Case Study

Discover Data Breach Risks

Industry: Consumer Technology

Silicon valley consumer technology company with over 100M users.

The Problem - Identify breach risks in massive streams of user data

A consumer technology company with a 100M user base had a large volume of user data flowing into a cloud data lake. They wanted to analyze their data to understand breach risks. The data lake contains data streams flowing from millions of devices and consumed by different teams in operations, security, marketing, and finance.

Any breach could result in millions of dollars in penalties and loss of consumer trust. Their analytics team believed their data lake contained only a few tables, but the CTO wanted to understand what personal data they actually hold. Additionally, they wanted to make sure the data is secure and has the right access controls.

100 MILLION USER DATA STREAMS

1+Billion PERSONAL DATA ELEMENTS

The Challenge

The data has been accumulating in a cloud data lake over the years. The analytics team hasn't done an audit on the number of databases and access controls on the data. Any manual audit effort might take several weeks, consuming time and resources. Moreover, with a large volume (TBs) of data, locating personal data was complex and required advanced AI/ML techniques to identify personal data.

Protecto Found PII, Critical Risks and Exceeded Customer's Expectations

The Solution

Protecto offered a turnkey solution that came with pre-built integration to the customer's cloud data source. Within a few hours of the deployment, Protecto's technology analyzed TBs of data to produce the results, which would have taken weeks for the customer's team to analyze.

They were expecting to find ~20 data tables, but our solution found 5700 tables across their data lake. The team was startled to see the amount of data that they had. Notably, 2800 tables of the 5700 tables had personal data on them.

The data analytics team was surprised to see 92% of the data was stale that no one in the company was using. Many of the stale data tables had personally identifiable information, posing a significant risk.

Additionally, many former employees' accounts were still active inside the data lake. Fortunately, there were no breach incidents, but any of these stale accounts could have triggered a breach.

The Results

Protecto saved 100s of hours that would have been spent in analyzing and discovering issues. Protecto monitors their data and provides early warning for any emerging issues.

100%
PERSONAL DATA

Discovered and mapped all the personal data. Continuously monitors the data for emerging risks.

$500K
UNUSED DATA

75% of the stale data, which no one was using, was deleted, reducing the attack surface and cost by $500K/year.

0%
STALE USERS

The IT security team pruned all the inactive users who still had access.

"Protecto gives us peace of mind that our users' personal information is well protected."
- Chief Executive Officer