



PARTNER BRIEF

Data Integration without Compromise with the AWS ecosystem

StreamSets delivers enterprise-proven capabilities for modern data integration as organizations migrate to the AWS cloud for cost effective, scalable on-demand processing and storage.

Traditional data pipelines break when the unexpected happens, and they are hard to move to new data processing and cloud platforms without complex refactoring.

StreamSets enables continuous data with centralized management and monitoring for all your data, whether you're in the cloud, on-premises, or in hybrid environments. StreamSets orchestrates pipelines for moving data into AWS: Whether data is going into the storage, processing, or consumption layers, StreamSets captures the metadata while maintaining an audit trail. By capturing data that is ingested into AWS, StreamSets enables organizations across a variety of use cases that include enterprise analytics, customer 360, fraud detection, drug research & discovery, and many others.

StreamSets is proud to be an AWS Data & Analytics Competency holder and Advanced Technology

Partner. That means StreamSets has proven technology and customer success with AWS by providing native integration with AWS Aurora, Amazon Redshift, Amazon Kinesis, Amazon S3, Amazon EMR, and Amazon Serverless EMR. StreamSets' smart data pipelines detect and handle schema and data changes* to prevent data loss and corruption in your data landscape and reporting environment.

The Advanced Technology Partner ranking demonstrates StreamSets' ability to solve real-world use cases and challenges in enterprise data management. StreamSets is the only data integration platform that brings enterprise-proven capabilities to modern hybrid and multi-cloud data architectures.

Accelerate AWS Use Cases



Migrate from On-prem to AWS

Simplify your migration to AWS and keep your environments in sync using StreamSets' pre-built connectivity to 100s of data sources, powerful data

transformation, change data capture (CDC), and an operations console to view all data movement. This enables you to cleanly perform your migration without disrupting your overall business operations.



Analytics with AWS

StreamSets provides a single, easy-to-use platform to integrate unstructured, semi-structured, and multi-structured data to Amazon Redshift, Amazon Kinesis, and Amazon S3 using both synchronous and asynchronous ingestion methods.

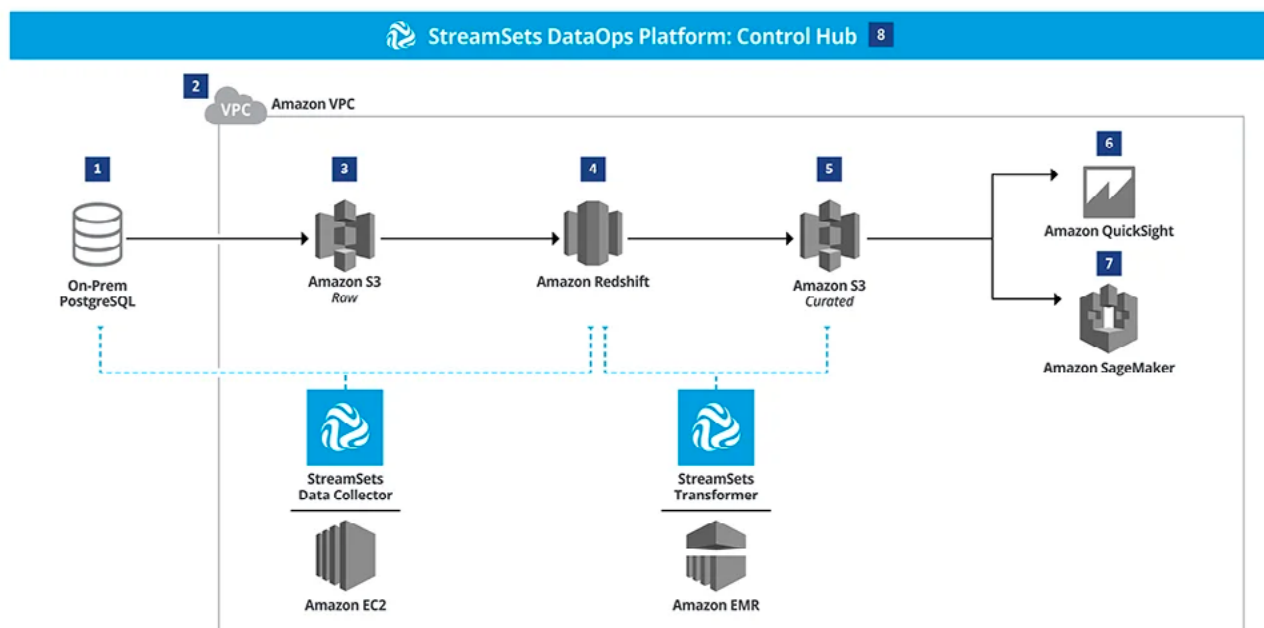


Extend to Databricks and Snowflake

Leverage Databricks and Delta Lake or setup a data warehouse and data marts with Snowflake. StreamSets manages the data into Amazon S3 where users can load into Amazon Redshift, Amazon EMR, and more to perform analytics, ETL, and data science. Connect those services to platforms hosted on AWS.

Reference Architecture:

Integrating Data from a PostgreSQL Database to AWS S3 and Redshift



Overview

The key data source is a PostgreSQL database. StreamSets Platform's Data Collector engine, which is deployed on a VPC in AWS, moves the data from PostgreSQL into Amazon S3 as a staging area and then subsequently to Amazon Redshift. Redshift supports high availability and

high volume analytical workloads. The StreamSets Transformer engine is used to further cleanse and curate the data utilizing Amazon EMR. The cleansed and curated data is staged in Amazon S3 and then available for analytics and data science using tools such as Quicksight and Sagemaker.

Benefits

Detect and Respond to Data Drift

Traditional data pipelines break when the unexpected happens, and they are hard to move to new data processing and cloud platforms without complex refactoring. Only StreamSets Platform features smart data pipelines with built-in data drift detection and handling, in a hybrid cloud architecture, so that your operations run smoothly despite constant change.

Design-Deploy-Operate Continuously

In a continuous data world, operations are everything. StreamSets runs natively in AWS so you can design, deploy, and operate your pipelines all in the cloud. StreamSets monitors data in flight to detect changes and predicts downstream issues to ensure continuous data delivery without errors or data loss as part of your modern data integration solution.

Go Fast and Be Confident

When your business moves fast on a traditional architecture, things break. But when you take your time, you fall behind. StreamSets Platform on AWS gives you end-to-end transparency across your data infrastructure, so you can detect emergent patterns and designs.

ABOUT STREAMSETS

At StreamSets, a Software AG company, our mission is to ensure data engineering teams thrive in today's world of constant change. The StreamSets Platform empowers engineers to build and run the smart data pipelines needed to power data integration 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 analytics, AI/ML, smart applications and hybrid integration.

To learn more, visit www.streamsets.com

