IMMUTA®

2023 State of Data Engineering Survey

Uncovering Data Security and Access Blind Spots

Table of Contents

Introduction & Executive Summary

Modern organizations want to be more data-driven than ever before. By attempting to squeeze every ounce of value out of their data, they can more quickly adjust to market conditions, spur innovation, improve customer experiences, and ultimately increase revenue.

It is alarming, then, that our survey of 600 data professionals shows organizations are only using an average of 58% of their data in decision making. The culprit is often data access challenges. And those challenges are impacting the bottom line, with 89% of organizations reporting that they have missed business opportunities because of data access obstacles.

Improving data access and policy management is critical to deriving value from data, but it must be done securely to minimize risk. While the importance of data security is widely understood, most organizations do not place enough emphasis on one of the root causes of data security risk: mismanaged, misconfigured, or simply outdated data access controls. Nearly all survey participants (97%) face challenges with implementing data access controls, with over half (54%) reporting that securing data with appropriate access rights is one of their biggest hurdles.

The challenge is also a human resources one, as organizations struggle to equip data professionals with the skills and tools required to effectively manage access controls. The productivity of data engineers—one of an organization's most valuable resources—is also negatively impacted by poor data access practices. On average, 69% of respondents are spending 6–10 hours per week responding to, managing, and resolving data access issues. Considering the average hourly compensation for a data engineer is $87.98[1], this could add up to nearly $28,000 per data engineer per year in wasted salary costs alone.

On top of this, organizations face a perfect storm swirling around the exponential growth in data volumes and a never-ending stream of privacy and compliance regulation requirements. Data professionals are tasked with balancing data access and security measures with the need to fuel insights that power their business and provide a competitive edge. This survey uncovers some of the top challenges and blind spots these data professionals encounter in the pursuit of helping businesses become more data-driven.

1 Immuta: GigaOm Data Access Control Report: Immuta ABAC vs. Apache Ranger RBAC, July 2021

Key Insights

01

There is a Disconnect Between Data Security and Data Access

Most data professionals report lacking visibility into data access controls and how they correlate with data security. Nine in ten (90%) said that they could improve their understanding of the association between data access and data security.

02

Data Teams Face a Lack of Resources to Manage Data in the Cloud

Data professionals expect that their organizations will house more than two-thirds (68%) of their data in the cloud by 2024, however most do not feel equipped to control it. 41% note that they don't have enough people to manage or analyze their data, while 36% report having too much data.

03

Data Access Challenges Result in Missed Business Opportunities

Almost nine in ten (89%) data professionals report that their organizations have experienced challenges with data access and that their department has missed business opportunities as a result.

04

Policies Are Limiting Organizations' Ability to Scale Data Access

As the amount of data produced continues to grow, so do the number of policies that organizations must implement to manage it — and it's impacting accessibility. More than half (51%) of data professionals claim current data access control policies limit their ability to scale secure data access.

05

Without the Right Resources, Data Teams' Performance is Impacted

Data teams lack the appropriate resources to scale, manage, and implement access control policies. Nearly half (46%) of respondents report that their organization's current data access control policies make it difficult for people to do their jobs.

Methodology

Immuta commissioned independent market research agency Vanson Bourne to conduct the 2023 State of Data Engineering Survey. The study surveyed 600 data professionals from the US, UK, Germany, and Nordics. Respondents represented companies of 500 or more employees across the public and private sectors, including financial services, healthcare, technology, retail, disruption and transport, and manufacturing, among other sectors.

All interviews were conducted using a rigorous multi-level screening process to ensure that only suitable candidates were able to participate.

The Disconnect Between Data Security & Data Access

Managing and automating control of who has access to what data is critical to achieving both compliance and high performance of data-driven processes, such as those for analytics, artificial intelligence (AI), and machine learning (ML). Most (63%) data professionals report lacking visibility into data access controls, highlighting a disconnect between data access and data security that results in coverage blind spots and creates opportunities for exposure. More than half (57%) believe they should be placing emphasis on data security, but only 39% consider data access to be part of data security.

