WHITE PAPER

Automated Data Access Control 101

New Ways to Automate Data Governance

IMMUTA™

Table of Contents

Why?

When it comes to the enterprise, data science and data governance programs are built on competing objectives, and organizations have no choice but to try to satisfy both demands. Compliance and privacy programs, for example, require increasingly strict limitations on who can use what data and why, thanks to a growing number of rules and regulations dictating data use. On the other hand, data scientists need access to as much data as possible, as quickly as possible — otherwise, analysis could stall entirely.

This is a problem that many businesses simply don't know how to solve.

In today's market, organizations have no choice but to become data-driven. From finance and healthcare to retail and more, enterprises are investing in more data science, machine learning, and advanced analytics to strengthen their competitive edge. Case in point: In a recent survey of data professionals, 72% reported that the number of data consumers in their organization is increasing over time, and 73% said that more data consumers – humans and machines – will need to access and use data in the near future.

At the same time, governments are placing stringent limits on when and how data is collected, used, and deleted to ensure citizens are fully protected. This is true from California to India and the European Union (EU), as we'll discuss below. At the same time, consumers are demanding more guarantees that their data is being used responsibly. Enacting stricter, standardized controls on data across the enterprise – and auditing their effectiveness – is no longer optional.

Any way you look at it, we're at a crossroads: governing data and putting it to use are dueling objectives, and businesses are stuck in the middle.

Can this problem be solved? In a word: yes. This white paper will show you how.

Our founding team spent more than a decade working in the U.S. Intelligence Community, tackling some of the most complex and sensitive data problems in the world. Our team of experts in distributed systems, cryptography, enterprise analytics, and law are committed to enabling the legal and ethical use of data. This paper outlines lessons we've learned about how data science and governance programs can, if implemented properly, reinforce each other's objectives – and why data access control is a critical first step in bringing data initiatives to life.

What's in a Name?

"Data governance" and "data access control" are loaded terms. The two are also often used interchangeably, which only adds to the confusion. Where do we draw the line?

The way around this dilemma is to think about **data governance** as a strategic framework for protecting data, and **data access control** as the enforcement arm of that framework.

Data governance seeks to solves three key issues:

Privacy: Ensures data is put to use in accordance with the rights and expectations of those to whom it belongs.

Security: Safeguards the confidentiality, integrity, and availability of data and the systems in which it's used.

Compliance: Requires aligning how data is used and stored with legal and policy mandates.

Data governance is the process of addressing these three needs, and requires ongoing collaboration across different teams with different sets of expertise.

Data access control, on the other hand, is a subset of data governance. While data governance defines the standards and processes for protecting data, data access control focuses on proactive policy enforcement that selectively restricts or grants access to data. Put simply, data access control puts data governance principles into practice.

The primary goal of effective data access control is to help ensure that the right people are able to access the right data at the right time, and for the right reasons, so organizations can avoid unauthorized access and usage that leads to costly data leaks and breaches. Therefore, without data access control, there is no data governance.

THE OLD APPROACH TO DATA ACCESS CONTROL

A Passive System That Can't Scale

Passive processes hold your data and your business back.

What's a passive process? One that's entirely reactive. In the world of data access control, passive processes involve waiting to evaluate requests for data until they're actually made, and managing those requests manually. Traditional signs of passive processes include time-consuming meetings, long policy memos, custom permissions, policies that vary per database, or the creation of new copies of data to satisfy compliance or privacy concerns.

Here's one way to tell if a passive approach to data access control is plaguing your organization:

How long does it take between:

1) when your organization collects data, and
2) when that data can be accessed and used?

☐ Days? ☐ Weeks? ☐ Months?

For most businesses, the answer is almost always "way too long." Every increment of time that passes between collection and use makes your organization's data less reflective of the present (and therefore less valuable). In fact, research shows that 63% of data consumers say data is stale or outdated by the time it's consumed or analyzed. The more passive processes you use to control data access, the less useful your data is – and the problem will only grow as data use scales.

