

CASE STUDY

Shell Accelerates Energy Transition Forecasting with Dremio's Lakehouse Platform

At a Glance

The Customer



Challenge

Shell's power retail organization needed to process massive volumes of distributed data across multiple sources to build advanced electricity consumption forecasting models, but faced bottlenecks from complex ETL processes and performance issues with large-scale data operations.

Solution

Shell implemented Dremio's data lakehouse platform as a unified access layer that enabled self-service analytics, eliminated complex ETL requirements, and provided the compute performance needed to process billions of records for real-time forecasting models.

Results

- Enabled processing of 6-8 billion records within minutes for production inference models
- Eliminated weeks-long ETL development cycles through self-service data access
- Successfully operationalized 100+ concurrent forecasting models in production
- Created a scalable data mesh architecture supporting cross-functional collaboration

The Customer

Shell is a global energy company at the forefront of the world's energy transition, helping drive the shift toward clean energy while meeting growing global energy demand. Through their "Powering Progress" strategy, Shell focuses on four key pillars: generating profitable value for shareholders, partnering with customers and governments to achieve net zero emissions, respecting nature through waste reduction and biodiversity protection, and powering lives and livelihoods during the energy transition. As the energy landscape becomes increasingly decentralized with renewable sources like solar, wind, batteries, and electric vehicles, Shell recognized the critical importance of accurate electricity consumption forecasting to navigate this complex new environment.

The Challenge

Shell's power retail organization faced a fundamental challenge as the energy sector transformed from a simple centralized model to a complex decentralized ecosystem. While electricity forecasting had traditionally been a compute problem, the proliferation of millions of smaller renewable energy sources—from rooftop solar panels to electric vehicle charging stations—transformed it into a massive data challenge requiring unprecedented scale and speed.

The organization needed to process enormous volumes of data residing across different sources without building complex ETL pipelines that would significantly delay their forecasting capabilities. Data analysts, engineers, and product teams required immediate access to distributed datasets to support data scientists in developing sophisticated forecasting algorithms. The urgency was compounded by real-world scenarios where data scientists encountered critical bottlenecks, such as Jupyter notebooks crashing when attempting to load 80 gigabytes of data into memory or dealing with complex 15-step joins involving non-unique identifiers.

Perhaps most critically, Shell needed to operationalize approximately 100 forecasting models running concurrently in production, processing 6-8 billion records within just a couple of minutes. This demanding performance requirement meant that traditional data processing approaches simply couldn't meet the speed and scale needed to support real-time electricity consumption forecasting in the new decentralized energy landscape.

The Solution

Shell selected Dremio's data lakehouse platform as their unified compute engine and access layer, recognizing it as the natural choice to address their complex data challenges. Dremio enabled Shell to tap directly into multiple distributed data sources without requiring extensive ETL development, dramatically accelerating their time-to-insight for forecasting model development.

The implementation leveraged Dremio's virtual datasets (VDS) and physical datasets (PDS) to create rapid iterations and prototyping capabilities, allowing data teams to quickly explore and prepare data before committing to formal ETL processes. Shell's teams utilized Dremio's reflections feature to optimize performance for frequently accessed datasets, though they learned to carefully manage and isolate these reflections as they became more heavily utilized across the organization.

Dremio's fine-grained access control capabilities proved essential for Shell's governance requirements, enabling secure provisioning of data spaces that could be safely assigned to Azure Active Directory groups. This approach maintained strong security standards while enabling self-service analytics across different teams and functions. The platform's sophisticated compute engine provided the performance needed to handle Shell's demanding workloads, including the ability to process billions of records within the tight timeframes required for production forecasting models.

Results

The implementation of Dremio transformed Shell's ability to develop and operationalize electricity forecasting models at the scale and speed required for the modern energy landscape. The platform successfully enabled the processing of 6-8 billion records within minutes, meeting the stringent performance requirements for running 100+ concurrent forecasting models in production.

Shell achieved significant acceleration in their data product development lifecycle, eliminating the weeks-long delays previously associated with complex ETL development. Data analysts, engineers, and product teams gained self-service access to distributed datasets, enabling rapid collaboration and iteration that dramatically reduced time-to-market for new forecasting capabilities.

The solution evolved into what Shell recognized as a comprehensive data mesh architecture, with Dremio's unified access layer providing elegant abstraction from underlying data complexity. This architecture enabled dynamic evolution of data products while serving multiple customer layers, including both internal data scientists and external customers consuming forecasting results.

Beyond the immediate performance gains, Shell discovered they had generated valuable reusable datasets that became broadly consumed across the organization. The visibility provided by Dremio's lineage capabilities enabled rapid identification and resolution of data quality issues, while the platform's iterative data modeling capabilities supported continuous refinement of forecasting accuracy and operational efficiency.

ABOUT DREMIO

Dremio is the intelligent lakehouse platform for the business, serving hundreds of global enterprises, including Maersk, Amazon, Regeneron, NetApp, and S&P Global. Based on open-source technologies like Apache Iceberg and Apache Arrow, Dremio provides an open lakehouse architecture enabling the fastest time to insight and platform flexibility at a fraction of the cost.

Learn more at www.dremio.com.

Dremio and the Narwhal logo are registered trademarks or trademarks of Dremio, Inc. in the United States and other countries. Other brand names mentioned herein are for identification purposes only and may be trademarks of their respective holder(s). © 2025 Dremio, Inc. All rights reserved.