

Zalando Powers Data Analytics and Reduces Costs with Starburst and AWS

 **zalando** To extract more value from its growing AWS

Amazon Simple Storage Service (Amazon S3) data lake, European e-commerce giant leverages Starburst for security, enterprise support, and superior performance

Zalando SE is Europe's leading online platform for fashion and lifestyle. Europe's most fashionable technology company boasts \$5B in annual revenues, largely from selling brand-name clothing and footwear online. The massively popular company has seen both its business and its data grow steadily.

Making smart, analytics-based business decisions is more important than ever in this uncertain global economy. When Zalando decided to transition from legacy data warehouses to an AWS Amazon S3 cloud data lake, the company needed to give its buyers and business analysts a more efficient way to extract value from this distributed data. Zalando deployed Starburst Enterprise for superior stability, fine-grained security, cost savings, and more.

Driving Better Business Decisions

Siloed data warehouse platforms, costly to scale and analyze

Analytics, data science, and machine learning drive the recommendation engines that increase Zalando's customer sales and help the company's buyers determine what to acquire and sell. When business analysts suggest that buyers purchase 10,000 pairs of jeans from a certain vendor, at a specific price, this decision cannot be based on a hunch. It has to be driven by data.

Zalando uses Spark for its data transformation and data science activities – to build and train the machine learning models that drive its recommendation engines. But the company also has thousands of business and marketing analysts who need a fast, simple, reliable way to query Zalando's data through their chosen Business Intelligence tool. The company has 3,000 MicroStrategy users alone.

Trino, the world's fastest distributed SQL query engine, was the initial choice, but the open source deployment couldn't deliver the efficiency, fine-grained security, and enterprise-grade features Zalando required.



"The decision to deploy Starburst Enterprise was made simpler because it has proven to be a reliable, fast, and stable query engine for S3 data lakes."

Alberto Miorin
Engineering Lead, Zalando

Solution

Zalando selected Amazon Simple Storage Service (Amazon S3) as the place to build its data lake due to its scalability, cost-effectiveness, and durability. "We can discover, access, and analyze data in our data lake with our preferred tools, and leverage it for business intelligence and data science," says Engineering Lead Alberto Miorin of Zalando. "This streamlined workflow helps our executives make the right decisions on time, and fosters innovation through machine learning."

Efficiency, Security, Support



Enterprise Support

The Starburst platform includes 24/7/365 support from true Trino experts, which has proven to be enormously valuable. "Enterprise support is a killer feature," says Miorin. "We actually had meetings with the creators of Trino."



Platform Stability & Performance

The data science and analytics teams at large enterprises don't have time to tolerate buggy releases. The Starburst platform deploys fully tested, stable releases that just work. Open source Trino was not underperforming for Zalando, but Miorin says that the Starburst deployment is more efficient.



50% Savings via Autoscaling

Starburst allows companies to scale their Amazon Elastic Compute Cloud (Amazon EC2) resources up to many machines during peak usage, and scale down at other times to save on costs, giving them unprecedented flexibility. "Auto scaling with graceful shutdown was another attractive feature," Miorin notes. "We are saving 50% on our AWS compute costs."



Security & GDPR

Due to GDPR, Zalando needed Starburst's fine-grained security and role-based access control features, which comes with Kerberos and LDAP integration, data masking, and encryption. Hot fixes and security patches are included as well.



Self-Service BI

The business intelligence and analytics teams can join and access data they couldn't query in the past without costly and time-consuming ETL.



Spark. vs. Trino

Zalando offers yet another example of how Spark and Trino serve different use cases within large enterprises, and are not viewed as competitors by the people who use their products on a daily basis. "Trino and Spark are complementary for us, not competing solutions," Miorin explains. "Spark prepares the data. Trino serves the data with fine-grained access control."

Ongoing Business Value

As the fully supported, production-tested, enterprise-grade distribution of open source Trino, Starburst Enterprise is constantly adding connectors, demonstrating Starburst's ongoing commitment to provide fast access to customer data no matter where it resides. With AWS and Starburst Enterprise, Zalando has the fast, affordable, secure data access the company needs to maintain its position as Europe's e-commerce leader in a difficult economy.



"We have gotten very good feedback from our users. With Trino, queries are now faster, and users can see the difference," says Miorin. "I cannot share the number, but our daily queries are increasing over time. We have observed that users are gravitating toward Trino."

Alberto Miorin
Engineering Lead, Zalando