NEW REGULATIONS

A Forcing Function for Access Control

A passive approach to data access control doesn't work well — it's reactive, time consuming, and inhibitory to the speed and value of enterprise data science initiatives. But for some businesses, being passive has worked "well enough." Why spend time and resources fixing something that's not entirely broken, right? Wrong.

New data privacy regulations are making data access control and management even more complicated. Nearly every day, a new regulation is enacted or proposed that increases penalties associated with poor data protection measures. These regulations go beyond federal laws, and include internal company rules, industry standards, and data use agreements, among others.

By and large, however, headlines about data leaks and breaches involve well-known federal regulations, including:

1. The EU's ***General Data Protection Regulation*** (GDPR), which came into force in May 2018 as the first law in a new wave of global privacy regulations. With fines of up to four percent of global revenue, the GDPR has driven many global companies to rethink how they collect and use their data.
2. The ***California Consumer Privacy Act*** (CCPA) was passed in 2018 and its second iteration, the ***California Privacy Rights Act*** (CPRA), passed just two years later. Frustrated with stalled federal efforts to create a national privacy standard in the U.S., state legislators in California implemented some of the strictest standards on consumer data in the nation, potentially affecting any business that collects California residents' data.
3. The Chinese government has recently ramped up its data protection regulations, after first enacting its ***Cybersecurity Law*** in 2017. The ***Personal Information Protection Law*** (PIPL) passed in 2021 and put tough restrictions on the use of Chinese citizens' data. Potential penalties go beyond even the GDPR – organizations are subject to fines up to five percent of annual revenue for noncompliance.

In the U.S., Congress has proposed dozens of bills in recent years to enforce new national standards for privacy, some of which could cost organizations upwards of $122 Billion USD per year.[1] Hundreds of similar proposals have been made and passed at a state level, from an industry-wide data protection act in Ohio, to restrictions on biometric data in Illinois, to limitations on retail data in New Jersey. Brazil, the eighth largest economy in the world, has also begun enforcing its own version of the GDPR, the ***Lei Geral de Proteção de Dados*** (LGPD). Throw a dart at a map and chances are you'll hit an area where new privacy regulations are making it harder for businesses to collect and use data.

How can data-driven businesses keep up with so many different regulations on data? A passive approach to data access control will quickly break down under these circumstances. Manual, labor-intensive approaches to applying access control policies cannot keep up with a regulatory environment that's increasing in complexity and intensity.

Businesses need a better way – that's where automated data access control comes in.

