



How-to Guide

Unlock the full potential of Trino



Introduction

Wherever you are on your data journey, let's face it: **Data is scattered, complex, and ever-evolving.** Data silos continue to be a challenge despite decades-long efforts to create a single source of truth.

Legacy data lakes and data warehouses such as Hadoop and Teradata attempted to address this problem by requiring all data to be landed in their systems. But it was hard for data analysts to run interactive queries on such large datasets without them failing.

Trino, formerly known as Presto SQL, was created in 2012 at Facebook to address this problem. It enabled Facebook to run analytics on their Hive/Hadoop data lake at petabyte scale without the need for unnecessary, costly data movement.

Today, Trino is a widely used, massively parallel processing (MPP) query engine for rapid, SQL-based analytics of big, distributed data. Companies such as **LinkedIn, Lyft, Netflix, GrubHub, and many others** have embraced Trino as a critical data architecture component to accelerate their access to data.

Trino at Scale



25PB on S3



1 Exabyte of Data
100PB weekly data
1200 nodes
2.5M queries/week



600PB on S3
1000 nodes



10PB daily read data
250k queries per day



300PB data lake



Enter Starburst

Like other open-source products, Trino requires resources and ongoing technical support to deploy at scale. The creators of Trino founded Starburst to help organizations extract the most value from their Trino investments.

Starburst offers a full-featured data lake analytics platform built on open source Trino. With Starburst, teams can access all of their data, lower infrastructure costs, use the tools best suited to their specific needs, and avoid vendor lock-in. Data teams from the largest enterprises to recently minted digital native businesses trust Starburst. With Starburst, you can evolve your Trino deployments to include enterprise-ready enhancements for improved performance, better connectivity, more security controls, and support from the experts.



**Richard
Teachout**

Chief Technology
Officer, El Toro

“ The needs we were struggling with was exactly why we ended up with Trino, and Starburst was a very natural next step for us as an Enterprise company. [Starburst] makes data access easier, better, more supported, more stable, and more developed without needing to put the resources in place. ”

How El Toro improved query performance by 300%

El Toro started on its path to democratize data access with open-source Trino, but found that the required resources and ongoing technical support to deploy at scale could be easily eliminated with Starburst, the enterprise edition of Trino.

Starburst would provide easy configuration on any environment, automated deployments anywhere, advanced monitoring, and 24x7 support by Trino experts to take the burden of deploying and maintaining the platform off of the data engineer's shoulders.

As a result, the data engineering team gravitated to Starburst Enterprise to further accelerate data access with features such as enterprise-grade performance, connectivity, security, management, and support.

"The needs we were struggling with was exactly why we ended up with Trino, and Starburst was a very natural next step for us as an Enterprise company – it makes data access easier, better, more supported, more stable, and more developed without needing to put the resources in place," says Richard Teachout, CTO at El Toro.

While the team did evaluate other options for data access, such as Snowflake and Databricks, the alternatives did not support their technical requirements at the scale the business demanded.

"Those options wouldn't work for our data and our volume. We could've made it work for an extra couple million dollars, but it wasn't worth the effort," Teachout commented.

[Read the case study >>](#)



el toro

Query federation made simple at Comcast

Faced with the problem of the need to access and perform analytics on a variety of data sources, Comcast decided that they needed to find a way to get multiple tools, platforms, and data storage systems to speak the same language, while still maintaining enterprise access policies, identity management, security standards, and performance.

Their data team built its first on-prem query engine using Trino to give its end users the ability to read data in Teradata and Hadoop without having to learn different query languages or request an ETL project. As a result, the end user experience improved, Comcast avoided data duplication, and their storage costs lowered. As more users started adopting the platform, that meant more teams wanted to access more datasets, too, including Delta Lake, Snowflake, Elastic, and others.

That's when Comcast decided to partner with Starburst. With Starburst, Comcast strengthened its hybrid analytics platform by adding enterprise-ready, high-performance connectors to more data sources, 24/7/365 support from the Trino experts at Starburst, a level of scalability that helps Comcast pull 250-300 TBs daily into the platform.

Now, this hybrid query fabric platform, which includes both Trino and Starburst implementations, provides hundreds of Comcast users with access to thousands of datasets, without data duplication. The data resides in its source system, and the end storage platforms have been abstracted away. Users can access everything through standard SQL.

[Watch Bryan Aller's talk, "Hybrid Analytics: Ground to Cloud, Cloud to Cloud" >>](#)

**"This doesn't just work.
It's fast."**

Bryan Aller

Director, Software Development
& Engineering, Comcast



Comparing Trino and Starburst

Trino is a powerful engine that requires extensive resources to manage and scale.

