Customer Case Study
ShopRunner – Personalizing retail experiences

At-a-glance

Industry
- E-Commerce Network for leading retailers and online shoppers

Products used
- Databricks Unified Analytics Platform
- Snowflake Data Warehouse

Customer Challenges
- Needed a powerful data pipeline for processing complex JSON log structures
- Intensive data processing and storage demands with 1 TB/day of new data
- Minimizing specialized DevOps knowledge requirements

Solution
- Decoupling storage and compute with Snowflake and Databricks
- Simplified onboarding and collaboration with Databricks
- Enhanced ability to run ad hoc and regularly scheduled jobs

Benefits
- Intelligent suggestion of additional products using behavioral analysis
- Support for daily trending products using scheduled analytics jobs
- Visual similarity-based recommendations based on Spark

ShopRunner gets shoppers what they love, faster

ShopRunner connects the best online shoppers to top retailers and brands. ShopRunner's exclusive member benefits and highly targeted, data-driven marketing help high value customers find their desired retail brands while increasing existing customers' purchase frequency and spend. ShopRunner delivers a highly personalized and customized retail environment so that shoppers have the best experience possible. The key is being data driven – on the front end customers are only a few clicks away from their favorite brands and items, while on the back end ShopRunner is leveraging Databricks and Snowflake to tackle personalization, recommendation, targeting, and analysis of text and images.

The challenge: achieving high-quality, fast, and efficient data management

ShopRunner's challenge was ingesting data from over 100 retailers and their customers including product feeds, logistics, and behavioral data while maintaining efficiency for online shoppers. Their solutions needed to maintain high data quality for analytics, meet the speed requirements of data science, and do it all at a low cost. They needed high-volume data storage with competitive pricing, and to run many scheduled batch Spark jobs in order to constantly update recommendations, while serving the objectives of multiple teams. Given the amount and variety of data they were handling from a variety of retailer sites and sources, ShopRunner needed to streamline their data pipeline.

Delivering more value with Databricks and Snowflake

With Databricks and Snowflake, ShopRunner was able to enhance the strength of their product recommendations and machine learning experiments by simplifying their data pipeline. With the basics of data management out of the way, they can focus on deeper experiences like daily trending products emails for each user, personalized feeds and product discovery, all targeted to gender, age, location, and what the person already owns. The best part is ShopRunner's jobs fail less often, increasing the productivity of their data science team.

"The combination of Snowflake and Databricks gives our team both the scaleability and the agility to build really innovative products on top of our valuable data assets"
– Greg Ball, CTO

Results

Databricks and Snowflake helped ShopRunner achieve significant benefits:

- Simplified ETL and feature generation tasks with Databricks notebooks
- Leveraged Spark deep learning pipelines for visual similarity of product images
- Shared custom libraries and better collaboration

- Self-service access to data and data pipeline management for greater insights
- Dedicated compute for different processes and loads, enabling efficient and simultaneous querying of data
- Ability to ingest raw data from structured and unstructured file types and query alongside other data sources for deeper exploration

Solution Architecture
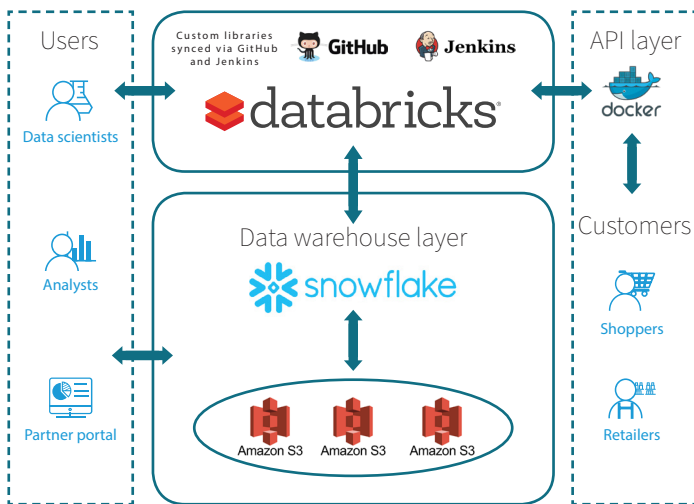
Solution Overview

ShopRunner has scaled up their ability to store and process data using Databricks and Snowflake. The Snowflake Databricks native connector enables data to move easily between the Spark engine in Databricks and dedicated compute and storage in Snowflake. ShopRunner can collect behavioral pageview data using Javascript plugins, correlate it with product feeds from retailers, store that data across multiple clouds in Snowflake Data Warehouse, and apply Spark for advanced machine learning and product recommendation. ShopRunner syncs their custom python libraries to Databricks via GitHub and Jenkins using an open-sourced package "apparate"

Their architecture includes:

- Databricks
- Snowflake
- Snowplow
- docker
- Amazon RDS

- Amazon Dynamo DB
- Amazon S3
- Amazon Kinesis
- Cloudfront
- MLlib

Visit ShopRunner and see what being data driven has done for their customers here

About Databricks

Databricks' mission is to accelerate innovation for its customers. Using the Databricks Unified Analytics Platform powered by Apache Spark, data science teams can collaborate with data engineering and lines of business to build data products.

About Snowflake

Snowflake is a modern SQL data warehouse architecture built for the cloud to make it easy to amass all your data, enable rapid analytics, and derive data-driven insights for all your business users.

Evaluate Databricks for yourself

Visit us at databricks.com and try Databricks free!

Learn more about Snowflake

Go to snowflake.com and enable your organization to be data driven