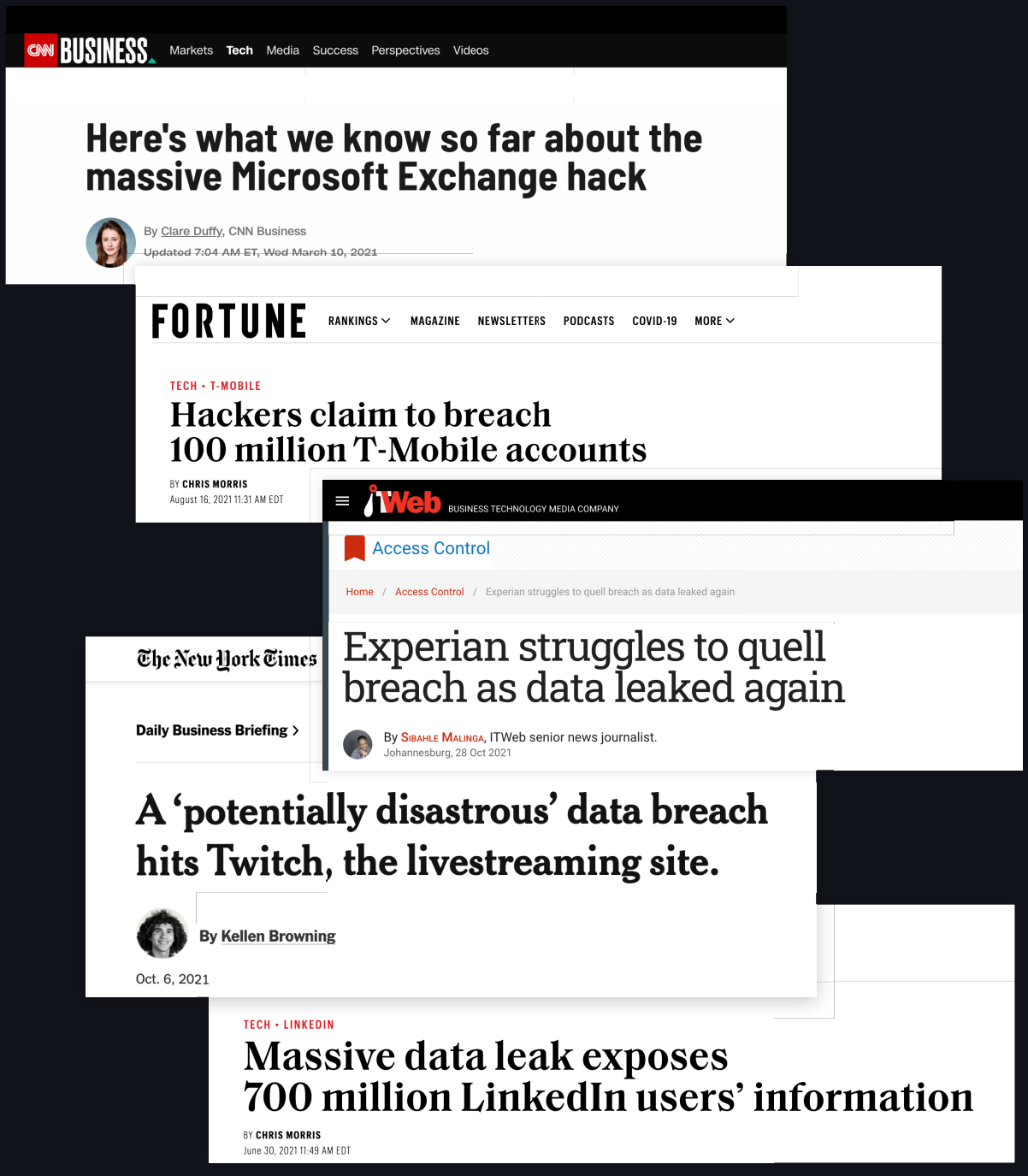
"Cybersecurity Control Failure" Was Named the #1 Executive Concern According to Gartner's Emerging Risks Monitor Report, with 67% of Leaders Citing It As a Primary Concern

Gartner

1 https://itif.org/publications/2019/08/05/costs-unnecessarily-stringent-federal-data-privacy-law

The Consequences of Getting Data Access Control Wrong

Governments are stepping in to ensure that businesses are governing their data correctly, and big tech companies like Facebook and Google aren't the only ones being fined. From large enterprises to small- and mid-sized businesses, companies across all industries and geographies are being held accountable for poor data protection practices.



AUTOMATED DATA ACCESS CONTROL

The Only Way to Control Data Access

How can business replace the old passive approach to data access control? The answer is through automated data access control, which introduces speed, agility, and precision into the process of applying rules on data.

Automated data access control is built on five pillars:

1 Any tool

Data science, analytics, and BI teams should not be forced to use specific tools for data access control purposes – they're too independent and their technology landscape moves too quickly to build governance efforts around any single tool or application. Instead, data consumers should be able to use their tools of choice to access the data they need. As a result, ***automated data access must be able to support any tool a data scientist or analyst uses*** – including tools that haven't yet been developed.

2 Any data

Data comes in all shapes, sizes, and formats, and there's no way to know what type of data a project will need – or even where that data will sit. Some data might be in the cloud, some on-premises, and some projects will require a hybrid approach. ***Automated data access control must be able to support ALL data, regardless of where it's stored or the underlying storage technology.***

3 No copies

A passive approach to data access control often relies on creating new copies of data, usually with sensitive identifiers removed or obscured. New copies of data are given to a group of data scientists, who then have access to that data for a specific period of time. Not only does this passive approach create inefficiencies — between determining what data to copy, how to copy it, and where to store the copy — but it creates new volumes of data that become harder to track over time. Yet again, a passive approach to data can't scale. Instead, ***automated data access control requires direct access to the same live data across the organization*** — data must never be copied for governance purposes.

