CASE STUDY

Mitigating Privacy Risks Within Customer Data

Global Asset Management Company Reduces Millions of Dollars In Privacy Risks And Ensures Data Governance

Global Asset Management Company

$750 Billion+
Assets Under Management

Volume of Data Scanned

2 Terabytes of Data
950 Tables
44 Billion Rows

Protecto Value

Monitored **6 Million** Database Activities

Scanned **25,000** Activities Per Day

Audited **174** User Accounts

Found **652 Tables** of Unused Data

Benefits

Reduced Overhead Costs By **$500K**

Saved **$100MM** In Data Risks

Saved **$50K** In Storage Expenses

Saved **$40K** In Migration Costs

Major Challenges

A global and publicly traded asset management company with over $750 billion in holdings—including real estate, renewable power, infrastructure, credit, and private equity—sought advice and assistance with managing and protecting their extremely valuable customer data.

Specifically, they were looking for a well-developed product or solution that would provide for better enforcement and governance as it related to data access and privacy compliance.

Moreover, they wanted to further improve their data hygiene processes while streamlining the migration of this data to AWS Databricks—a simple, unified data platform that seamlessly integrates with Amazon Web Services (AWS). Basically, Databricks on AWS allows for the cloud storage and management of all data on an open lakehouse platform combining both data warehouses and data lakes to unify all analytics and artificial intelligence (AI) workloads.

Currently, this asset management company was using Azure Databricks as their data lakehouse for storing and processing their customer data. The company's CIO wanted to improve their data hygiene prior to migrating the data to AWS Databricks to help them manage all of their data on a single, open lakehouse platform and unify all analytics and AI workloads. However, with the company's growth came increased scrutiny of data risks prompting the CIO to gain a better understanding of the company's data posture, privacy risks, and to ensure they were meeting all compliance requirements.

Major Challenges

1. **Obtain Insights Into Data Posture**

2. **Ensure Ongoing Compliance**

3. **Improve Data Governance/Hygiene**

4. **Accelerate Data Migration Between Clouds**

The CIO prioritized four specific use cases that any potential solution must properly address: obtain insights into data posture, ensure ongoing compliance, improve data governance and hygiene, and accelerate data migration between clouds.

Obtain Insights Into Data Posture

Here, the CIO wanted to gain a comprehensive understanding of who has access to which data, the frequency of access to data, and what data is most commonly used. The company also wanted to know the extent of risk as measured by which users are executing various queries against their datastores.

Ensure Ongoing Compliance

This would involve obtaining accurate insights on the extent of personally identifiable information (PII) that existed across their entire datastore to ensure governance and controls around who has access to this data. Since this company had a large amount of sensitive data shared across their internal systems, they wanted to tighten controls and be compliant

with requirements such as SOC 2 that defines criteria for managing customer data based on five "trust service principles."

Improve Data Governance and Hygiene

Over its many years of operation, the company prodigiously increased the size of its customer data immensely. Currently, it was sitting on over 2 terabytes of data across several hundred tables in their lakehouse. The company and its CIO wanted to keep a tighter tab on ballooning storage costs. This would entail cleaning up their environment while moving unused data to inexpensive, temporary storage.

Accelerate Data Migration Between Clouds

The company absolutely wanted to prioritize and accelerate the migration of data from Azure Databricks to AWS Databricks. However, privacy and compliance requirements created a bottleneck that was slowing their migration process down significantly and this needed to be immediately addressed.

The Protecto Solution

To address these issues and use cases, the asset management company evaluated several solutions before ultimately choosing Protecto's *Data Posture Intelligence* product. Protecto was selected because it was the only product that looked at all three critical aspects of the challenge: who has access to the data, what data is overexposed/unutilized, and how the data is being used.

Protecto's *Data Posture Intelligence* solution helped this asset management company gain a much deeper

understanding of the data residing within their Databricks lakehouse that contained over 44 billion rows of data across 950 tables. In fact, Protecto monitored over 6 million activities across the company's datastores. This often involved scanning over 25,000 activities every single day. By doing so, Protecto was able to successfully identify all personal data across multiple tables.

As part of the data access audit, 174 user accounts, roles, and access permissions for users were audited. When doing this, Protecto was able to detect that some of the users did not have the appropriate level of roles and permissions. Several users had unnecessary access to all data which only increased security and compliance risks.

Furthermore, by using Protecto, the company was provided with insights into the extent of unused and stale data that currently existed across all datastores. Protecto uncovered 652 tables of data that were unused in the last 90 days. Accordingly, the company decided to move this data to less expensive, temporary storage and lower overall costs. The company estimated that they saved $50,000 as part of their storage costs and reduced their overhead costs by $500,000. Also, by removing this unused and sensitive data that was granted excessive access privileges, the company reduced over $100M in potential data protection risks. By doing so, they were also able to accelerate the generation of reports to establish compliance with SOC 2.

Lastly, the company wanted to migrate data from Azure Databricks to AWS Databricks as part of its overall data

consolidation strategy. Protecto simplified and accelerated the data migration efforts since it made it possible for the company to quickly identify which datastores had the highest number of users—so those data stores were migrated first.

By identifying the top datastores being used and those not being used, the company ultimately saved two months of effort during the migration to AWS Databricks which translated into $40,000 in cost savings.

It should be noted that had the above tasks been performed using a team of data engineers, the company estimated that they would have spent a total of 6-9 months, requiring a team of 3 personnel. Instead, by using Protecto, they were able to accomplish this using less resources in just 2 months, resulting in an overall cost savings of $300,000.

ABOUT PROTECTO

Protecto applies privacy engineering and AI to give instant visibility into privacy risks across data sources, who has access to it, and how it is used. Protecto's Data Posture Intelligence provides actionable insights to reduce privacy compliance efforts from months to days, eliminate millions of dollars in data security risks, and save millions in overall privacy-related overhead costs.

CONTACT US

📞 help@protecto.ai

✉️ +1 (866) 540-0073