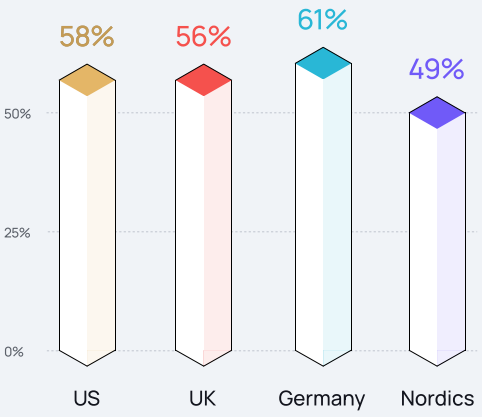
> Nine in ten (90%) data professionals agree that they could improve their understanding of the correlation between data access and data security.

The rapid shift of data from on-premises to the cloud is introducing one of the greatest cybersecurity challenges to date, and the disconnect between data access and data security puts organizations at increased risk for data exposure incidents and breaches. Without considering data access as part of a larger data security strategy, there are more opportunities for unauthorized data access and potential leaks in the data pipeline.

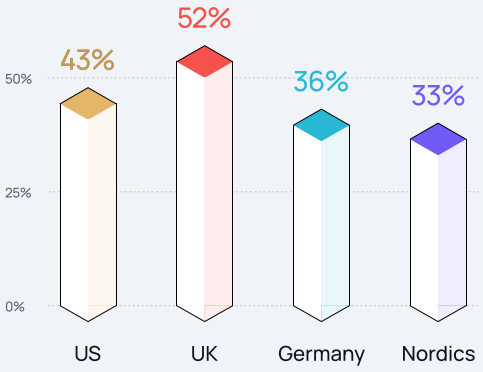
Despite most data teams having a full arsenal of tools for protecting data in the cloud, the proliferation of cloud players such as Snowflake, Databricks, Google BigQuery, Amazon Redshift, and cloud-based SaaS solutions, has created a significantly bigger attack surface. This environment involves more data, more data sources, more data sharing, and an increased risk tolerance. As organizations migrate to the cloud, it is not surprising that over half (54%) of respondents indicate that securing data with appropriate access rights is among their biggest challenges.

When asked about which capabilities they believe their organizations should prioritize to support a modern data architecture, data security rose to the top for respondents across regions:

Data Security



Data Access



As enterprises embark on cloud migrations and adopt modern data stacks, they must ensure customer and mission-critical data remain secure and protected. At the same time, regulatory compliance and data utility must also be considered in order for data to deliver scalable, timely, and valuable insights.

PREDICTION

In 2023, organizations will continue tightening the relationship between product and customer success. This will help mitigate some of the misalignment and ownership gaps around data policy and access by empowering anyone who touches the customer with the ability to make key decisions.

– **Will Rahim,** Chief Customer Officer, Immuta

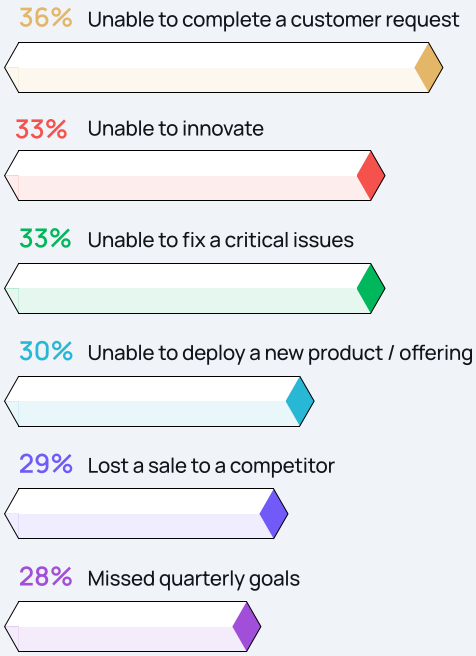
Data Access is a Complex Problem that is Difficult to Scale

The cloud adds another more urgent layer to data access challenges. Many organizations accelerated adoption of and migration to the cloud at the start of the pandemic[2] to accommodate distributed work environments. Now, the decentralized nature of the cloud complicates data access management within these platforms, making it harder to gain visibility into what data exists and who is accessing it.

