



CASE STUDY

Transforming Inspire Clean Energy: Streamlined Data Processing and ML Ops with Spark

Executive Summary

Inspire Clean Energy, a US retail energy company specializing in renewable wind, solar, and hydro power, sought assistance from New Math Data (NMD) to enhance their development process and scale their workloads. NMD recommended, designed, and implemented a Spark processing platform using Databricks, integrated with Inspire's existing resources and AWS accounts. The platform was securely integrated with Inspire's SSO system, providing streamlined access for their data scientists and engineers. Additionally, NMD improved Inspire's ML Ops capability, ensuring reliability, reproducibility, and portability of ML models across their dev, staging, and prod environments. The deployment of Spark infrastructure resulted in significant dataset scalability and over 50% reduction in model training time.

Customer Description

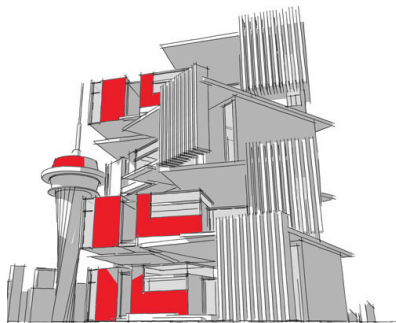
Inspire Clean Energy is a US retail energy company, specializing in renewable wind, solar and hydro power for its clients, while providing a predictable pricing model.



Description of Service

Inspire currently has a sophisticated data science and engineering team and has built a data platform comprising Snowflake, dbt pipelines, Airflow, Kubernetes. Inspire hoped to streamline their development process and scale their workloads significantly and requested New Math Data's assistance to assist with their adoption of Spark.

NMD recommended, designed, and implemented a Spark processing platform utilizing Databricks and integrating it with their existing resources and AWS accounts. The platform was also integrated to Inspire's existing SSO system to provide secure and streamlined access for their team of data scientists and engineers.



The next phase of the engagement focused on improving Inspire's ML Ops capability in terms of reliability, reproducibility, and portability of ML models between their tiered account structure (i.e. dev, staging, and prod environments).

Description of Solution

NMD designed and implemented/integrated a Spark processing environment based on the Databricks platform into their existing data platform and operations, which included – infrastructure definition using Terraform, integrated with Snowflake, GitHub, and Airflow running on Kubernetes in existing AWS accounts.

NMD also designed and implemented a robust ML Ops pipeline that enables Inspire to have reproducible and scalable model training environments that streamline their model training, validation, and promotion processes.

Description of Outcome

Spark infrastructure is successfully deployed across all Inspire environments enabling data scientists to scale the size of their datasets. Another direct result of this migration is that model training time has been reduced by more than 50%.