4 Any level of expertise

Privacy and compliance personnel know the rules that must be applied to data, but they don't necessarily know the technology. (Chances are, they can't explain the difference between an Amazon S3 bucket and a Spark cluster, nor should they have to.) Instead, ***automated data access control requires that anyone, with any level of expertise, can understand what rules (e.g., privacy policies) are being applied to enterprise data.*** Data access control must empower both those with technical skill sets and those with privacy and compliance knowledge, so all teams can play a meaningful role controlling how data is used, without creating bottlenecks.

5 One policy, in one place

Policies can't live in different formats and in different places. It's all too common for different policies to be expressed in different ways, varying by the database and the underlying storage technology. This causes policy bloat and rules that no one organization, team, or employee will understand, let alone realistically manage. Instead, ***automated data access control requires that data privacy policies live in one central location***, so they can be easily tracked, monitored, and updated over time.

> **Without these five pillars, organizations will struggle to break free from the passive approach to data access control. Their governance and data science programs will always be at odds, which can stunt growth, increase risk, and dilute the power of their deepest sources of technical expertise.**

It may sound difficult to combine each requirement into a functional approach to data access control, but it can be done. That's why we founded Immuta.

policy bloat

[**pol**–*uh*–see bloht]

1. The increasing complexity of policies governing data over time. The longer policies have existed and the more updates they have required, the more bloat causes them to be harder to manage and understand.

Our Approach to Automated Data Access Control

Immuta was founded in 2015 by experts in distributed systems, cryptography, enterprise analytics, and law, with decades of combined experience working in the U.S. Intelligence Community – which comprises some of the world's most heavily monitored, secured, and high-stakes data environments.

> **When a mistake could compromise a project — or even a life — speed, efficiency, and compliance mean everything.**

And yet, our team encountered all of the mistakes described above. Passive approaches to data access control are a trap – they're intuitive and easy to fall into, even for the most data-driven organizations. We knew there was a better way – that's why we integrated each of the five pillars of automated data access control into our platform.

How Did We Do It?

We started with an approach to data based on a single point of policy enforcement, meaning that all data policies exist in one place, in one easy-to-manage format that even the least technical personnel can understand. This approach allows any tool to access any data through Immuta, no matter where or how that data is stored. Because Immuta sits between the raw data and the end users, it also means new copies of that data are never required, and that the right policies are applied to the query results in real-time. Immuta's approach to centralizing policy management is critical as organizations increasingly adopt multiple cloud data platforms to power their data initiatives, and it helps enable self-service data use without burdening IT teams.

We also rely on dynamic policy enforcement to grant or restrict access to data. This decouples policy from platform, and means that policies are applied to data as it's accessed by individual users. As a result, users are freed from the passive notion that entire databases must be compliant – Immuta ensures the right policies are applied to the right data at the right time. In practice, a user working in one project might have one view of a data set – sensitive IDs might be completely removed, for example – but that view might change if they switch to another project, which requires different levels of access. In this way, Immuta's highly powerful, granular policies take advantage of context in order to grant appropriate access to data.

PURPOSE CONTROLS

No Automation Without Context

There's no factor more important in automation than context: To apply the right policy without slowing down for manual intervention, rules must take into account who's using what data, for what purpose, and when. That's why purpose restrictions are a central element of Immuta's approach to automated data access control.

With purpose restrictions, governance personnel can create and manage rules based on how and why data is being used, while users must declare the project under which they're using any given data source. Some purposes might require enhanced access and increased monitoring, and others might call for diminished levels. Tying rules to context enables more powerful policies, which leads to more powerful automation.

Immuta policies are built through a ***policy-as-code policy builder***, which means that anyone can understand and create policies on their organization's data, all in one place. Our team of legal and technical experts created a policy engine that can accommodate the strictest requirements of nearly any data rule or regulation — both internal and external – without requiring any technical coding skills. If you can write a memo, you can write a rule in Immuta. And that allows for real, lasting data access control efforts. Expertise among privacy, security, and compliance, and data science teams is no longer siloed — everyone can use Immuta to understand how data is being accessed and controlled.

