Introduction

When you are on a mission to change the way that an entire country powers itself, innovation is the engine driving your business. That describes Vandebron, an online energy platform that allows consumers to pay for energy in a smarter way.

Essentially acting as an energy supplier, Vandebron provides a digital marketplace that enables consumers and local renewable energy producers to directly trade energy. As of early 2019, Vandebron had around 180,000 customers in the Netherlands, and partnerships with approximately 220 producers of renewable energy.

Challenge:

Building and Deploying a Secure Machine Learning Pipeline

As Dr. Giuseppe Procaccianti, Head of Big Data at Vandebron, explains, Vandebron can realize enormous value by accurately predicting how much renewable energy its partners will produce. However, by its nature, renewable energy is unpredictable. While this is a large-scale problem for a grid operator, it’s still problematic for a company like Vandebron because an energy imbalance means the company incurs costs.

For this reason, Vandebron decided to build an in-house wind forecasting pipeline

Vandebron Project At-a-Glance

Customer
Vandebron – renewable energy marketplace

Challenge
Enable a real-time data-intensive machine learning pipeline

Solution
DC/OS

Impact
Rapidly going live with robust, real-time data processing
that predicts the anticipated energy output 24 hours in advance since the Dutch energy market works on a day-ahead basis. The goal was to generate an energy plan for each wind producer showing how much they would produce every hour. To build the pipeline, Vandebron would need to ingest and process a few gigabytes of data every six hours. Ingesting and processing this much data in a short time frame meant Vandebron needed a high-performance distributed system to analyze and take action upon huge datasets.

Solution:
Establishing a Mature Modern Architecture with DC/OS

Vandebron had been using DC/OS since 2017 to put in place a modern architecture and host its Smart Charging pilot promoting electric transportation—one of their most innovative projects. The pilot was a success, so it was a natural extension for DC/OS to enable Vandebron’s new big data project.

In addition to using a SMACK (Spark, Mesos, Akka, Cassandra, Kafka) stack to enable real-time data processing for its wind forecasts, Procaccianti planned to use Apache NiFi. This would make it possible to quickly ingest and transform data from different sources, and graphically build its wind forecast pipeline. “We needed a more mature infrastructure setup for this project, which included setting up DC/OS in a production environment and migrating from Microsoft Azure to AWS,” Procaccianti says.

As he evaluated considerations when moving to a container-based architecture, he says it was clear that DC/OS was the platform his company should continue using. “We considered Kubernetes-as-a-Service solutions like AKS or GKS, but decided there was no real alternative to DC/OS for our SMACK stack and data-intensive processes,” continues Procaccianti.

As it prepared its production instance of DC/OS, Vandebron called upon many services and packages in the DC/OS Service Catalog. “D2iQ has invested in a lot of technologies that we find valuable. We chose to leverage as many certified packages as possible since we would get the added benefit of support, along with active maintenance and testing,” Procaccianti explains.

For security reasons, Vandebron set up separate clusters for its test, acceptance, and production environments. To maintain solid control of its forecast pipeline and infrastructure, Vandebron also deployed a monitoring tool, like Grafana, which it runs in yet another service cluster dedicated to operational monitoring.

The company runs various tools to gather operational metrics associated with Cassandra, its Kafka nodes, disk capacity, CPU usage, and more. These metrics are fed from the monitoring cluster into Elastic Search and streamed using Grafana to create monitoring dashboards. “D2iQ offers great tools and packages that enable monitoring, so you are confident that your infrastructure is running as planned,” says Procaccianti.

Impact:
Quickly Reaching Production Readiness to Unlock Data Value Faster

While the SMACK stack enables real-time data processing and a pipeline processing model, it also introduces new levels of operational complexity running multiple distributed systems such as Spark, Kafka and Cassandra. However, DC/OS tames that complexity.

“DC/OS allowed us to set up a sophisticated IT infrastructure to manage the size and complexity of our pipeline. It surprised us that we went from planning to execution of the project in just three months,” says Procaccianti.

According to Procaccianti, Vandebron went live in such a short time partly because of the packages available through DC/OS, such as Apache NiFi and Jupyter, which Vandebron added so it can quickly write Python code, test models, and explore data. “These are valuable tools enabling data scientists to move data, perform queries, quickly conduct analysis, and prepare reports. D2iQ gave us the right tools so our team could be productive quickly,” he continues.

Since going live with its wind forecasting pipeline, Vandebron has been incredibly happy with the results. “By leveraging the full power of DC/OS’s robust model processing technology, and optimized code, we are able to process huge datasets
in just 3-4 minutes.” explains Procaccianti. Going forward, Vandebron wants to expand its service offering so energy producers can visualize production in real time and know how much they need to produce and how much revenue they can expect.

Also thanks to DC/OS, Vandebron’s deployment process is easy and fast. “We’ve significantly shortened time to market by being able to quickly deploy new services to production. Before it would take weeks to install and run a service like Elastic Search or Spark. Now we’re operational in hours because we don’t have to take time reading documentation and manuals,” says Procaccianti. In fact, he says that having DC/OS in the Vandebron ecosystem makes it easy for the company’s developers to try new technology and upskill.

“We’ve significantly shortened time to market by being able to quickly deploy new services to production. Before it would take weeks to install and run a service like Elastic Search or Spark. Now we’re operational in hours because we don’t have to take time reading documentation and manuals.”
—Dr. Giuseppe Procaccianti, Head of Big Data, Vandebron

Now that other teams within Vandebron see how Procaccianti’s team is able to work with data efficiently, they are interested in migrating services and other parts of the company’s infrastructure to DC/OS. As Procaccianti explains, “Any new project goes inside DC/OS. It’s enabled a great transformation within our company, automatically elevating us to a higher technology level. We don’t need to use DC/OS for everything but can use it to extract some workloads and run them efficiently in an isolated and controlled environment.”

While the plan is to migrate its customer-facing marketplace, the company has already migrated the backend layer of its website into a microservices architecture inside DC/OS. According to Procaccianti, “This was a complex project involving domain expertise from across the organization. Yet it took less than three months. Not only did backend performance improve, we can now deploy the backend in just 10 minutes whereas it used to take about an hour.”

As Procaccianti says, the energy sector is being disrupted by technology in so many ways, such as through big data and data science, that the utility market is struggling to keep up with the transformation. “But if you can hire smart people and equip them with tools to quickly unlock the potential of technology, you gain an edge. People are key but tools can make a difference, and D2iQ has positioned us for success in that way,” he concludes.