

Feature comparison

This guide is designed to help you understand the similarities and differences between Starburst and Amazon Elastic MapReduce (EMR)

Starburst and Amazon Elastic MapReduce (EMR) share common big data frameworks in their use of open source engines but differ in their deployment, features, flexibility, scalability and ease of use.

ABOUT STARBURST

Starburst offers a purpose-built SaaS platform for building and scaling production-grade data workloads in and around the data lake. It provides data lake analytics modules that support the full data lifecycle from discovery to organization to security and consumption.

Starburst's flexible engine allows users to perform both ETL and interactive queries with intelligent auto-scaling that can be tailored to the performance needs of individual data workloads. Starburst lets customers choose between its native features and third party providers for data cataloging, security policies and log-in.

For organizations with multi-region or multi-cloud architectures, Starburst supports cross-cloud query capabilities that maintain the highest levels of compliance without the need for data duplication.

ABOUT AMAZON EMR

Amazon EMR offers users provisioned and configured software including Apache Hive, Presto, Apache Spark, and Trino, delivered on AWS's managed Hadoop cloud infrastructure. With Amazon EMR, customers can execute large-scale analytics jobs, primarily on AWS S3, at a lower cost than traditional on-premise solutions. By decoupling compute and storage with Trino, users manually scale resources to meet the SLA of their data workloads. Amazon EMR secures data workloads through data encryption and table-level permissions. It relies on the underlying OS communities to advance its core set of features for performance, scalability and security.

Feature comparison



AMAZON EMR TRINO



Starburst

STARBURST GALAXY

PERFORMANCE

Trino OSS version

Manual upgrades

Automatic version updates

Smart indexing and caching

✓
via accelerated clusters

Fault tolerance

✓

CLUSTER MANAGEMENT

Deployment model

AWS-only

Run on AWS, Azure, or GCP

High-availability

Not guaranteed

99.5%

Autoscaling

Manual

✓

Idle shutdown

✓

Cluster scheduling

Manual

✓

API support

Limited

✓

CONNECTIVITY

Data source connectors

AWS-specific sources

Wide ecosystem of connectors including ADLS, GCP, Tabular, and Snowflake

Connector enhancements

Exclusive and improved connectors

BI tool connectors

Many
(Tableau, PowerBI, Thoughtspot)

Built-in query editor

3rd Party

✓

SECURITY & GOVERNANCE

RBAC

✓

ABAC

✓

AWS PrivateLink

✓

Audit logs

✓

ADDITIONAL PLATFORM CAPABILITIES

Graphical user interface

✓

Global search

✓

Data catalogs

✓

Data products

✓

Query sharing

✓

Cluster utilization graphs

✓

EXPERT SUPPORT

Support SLAs

Ticket-based

24/7 live support

Help and support from founders of Trino

✓

Feature comparison

SCALABILITY

Using Amazon EMR to run Trino as a big data framework lacks the components necessary for an enterprise-ready auto scaling experience. Additionally, the process of configuring auto scaling using 50+ enterprise data sources ranging from data lakes and warehouses to streaming systems, relational database systems, and more. EMR is limited and technically challenging.

In Starburst Galaxy, autoscaling is as simple as choosing the range of workers in a cluster with an intuitive user interface. There are no technical hurdles to address. This capability allows your organization to run a more efficient, cost-effective cluster.

With Starburst Galaxy, you can run both your long-running and interactive queries with a single engine. There's no need to use other big data frameworks for different workloads. An in-memory buffer service enables an exchange polling up to 60TB of data to support your longest running workloads.

EASE OF USE

Everyday management of Amazon EMR is technically challenging and requires dedicated resources. Starburst provides a management console to support your deployment. Starburst Galaxy features data products which enables data producers to create, publish, discover, and manage curated datasets. Data products enable increased business agility, and unlock domain-specific insights.

Starburst Galaxy abstracts away the complex management tasks traditionally associated with open source big data frameworks like Trino and Apache Spark. Starburst Galaxy manages all aspects associated with cluster management providing you with a simplified user interface. This allows you to connect to your data source, start a new cluster, and query your data in only a few minutes. Complex technical challenges like autoscaling and configuring connectors to data sources are all handled by Starburst Galaxy.

STARBURST ENTERPRISE

As companies start to scale up their query usage they need to address some core challenges with Amazon EMR. Management of open source Trino can be difficult without a user interface and autoscaling. This impacts resources and efficiency. In terms of stability, queries on EMR can fail without explanation or remedy. With regards to connectivity, EMR lacks enterprise data platform systems like Snowflake, Teradata, and Oracle. Companies often transition to the self-managed Starburst Enterprise platform as a result of these challenges, and to obtain support and expertise to optimize for larger deployments. They also transition to create and share data products that contain high quality datasets from all of Starburst's connectivity.

DOMAIN EXPERTISE

Amazon EMR does not offer support for Trino. There is a level of support that may include updates and patches. Starburst offers 24x7 support from the foremost Trino experts, project creators, and project maintainers. As a company, Starburst is invested in each customer's success, and our dedicated team of Trino experts help guide your team to success.

Get started with Starburst Galaxy

The analytics platform for your data lake
starburst.io/platform/starburst-galaxy

Feature comparison

CONNECTORS

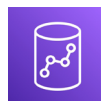
Please visit www.starburst.io/platform/connectors/ for complete details.

Starburst includes all OS Trino connectors as well as exclusive and improved connectors.

Starburst exclusive and improved connectors



*Includes Hive Metastore and HDFS (S3, ALDS, GCS, etc.)



The Data Lake Analytics Platform