

# Severstal Reduces Plant Downtime with Real-Time Streaming Data



## Headquarters

Cherepovets, Russia

## Industry

Manufacturing

## Challenge

Make use of the multiple terabytes of time series data generated weekly by industrial equipment to reduce downtime and increase efficiency.

## Solution

Use Confluent Platform to feed machine learning models and data analytics algorithms with near real-time data streams of plant data.

## Results

- Reduced plant downtime
- Completed initial deployment quickly
- Received support for securing in-transit data
- Achieved one-second latencies



Confluent Platform



Schema Registry



Kafka Streams API



Kafka Connect API

Severstal is among Russia's largest integrated steel and mining companies, producing more than 11 million metric tons of steel and close to \$8B in revenue annually. The company's key performance indicators – including revenue, profitability and dividend payouts – all improved last year, in part due to Severstal's strategy of defensive growth, which is focused on increasing earnings via enhanced efficiency and product quality rather than increased scale of production. To further solidify its position as a global leader in value creation, the company has embarked on the next phase of this strategy, a digital transformation in which resources are invested in big data, the Internet of Things, predictive maintenance and machine learning initiatives.

As part of this digital transformation, Severstal is using Confluent Platform to stream data from manufacturing sites, integrate microservices and feed machine learning models for predicting problems before they occur. "With Confluent Platform we have achieved low-latency transportation of time series data from manufacturing equipment and reliable distribution of that data throughout the company," says Donat Fetisov, Principal Architect at Severstal. "The solution we developed is being used to alert engineers to potential equipment downtime and failures, to optimize equipment operating modes and to suggest changes to key manufacturing parameters. Further, it is scalable to support not only the data we are generating today, but also the data we will generate in the future as the number of sensors used in manufacturing increases."

## Challenges

Much of the industrial equipment at Severstal's mining and production plants continuously produces data on the status of the equipment and on the quality of the product being produced. In the past, analysis of this low-level time series data was typically limited to vendor-specific software applications. Severstal wanted to make the more than 9TB of data generated weekly widely available across the company to analysts and engineers developing advanced machine learning models. In order for these models to deliver actionable information that could be used to minimize costly downtime and potentially catastrophic damage, the models needed data delivered to them in near real-time. "Delivering large flows of data with low latencies was essential for us to get a detailed picture of the equipment's current state and to leverage all of the mathematical tools that are in the arsenal of data scientists," says Fetisov.

Fetisov worked together with the company's chief digital officer and chief data scientist as well as analysts, engineers and plant teams to architect a solution that would meet the company's requirements for performance, reliability and availability. The group considered a range of possible solutions, including message queues such as RabbitMQ and ZeroMQ as well as open source stream processing systems such as Flink, Flume, Spark Streaming and Storm. Ultimately, the group found that each of these potential candidates were less than ideal in key areas such as support for numerous publishers and consumers, resource management and overall scalability and reliability. "Since our project requirements were not entirely clear when we started, we needed to choose a

solution that would be reliable and scalable enough to meet future, unforeseen challenges," says Fetisov.

Lastly, Fetisov and colleagues viewed support as an important component to ensuring the successful implementation of any data streaming solution they would adopt. "Taking into account that English is a foreign language for many of the engineers on our teams, it can be difficult for them to find answers when something goes wrong," Fetisov notes. "We wanted some insurance in the form of being able to communicate with a dedicated support team."

## Solution

Severstal selected Confluent Platform as a foundational component of its digital transformation strategy.

Working with Confluent engineers during the initial deployment and configuration of Confluent Platform, Fetisov and his team began by defining a set of standard event schemas and by setting up connectors to stream data into and out of Apache Kafka®. They used Schema Registry to define the schema as well as manage them and share them across the organization to help ensure compatibility between producers and consumers. The team also configured Kafka Connect to reliably store data passing through Kafka into HDFS.

The first producer that the team created generated messages based on time series data coming from manufacturing equipment and published the messages to topics in a Kafka cluster the team had set up.

*"The first steps with Confluent Platform proved to be of crucial importance as they involved changing of mindsets and transforming well-established processes."*

— Robert Christ, Principal Engineer at TIVo

On the consumer side, the first project was focused on using machine learning models with the available data to predict when manufacturing equipment would start heating beyond an acceptable rate in order to minimize product line stoppages and reduce the risk of running machinery under hazardous conditions. Later, the team began using Kafka Streams and the Kafka Streams DSL (Domain Specific Language) to build applications that aggregated stream data.

Since the initial deployment phase, Severstal has continued to add producers and consumers while building out its Kafka cluster. Their platform handles about 30,000 messages per second and has connections to Elasticsearch, Oracle, InfluxDB and a variety of applications running on Kubernetes. About 3TB of data is streamed each week through each of the cluster's brokers.

Looking ahead, Fetisov and his colleagues plan to move further onto Kubernetes deployments. In addition, they are planning to develop an aggregation layer that combines raw data into more meaningful entities as well as a toolset for end users based on KSQL. The team is also exploring the use of log compaction for certain topics to retain data in Kafka.

"At Severstal, we are at the beginning of our digital journey. The first steps with Confluent Platform proved to be of crucial importance as they involved changing of mindsets and transforming well-established processes," says Fetisov. "Early results showed the benefits of enabling low-latency transportation of data as a way of improving overall efficiency by reducing human error in day-to-day decision-making with machine learning."

## Learn More About Severstal

<https://www.severstal.com/eng/>

## Results

- **Reduced plant downtime:** "In the first six months since implementing our new system based on machine learning models fed by data streams with Confluent Platform, we saw several times fewer incidents involving mission-critical production equipment," Fetisov says. "Other factors may have been involved, but it's clear that the system contributed to significantly reduced downtime, as well as fewer quality issues and lower energy consumption."
- **Completed initial deployment quickly:** "The simplicity of the launch of our Kafka cluster and the reliability of its basic features played a big role in our initial success as did support from Confluent during the deployment phase," says Fetisov. "Even with a small team of just two or three engineers it was easy to build up the cluster – it didn't have to be tuned precisely for us to start seeing the benefits."
- **Received support for securing in-transit data:** "It is not easy to find qualified engineers with expertise needed to secure communications between brokers and consumers," says Fetisov. "As a result, this is one area in which we have relied more heavily on Confluent for support. If we face some difficulties configuring a new security feature, for example, then we simply open a support ticket and we know we will get the support we need."
- **Achieved one-second latencies:** "With Confluent Platform, we have latencies of about one second from raw data being produced by our equipment and a prediction being generated by our models," says Fetisov. "That's very good for us because low latency was one of our key requirements."