

DRIVING LOGISTICAL EFFICIENCY AND SUSTAINABILITY WITH BIG DATA

300K

Connected vehicle fleet

Impact

- Data processing from a 300,000 connected vehicle fleet
- Continuous model of analysis, development and improvement
- Much more efficient and sustainable model powered by data analysis

Scania is a world leading provider of transport solutions and is leading the shift towards sustainable transport systems. In 2018 it delivered 88,000 trucks, 8,500 buses as well as 12,800 industrial and marine engines to customers. Research and development is in Sweden, with branches in Brazil and India. Production takes place in Europe, Latin America and Asia, with regional production centres in Africa, Asia and Eurasia.

From its creation in 1891, Scania has seen the transportation industry mature and is anticipating the next phase of evolution as companies seek to be more sustainable in the systems they use. A by-product of this is the appetite to make operations as efficient as possible, but often those within the business struggle to understand exactly where the inefficiencies lie and consequently how to solve the problems causing them.

Having already set out a clear strategy for becoming a connected business, Scania was ahead of the curve in terms of gathering insights through data analysis. The company now needed to move towards a data-driven mindset to make use of the information coming from the 300,000-strong connected fleet.

Harvesting the potential

Although collecting terabytes of position data is a good idea it will not automatically create value. Scania saw the need of creating a data product with the ability to transform position data into business insights in a reproducible, reliable and scalable way. This was beyond the capability of traditional ways of working with data and new processing paradigms and tooling were needed.

"The process of creating a data product has a lot in common with the process of software development. We work in iterations and the ability to reprocess the data quickly and reproducibly is crucial in order to fail fast and move forward with speed," said Gustav Rånby, Senior Data Scientist at Scania.

Going from query to solution, repeatedly, to scale

The streaming paradigm in combination with the processing power of the data lake proved to be a success factor. Also, a common data platform facilitated cross-functional collaboration between the business, data scientists and data engineers.

Cloudera and Scania worked together on the data processing which was done using Apache Kafka and Apache Spark to translate the desired business outcome into data language. By using practices from DevOps and Continuous Delivery an end-to-end data pipeline was created.

Tools like Ansible allowed Scania to translate the process from one environment to another; with a structure, playbook, roles and task lists. With Jenkins as a glue to chain the projects and processes together in a truly DevOps workflow, Scania has a verifiable, reliable and repeatable streaming big data application.

"Data has become the vehicle that takes us from problem to solution, and the data lake, processing and data science team have worked with us to build a 'map' that can turn this into a reality. The ability to tap into data, trends and insights is truly powerful and promises to be the foundation for our sustainability mission long into the future."

Sara Björklund, Strategy and Business Development Manager, Scania Group

"By processing the position data from the trucks in a reproducible and scalable way and decorating it with other data sources in the data lake we were able to show various stakeholders what true business value looks like," said Sara Sylvan, Data Scientist at Scania. "A data product based on truck position information creates a foundation for data scientists to produce actionable insights and ultimately value across different parts of Scania."

Generating real business value by painting a picture with the data

There is no doubt that experience combined with trial and error can overcome business problems, but this is impossible to repeat or scale. Similarly, while data can be analysed and processed, it doesn't mean much without the context of why it has been selected and what is happening on the front line. Scania was able to gather those within the business that understood the problem and outcome they needed, with those that could manipulate the relevant data to achieve mission-critical business outcomes.

For example, while a coordinator might find it useful knowing where all the fleet is at one time, this doesn't necessarily show where efficiencies can be made. But, if all the fleet is mapped, alongside other data, such as wait time, time of day, day of the week, and type of truck, it can focus a team on where to look for those efficiencies.

In practice, for example, the larger trucks may consistently find it hard to cross one particular bridge during rush hour traffic, slowing down their journey and ultimately producing more emissions. By jointly analysing seemingly disparate datasets such as truck size, location and time for example, Scania can paint the most accurate picture to adequately inform prompt decision making, whether it is a case of deploying a certain type of truck or creating a diversion to the route based on live streaming traffic information.

"Data has become the vehicle that takes us from problem to solution, and the data lake, processing and data science team have worked with us to build a 'map' that can turn this into a reality. The ability to tap into data, trends and insights is truly powerful and promises to be the foundation for our sustainability mission long into the future," concluded Sara Björklund, Strategy and Business Development Manager at Scania Group.