

CASE STUDY

Global Biopharmaceutical Company Embraces Data Modernization

Overview

For 130 years, this well-known global biopharmaceutical company, has been inventing medicines and vaccines for many of the world's most challenging diseases in pursuit of their mission to save and improve lives. They demonstrate their commitment to patients and population health by increasing access to health care through far-reaching policies, programs and partnerships. They're at the forefront of research to prevent and treat diseases that threaten people and animals – including cancer, infectious diseases such as HIV and Ebola, and emerging animal diseases.

Challenge

The Enterprise Platform Group needed a tool to spearhead its Next Gen Analytics initiative to support its time to market for data and decisioning. With rigid on-premises data centers to support scientific research workflow, siloed data

from various groups prevented collaboration. The new platform needed to provide an economic way to move, store, and analyze hundreds of terabytes of clinical trial and genomic data. The data would be used to guide business intelligence, improve compliance and management reporting, and solve advanced analytics problems using AI and ML techniques. The team needed an answer for their data infrastructure that would support scalable queries and enterprise data integration.

Solution

The team decided a move to the cloud was their best option, and chose StreamSets and Amazon Web Services (AWS) as a perfect fit for this transition. Collaborating with StreamSets, they moved their data from a legacy environment composed of Hortonworks Hadoop, Teradata and Informatica along with custom-built data tools to a next-gen data platform with AWS Redshift. StreamSets Transformer allowed them to create data processing pipelines to execute on Spark without requiring a technical understanding of the platform. With StreamSets simple user interface, they were able to see exactly how long every operation takes, how much data gets transferred at every stage, and had the ability to quickly solve core business problems as they arise. Additionally, Redshift provided a simple and cost effective solution to run quick queries

of structured and unstructured data from a central source. With its ease of use, they could effortlessly pull data to generate reports, dashboards, and use business intelligence data-science tools to give their researchers real-time operational insights on their clinical trials and genomic data to make informed decisions. The alliance with StreamSets and AWS allowed the team to deliver the business outcomes that they're focused on and make an impact in saving lives through the analysis of their clinical data.

ABOUT STREAMSETS

At StreamSets, our mission is to make data engineering teams wildly successful. Only StreamSets offers a platform dedicated to building the smart data pipelines needed to power DataOps 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 business intelligence, data science, and AI/ML. With StreamSets, data engineers spend less time fixing and more time doing.

To learn more, visit www.streamsets.com

