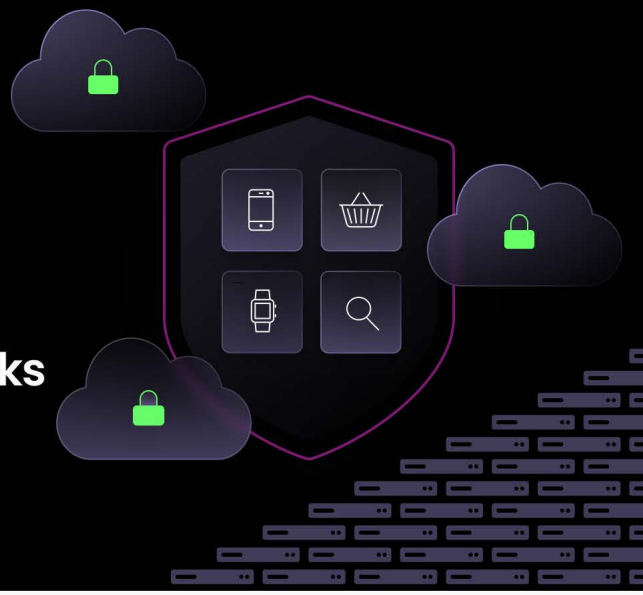
sentra

# Accelerating Cloud Data Classification: How a Consumer App Company Secured Over 130 Petabytes in Just Weeks



A global Consumer App company manages vast, complex cloud environments spanning multiple continents and hundreds of petabytes of sensitive customer and operational data. But their legacy data classification tools were not designed for the massive scale and speed of their cloud data, especially when it came to identifying sensitive information buried deep in complex file formats like JSON and Parquet.

Faced with multiple, complex compliance requirements and ballooning data security costs, the company turned to Sentra.

By adopting Sentra's AI-powered Data Security Posture Management (DSPM) platform, they accelerated and scaled their data security strategy—achieving **98% classification accuracy and full visibility across cloud-scale infrastructure**, and enabling faster compliance — all while reducing operational overhead and **cutting cloud costs**.

## The Challenge: Massive Data, Complex Formats, and Untenable Costs

The data security team's existing classification tools were never built for the scale and complexity of a data estate over 130 petabytes. As regulatory requirements increased, and data structures became more nested and dynamic, manual tagging and legacy solutions became expensive, inaccurate, and unsustainable.

aws PARTNER
- Amazon RDS Ready
- Security Software Competency

The team also faced an immense data security challenge: how to accurately classify sensitive information across an enormous cloud environment, while keeping operational costs in check. Their existing legacy tools lacked the precision and scalability to handle complex, nested file formats like JSON and Parquet, which are common in modern data engineering pipelines. Manual tagging was not only time-consuming but also inaccurate, resulting in low coverage and high compliance risk. With regulatory deadlines rapidly approaching, the security team needed a way to gain complete visibility into sensitive data, improve classification accuracy, and implement a scalable architecture that wouldn't break the budget.

> Our previous solutions simply couldn't keep pace with the sheer volume and complexity of our cloud data. We needed a robust, cloud-native approach that was both effective and economically sound across our entire digital footprint.
>
> — Deputy CISO

## Why Sentra:
## Accuracy and Efficiency at Cloud-Native Scale

After evaluating multiple vendors, the company selected Sentra for its unique combination of deep technical sophistication and practical efficiency.

### What stood out:

- **AI-Driven Classification at Scale:** Sentra's multi-model architecture, including GLiNER for Named Entity Recognition and embedding-based contextual detection, enabled **granular, column-level classification**, even inside deeply nested Parquet structures.
- **Cost-Efficient Ephemeral Scanning:** Unlike always-on tools, Sentra's ephemeral EC2 architecture scales to zero when not scanning. Combined with S3 inventory-based change detection and AI-driven smart sampling, it enables fast classification across hundreds of petabytes, at a fraction of the time and cost, and without impacting performance.
- **Seamless Terraform Deployment:** Rapid deployment via infrastructure-as-code made it easy to scale Sentra across multiple environments while enforcing least-privilege access through dual-role AWS authentication.

Sentra accurately uncovered mislabeled sensitive customer data, enabling rapid validation and remediation. It is now an indispensable element of our data protection strategy allowing us to stay compliant and keep our data protection promise to millions of customers around the world.

— Deputy CISO

## Implementation:
## From POC to Petabyte-Scale Coverage

Sentra was deployed and delivering results in the customer's environment in just 12 days. During the initial proof of concept, the data security team was able to select where they wanted scanning to begin and easily configure the platform, allowing the solution to scan 1 terabyte of high-risk data across complex file formats to achieve over 98% classification accuracy. Sentra's smart sampling approach prioritized the most sensitive and high-impact datasets, optimizing performance without sacrificing precision. The platform was deployed seamlessly using Terraform, integrating directly into the customer's existing AWS architecture. A secure two-role access model, one for metadata access and another for scanning, ensured strict least-privilege control throughout the process.

Following the successful POC, the security team decided to continue scaling Sentra's coverage across their vast data estate to cover  hundreds of petabytes. The data security team was able to easily roll out Sentra according to their data priorities and leverage automation to minimize manual effort and dramatically accelerate risk remediation.

