CLOUDERA

CUSTOMER SUCCESS STORY

STREAMING DATA WITH REAL-TIME ANALYTICS IMPROVES CYBERSECURITY DETECTION AND RESPONSE TIME

IMPACT

- Mean time for detection for cybersecurity threats has gone from 70 mins to 7 mins
- Ingesting log data from 130,000 PC's around the world in real-time
- Accelerated search applications by 55%
- 60% log reduction - leading to a $2 million license cost reduction over five years and 30% reduced infrastructure cost

Challenge

This multinational oil and gas corporation had a goal to build a manufacturing data lake, encompassing its considerable refinery, historical, and sensor data. Prior to rolling out this project, the company was unable to get a consolidated view of its operations. Data was evolving at such a high frequency that relational databases couldn't handle it - resulting in a data pipeline challenge.

The manufacturing data lake was needed for the company's log analytics application, used for ingesting log data from multiple environments - everything from personal computer workstations to the cloud. The data lake also needed to generate real-time alerts on events across different teams throughout the organization.

The cost of the log analytics application continued to increase, and it wasn't able to filter, parse, or distribute data across other applications. The initial data lake was built for only one application, and the company needed to be able to reduce licensing and infrastructure cost by moving some data into a less expensive data lake for storage, avoiding being locked into using one application for one purpose. The company determined it needed the proper connectors to collect data from the edge and from different operating systems, such as Windows and Linux. It also needed a data flow pipeline that could process and distribute data to specific applications. While the oil and gas corporation initially tried other technologies as a stopgap solution they soon determined they needed to find the right tool for the job.

Solutions

To address these challenges and gain additional flexibility, the oil and gas corporation installed a hybrid, multi-cloud environment to power analytics in real-time. Establishing a logging data pipeline is what prompted the company to adopt CDP Public Cloud on AWS, replacing Amazon Elastic MapReduce (EMR) with Cloudera Data Engineering (CDE) and Cloudera Data Warehouse (CDW). The company found that CDE autoscaled much faster than EMR, while providing rich insights into application level tuning. By choosing CDP on AWS, the company could utilize Cloudera's NiFi support and expertise. Integration, security, and governance were also huge considerations. The team that was processing and using data from cybersecurity and forensics logs required a pace of innovation much faster than what the on-premises team could provide. NiFi and Kafka solved this data ingestion challenge.

Now the company has assets developed initially for on-premises deployment, with options to deploy anywhere, while keeping a single code based pattern for development. CDP provided ease of use, letting the company avoid having to manage data in multiple locations, while providing consistent security and governance. Some of the current use cases in production include mean time to detection on instrumentation failures, crude erosion, and performance monitoring. Windows workstation logs are being sent directly to CDP no matter where in the world they reside.

90%

Faster response time for cybersecurity threats

About Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.
Learn more at cloudera.com

With this new data architecture, cybersecurity and log analytics can be consumed across the company. Apache MiNiFi is used to capture logs from their multi-cloud environment. MiNiFi proved to be an effective choice for collecting data out of Windows operating systems, while NiFi was used to parse, process and distribute log data from various connectors to different applications. Cloudera Data Science Workbench (CDSW) is used for machine learning, and the corporation will soon be moving to Cloudera Machine Learning (CML) to have the same tools in the cloud. Cloudera's Professional Services played an integral role in helping to establish the data architecture and supported the team throughout.

Results

With the streaming data architecture put into place, the oil and gas corporation is now ingesting log data from 130,000 PC's and all different types of cloud platforms around the world in real-time. This massive and global scalability from multiple environments has created a unified data downstream that many applications can consume for analytics, providing millions in ROI.

Keeping a strong brand is a high priority for the company, and why they will continue to invest in protecting customer data and keeping their operational dataset secure. One critical result achieved through this project is the mean time to detection of cybersecurity threats has gone from seventy minutes down to seven minutes - a 90% faster response time.

"Previously, our resources were constrained and the biggest queries could only happen once per hour. If malware or another threat was detected it was only tied to one host. Now scoring is happening across multiple hosts instead of being strictly siloed per machine. This is increasing visibility across the enterprise in how we view the data. Investigators can now compare across dozens of machines, and increase the priority level as they see fit. Instead of dozens of tickets being raised, it's combined into one and is allowing us to focus our cybersecurity efforts," said a cybersecurity expert from the oil and gas corporation.

Overall, the oil and gas corporation can now secure assets more affirmatively. The original goal was for fifteen minutes, as this enabled an environment in which they could protect their asset base. At the seven minute mark, it's well within the range that if a suspicious behavior is detected it can be reacted to very quickly. Previously, the company was being so inundated with false positives that it was taking too long to address the actual threats. With NiFi, Kafka and accelerated processing, now they are detecting true positives and filtering out the false.

This also allows the cybersecurity team to control where the data flows and stay within their budget. Using NiFi to transform XML data to JSON has accelerated search applications by 55%. Additionally, it has resulted in 60% log reduction - leading to a $2 million license cost reduction over five years and around 30% reduced infrastructure cost. By adding Apache Flink in the near future to create a complete CDF platform, and utilizing Professional Services, the mean time for cybersecurity threat detection will be reduced even further - with the new goal being within 30 seconds.

© 2021 Cloudera, Inc. All rights reserved. Cloudera and the Cloudera logo are trademarks or registered trademarks of Cloudera Inc. in the USA and other countries. All other trademarks are the property of their respective companies. Information is subject to change without notice.