

WIDELY-ADOPTED COMMUNICATION PLATFORM IMPROVES QUERY PERFORMANCE & REDUCES COSTS WITH STARBURST

Leading Business
Communication Platform

Upgrading to Starburst Enterprise Presto has improved the user experience, strengthened security, and allowed the messaging leader to extract more value from its data lake.

This customer, provides a platform that effectively replaces email within organizations, allowing employees to easily exchange messages, share files, and collaborate individually or as teams. The technology has become an essential component of the daily way of doing business within companies across many industries. Their popularity has also driven a steady and tremendous flow of data. The company manages:

- 270M daily messages
- 215B daily logs
- 700B records, including messages

The company stores its data entirely in the Amazon Web Services (AWS) S3 cloud, but querying this data with open-source AWS EMR Presto had proven troublesome. Queries failed regularly, leading to increasing costs and a loss of confidence among various internal teams looking to glean insights from the data. This case study reviews the company's switch to Starburst Enterprise Presto, and how they're now enjoying improved performance, lower costs, enhanced security, and more.

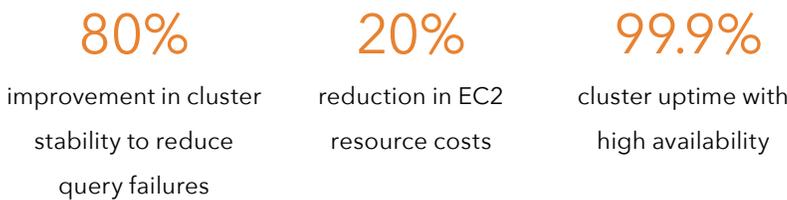
EMR Presto vs. Starburst Enterprise Presto

Prior to working with Starburst, this company's engineers were relying on EMR Presto, the open-source version of the world's fastest distributed query engine. The older version of open-source Presto that EMR uses lacks many performance improvements, security patches and other features offered by the Starburst Presto distribution. Query response times were slow, and queries were regularly failing before completion. Yet they still had to pay for the compute behind each query, even if it failed, which drove up costs.

“It was fine for about two years and then we ran into a situation where we could not scale the cluster anymore. Adding an extra node didn’t really help,” recalls a data engineer. “So we had to make a radical decision. That’s when we switched to Starburst.”

The Fastest Distributed Query Engine

Starburst Enterprise Presto is an enhanced version of the world’s fastest distributed query engine, purpose-built for the enterprise, with added features and bug fixes that drive efficiency, stability, performance, and security. The upgrade to Starburst Enterprise Presto resulted in:



After upgrading to Starburst Enterprise Presto, their data engineering team saw immediate changes. Previously, the number of queries had dropped from 30,000 daily to 20,000 per day, but many of these were repeats. Analysts would try their queries over and over after repeated failures, driving up EC2 costs. Once they switched its query engine to Starburst Enterprise Presto, this pattern stabilized, and Slack began seeing improvements in four key areas.

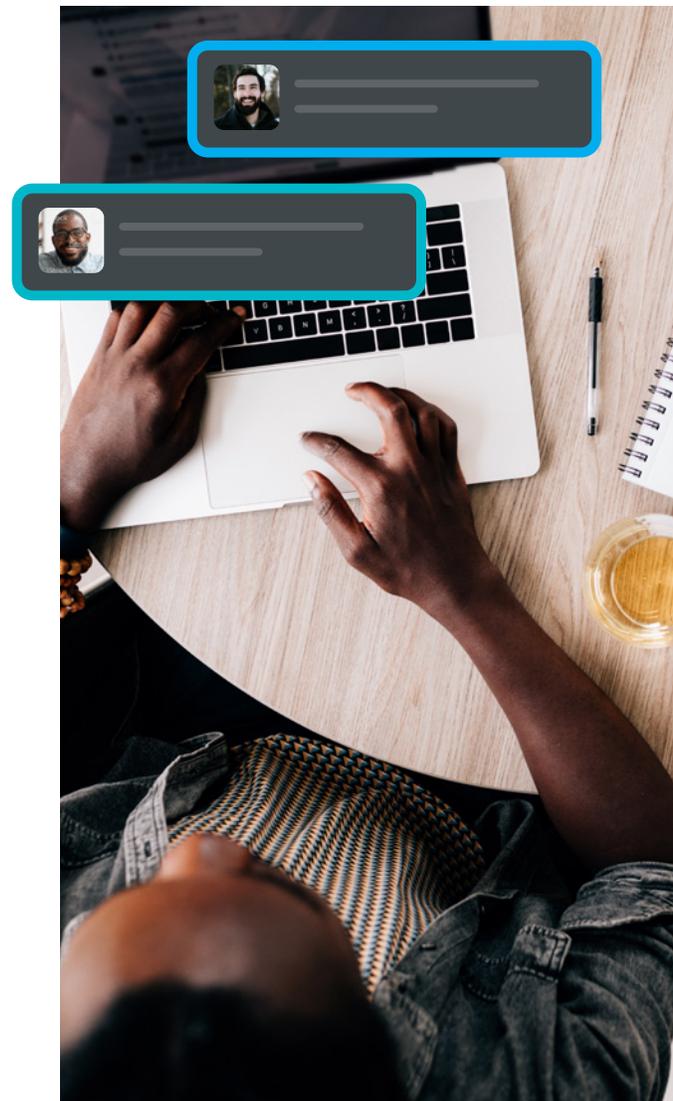
Performance & Uptime

The switch from EMR Presto to Starburst Enterprise Presto improved the query success rate and efficiency for the business intelligence team. Previously the query success rate was volatile, fluctuating between 50-90%, and now is consistently above 99% with Starburst. Regarding efficiency, Starburst’s cost-based optimizer ensures that queries are as fast and efficient as possible, with 99.9% cluster uptime.

Cost Savings

The huge improvement in query success rate and cluster reliability translated into a large reduction in EC2 costs. “A lot of people were re-trying the queries again and again until they worked,” says a data engineer. “We no longer have to do that, and this basically saved us about 20% of our resource cost.”

“A lot of people were re-trying the queries again and again until they worked,” says a data engineer. “We no longer have to do that, and this basically saved us about 20% of our resource cost.”



Cluster Tuning

This company needed to scale its capacity as additional users and workloads began using the platform. The Cost-based Optimizer automatically optimizes queries, but Starburst advised this company to further tune its approach by deploying multiple clusters—one for Ad Hoc queries and another for Data Science and ETL operations.

“Having a single cluster and trying to run every query on it was extremely hard,” explains a data engineer. “If you’re trying to optimize for one query you might make another query fail.”

With Starburst Enterprise Presto and the ability to separate storage from compute, customer was able to optimize one cluster for fast, simple queries, and the other for more complex queries and more users using the cluster simultaneously.

The Center of the Data Engineering Universe

The Data Engineering team’s success with Starburst Enterprise Presto has prompted the group to center its business intelligence operations around the platform. “We are the custodians of all the data that the product generates, and Presto is more or less at the center of our universe,” says a Senior Engineering Manager.

This company’s partnership with Starburst also gives the company input into future features and enhancements, along with enterprise-grade support from leading Presto experts. “In data infrastructure, we are about six engineers,” adds the Senior Engineering Manager. “We manage all of the distributed systems that make up the data warehouse. Our partnership with Starburst means we can leverage our limited engineering strength to solve our problems.”

“We are the custodians of all the data that the product generates, and Presto is more or less at the center of our universe”

“Our partnership with Starburst means we can leverage our limited engineering strength to solve our problems.”