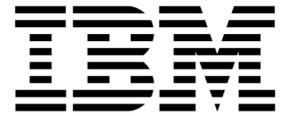


CASE STUDY

IBM: How Self-Service Data Supports Operational Excellence



Overview

With over 400,000 employees across the world, IBM has one of the largest corporate networks in existence. This network acts as a foundation service with many unique and complex technologies running across it. Network health is critical for both ongoing operations and IBM's financial health. Charged with the health of the network, the CIO Network Engineering team delivers automation and tooling services that provide visibility into and reliable operations of the environment. To do this they need continuous, reliable, and transparent operational data that teams around the world can use in real-time.

Challenge

IBM's global network has over 1000 sites in more than 160 countries and processes terabytes worth of data and metrics each day, with some devices generating 10K-12K records per second—that's high volume and high velocity data that the team needs in real time. In addition, the global network is heterogeneous; there are thousands of different hardware and software systems from hundreds of vendors, with hundreds of different data formats.

To parse and get meaning out of all of this network equipment and operations data, the team needs to ingest, transform, and move the data to the nearest data lake. Local teams work with data on regional data lakes to reduce bandwidth costs, while corporate visibility ensures oversight and reporting.

The team uses a lambda architecture, which allows for both batch and stream processing. Initially, they ran an ELT-based log collection platform, using Elasticsearch, Logstash, and Kibana. However, Logstash's code-intensive interface, with no visualization component, made integration with various sources and destinations exceedingly difficult. The team began a POC to look at alternative tools to replace Logstash "because it clearly lacked DataOps capabilities"—in IBM's case, the ability to enable local teams to easily build and manage their own pipelines by operationalizing continuous data management and integration.

Solution

"StreamSets was one of the tools we looked at and, in the end, it naturally filled the space for us. People had already started using it; the adoption rate was very high, and—guess what—they built pipelines which actually

worked. So, it became a de facto ingredient in our DataOps practice,” explained Divya Yashwanth, software developer, CIO network engineering, IBM.

The team is using StreamSets to develop pipelines, create jobs, and make sure the jobs are running on specific StreamSets pods. They have done a lot of work in automating and operationalizing their StreamSets deployment, setting up automatic provisioning agents, pipeline template fragments and topologies, as well as doing StreamSets CI/CD integration with Docker and GitHub.

Result

“StreamSets pipeline template fragments and topologies provide great value in operationalizing the process and enabling teams around the globe.”

Divya Yashwanth, Software Developer, CIO Network Engineering, IBM

By using StreamSets, Divya and team are able to empower colleagues around the world to innovate from the edge. There are no waiting lines for data. Smart data pipelines enable self-service data for everyone. “StreamSets easy-to-use, drag and drop UI helps people with less coding skills gain confidence by allowing them to easily build pipelines and deliver something better than they could with traditional programming.”

The HQ team now has global visibility into 20,000+ data pipelines and billions of streaming records, while local teams can immediately access the data they need to keep local network operations running smoothly.

“We use StreamSets because it’s the only technology that handles volume at scale,” shared Stephan Barabasi, big data, cloud architect, and data scientist, IBM, who has plans to scale StreamSets beyond the CIO Network Engineering team.

ABOUT STREAMSETS

StreamSets built the industry’s first multi-cloud DataOps platform for modern data integration, helping enterprises to continuously flow big, streaming and traditional data to their data science and data analytics applications. For more information, visit www.streamsets.com.