

CASE STUDY

SamTrans: Driving Better Decision-Making with StreamSets and AWS



OVERVIEW

SamTrans, the San Mateo County Transit District, is a public transportation agency in and around San Mateo, California, in the San Francisco Bay Area. It provides bus service throughout San Mateo County and portions of San Francisco and Palo Alto. In addition to fixed-route bus and paratransit operations, the district participates in administering the San Jose-San Francisco commuter rail line Caltrain. The IT team uses data to understand ridership trends to best serve their customers.

CHALLENGE

SamTrans IT spans several teams working in a decentralized ecosystem. Team member skill sets, use cases, data sources, and infrastructure vary widely between agencies, departments, and vendor solutions. This led to disparate processes for integrating data that were bespoke, non-sharable, and prone to failures.

When SamTrans initiated a Business Intelligence reporting project to better understand ridership trends, the team knew they needed to find a way to replace distributed and isolated data with more streamlined processes. This would allow for more self-service and rapid development of data analytics.

SOLUTION

The first step SamTrans IT took in supporting the decentralized organization and streamlining processes was finding a common data integration platform that was easy-to-use and could support any data source and infrastructure environment —on-premises, hybrid, or in the cloud.

They chose StreamSets for its ability to support their multiple on-premise source environments (Oracle, SQL Server, CSV, etc.) and cloud destinations like PostgreSQL and AWS S3. Using StreamSets, they created multiple smart data pipelines that gathered data from previously isolated sources and landed it reliably and continually in S3 and AWS PostgreSQL.

RESULTS

“With StreamSets and AWS, we’re able to consolidate distributed and isolated data and present it in a dashboard for better decision-making. This data became especially critical during the Covid period, as it impacted overall ridership.”

- Mehul Kumar, Executive IT Officer

Where it used to be impossible to manage and ingest data from various siloes into the cloud, that’s no longer the case. Kumar credits StreamSets with helping SamTrans embark on their DataOps journey. “Having one unified tool with a graphical UI that lets us easily interact with and manage multiple sources and destinations is a big step in the process of streamlining.”

It’s also aided in feeding data to SamTrans AWS partners in real-time, especially around ridership for management decision-making around budgets and lowering costs. And as the team churns high volumes of data, they’ve begun introducing a new cloud data warehouse and will subsequently be venturing into AI-related technologies when they move on to analytics.

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

