

CASE STUDY

Speedgauge Innovates Fleet Management with Advanced Data Handling Solutions

MINUTES

to process data, down from days, enabling rapid experimentation

PREDICTIVE

models forecast events like a driver's likelihood of a ticket next week

SAFER

fleets, with stronger safety measures and risk management across drivers

COMPANY

SpeedGauge

INDUSTRY

Driving analytics and performance

PROFILE

Driving analytics and performance company focused on effective solutions help fleets manage and improve driving behaviors to transform a company's approach to driving, reduce risk, strengthen business operations and enhance financial results.

"Big data is characterized by three Vs: volume, velocity, and variation. At SpeedGauge, we handle a moderate volume of data, but it's highly varied. The Shakudo platform has enabled me to experiment with various methods for managing nested JSON and missing elements in our S3 archives. The ability to use Jupyter notebooks and then convert them into scheduled jobs is a significant upgrade from our previous developer workflow, which used AWS EMR and PySpark."

— Matt Moehr, Ph.D., Senior Data Scientist, SpeedGauge

INTRODUCTION TO COMPANY AND INDUSTRY

As a key player in the transportation and logistics industry, SpeedGauge focuses on providing advanced fleet safety improvements and risk management solutions. One of the prominent challenges in this industry is the management of large and diverse data. Companies often encounter data in various non-standardized formats, a complexity that can hinder accurate logistics tracking and the extraction of valuable insights.

To solve this challenge, SpeedGauge has continually updated and adjusted their technology stack. Their journey began with traditional tools for handling their extensive and complex data. As the industry and technological landscape advanced, SpeedGauge recognized the necessity for more efficient and economical data processing solutions, spurring their exploration and adoption of innovative, next-generation tools.

PROBLEM IDENTIFICATION

SpeedGauge's challenge centered around the efficient and accurate management of their non-standardized data. The data that they had to deal with came from a variety of Telematics Service Providers (TSPs) which provides foundations for driver behavior analysis which has wide use cases around risk management, driver safety, insurance claims and CRA score improvement, each with their own unique data format. With no standardized logging methodology in the industry, trucking companies found it increasingly difficult to accurately track logistics data.

The problem was clear: the company needed a streamlined, cost-effective, and efficient system for handling and processing their diverse and complicated data.

PAST SOLUTIONS AND THEIR SHORTCOMINGS

The company's previous solution was reliant on EMR Spark clusters to manage their JSON-based geospatial temporal data. This system was effective but not without its drawbacks. The data processing tasks required considerable resources and engineering time, often taking hours to complete. This extended processing time led to inefficiencies, limiting the team's productivity and their ability to swiftly derive valuable insights from the data.

Shakudo introduced a user-friendly one-liner code to spin up Spark and other distributed computing clusters with preemptible nodes. This new approach was a substantial improvement over the former EMR setup, removed the Spark cluster management effort and allowed SpeedGauge to migrate existing Spark based pipeline as is to run on a lower cost setup. While the pipelines are running, SpeedGauge explored the possibility of migrating the code to using Dask. However, they found that their data processing required frequent group-by types of operations, which is not one of Dask's core strengths.

In their journey to find a more suitable solution, SpeedGauge explored various data processing tools including Spark, Dask, Polars, and ultimately DuckDB. One of the key advantages of their experimentation process was the flexible environment provided by Shakudo. With Shakudo's [data stack integrations](#), SpeedGauge had the ability to swap out one tool and plug in another without the need to leave their notebook environment or refactor their existing pipelines.

PROPOSED SOLUTION AND IMPLEMENTATION

The freedom to experiment with various tools directly in the notebook, coupled with Shakudo's robustness to handle these changes without disrupting existing workflows, provided a significant boost to SpeedGauge's quest for the right tool. As a result, they

were able to identify and adopt the solution that best met their needs – an in-memory database system, DuckDB, which provided a much-needed increase in efficiency and transparency in their data processing.

The introduction of PyArrow and DuckDB, supported by Shakudo, not only streamlined their operations and improved their driving analytics capabilities, but facilitated the deployment of their solution with [Shakudo Jobs](#).

Another benefit of the transition was the significant improvement in error logging. Prior to this, the debugging process was complicated due to the vague error messages from PySpark and Dask and long waiting time for workers to return logs. Now, with DuckDB, logs are instant and local, leading to substantial time and cost savings and highly reduced time to value .

LOOKING AHEAD

Now that SpeedGauge has an efficient and lightweight data pipeline to ingest all the GPS data into one unified platform, it built the foundation of advanced analytics and machine learning use cases with the rest of the Shakudo stack.

SpeedGauge and Shakudo's future collaborations revolve around developing machine learning models that categorize different types of fleets and driving styles. The objective is to enhance safety and suggest coaching methods to reduce speed violations. By leveraging historical driver data, SpeedGauge will be able to predict events such as the likelihood of receiving a ticket the following week. These predictions will lead to lower operational costs, and help support customers with on the ground recommendations.

To facilitate this, Speedgauge will extend the stack to include model training and MLOps tools for model deployment and production lifecycle management on Shakudo.

LEARN MORE

Are you considering modernizing your data processing operations, to improve the speed and lower the cost of your data stack? If so, learn more about [Shakudo](#) and the benefits of running a data stack on our platform.

GET IN TOUCH WITH US

Ethan Brooke

Senior Account Executive

ethan@shakudo.io

(437) 344-4349