Real-time machine learning with Swarm64 DA-accelerated PostgreSQL

Wind turbine-generated SCADA data

Wind turbines generate a large amount of SCADA (supervisory control and data acquisition) data, outputting values for approximately 500 different metrics each second. Turbine SCADA metrics include energy output, weather conditions, nacelle positions, blade angles, gearbox acceleration, fault codes, and other system statuses and control values.

Berlin-based **Turbit Systems** develops condition monitoring software for wind turbines. The software runs machine learning algorithms against the turbine-generated SCADA data in real time, comparing measured data with expected behavior. Relying on its large database of failure events, Turbit automatically detects the slightest technical fault and recommends corrective measures. This insight helps wind farm operators and service providers avoid costs that are related to damage or operational stops, and can improve wind-farm energy output by up to 5%.

In order to meet the data management challenges of real-time condition monitoring, Turbit chose Swarm64 DA-accelerated PostgreSQL over MongoDB for scalable, sub-second query performance.

Real-time machine learning for wind turbine data

The industry standard is to aggregate turbine SCADA data at the source and report 10-minute minimums, maximums, averages, and standard deviations. Turbit built their original machine learning system to work at the industry-standard pace, collecting 200 data points per minute per wind turbine in a MongoDB database and re-running their machine learning algorithms on an hourly or daily basis.

In 2019, Turbit began work on a real-time version of their system, capable of creating a learning-feedback loop, that feeds machine learning outputs back into the turbine control algorithms to continuously improve turbine behavior.

"Real-time condition monitoring and control offers the ability to detect a gust of wind at one end of a wind farm, and precisely adjust the position of the turbines to capitalize on that gust before it reaches the other end."

Michael Tegtmeier, CEO & Founder, Turbit Systems.

Real-time condition monitoring on a sub-second level is a challenging data management task:

- **150x more data.** Real-time monitoring collects 30,000 data points per minute per turbine, or 150x more data than processing 10-minute aggregates.

- **300x faster processing.** Algorithms must run within two seconds...a 300x smaller response-time window than the competitors' 10-minute cycles.

- **Data quality.** Missing data gaps arising from data lag and connectivity problems (e.g., from a turbine at sea) must be detected and filled on the fly in order to avoid skewed insights.

- **Hundreds of customers, thousands of turbines.** As Turbit grows, they need to provide fast analytics to many concurrent users looking at data generated by thousands of wind turbines.

Swarm64-accelerated PostgreSQL for real-time machine learning

MongoDB was not able to meet Turbit's real-time data requirements, so they explored alternative database options. Turbit wanted to stay on a free, open-source database, so they decided to try PostgreSQL accelerated by Swarm64 DA.

Swarm64 DA extends PostgreSQL with performance-acceleration features that enable PostgreSQL to analyze data orders of magnitude faster than usual, even while data is streaming into the database. Swarm64 achieves its acceleration by enhancing PostgreSQL with:

- Greater parallel processing
- I/O reduction via compressed columnstore indexing
- Faster SQL JOIN algorithms
- Much more efficient query planning and execution

When Turbit compared Swarm64 DA performance to MongoDB, they found that Swarm64 DA-accelerated PostgreSQL delivered faster and more reliable performance:
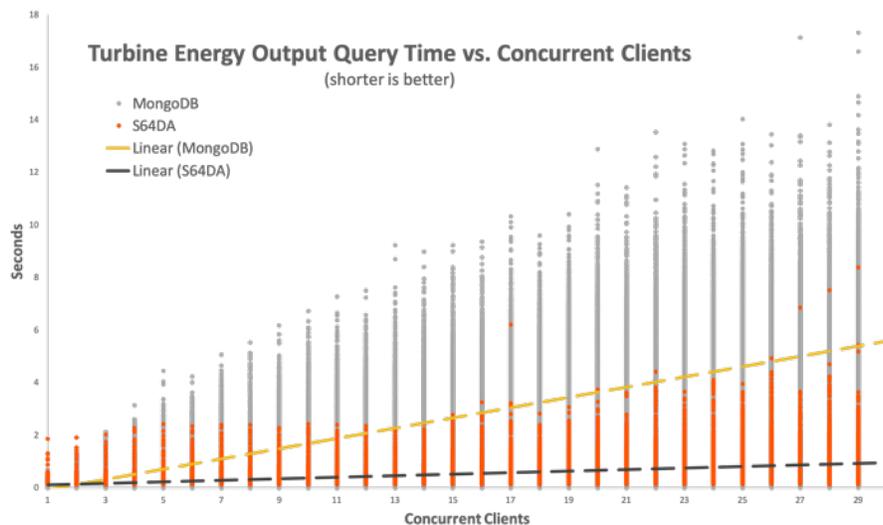
Swarm64 DA-accelerated PostgreSQL provided sub-second response more consistently than MongoDB at scale.

- **Sub-second response at scale.** Swarm64 DA provided sub-second response, even as the number of concurrent users increased from one to 30. Sub-second response is essential for real-time control of turbine performance.

- **2x - 6x faster performance.** As the number of concurrent clients increased, the Swarm64 DA performance advantage over MongoDB increased from 2x to 6x.

- **More reliable performance.** As the number of concurrent users increases, MongoDB query performance deviates greatly. Swarm64 DA performance is more consistent at scale.

> *"If a turbine manufacturer needs real-time data for their turbines' control system, with Swarm64 DA, I can say 'Yes, no problem.'"*
>
> *- Michael Tegtmeier*

Besides the real-time performance, Swarm64 DA-accelerated PostgreSQL provided other benefits.

SQL support - greater tool compatibility and ease of hiring

PostgreSQL is one of the most widely used databases in the world, and SQL is so well known, the technical talent pool from which to hire is larger than that of NoSQL databases like MongoDB. The SQL tools ecosystem also offers many more choices for compatible data pipelining and data visualization.

Data quality - healing data gaps and time shifts in the data stream

Turbines are often located in remote areas, and receiving a steady stream of SCADA data from them can be unpredictable. As data streams in, Swarm64 DA detects missing data gaps, and raises connectivity issue alerts to system operators. Swarm64 DA was also able to fill in missing data via interpolation.

Net result: fast time to market for a sustainable-energy breakthrough

In less than two months, Turbit was able to develop a working prototype of their real-time, condition-monitoring system, running on Swarm64 DA-accelerated PostgreSQL.

When asked for a closing thought on his experience with PostgreSQL and Swarm64 so far, Michael Tegtmeier replied: "Swarm64 DA gives us answers within seconds, which gives our data scientists the ability to test new algorithms more quickly and against larger amounts of data. It's a nice competitive advantage for us."

Learn more

For more information about Turbit, go to turbit.de. If you would like more information about Swarm64 DA, or to download a free development license, visit swarm64.com.