Finally, Immuta satisfies the need to prove compliance through ***unified audit logs and reports***. Because the platform acts as a single access and control layer to all underlying data, audit logs are standardized across storage technologies and projects, making it easy to build reports and track data activities. This is increasingly important as data regulations become more stringent and widespread.

How Organizations are Succeeding with Automated Data Access Control

Global Financial Institution

One of the world's largest multinational banks has a team of 5,000 data analysts who needed real-time access to data for strategic planning, but manual processes meant they were spending 35% of their time waiting for data access.

Not only did this slow speed to insights and hinder decision-making, but it also led to risky workarounds and friction between teams. After receiving a multimillion dollar fine, the bank couldn't afford to face another data violation.

With Immuta's automated data access control, the bank has:

- Saved more than $50M in DataOps resources as a result of automating 95% of data access control requests and relieving the burden on DataOps teams.
- Allowed the data team to decouple its on-premises data warehouse from users, accelerating cloud adoption across lines of business.
- Scaled self-service data access to more than 5,000 users in just six months.

The Center *for* New Data

To help inform public policy decisions relating to issues like COVID-19 community spread and voting rights, The Center for New Data relies on commercial, donated, and public-access data sets, and a team of volunteer data scientists and academic researchers.

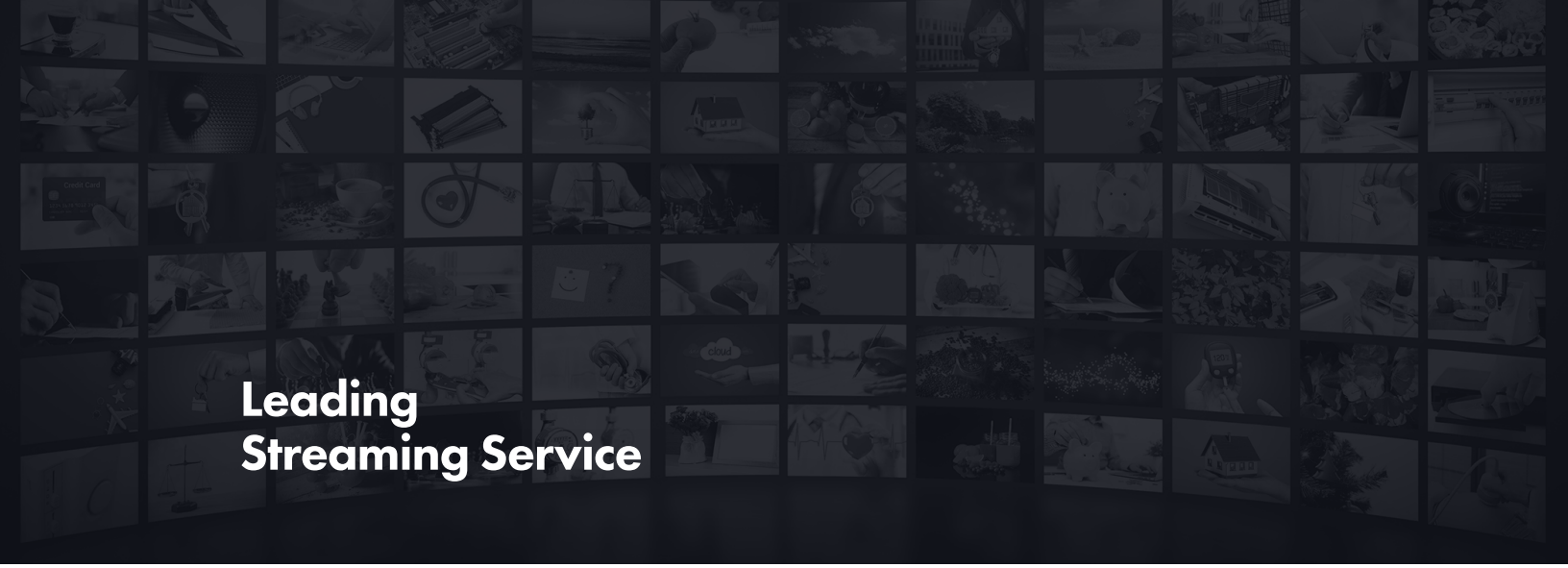
Enabling secure external sensitive data sharing without delaying insights was a critical need – one that could trigger lifesaving public health decisions or stop them in their tracks.