## Real Business Impact:
## Accuracy, Efficiency, and Security by Design

With Sentra, the consumer technology company gained visibility into sensitive data across the cloud estate and fundamentally changed how they govern data and scale their 'Security by Design' culture.

### What stood out:

- **98% Classification Accuracy**
  Unidentified sensitive data reduced by 92%
- **Reduced Over-Tagging and Related Costs**
  Ephemeral architecture and efficient scanning enabled classification at a fraction of the cost compared to other vendors evaluated.
- **Increased Confidence and Precision in Data Governance Enforcement**
  Automated classification replaced manual tagging process and eliminated inconsistencies, over-tagging, and human error.
- **Seamless Alignment and Velocity Across Cross-Functional Teams**
  Sentra's implementation spanned multiple stakeholder groups—including security, privacy, and cloud operations—enabling unified deployment across over 130PB of diverse data classes and formats (personal, financial, proprietary, etc.)
- **Scaled for Long-Term Coverage Across Thousands of Buckets**
  The initial rollout started with targeted high-priority data and a strategic plan to expand classification coverage to thousands of buckets across multiple AWS accounts. Sentra's smart sampling and scalable architecture made it possible to ramp fast without compromising accuracy or performance.
- **Cloud-Scale Deployment**
  AI-driven insights now power DSPM across one of the world's largest cloud-native datasets

## Security by Design at Scale

The consumer app company's journey with Sentra demonstrates how large-scale, cloud-native organizations can protect sensitive data **without sacrificing speed or cost-efficiency**. By pairing **cutting-edge AI** with **ephemeral scanning infrastructure**, they've shifted from reactive compliance to proactive governance.

Together Sentra and 'Security by Design' customers are building the blueprint for a new era of AI-powered classification at massive scale and cost-efficient scanning architecture.

Gartner Peer Insights Customers' Choice 2025

Gartner Peer Insights™

**4.9** ★★★★★

By Sentra in Data Security Posture Management (DSPM)

# Setting a New Standard in Data Security

## >95% Accuracy

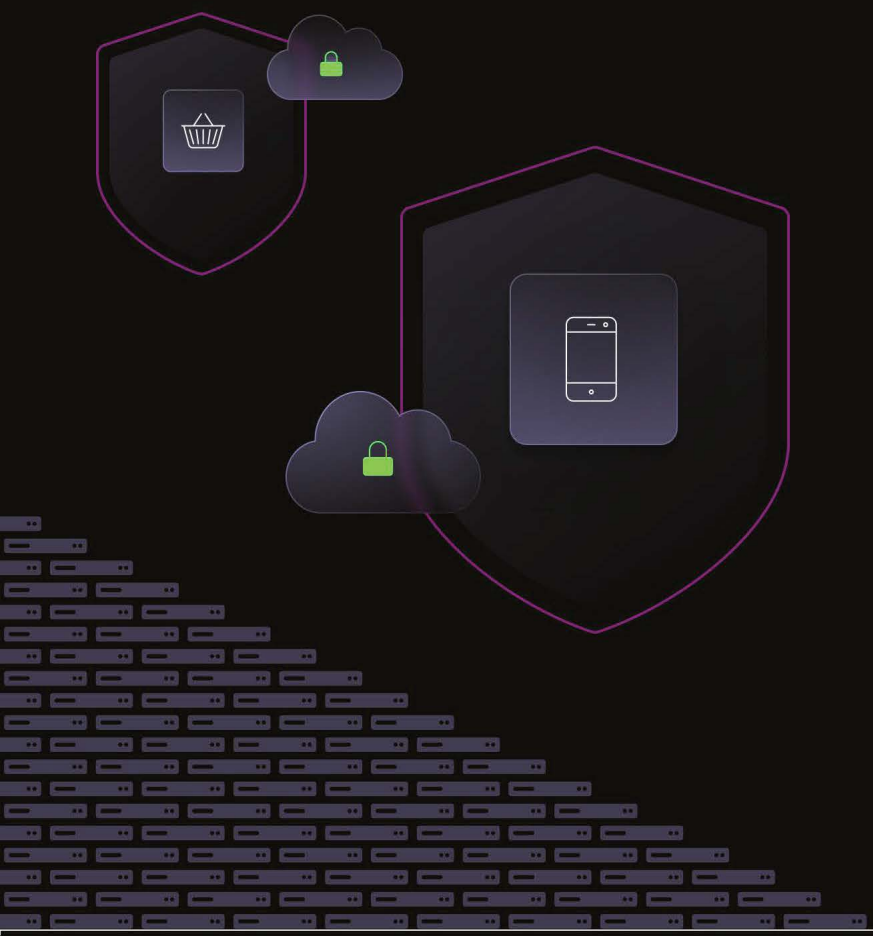
AI-powered classification

## 10x more efficient

In scanning compared to industry

## In less than 1 week

Discover and assess data risks @ PB — scale



available in aws marketplace

sentra

Visit www.sentra.io | Watch a demo

For more information, please contact us at info@sentra.io