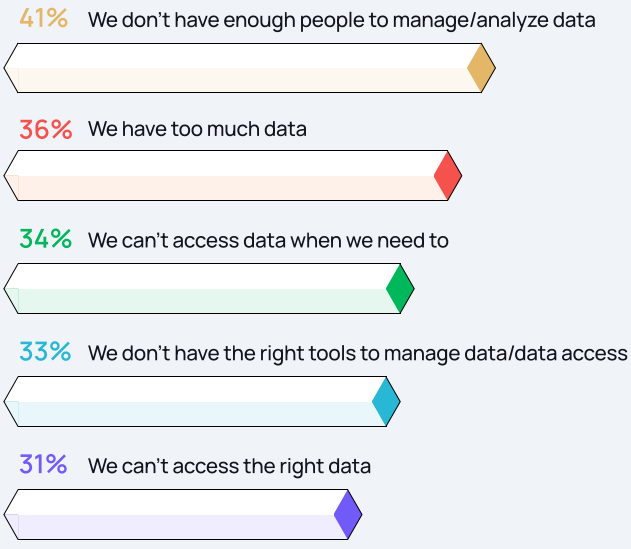
Still, cloud adoption will continue to grow, with data and IT teams expecting their organizations to house more than two-thirds (68%) of their data in the cloud by 2024, and nearly half (46%) expecting to move to a "modern data architecture" for greater flexibility and scalability.

Meanwhile, 41% of data and IT teams are understaffed and don't have enough people to manage or analyze their data. As a result, 36% report having too much data to handle. Additionally, nearly nine in ten respondents (89%) said that their organizations missed business opportunities because of data access challenges, with almost one in four noting that this missed opportunity stemmed from the inability to complete a customer request.

Top Six Missed Business Opportunities Due to Data Access Challenges:



Top Five Challenges Preventing Organizations From Being Able to Utilize Data to Advance Broader Business Decisions:



2 Synergy Research Group: COVID-19 Boosts Cloud Service Spending by $1.5 Billion in the Third Quarter, November 2020

In addition to not having enough people to manage data, many teams lack the right tools to streamline and simplify the management process. Modern cloud data infrastructure from the likes of Snowflake and Databricks have made it much easier for organizations to move data into a data lake or warehouse. As a result, more data is accessible to more users, and data teams can immediately start joining data sets to gather insights. However, this creates a complex and slippery slope when it comes to securing access — 69% of participants report spending an average of 6–10 hours per week responding to, managing, and resolving data access issues. That's 24–40 hours a month, 288–480 hours a year.

Challenges with data access and resulting blind spots are also preventing organizations from keeping pace with growing data volumes, data policies, and cyber threats. More than half (51%) of respondents claim current data access control policies limit the ability to scale secure data access. While this could be overcome by automatically enforcing policies across various data sets and users, only 26% of organizations have fully automated their data access control systems, with 44% relying on native controls. Native controls are notoriously hard to implement and involve lots of manual coding, which likely contributes to the number of hours being spent on data access issues.

It is promising to see that 45% of organizations have a vendor-agnostic approach to data access management; however, 54% still experience difficulties managing cloud data access, and 61% face challenges securing data in the cloud. This is a deadly duo for already overburdened IT, security, and data teams that are responsible for wrangling ever-increasing data streams to power modern analytics strategies, while avoiding running afoul of regulators or suffering compliance penalties and reputational damage.

> The proof is in the data, with 43% of organizations believing data access challenges are slowing down analytics processes.

PREDICTION

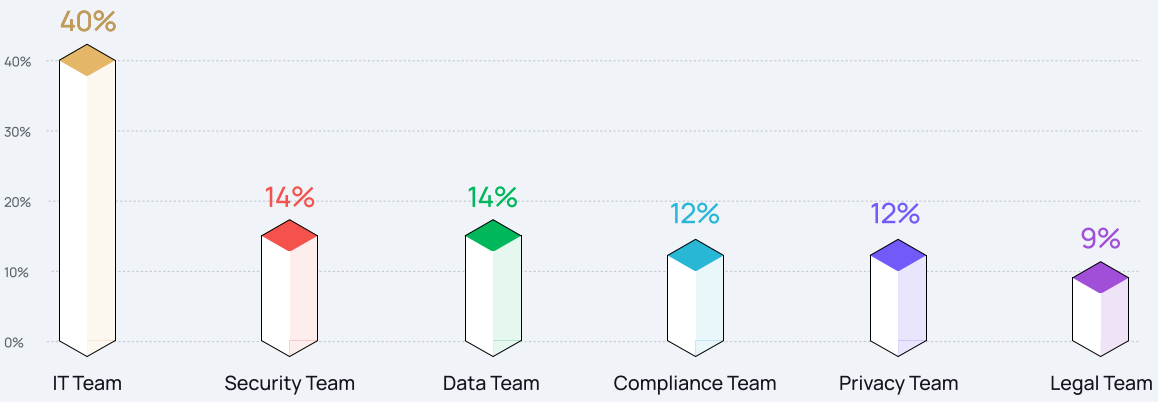
In 2023, as data sharing continues to grow, and data and IT teams are strapped to keep up, no-copy data exchanges will become the new standard. As organizations productize their modern data stacks, there will be an explosion in the size and number of data sets. Making copies before sharing just won't be feasible anymore. With Snowflake's Data Exchange and Databricks' Delta Sharing protocol, enterprises will flock to these exchanges to make it easier to securely share and potentially monetize their data.