With Immuta's automated data access control, The Center for New Data has:

- Saved more than $1M annually in data engineering costs.
- Simplified the data request process for researchers, reducing the time to data by 30x, from 90 days to 3 days.
- Eliminated the need to manage hundreds of individual data access policies by moving from RBAC to ABAC.
- Reduced the number of access control policies to fewer than 10, which can be easily authored by non-technical data governance experts.

"As we add more components in a cloud database environment, it's much cleaner than any sort of on-prem situation we've had before. The ability for us to manage access controls, deploy privacy-enhancing technologies, and rapidly implement novel frameworks of governance for our research teams has been a breath of fresh air, with no management or overhead costs for adding additional cloud database solutions."

– **Ryan Naughton**, Co-Executive Director & Founder, The Center for New Data



Leading Streaming Service

One of the largest streaming services, with millions of subscribers worldwide, uses customer analytics to generate individualized recommendations that personalize and enhance the customer experience.

But, in a category with several major players, providing a top-notch, trustworthy customer experience is essential to staying competitive. The streaming service's data team needed the ability to collect and analyze subscriber data in real time, while remaining compliant with data use rules and protecting personal data across a diverse data stack.

With Immuta's automated data access control, the streaming service has:

- Increased data engineering productivity by eliminating time-intensive manual processes, allowing data to get into analysts' hands faster.
- Reduced risk of data leaks and breaches by implementing dynamic access controls and advanced privacy-enhancing technologies (PETs).
- Scaled efficient data use to keep up with exponential subscriber growth, which vastly outpaced projected timelines.
- Achieved compliance with data use rules and evolving state and federal regulations, including COPPA, CPRA, and GDPR.

LMĨ

Immuta's automated data access control is an essential component of an integrated data analytics platform designed and managed by LMI for the Office of the Secretary of Defense.

Its Maintenance and Availability Data Warehouse (MADW) contains availability, cost, inventory, and transactional data on nearly every Department of Defense (DoD) weapons system and readiness reportable piece of equipment — more than one billion maintenance records from 46 authoritative data systems. The integration of availability, cost, inventory, maintenance, and supply data makes numerous analyses available to leaders across the DoD enterprise.

"Partnering with Immuta has enabled data scientists to leverage government data in compliance with appropriate governance regulations and privacy safeguards more quickly. Given the size of these data sets, Immuta has been a force multiplier in accelerating the delivery of insight to LMI's clients."

— Joseph Norton, Ph.D., Director of Data Visualization and Product Development, LMI

J.B. Hunt is a leading freight transportation provider, employing more than 1,000 people across North America.

As part of a large-scale cloud migration, its HR team needed to run models to predict resourcing for their fleets. However, J.B. Hunt's compliance team required anonymization in order to leverage the full data sets, in addition to full auditing for every use case deployed. The data team needed a way to meet both requirements while scaling and providing secure self-service data access across its Databricks Lakehouse Platform.

With Immuta's automated data access control, J.B. Hunt has:

- Saved $2.7M in infrastructure and productivity costs, with a projected additional $4.4M in savings over the coming year.
- Accelerated delivery of freight recommendations to drivers 99.8% faster than before.
- Grown permitted cloud data use cases by 400% by leveraging Immuta's native integration with Databricks to enable self-service data access.
- Enabled 100% auditability of every query with Immuta's automated audit logs and reporting.

> "Databricks opens up many opportunities for self-service data analytics, data science, and enterprise reporting. Paired with Immuta, we can make all our data available to all types of business analysts, data scientists, and data engineers."
>
> — **Ajay Sahu**, Director of Enterprise Data Management, J.B. Hunt

To learn more about how automated data access control can help your organization maximize data use, stay ahead of the evolving regulatory landscape, and increase business results, book a capabilities briefing with our team.

IMMUTA™

25 Thomson Place, 4th Floor, Boston, MA 02210 immuta.com (800) 655-0982