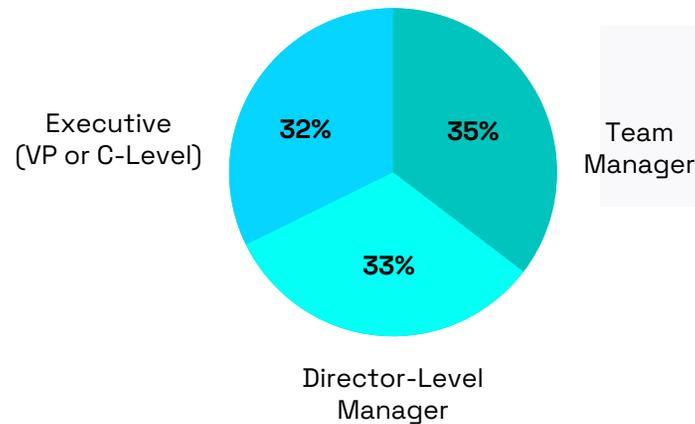Responses were captured between September 17 and September 24, 2021.

Company size (# of employees)

- More than 10,000 — 41%
- 1,000-5,000 — 37%
- 5,000-10,000 — 22%

Company ownership

- Public — 49%
- Private — 47%
- Other — 4%

Job level

- Team Manager — 35%
- Director-Level Manager — 33%
- Executive (VP or C-Level) — 32%

Role

- IT — 67%
- Line-of-Business (LOB) — 33%