

CASE STUDY

Shell Delivers AI at Enterprise Scale



Overview

Royal Dutch Shell PLC is a British-Dutch multinational oil and gas company, headquartered in the Netherlands and incorporated in England with 86,000 employees in more than 70 countries. Shell is one of the oil and gas “supermajors” and the third-largest company in the world measured by 2018 revenues. In the 2019 Forbes Global 2000, Shell was ranked as the ninth-largest company in the world (and the largest outside China and the United States), and the largest energy company.

Shell recognized that they could turn massive amounts of data into insight by building a platform-as-a-service to process data at scale for self-service analytics: the Shell.ai Network.

Challenge

In an effort to modernize, Shell needed a platform that could automate data collection to a public cloud for processing and make it available to applications used by groups with different analytic requirements without creating duplicate data storage. The scale and variety of this data made a single approach to data

management and processing problematic. The goal was to provide a secure multi-tenant architecture that allowed for self-service data access and the ability to transform the data into the format that the data scientists require. Looking to the cloud, Shell found Amazon Web Services (AWS) and StreamSets as a perfect fit for this shift to the cloud.

Solution

AWS provided a burstable, elastic cloud that could meet the evolving demands of data workloads for AI. The StreamSets DataOps Platform helps automate data collection from a vast array of sources—from sensor data on submersibles to retail data at point of sale locations—into a public cloud and ensures the reliability of the data.

Result

Shell used AWS S3 data lake to hold raw data in native format until required for analysis. Taking advantage of AWS S3, gave Shell the flexibility to manage a wide variety of data at scale with reduced costs, improved access control, and strong regulatory compliance. Beyond the natively

supported governance and security features of AWS S3, Shell integrated its internal services for authentication, authorization, and data governance. It also developed a metadata service to simplify dataset discovery, which allows data consumers to easily search, sort, and identify datasets for analysis.

Shell is building a center of excellence for data movement and needs to scale quickly to meet the demand for self-service analytics with existing resources. The COE combines the existing data pipeline logic in a consumable, shareable, and searchable interface to help end users receive reliable data in a self-service manner. The function facilitates 2000+ members of the Shell AI network and more than 600 self-service users from multiple teams across Shell for traditional analytics functions to experimental data science efforts. The platform also provides pipelines for real-time model training and scoring. Each team at Shell found value in working with AWS and StreamSets with scalability and elasticity. The combination of AWS and StreamSets DataOps Platform has allowed Shell to scale up and down to meet the data processing needs and optimize the cost of their environment.

“StreamSets allows me to provide stable, sustainable data operations on top of both a self-service and professional platform and to operate this at scale.”

*Dan Jeavons
General Manager of Data Science
Shell*

Next Steps

Shell has identified 23 use cases spread across upstream and downstream production, including retail and trading, that they will power with their new AWS data lake. From a technician who knows when to service a piece of equipment to the customer who can better optimize sustainable energy use. As more data is piped into their AWS data lake to power Shell AI, they are able to adopt new use cases using StreamSets and AWS.

ABOUT STREAMSETS

StreamSets built the industry's first multi-cloud DataOps platform for modern data integration, helping enterprises to continuously flow big, streaming and traditional data to their data science and data analytics applications. For more information, visit www.streamsets.com.

