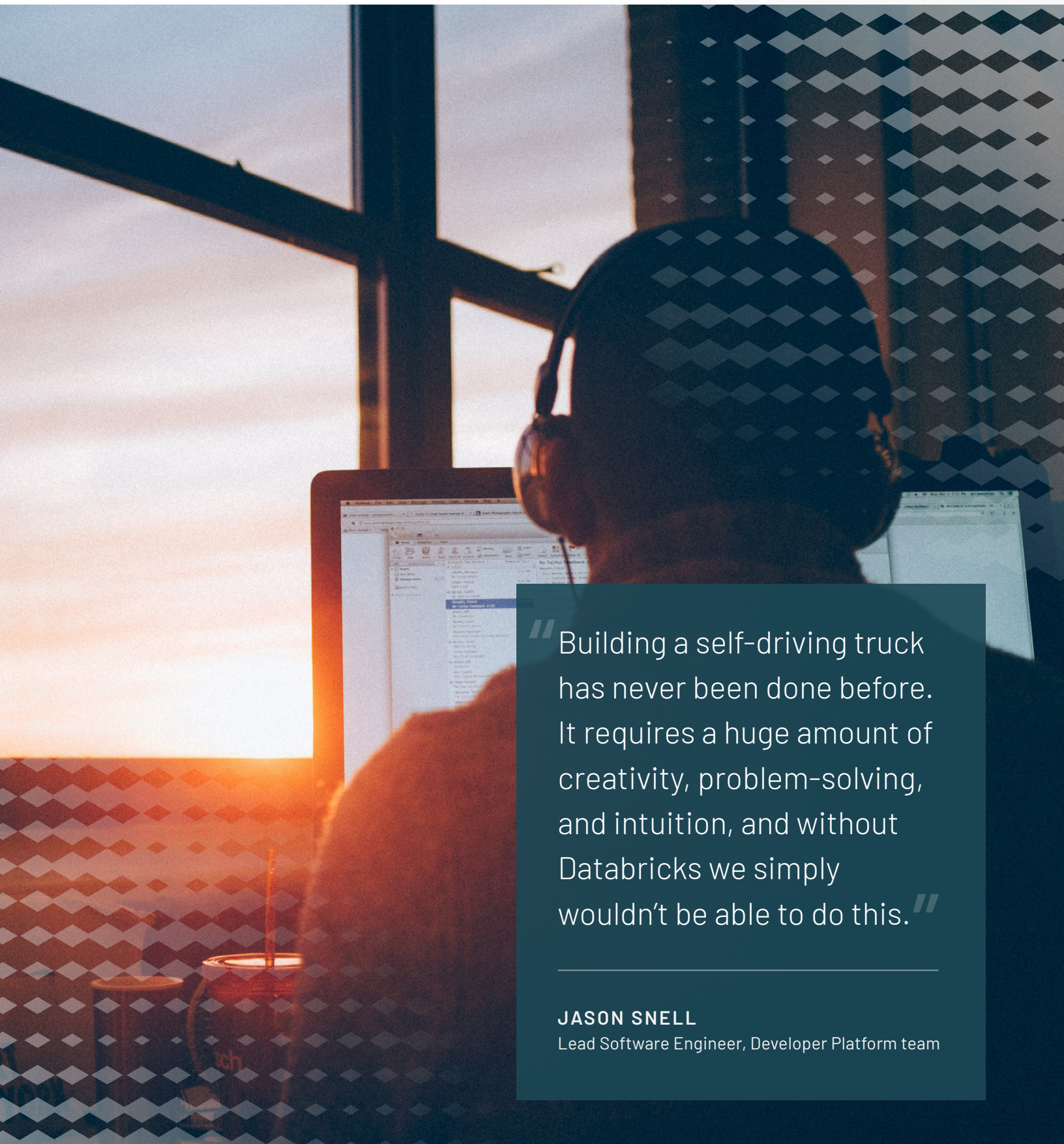


Databricks Customer Story: Embark





“ Building a self-driving truck has never been done before. It requires a huge amount of creativity, problem-solving, and intuition, and without Databricks we simply wouldn't be able to do this. ”

JASON SNELL

Lead Software Engineer, Developer Platform team

Ensuring long-haul truck safety with data and machine learning

Embark safely ushers freight trucking into a self-driving future with the help of Databricks

Embark is on a mission to move the entire trucking industry forward. They're building self-driving truck technology to make highways safer and the transport of goods more efficient.

Databricks is a fundamental partner in making Embark's vision come to life. With a unified approach to data and machine learning, the Embark team is now able to analyze their uniquely dense sensor data to safely and swiftly iterate on the models that lend the trucks their autonomy. According to [NHTSA](#), approximately 94% of vehicular accidents can be attributed to human error. By leveraging autonomous systems that never get tired or distracted, Embark's trucks, trained with the help of Databricks, can pay attention to the scene around them 24 hours a day, ultimately moving freight safely and more efficiently.