Inadequately resourced teams are challenged with deploying and scaling infrastructure and managing day-to-day operations. Data teams turn to Starburst to ease that burden and help reduce their total cost of ownership (TCO) by up to 50%.

To illustrate the full business value of Starburst's data lake analytics platform, we look at the following benefits:



Performance: Get access to enhanced and additional connectors for data sources like Snowflake, Salesforce, Teradata, and others, as well as performance improvements, such as Starburst Warp Speed.



Scalability: Easily manage and scale your architecture to support ever-growing demand, whether you manage Starburst in-house or let us do the work for you with [Starburst Galaxy](#).



Simplicity: Operate more efficiently with an intuitive interface and out-of-the-box query editing, data discovery, and data product capabilities.



Security and governance: Leverage built-in access controls with role as well as attribute-based configurations, and guarantee enterprise-grade security from the client to the underlying data source.



Support: You are backed by a professional services team and the largest group of Trino experts in the world who also just happen to be the leading contributors to the project.

Performance

Starburst has made a series of enhancements on top of open source Trino to significantly improve query performance.

Starburst provides improved and exclusive connectors which manage and access statistics allowing the query engine to generate faster and more optimal query plans. Starburst offers proprietary features like Starburst Warp Speed which provides autonomous workload acceleration for your data lake queries with its proprietary indexing and caching technology.

	 trino	 Starburst
Petabyte scale	✓	✓
Accelerated Parquet	✓	✓
Materialized views	✓ for Iceberg	✓ for Iceberg and Hive
Complex expression pushdown		✓
Warp Speed (smart indexing and caching)		✓

“With Starburst, we have accelerated data discovery, simplified data pipelines and have a unified query layer across all data sources. These three points are critical to what we do.”


Patrice Linel, Senior Manager of Data Science and Data Engineering, Genus



Scalability

As your Trino usage scales, it becomes a larger and larger task to guarantee system reliability. Questions arise about machine sizing, scaling logic, and query optimization that become harder to answer. Starburst offers best in-class support from Trino experts to guide you along your scaling journey.

Need an extra hand? Starburst abstracts all the overhead away with Starburst Galaxy, a fully-managed data lake analytics platform. Galaxy handles operational tasks like machine sizing and autoscaling for your team to save you time and minimize downtime risks with 99.9% uptime guarantee.

	 trino	 Starburst
Resource groups	✓	✓
A single end-to-end platform		✓
Autoscaling and cluster scheduling		✓
Fully-managed offering		✓
Flexible deployment and billing models		✓

“Starburst gives us a single platform to explore more data through connectivity, maintain data quality and governance, and provide the data to all of our employees using their visualization tools of choice. These are the big differentiators of the platform.”

André Gortari, Data Engineering Manager, Banco Inter



Simplicity

When using Trino, complexities can arise due to intricate architecture, extensive configuration options, and the need for specialized knowledge to effectively utilize its advanced features.

Starburst provides an easy-to-use interface, simplified installation process, extensive documentation, intuitive query tools, and optimized performance features, enabling users to interact with and query data efficiently without requiring extensive technical expertise.

	 trino	 Starburst
Trino command line interface	✓	✓
GUI or API-based management		✓
Query editor and sharing		✓
Data products		✓
Global search		✓
Schema discovery		✓

“The needs we were struggling with was exactly why we ended up with Trino, and Starburst was a very natural next step for us as an Enterprise company – it makes data access easier, better, more supported, more stable, and more developed without needing to put the resources in place.”


Richard Teachout, CTO, El Toro



Security and governance

Trino has numerous security features, but lacks a powerful and easy to use access control system.

Starburst provides a built-in role-based access control system that meets the requirements of some of the most security-conscious organizations of the world. It integrates with numerous authentication providers and open source projects like Apache Ranger.

	 trino	 Starburst
Kerberos support	✓	✓
Immuta AC support	✓	✓
Built-in role base (RBAC) and attribute-based access control (ABAC)		✓
Password credential passthrough		✓
User impersonation		✓
Query logger and auditor		✓

“We just want one place to configure security for all data access and Starburst Enterprise allows us to do that.”

Mike Prior, Principal IO Engineer, Optum

Support

Starburst has proven experience with deploying and supporting the largest enterprises in the world. With Starburst, you're backed 24/7 by the world's largest group of Trino experts.

	 trino	 Starburst
Online community	✓	✓
24x7x365 support (<30 min SLA)		✓
Fully-tested, stable releases		✓
Technical account manager		✓

“We get great support from the Starburst team on questions like how to make a query run quicker. That’s important for us and our customers, because the quicker we can run our queries, the quicker a clinician can get to a new insight, and the quicker a patient can receive better care.”