– **Matthew Carroll,** Co-Founder & CEO, Immuta

Lack of Ownership for Aligning Data Access Policies with Compliance Regulations

To ensure a successful balance between data access and data security, data policies must align with regulations. However, this task seems to be a hot potato for many organizations, as there is often no clear owner for validating policies across compliance, privacy, and legal teams.

Which Groups Validate Policy Alignment with Regulations:

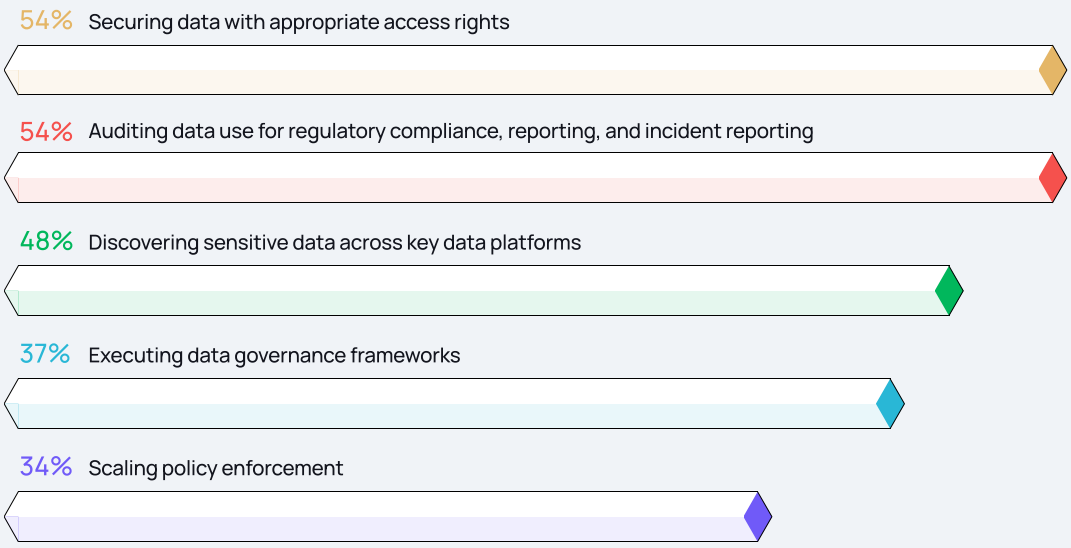


The impact of this lack of ownership is evident in the fact that users cannot get access to the data they need to do their jobs. 47% of respondents say they don't have all the data they need, and the other 53% say they are actually getting over-provisioned access to data. In both scenarios, there is a lack of visibility, as more than half (63%) reported not having complete visibility over who has access to what data. Better visibility into data access and usage is critical for both achieving and proving compliance, as well as uncovering and limiting the impact of potential security breaches.

Still, it's clear that data teams are prioritizing data delivery to gain a competitive edge, instead of focusing on data access and policy enforcement practices that could save them from legal and ethical consequences. A recent study from 451 Research supports this trend, with more than half of survey respondents (52.5%) stating they are primarily motivated to adopt or improve data policy management efforts by the need for business intelligence and data analytics insights, while just 43.5% pointed to compliance as the primary motivator.

When it comes to the biggest challenges data teams face when implementing data access controls for improved data security, securing data with appropriate access rights and auditing its use for regulatory compliance and incident reporting were the top two responses. Clearly, visibility is also critical to overcoming these challenges.

