STREAMANALYTIX

Real-time Insider Threat Detection using Machine Learning

5x Scope Improvement for a Bank

Insider threats are one of the biggest cybersecurity risks to banks today. These threats are increasingly becoming more frequent, more difficult to detect, and more complicated to prevent. Information security breaches originating within a bank can include: employees mishandling user credentials and account data, lack of system controls, responding to phishing emails, or regulatory violations.

Ignoring these internal security breaches, poses as much risk as an external threat such as hacking, especially in a highly regulated industry like banking.

Identifying and fighting insider threats requires the capability to detect anomalous user behavior immediately and accurately. This detection presents its own set of challenges such as appropriately defining what is normal or malicious behavior, and setting automated preventive controls to curb predicted threats.

About the Customer

A large US-based financial services corporation known for its extensive credit card business

This large bank chose StreamAnalytix to identify and prevent insider information security threats across sensitive applications in its retail banking and wealth management divisions. StreamAnalytix enabled the use of predictive analytics and machine learning on a large data set from highly sensitive applications to automatically and effectively detect previously unknown threat scenarios and patterns and raise appropriate alerts and actions to prevent predicted breaches.

Challenge

Simple rule based alerts proved inadequate for accurate and timely threat detection

The bank's traditional threat detection relied on setting static rule- based alerts on users' activities to define and identify indicators of compromise. But when applied to thousands of users this model generated a high number of irrelevant flags, which resulted in un-timely action on real threats vs. false positives.

Also, savvy attackers went unnoticed by keeping their malicious activities within the defined set of rules.

An expensive and inflexible technology stack limited threat detection to only a few applications, exposing the bank to vulnerabilities

The bank wanted to expand detection of anomalous behavior across all sensitive applications. However, the bank's current relational technology stack was proving to be too expensive and inflexible, limiting the bank to processing data from only 15-20% of hundreds of sensitive customer-facing and operational applications.

The existing solution was taking too long to develop and move use cases into production

The bank was struggling to deploy timely threat detection use cases with its existing solution. It took almost 2 years for the solution to move a single use case to production, making it difficult for the bank to scale out.

The StreamAnalytix Advantage

Solution

Ingestion and data processing from 5x more applications, at a fraction of the cost

The new threat detection application enabled by StreamAnalytix could now ingest data from 80-90% of customer-facing and operational applications. StreamAnalytix used network attached storage systems and Apache Kafka a fast message queue; to ingest data at a ten times lower infrastructure cost and at a speed of 98,000 events per second, four times the speed of the older technology stack.

Data transformation in real-time

In-memory data transformation allows faster data quality scoring, data cleansing, and data enrichment.
The platform enables:

- Real-time data quality scoring and auto-cleansing

- Data deduplication over seven days of history. This helped curb a high number of false positives, narrowing the flags to relevant suspicious behavior and activity

- Enriching event records with employee and application data, such as:
    - First name, last name, employee ID
    - Employee details such as department, role, access permissions, and online activity
    - Details of applications each employee has permission to access

-Executing data transformations such as:
    - Lookup cache and set record type as first name and populate “standard ID”
    - Lookup on, first name/full name/person ID caches to set the user ID
    - Lookup on cache to set user ID
    - Generate unique sequence numbers #
    - Move all the non-schema fields into the EXTRA field

Use of machine learning models on log and complex event data for automated, continuous, and accurate anomaly detection

StreamAnalytix enables the use of machine learning to move away from static rule-based alerts to dynamic models. These models periodically learn normal baseline behavior and detect anomalies based on both dynamic and static factors such as identities, roles, and excess access permissions; correlated with log and event data.

Models developed using built-in machine learning operators in StreamAnalytix include self-learning and training behavioral profile algorithms. This helps in processing each new transaction in real-time to build risk scores and dynamic thresholds for various risk factors. Which leads to automated, accurate, and timely identification of suspicious behavior.

For instance, the algorithm groups users based on roles and access permissions and identifies anomalous activity levels for individual users in relation to what is usual behavior for the whole group. This enables identification of specific people compared to only identifying anomalous events.

This approach proved highly effective in reducing false positives and highlighting behavior that truly accounted for malicious activities; positives were reduced to tens per day as opposed to hundreds or thousands per day.