Richard Jarvis, CTO, EMIS

Upgrading from Trino to Starburst

As a first step, choose which Starburst offering is right for you:

Starburst Enterprise

Starburst offers **Starburst Enterprise** as a distribution of Trino with enhanced and additional features. You can install it just like Trino, and query your data wherever it lives – on-premise in your own data center, in the cloud, metal, virtual machines, or containers – just like Trino, Starburst Enterprise offers a wide range of options and is flexible in terms of where and how you run it.

Starburst Galaxy

Starburst Galaxy is a fully-managed data lake analytics platform built on top of Trino. It takes care of the heavy lifting of designing, provisioning, maintaining, and securing your data infrastructure while making it easy to discover, govern, and consume your data.

Both choices include features for your cluster management and operation as well as query editor, data products and other enhancements for your analytics users.

Upgrading from Trino to Starburst



When you upgrade to Starburst Enterprise from Trino, you can use all the same infrastructure you used for Trino. In addition, Starburst Enterprise requires a backend service database that you have to provision and manage.

If you run Trino from the RPM or tarball, in the cloud or on-premise, you now have access to [starburst-admin](#) to automate the deployment of your configuration to all your nodes in a cluster. All your catalog and other configuration files seamlessly work with Starburst Enterprise.

Here are the high-level steps you will take:

- Provision your cluster and configure SSH access as needed by Ansible used inside of starburst-admin
- Configure starburst-admin on your workstation or separate server that you use to manage your clusters
- Add the catalog and other configuration files from Trino
- Add your Starburst Enterprise RPM or tarball archive
- Add your license
- Adjust the configuration files to add the backend service database details
- Use starburst-admin to install Starburst Enterprise on the coordinator and all workers, and start up the cluster.

Now you are ready to test the cluster, and start refining your configuration.

If you are already using a Kubernetes-based deployment with Trino, you get access to more [powerful Helm charts](#), and you also get Helm charts to manage your Hive Metastore Service. When you upgrade you need to transcribe your global configuration and your catalog configuration into the new configuration files for the Starburst Helm charts, and start a new cluster with the same configuration.

In both cases you can start your cluster and have access to additional features like a more powerful web UI, more connectors, built-in access control, performance improvements like Starburst Warp Speed, inter-cluster connectivity with Starburst Stargate, and other features.

A successfully running cluster allows you to perform next steps such as activating the built-in access control, adding more catalogs, adding more users and starting to explore all the features of the web interface.

Find more details about all features and more resources in the [documentation](#).

Upgrading from Trino to Starburst



Before you begin your upgrade, confirm that:

- all your data sources are available in the cloud, and supported by the [available catalogs](#)
- your organization is able to use a software-as-a-service running in the cloud

To start, create an account, and gather the necessary connection details for your data sources. This means you need to look at each catalog properties file from your Trino deployment and gather details such as data source host URLs, username, password, and also where the data source is located.

As a next step you use those details to create one or more catalogs in the web interface of Starburst Galaxy following the detailed steps in the documentation. Make sure you use the same name for the catalog in Starburst Galaxy, that you had as the filename in Trino. For example, if you had a file `etc/catalog/bigdatalake.properties`, the name of the catalog must be `bigdatalake`. Following this practice allows you to use the SQL queries from your Trino cluster without changes.

Once you have created all the catalogs, you must add them to one or more clusters. For starters, use the default cluster execution mode. With all the catalogs added, start the clusters.

Now you can query your data sources right from the web interface with the built-in query editor. Also access the cluster connection details and connect the same BI tools you used with Trino. Alternatively try one of the additionally supported client tools.

With the cluster operating and catalogs configured you can explore additional features. Run the discovery service to find out details about your data lake, save queries, or create a data product. At this stage everything is up and running and you can take the next step and create additional users and configure security with built-in access control.

Just like that, you are well on your way to faster analytics with Starburst Galaxy. Find more details about all features, demo videos and more resources in the [documentation](#).

Conclusion

Even if this is your first time hearing of Trino, formerly known as Presto SQL, Starburst can be the foundation for your modern data lake strategy, enable you to scale affordably, and help you cut down on wasted time and resources. With Starburst, teams build new features faster and spend less time on maintenance and overhead.

Take your Trino usage to the next level and sign up for a free **Starburst Galaxy** account to see what you are missing. And if you are interested in a custom evaluation tailored for your specific Trino use case, **contact Starburst** for help with Starburst Galaxy or Starburst Enterprise.