Biggest Challenges Organizations Face When Implementing Data Access Controls for Improved Data Security:



PREDICTION

Getting access to data does not necessarily mean being in a position to derive useful insight. In this data deluge, the successful organizations will be those who will be able to crack the data governance dilemma by leveraging both self-executing policies, such as access control and obfuscation, and auditing capabilities, with a view to reduce time to data. They will discard meaningless pre-approval workflows and federate data governance by making data owners the key players: data owners will be both domain experts and data stewards.

— **Sophie Stalla Bourdillon,** Senior Privacy Counsel and Legal Engineer, Immuta

While Securing Data is Critical, Managing Access Policies is Burning Users Out

The good news is that the individuals handling data every day understand the importance of data security. In fact, almost 9 in 10 (89%) of respondents agree that data security training is critical in advancing their careers. However, this focus on data security is often at odds with the need for data access. Access control policies impact day-to day-responsibilities, with nearly half (46%) of respondents reporting that their organization's current data access control policies make it difficult for people to do their jobs.

Therefore, it's not surprising that although data engineering jobs are among the most sought-after careers in today's employment market, burnout is rampant.

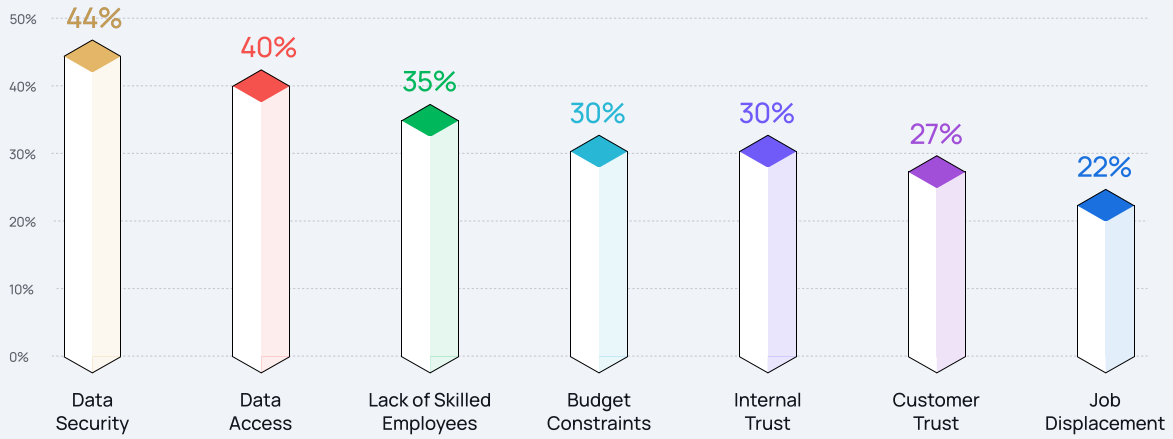
> Nearly 4 in 10 (39%) respondents claim managing data access makes them feel burnt out and consider looking for another job.

Data engineers play a critical role in ensuring that their organizations are harnessing the value of data to make more informed decisions that directly affect business performance. This makes their job satisfaction critical, and organizations should take note of how important data security processes and resources are to improving this standard.

With just under half (49%) of respondents feeling overwhelmed by the number of data policies and regulations they must adhere to, burnout will likely only worsen with the increase of both data volumes and regulations. Another contributing factor to this burnout is the lack of automated data access implementation, with only 26% of respondents employing fully automated systems to improve data access.

Many teams are also implementing AI/ML as critical components of their modern analytics strategy. While a lack of skilled workers was the third biggest challenge that data teams face or may face when implementing strategies around AI/ML, data security and data access were ranked first and second, respectively. This could also be the reason why 57% of respondents claim data access processes are unable to keep pace with the demands of modern analytics. An AI model is only as good as the trust in the underlying data.

Top Challenges Organizations Face in Implementing AL/ML Strategies:



All of this underscores the importance of having the right people with the right skills and resources in place to ensure that data policy management can keep up with demand. It also highlights the delicate overall balance data professionals must strike when managing data access policies — a task that is only getting more difficult as their analytics needs become more sophisticated.

PREDICTION

In 2023, to support data teams in the criticality of these challenges and disconnects, the role of the CISO will shift by becoming the enabler, not the bottleneck, of the modern data stack. As the gravity of data shifts to the cloud and the modern data stack becomes mainstream, balancing security and access will become a major priority. This includes refining security policies with the transparency and automated controls required to confidently secure data, while still providing real-time access to data users across the enterprise.

- **Matthew Carroll,** Co-Founder & CEO, Immuta