



CASE STUDY:

Simplifying Data Transformations in Snowflake for Analytics

Challenge

StreamSets was transitioning from an open source infrastructure to the cloud and faced several challenges in our data integration and analytics processes. The existing relational database management system (RDBMS), managed by the DevOps team, was not suited for large volume analytical workloads or multiple data formats in one place.

Data sources were also limited as ingesting high-volume data sources, like cloud product telemetry and web interactions, can be costly and create database performance concerns. And since all database changes had to go through the busy DevOps team, the Analytics team was often delayed in responding to rapidly evolving business needs.

On top of these challenges, StreamSets was gearing up to launch a new 30-day platform trial but lacked the flexibility and sophistication required in our analytics capabilities. The Analytics team needed to understand how trial users were finding value in the product to provide timely and reliable insights across multiple business functions. The team also needed to provide insights on how product trial activity was creating or accelerating sales pipeline. There was also the challenge of not being able to instantly match trial user emails to corporate organizations, creating a manual process and bottlenecking the Marketing Operations team.

Finally, while the existing system allowed for effective historical trend reporting, the Analytics team needed to be able to better drill into those trends — moving



from “what happened” to “why it happened” and “what to do next. Line of business users needed actionable insights in real-time.

With one challenge piling on top of another, StreamSets decided it was time for a change to allow the Analytics team flexible self-service access to more data.

Solution

StreamSets decided to transition to the Snowflake Data Cloud for its flexibility, scalability, efficiency, and self-service capabilities. With Snowflake and StreamSets Data Collector, Software AG webMethods.io, and StreamSets Transformer for Snowflake for data integration and transformation, the Analytics team could overcome all of the challenges outlined and meet their trial analytics goals.

StreamSets migrated common data sets from the MySQL database to Snowflake using StreamSets Data Collector and Software AG webMethods.io. We ingested new data from sources like SDK telemetry data generated in Google. Using StreamSets Data Collector processors, the Analytics team created an architecture model to move terabytes of data into Snowflake.

Leveraging Snowflake native functionality, Python User-Defined Functions, and Transformer for Snowflake allowed the team to filter relevant data, enrich additional metadata, and add multiple dimensions and aggregations to create standardized

data models. The team then loaded standardized data models into dedicated Snowflake databases for core groups such as Product, Engineering, and Marketing.

The Analytics team also built a data model for the 30-day product trial using Transformer for Snowflake, merging product and marketing data and making crucial user information available to the core groups. To digitize the manual process of matching user emails to corporate organizations, Engineering wrote a program using Python User Defined Functions to reference external data sources to screen, filter, and match email domains. The Marketing Operations team can now easily and efficiently manage the data and enhance Salesforce with information leveraged by Sales and Marketing.

With product data now integrated into a Snowflake Table, the Marketing Operations team was also able to leverage the transformed data to identify meaningful product actions or events using

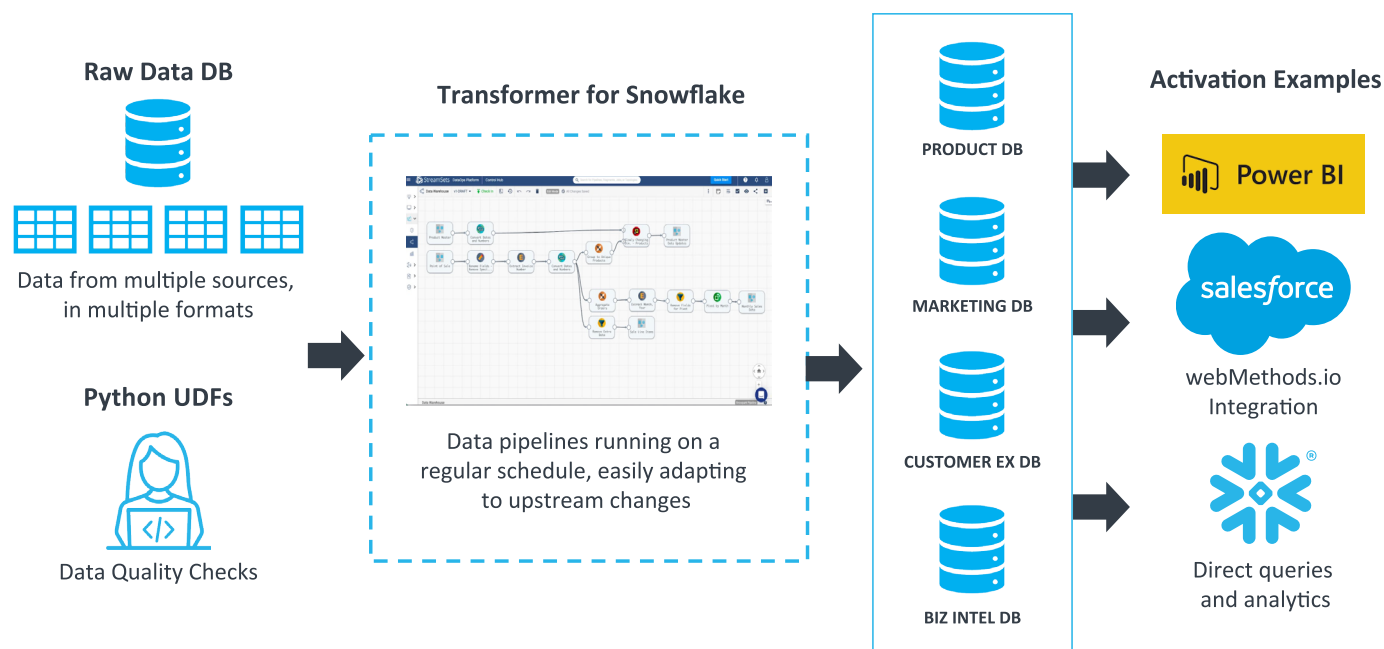
SoftwareAG webMethods.io. This enabled the team to map Snowflake fields to their respective Salesforce fields and push the transformed data into Salesforce. Product activities are now linked to Leads and Contacts and matched to Accounts in Salesforce. This gave StreamSets visibility into how many activities an organization has completed within the StreamSets Platform, which defines a Product Qualified Account (PQA) and a 30-day trial user's willingness to purchase. Sales and Marketing can now understand and use product engagement data through business applications such as Salesforce and DemandBase.

