

PREDICTING ELECTRICAL ENERGY DEMAND WITH MACHINE LEARNING

10K+

Sensor measurements per second captured

Industry

– Energy

Locations

– Headquarters: Luxembourg

“Through the Cloudera solution, we’re able to make decisions in order to save costs and improve the efficiency of our manufacturing process. Producing better pipe allows the company to deliver a better product to the final client, and this allows the company to be a market leader for this kind of product.”

Vincenzo Manzoni, Head, Data Science, Tenaris

Tenaris is a multinational manufacturer of seamless steel pipes for the oil industry. Its pipes are used for drilling and transporting oil from well to destination.

Challenge

Tenaris has always leveraged data to improve manufacturing processes and supply chain operations, but historically relied on small data samples to drive those decisions. Its legacy systems lacked the flexibility to combine data from various sources, and creating forecasts and predictive models was time consuming using traditional tools.

As their analytical questions became increasingly complex, the organization saw the need for a modern data platform.

“We were asked how target variables correlate with hundreds of possible features,” said Vincenzo Manzoni, head of Data Science at Tenaris. “We needed to answer questions like, ‘What is the best model that matches this feature with this target variable?’”

Solution

Tenaris implemented a Cloudera modern data platform to see how process controls are actually impacting manufacturing. The solution correlates data generated during the manufacturing process with process control data from the Department of R&D.

Using Cloudera’s platform, Tenaris ingests industrial log data from thousands of sensors at each of five manufacturing plants’ proprietary process control applications and databases, ingested via Apache Flume and Apache Sqoop respectively. The Flume ingestion leverages a Tenaris-contributed feature to the open source project that is now generally available in CDH.

The sensors’ time series data track electrical consumption, pressure, furnace temperature, and other variables throughout the process.

Tenaris uses Apache Spark to process data and makes it available for browser-based analysis through Apache Impala and Tableau. The same data is used to train models, leveraging Spark’s ability to scale machine learning across the Cloudera cluster.

Solution highlights

- Modern Data Platform: Cloudera Enterprise
- Workloads: Analytic Database, Data Engineering
- Key Components: Apache Spark™, Apache Spark MLlib, Apache Impala®, Cloudera Manager, Cloudera Navigator
- BI & Analytics Tool: Tableau

Applications supported

- Manufacturing process optimization
- Supply chain management

Data sources

- Thousands of furnace sensors from five manufacturing plants, sending time series data regarding temperature, pressure, and electrical consumption throughout the manufacturing process
- Parameters from Department of R&D's process control systems

Impact

- Higher product quality results in greater market share
- Manufacturing and energy consumption efficiencies save costs
- Data science results delivered in 2-3 days, not weeks

Big data scale

- Multiple gigabytes collected each day
- 10 measurements per 1,000 sensors captured every second

Implementation

Tenaris selected Cloudera because of the platform's maturity—demonstrated through its administration tools (Cloudera Manager) and integration with other ecosystem technologies (Tableau).

The Tenaris team put Cloudera into production within one month.

"Thanks to Cloudera Manager, we are able to monitor and tune the whole cluster in just a few clicks," said Andrea Rota, data engineer at Tenaris. "This is very useful. Since we are just a few people, it's very time effective for us."

Results

"Through the Cloudera solution, we're able to make decisions in order to save costs and improve the efficiency of our manufacturing process," said Vincenzo Manzoni. "Producing better pipe allows the company to deliver a better product to the final client, and this allows the company to be a market leader for this kind of product."

Tenaris can now deliver results in response to complex data science requests in a fraction of the time it took previously.

For example, Tenaris is optimizing its main power plant's manufacturing operations based on energy consumption predictions, fueled by machine learning models, that determine how much to produce and sell under up-to-the-hour market constraints. Those models are tested at scale on Cloudera's platform, delivering computations with better results in a fraction of a second. In comparison, the legacy spreadsheet-based model would take minutes to compute and glean less accurate results based on historical datasets.