

STREAMSETS FOR SNOWFLAKE

The StreamSets for Snowflake connector delivers streaming, batch, and change data capture (CDC) for the Snowflake cloud data warehouse.

DataOps

A single visual UI for designing all data ingestion pipelines, covering a wide variety of on-premises and cloud data sources, including comprehensive monitoring, alerting, automation, and handling of data drift.

Change Data Capture (CDC)

Enables CDC for continually syncing RDBMS origins (such as Oracle, Teradata, SQL) with Snowflake, with easy setup and monitoring.

Rich Transformations

Perform numerous pre-built or custom data transformations on data-in-motion.

Automatic Multi-Table Creation and Inserts

Automatic table and multi-table inserts to manage complex evolving fields, along with full table creation for migrations.

High-Performance Ingest

Faster record throughput than any solution in market for both synchronous and asynchronous workloads, regardless of schema.

Data Privacy

Policy-based detection and protection of sensitive data before it lands in the Snowflake data warehouse.

StreamSets for Snowflake Data Warehouse

Overview

As analytics moves to the cloud, a fast and scalable cloud data warehouse becomes essential. Even with the easiest-to-use cloud platform, however, users often struggle to ingest and move data from their systems to a cloud data warehouse. What's needed is a combined solution providing self-service business intelligence (BI) and a comprehensive DataOps platform for development simplicity and management control of all data pipelines.

Challenges

It's gotten to the point where many organizations are asking if EDW (enterprise data warehouse) is a 4-letter word. High costs, low performance, limited capacity, and poor support for unstructured data make these systems a headache to manage. Traditional EDW approaches also suffer from poor scalability, hampering data science and precluding advanced analytics. To top it off, these appliances become out of date almost as soon as they are deployed.

Thanks to the ease of use and deployment of the Snowflake cloud platform, users of the Snowflake cloud data warehouse enjoy a welcome reprieve from the difficulties of legacy EDW platforms. Even so, getting data into the cloud data warehouse can remain a frustrating and time-consuming process.

To get the most from their cloud data warehouses, companies need to:

- Develop high performance ingest capabilities.
- Enhance DataOps visibility and control of data pipelines.
- Automate bulk uploads and multi-table updates.

Ingesting, moving, and handling data, especially streaming data, with Snowflake and other cloud data warehouses is challenging because:

- Performance is often a limiting factor in successful workflow design.
- Not all data is in the cloud or in the cloud data warehouse.
- Single-use tools for cloud solutions often leave gaps in management, visibility, and security when applied to hybrid architectures.

To maximize the value of cloud data warehouses such as Snowflake you must design a system that delivers all the data into the platform; easily, securely and continuously.

KEY POINTS

- **Build high-performance dataflows from varied sources**
- **Gain visibility and control**
- **Handle change with style and grace**
- **Integrate Snowflake into the larger data architecture**
- **Create at-scale data pipelines for streaming data**
- **Build with ease, plus rich transformations**
- **Achieve central control across hybrid cloud architectures**

ABOUT STREAMSETS

StreamSets transforms how enterprises flow big and fast data from myriad sources into data centers and cloud analytics platforms. Its DataOps platform helps companies build and operate continuous dataflow topologies, combining award-winning open source data movement software with a cloud-native Control Hub. Enterprises use StreamSets to enable cloud analytics, data lakes, Apache Kafka, IoT, and cybersecurity.

Founded by Girish Pancha, former chief product officer of Informatica, and Arvind Prabhakar, a former engineering leader at Cloudera, StreamSets is backed by top-tier Silicon Valley venture capital firms, including Battery Ventures, New Enterprise Associates (NEA), and Accel Partners.

For more information, visit streamsets.com

Solution

StreamSets helps organizations build a DataOps practice that manages delivery and performance of data pipelines on premises, across public clouds, and with managed services such as the Snowflake cloud data warehouse.

Develop high performance ingest capabilities. StreamSets simplifies the building and operation of pipelines from myriad on-prem and cloud data sources into the cloud data warehouse. As a result, data engineers become more productive and gain continuous visibility and control over their data availability, integrity, and protection. StreamSets users will see increased and consistent performance for both synchronous and asynchronous workloads.

Enhance DataOps agility and control. With StreamSets, users take a centralized approach to the deployment, management, and monitoring of their data pipelines, spanning workloads and requirements—so they can remain collaborative and agile when working with specific cloud services.

Automate bulk uploads and multi-table updates. By automatically creating a table or multiple tables, StreamSets makes ingestion into a cloud data warehouse resilient to shifting changes in the table structure. That way, users can focus on getting data into the cloud data warehouse rather than worrying about the schema and structure.

StreamSets Benefits

StreamSets enables organizations working with cloud data warehouses such as Snowflake to:

- Quickly stream all types of data into a cloud data warehouse.
- Ingest data faster, including 'complex' files types such as JSON blobs and semi-structured data.
- Get end-to-end performance monitoring, performance alerts, and data protection.
- Move and share data into and within cloud data warehouses beyond the requirements of specific workloads such as data marts.
- Identify new tables that are created, and edit the upstream system to accept the data.
- Support multi-table creation, which is useful when the database structure is unknown.
- Perform in-line data transformations, automatic table creation, and multi-table inserts.

Closing

StreamSets amplifies the power of cloud data warehouses such as Snowflake by simplifying and automating the process of getting both structured and unstructured data into the cloud data warehouse platform—so analytics experts, data engineers, SQL developers, enterprise architects, and other users can concentrate on how to best use that data.

Find out more about how StreamSets can take the hassle out of data ingestion into cloud data warehouses. [Contact a StreamSets representative today.](#)