Results

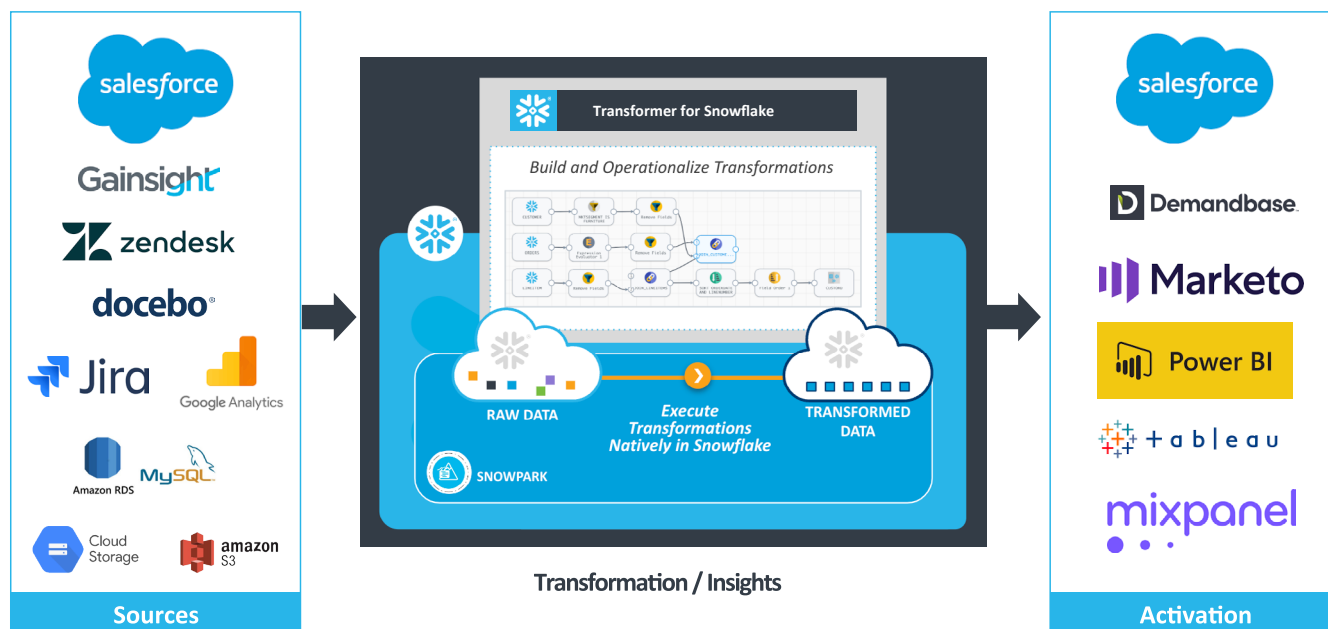
By providing a unified view of meaningful product events and interactions, different departments at StreamSets can now leverage actionable insights. A graphical user interface enables business users in the Marketing and Product teams. The Product team can

Snowflake with StreamSets & webMethods.io:

Transforming raw data to provide domain experts with direct and secure access to analytics-ready data



Full Circle Integration: Applications & Data to Insights to Applications



identify where users hit product roadblocks and relay this information to the Engineering team for future product releases and roadmap updates.

Sales and Marketing now have an additional dimension of data when evaluating opportunities for new business, renewals, and upsell. Marketing can respond in real-time to meaningful events and create personalized experiences and campaigns.

The Analytics team can now function as a highly efficient business operations team, providing enriched dashboards for marketing attribution, campaign performance reporting, and PQA scoring and alerts.

This is a pivotal step in StreamSets' journey to data-driven growth and agility across the business.

About StreamSets

StreamSets, a Software AG company, eliminates data integration friction in complex hybrid and multi-cloud environments to keep pace with need-it-now business data demands. Our platform lets data teams unlock data—without ceding control—to enable a data-driven enterprise. Resilient and repeatable pipelines deliver analytics-ready data that improve real-time decision-making and reduce the costs and risks associated with data flow across an organization. That's why the largest companies in the world trust StreamSets to power millions of data pipelines for modern analytics, smart applications, and hybrid integration.

To learn more, visit www.streamsets.com and follow us on [LinkedIn](#).