Simplifying reliable ETL on Delta Lake with Delta Live Tables

databricks

Simplifying reliable ETL development on Databricks Delta Lake

Shell optimizes trillions of rows of IoT sensor data with Delta Live Tables

Shell is at the forefront of creating a cleaner tomorrow by investing in digital technologies to tackle climate change and become a net-zero emissions energy business. Essential to the strategy is managing trillions of IoT sensor data coming from the cloud and field signals, that help drive insights to enable greener, cost-effective operations. Shell needed the ability to scale ETL to enable machine learning (ML) and AI across varying data sources and types. To meet key pipeline requirements, Shell chose to leverage Delta Live Tables, built on the Databricks Lakehouse Platform, for automated ETL development and management on Delta Lake.

The challenges of managing ETL at scale

Reliable, large-scale ETL can be complex and brittle. With trillions of rows of IoT sensor data, varying in type and source, Shell sought an efficient solution for processing that data. Sophisticated data pipelines are hard to build, but necessary for profit growth, analytics, data science and machine learning. Dan Jeavons, GM of data science at Shell said, "Digital technology is absolutely core to making our existing business more effective and efficient. As the industry continues to expand into new areas of energy that are more sustainable and reduce environmental impact, data and digital technology are now table stakes."

In order to deliver and leverage trusted data to all of Shell's workstreams — from predictive maintenance, drone missions, and digital twins, to hydrogen efficiency, the Gold+ loyalty program, and clean production optimization — Shell needed production-grade pipelines that can deploy worldwide and at scale. Shell identified six key requirements for its data pipelines: connectors, transformation and AI, quality, recovery, scale, and monitoring.

1.3 trillion
rows of sensor data processed

Millions
of dollars saved with data-driven operations

70+ AI-driven
transformation use cases

Meeting pipeline needs with reliability and automation

With the Databricks Lakehouse Platform as the foundation for their Shell.ai Platform, they looked to Databricks to help streamline and simplify data management and processing through ETL. With Delta Live Tables, Shell can now easily ingest data into Delta Lake and automatically build SQL endpoints for real-time queries and multi-layer insights into data. Equipped with a Delta-centric service, Shell has significantly reduced pipeline complexity, elevated software engineering standards, improved pipeline performance, and gained new data insights.
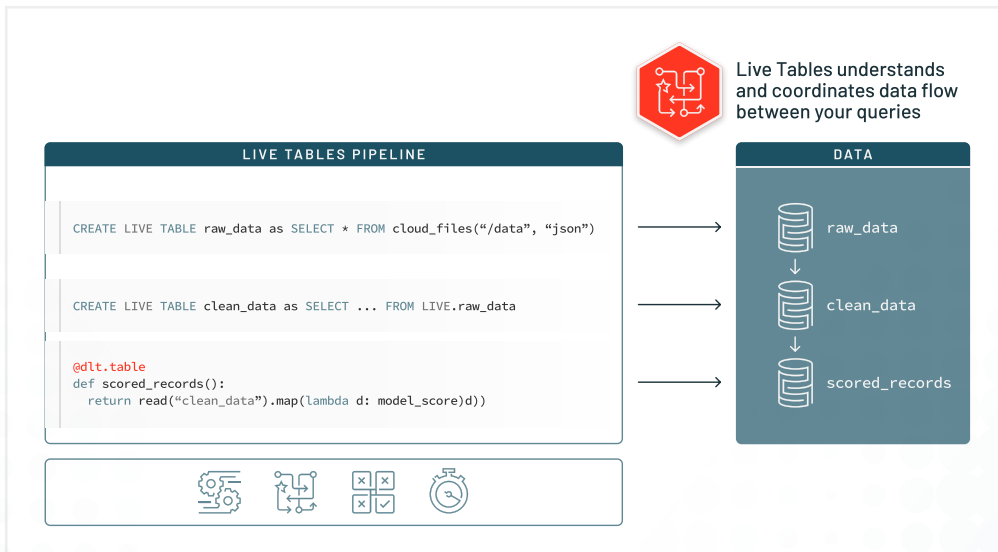
By building intelligence into the data ingestion process, Shell is saving millions of dollars with data-driven operations and innovative opportunities. By simplifying ETL development and management on Delta Lake, Shell data engineers can now build and leverage their own ETL pipelines, and gain insights and analytics quickly.

Beyond saving time and money, Shell's ETL pipelines can be reused regardless of environment, and data can run in batches or streams with incremental or complete computation per table. Delta Live Tables also increase the accuracy and reliability of the data with validation and integrity checks, and predefined error policies.

> **"With this capability augmenting the existing lakehouse architecture, Databricks is disrupting the ETL and data warehouse markets, which is important for companies like ours."**
>
> DAN JEAVONS | GM of Data Science, Shell

The end result is deeper pipeline visibility and data knowledge, reduced downtime with automatic error handling, and collaborative ETL between Shell's applications.

Delta Live Tables makes it easy to build and manage reliable data pipelines that deliver high-quality data on Delta Lake.

Millions in savings with ETL simplicity

Shell is taking advantage of data and ML availability for AI-driven transformation across 70+ use cases — including operations, product development, marketing, and customer experience. Delta Live Tables has made it possible for Shell to overcome the challenges of processing 1.3 trillion rows of streaming IoT data from millions of sensors. With clean data flowing freely downstream, Shell's data teams can now harness actionable insights and accelerate ML implementation.

Dan explains, "At Shell, we are aggregating all our sensor data into an integrated data store — working at the multi-trillion-record scale. Delta Live Tables has helped our teams save time and effort in managing data at this scale. We have been focusing on continuously improving our AI engineering capability and have an Integrated Development Environment (IDE) with a graphical interface supporting our Extract Transform Load (ETL) work."

Additionally, Delta Live Tables is driving better software engineering standards around pipeline development, improving pipeline performance and data integrity, providing real-time query and data insights, spurring innovation with facial recognition and ML libraries, plus, saving Shell millions of dollars with task automation, ETL simplicity, and data-driven operations.

Going forward, Dan says, "We are excited to continue to work with Databricks as an innovation partner." Utilizing Databricks to unlock the promise of data and AI, it's clear that Shell will more easily reach its goal of creating a cleaner tomorrow.

> Databricks is disrupting the ETL and data warehouse markets, which is important for companies like ours. We are excited to continue to work with Databricks as an innovative partner.

DAN JEAVONS
GM of Data Science, Shell

About Databricks

Databricks is the data and AI company. Thousands of organizations worldwide —including Showtime, Shell, Conde Nast and Regeneron — rely on Databricks' open and unified platform for data engineering, machine learning and analytics. Databricks is venture-backed and headquartered in San Francisco with offices around the globe. Founded by the original creators of Apache Spark™, Delta Lake and MLflow, Databricks is on a mission to help data teams solve the world's toughest problems. To learn more, follow Databricks on Twitter, LinkedIn and Facebook.

databricks

EVALUATE DATABRICKS FOR YOURSELF

WATCH VIDEO   **LEARN MORE**

Contact us for a personalized demo databricks.com/contact