Approximately 94% of vehicular accidents can be attributed to human error. This is the driving stat behind Embark's mission. There's never been any doubt as to how big and how complex of a project creating self-driving trucks is, but for Embark, the reward of reducing that egregious number is invaluable as they are able to make a significant contribution to improving the safety of our roads while transforming the long-haul trucking industry.

INDUSTRY

Automotive

SOLUTION

Self-driving vehicles

TECHNICAL USE CASE

- Machine Learning
- Deep Learning
- Data ingest
- ETL



100%

SPEEDUP IN THE ANALYSIS OF
NEURAL NETWORK PERFORMANCE



75%

INCREASE IN THE TIME WINDOW THAT
COULD BE ANALYZED ON THE FLY

Human error causes the majority of vehicular accidents, but autonomy is a tall order

Addressing this challenge began with capturing the world around the trucks, which translated to a massive amount of recorded high definition (HiDEF) video, LiDAR (which is a method for measuring distances by illuminating the target with laser light and measuring the reflection with a sensor), and RADAR (which uses radio waves to determine the range, angle, or velocity of objects) data to train their models. Unsurprisingly, this amount of uniquely dense data presented a variety of technical challenges. For starters, Embark was working with in-house machines and individual laptops to access the data and run queries and analysis. But because of its scale, they were limited to the analysis of only 30 seconds of data at a time. When you have 37k hours of recorded data, analyzing 30-second clips just doesn't work.

In addition to limited data access, the use of single laptops prevented the team from storing enough meaningful data in a single location, and transferring it to and from their office was slow and costly. Any small failure or change meant a ton of engineering time and multiple days of machine time just to get the results they needed.

"We could only view maybe 30 seconds of data at a time. And that made it really difficult to make informed decisions," said Jason Snell, Lead Software Engineer of Embark's Developer Platform team.



Leveraging data and ML to enable fully-autonomous driving

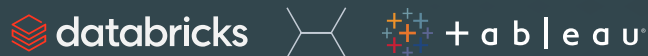
Once Embark implemented Databricks on their AWS infrastructure, everything changed for the better. The unified approach to data analytics made ingestion and ETL easy to manage, and the distributed platform was able to scale to meet their data needs. Automated cluster management simplified the provisioning of processing power without DevOps fuss, ultimately allowing data scientists to actually focus on data rather than infrastructure. Better yet, the Embark team went from single laptops with limited memory to hundreds of machines in the cloud – allowing them to analyze more data than ever before.

Cross-team collaboration also received a boost with interactive notebooks, which allowed the teams to write code and visualize all the results in one place, as well as easily share and collaborate with various cross-functional teams (data engineers, product management, data analysts and data science), offering complete transparency while boosting productivity. “Databricks has allowed us to unlock over 35,000 hours of recorded data from our trucks,” said Jason. “Our engineers can essentially access this data whenever they want, in whatever size they want, with whatever resolution they want.”



From an analytical standpoint, Embark’s data analysts are able to quickly generate dashboards – powered by Databricks – via Tableau to better understand how their software and sensors are performing.

This newfound freedom has allowed the Embark team to iterate much faster and more efficiently. It has also given them more confidence in the reliability and safety of what they are building, knowing that their models have been trained and verified by massive amounts of data.



Embark’s data analysts leverage internal metrics and generate dashboards via Tableau to better understand how their software and sensors are performing. Today, their data analyst team has a number of Tableau dashboards running – all powered by Databricks – including the performance of their software versions, sensor calibration performance, and more.

Scaling a moonshot: Speeding up a new industry with a unified approach to analytics

Databricks has ultimately enabled Embark to achieve the massive scale they need for data processing and model training, all while remaining as conscious of safety as ever. Before they deploy a new model on the road, for example, the Embark team measures its performance offline and ensures that it meets their safety criteria.

Prior to Databricks, every change would take days. Databricks has allowed Embark to scale this offline analysis to just minutes, powering significantly faster and safer iteration.

“Building a self-driving truck is an incredible challenge. It’s a moonshot,” added Jason. “It’s never been done before and it requires a huge amount of creativity, problem-solving, and intuition, and without Databricks we simply wouldn’t be able to do this.”

Looking ahead, Embark is excited about how Databricks will continue to play a pivotal role in increasing the pace of innovation for the self-driving industry. Together with Databricks, Embark is able to build a better freight company that delivers freight faster, more efficiently, and most importantly, safer.



“ Databricks has ultimately enabled Embark to achieve the massive scale they need for data processing and model training, all while remaining as conscious of safety as ever. ”

JASON SNELL

Lead Software Engineer,
Developer Platform team



About Databricks

Databricks is the data and AI company. Thousands of organizations worldwide—including Showtime, Shell, Conde Nast and Regeneron—rely on Databricks' open and unified platform for data engineering, machine learning and analytics. Databricks is venture-backed and headquartered in San Francisco with offices around the globe. Founded by the original creators of Apache Spark™, Delta Lake and MLflow, Databricks is on a mission to help data teams solve the world's toughest problems. To learn more, follow Databricks on Twitter, LinkedIn and Facebook.



EVALUATE DATABRICKS FOR YOURSELF

START YOUR FREE TRIAL

Contact us for a personalized demo databricks.com/contact