Custom alerts to curb fraud in real-time

StreamAnalytix enabled appropriate real-time alerts and actions to prevent predicted breaches. These included routine rule-based alerts like: off-hours activity, multiple-failed logins, multi-station logins and custom-alerts for 'suspicious' activity (based on a complex mix of factors deduced by the machine learning algorithms) which could be manually validated by security experts.

Results

5X expansion in scope

The bank went from processing data from 15-20% of applications to 80-90% of critical applications, processing 85M records per day

10x cost reduction

Realized a dramatic cost reduction compared to their traditional RDBMS stack

4x boost in performance

The data throughput went up to 98,000 events per second, four times the speed enabled by the previous technology stack.

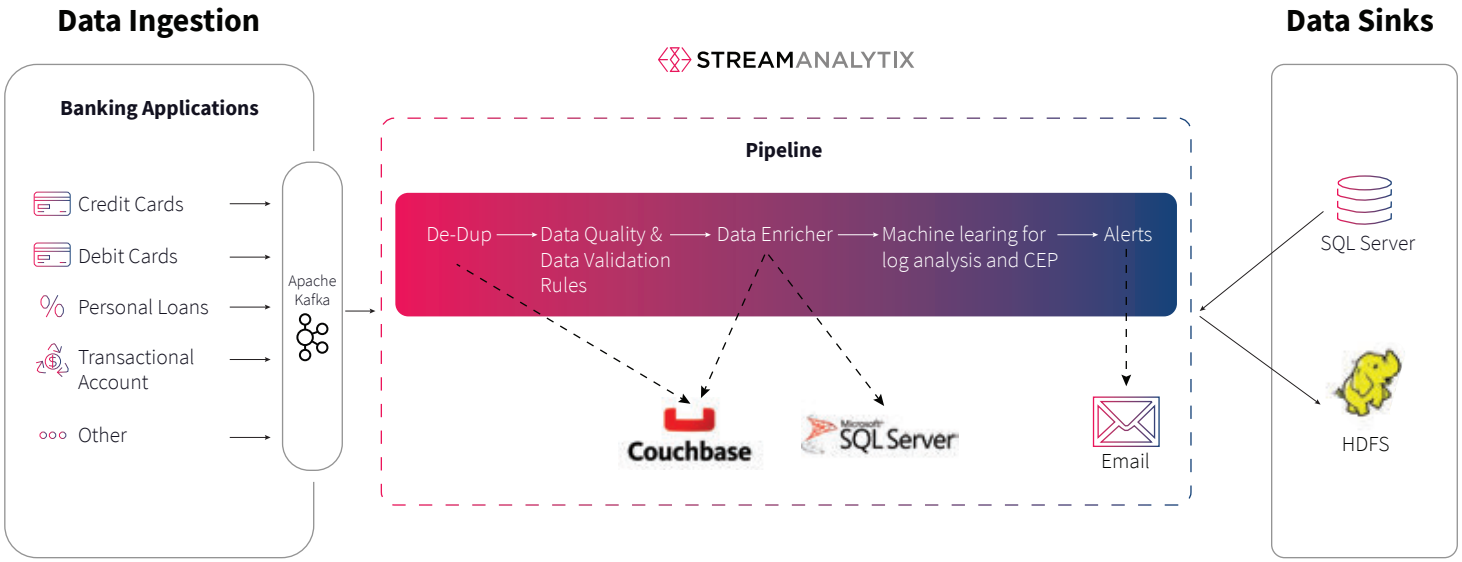
10x faster application development and production

The threat detection application was re-developed in three weeks and moved into production in eight weeks vcompared to nearly 12 months taken by the earlier solution

Enhanced threat detection accuracy and timeliness

Use of machine learning proved highly effective in reducing false positives and highlighting behavior that truly accounted for malicious activities

Insider Threat Detection Solution with StreamAnalytix



Technology Stack

hadoop | apache kafka | APACHE Spark | Couchbase | Microsoft SQL Server

StreamAnalytix is an enterprise grade, visual, big data analytics platform for unified streaming and batch data processing based on best-of-breed open source technologies. It supports the end-to-end functionality of data ingestion, enrichment, machine learning, action triggers, and visualization. StreamAnalytix offers an intuitive drag-and-drop visual interface to build and operationalize big data applications five to ten times faster, across industries, data formats, and use cases.

STREAMANALYTIX

Visit **www.streamanalytix.com** or write to us at **inquiry@streamanalytix.com**