CASE STUDY

BT Group's OpenReach: Democratizing Data to Drive Business Results

BT

Overview

BT Group is one of the world's leading communications services companies and the largest provider of consumer fixed line voice and broadband services in the UK, with operations in 180 countries. Openreach, a subsidiary of BT Group, runs the UK's digital network, connecting homes and businesses, large and small, to the rest of the world.

With the launch of Broadband Britain, an ambitious initiative to build full-fibre connections to 20 million premises by the mid to late 2020s, Openreach realized that to deliver on an undertaking of this scale would require a new way of working. Everyone in the organization would need the ability to make accurate decisions quickly and thousands of resources would need to be allocated with precision. The team decided they could accomplish their goals by upgrading to a modern data architecture. The goal was to move from scattered, siloed data in various on-premises systems to 'one true data source' in the cloud.

Challenge

"What it boiled down to was we needed 'one true data source.'"
Darren Delsol, Client Lead, BT Group

With the pressure on to rollout Broadband Britain, the Openreach team was running into several challenges that would delay their ability to deliver. Multiple versions of the same data from different systems led to incorrect operations reporting and analysis. A lag in data led to resource allocation inefficiency (i.e., without all the data, they couldn't put the right people in the right place, at the right time). In addition, they lacked a 360 degree view of the environment which further impaired decision making. And finally, when they got new data from source systems to improve any of these issues, the delay to delivering a better model could take months due to long release processes.

The BT technology team and their partner TCS determined that all of the challenges amounted to a lack of data availability. "What it boiled down to was we needed 'one true data source,'" said Darren Delsol, Client Lead for the BT technology team working on the Openreach project. In addition, they needed to find a solution that would make data sets available for analysis in a seamless and automated way across the organization. More than that, they wanted a single, zero-coding platform where data engineers, data scientists, and business analysts could all work together. And since they were moving from legacy on-premises systems to the cloud, they also required a solution that

supported both environments, multicloud, and cloud data migration. Finally, coming from a datamart environment where data drift was a consta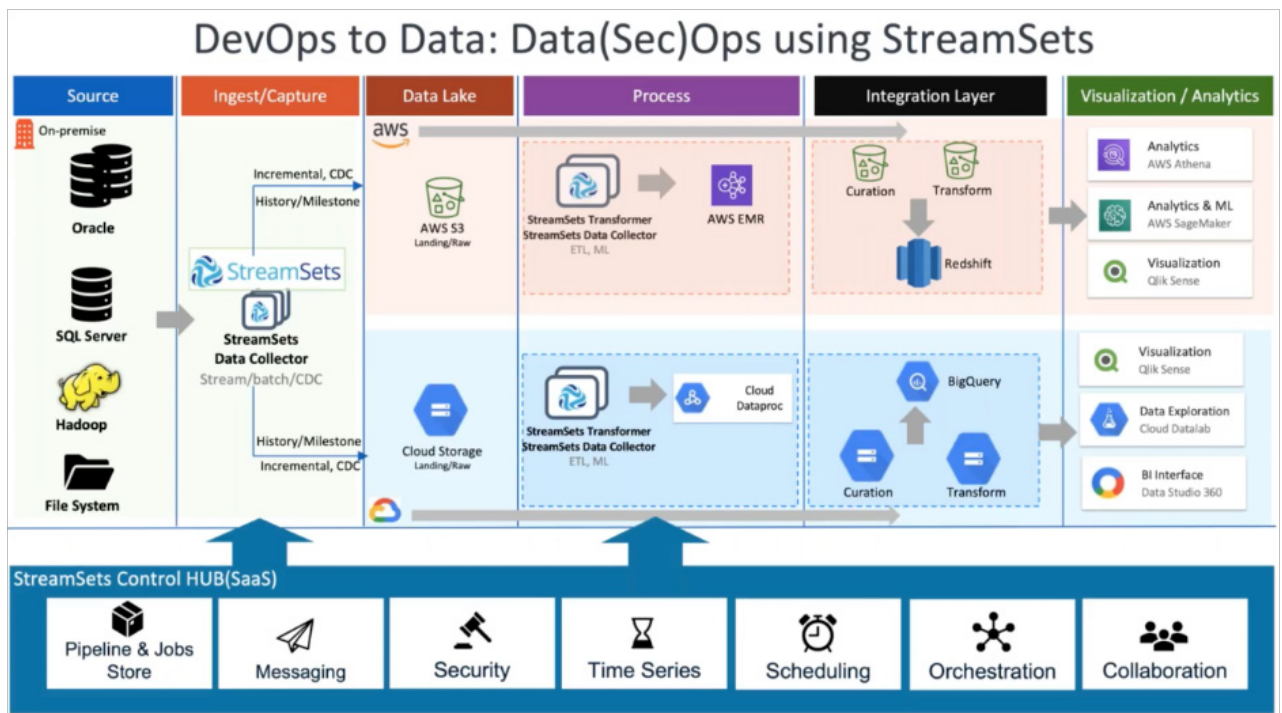nt challenge, they required a solution that could handle those unending, unexpected changes to data structure, semantics, and infrastructure that would be core to a modern data infrastructure.

