

Time Series in the Energy Sector: Successful Big Data Project of HAKOM

HAKOM, the technology leader for time series management in the energy industry, is responding to the market's demand for faster processing of a growing amount of data. In order to add more advanced big data analytics capabilities to the already impressive repertoire of solutions, the company tested its technology on a supercomputer. Little cluster, big data, high performance – the use case of HAKOM shows how high-performance computing brings data analytics to the next level.

Like virtually any other industry, the energy sector lives through a time of digital transformation. The sophisticated technology of IoT and Industry 4.0 drives this process and generates more data than ever before. Most of it is time series data – basically any data that has a time stamp on it.

Time Series

In the energy industry, time series data is generated by technical parameters such as power consumption and power fluctuations as well as by recording weather and meteorological conditions (e.g. temperature, wind strength, radiation, humidity).

In the past three years, the amounts of data that energy providers have generated and obtained from users have dramatically increased. Growing volumes of high-frequency data pose a challenge because all of them have to be stored, harmonised, aggregated, and analysed. In addition, the processing must take place close to or in real time. A timely analysis is now a necessity – for forecasting, capacity planning, improving energy efficiency, generating new revenue sources, as well as reducing power system blackouts.

Bracing for new data management

“Energy providers are overwhelmed with constantly growing data sets, and the lion's share of that data is time series,” confirms Stefan Komornyik, managing partner at HAKOM. For thirty years, the company based in Vienna has been providing solutions for managing, storing, and processing time series data, winning any challenge the industry has been facing. Today, it is evident that a paradigm shift in data management is necessary to control the data flood and make time series data valuable.

“Standardisation must be the goal in all areas of data management in order to offer innovative solutions,” says Stefan Komornyik. “Based on fully scalable base technologies, highly scalable applications for data of different sources and formats can be developed in a short time. Of particular importance here is a very special feature of HAKOM time series technology named Polyglot Persistence. It gives the possibility to combine data stored in different technologies in a consistent and performant way. This makes it possible to maximise the value of data that originates from different sources and to use the appropriate database technology for each use case without having to redevelop large parts of the application each time.”

Little Big Data Cluster

For companies that want to investigate new base technologies and handle computationally intensive tasks such as big data analytics, there is the possibility of receiving technical support from the network of EuroCC Austria. The EuroCC competence center referred HAKOM to the experts of the Little Big Data (LBD) cluster, a high-performance computing (HPC) system at TU Wien. On the LBD cluster, users can create their own environment and test various software. The system is primarily available to members of the TU Wien and since recently also to external users – those can request access through EuroCC Austria and explore the possibilities of HPC as part of a proof of concept (PoC).

Proof of concept: high-performance data analytics

During the PoC, HAKOM was able to work with a data set significantly larger than usual volumes of their clients: 5,300,000 observations of different time series in measurement intervals of 2 seconds each.

Based on the existing time series technology, HAKOM created a prototype that can use the transaction data from a distributed file system to store the time series data and calculate simple but very computationally intensive statistics.

Using the PostgreSQL database for the movement data with sequential storage, these statistics were practically impossible to calculate – the computing time was in the range of many hours. The breakthrough came with a distributed scalable file system and parallelised storage on multiple nodes. This enabled HAKOM to accelerate the processing of high-frequency time series data by a factor of 400 and bring it down to about six minutes.

This use case demonstrates one of the main advantages of supercomputing: the possibility to parallelise problems, i.e. to efficiently divide a large problem into several smaller ones and solve each of them separately, but simultaneously. Parallelisation helps speed up the code and shorten computation times. This way, companies benefit from HPC in many ways and can process and solve complex computationally intensive problems, optimise software, accelerate product development and shorten time-to-market.

What's next?

The successful integration of time series management (TSM) software on a highly parallel cluster enables HAKOM to further develop the tools for analysis of very large data sets directly through its TSM system. To this end, various modalities are being explored to offer distributed analysis both on HAKOM's own infrastructure and within the cloud. Tools for analysing and processing time series are already available for testing in the cloud as HAKOM TSM Hosted Lab. In 2022, the company will launch a cloud-based time series service for the energy market that includes both processing and storage of data in the cloud for simple use cases based on a few time series in a 15-minute grid and use cases based on a few hundred thousand time series from sensor data in a seconds grid.

EuroCC Austria - contact point for HPC

HAKOM's experiment with high-performance data analytics is yet another successful outreach project of EuroCC Austria, whose mission it is to provide access to HPC infrastructure and know-how to a wide range of users. Through cooperation with EuroCC Austria, companies can test and adopt HPC, big data analytics, or artificial intelligence and integrate them into their business models – for product optimisation and further development of advanced applications and technologies.

The services of EuroCC Austria are available free of charge to companies, researchers, and public institutions. For inquiries please contact info@eurocc-austria.at.