Solution

> **"We don't go anywhere outside of StreamSets for DataOps, because it has all of the capabilities we need."**
>
> *Anirban Chakraborty, Technical Lead at TCS*

After a brief pilot, BT realized that StreamSets could meet all of their streaming platform requirements and began their journey to a modern data architecture with StreamSets as their data engineering platform.

Data Migration from On-prem Big Data to Cloud Data Lake

First, they migrated on-premises data to an AWS S3 data lake. StreamSets drag-and-drop GUI allowed the BT/TCS team to easily pull data from any form or format into AWS S3, and change sources and destinations easily. "Believe it or not, with StreamSets it takes less than 10 minutes to create a data pipeline to bring your on-premises RDBMS data to AWS S3," said Anirban Chakraborty, technical lead at TCS. They began with Oracle and moved on to their massive Hadoop on-premises system soon after. The team has now migrated more than 20TB of data into their data lake. Anirban shared, "It doesn't matter which format or how big the data is. It's always easy to scale up our StreamSets capabilities to bring the data into our data lake."

Data(Sec)Ops in Practice

Next, they turned to bringing real-time data to end-users. StreamSets made it easy to

create a single pipeline from on-premises to the integration layer where data can be exploited, or visualized or analyzed. That smart data pipeline can be published, shared, and reused to democratize streaming data use through the management, administration, and orchestration layer.

BT uses the StreamSets platform as the basis for their Data(Sec)Ops practice. The StreamSets platform allows for role-based use, reuse, and collaboration on data pipelines and full monitoring of all data pipelines across on-prem, AWS or GCP—as well as data drift handling. "We don't go anywhere outside of StreamSets for DataOps, because it has all of the capabilities we need," said Anirban Chakraborty.

Multi-cloud Smart Data Pipelines

As the BT team completed their data lake migration to S3, they realized Google Cloud Platform (GCP) offered advantages for certain workloads. Instead of doubling the number of pipelines to manage and maintain or investing in months of change management, they simply added GCP as a new destination, extending the existing smart data pipeline. Because StreamSets data pipelines are decoupled from the architecture, stages can be added without pausing dataflow. Aniran explains it like this: "So StreamSets will multiplex the data on both a single source, and to both destinations—one on AWS S3, the other Google Cloud Storage.

We now have data lakes on both sides at the same time, with minimal effort and minimal disruption to data pipelines."

Results

"StreamSets gives us the ability to democratize data which has led to an explosion of user adoption, excitement around data that we've never seen before, and real business results."
Darren Delsol, Client Lead, BT Group

BT credits StreamSets with their ability to go quickly into multi-cloud, and test the full capability of their applications. That ease of use has real implications for the business. The faster data consumers can build pipelines to get the data they need, the better they are able to utilize data to drive business results.

Since StreamSets went live, user adoption has exploded from just 23 data engineers to 250+ users across job functions. "We were able to democratize the platform because we are able to create guardrails," said Anirban, "[Everyone can] bring their data into the data lake, explore that, create visualizations, consume it, and bring value."

The team is running over 7000 pipelines, and is excited about the 450 fragments—reusable components that can be shared with others to plug into any pipeline—built to date.

ABOUT STREAMSETS

At StreamSets, our mission is to make data engineering teams wildly successful. Only StreamSets offers a platform dedicated to building the smart data pipelines needed to power DataOps across hybrid and multi-cloud architectures. That's why the largest companies in the world trust StreamSets to power millions of data pipelines for modern business intelligence, data science, and AI/ML. With StreamSets, data engineers spend less time fixing and more time doing.
To learn more, visit www.streamsets.com.