Starburst for ELT data processing

## The reliability you expect, with the performance you desire

Trino was created primarily for interactive or ad-hoc analytics. The MPP SQL engine features best-in-class performance, support for standard ANSI-SQL, and the ability to query data from multiple data sources simultaneously. Its superiority as an engine for interactive analytics naturally led to interest in using it as a unified engine for all workloads, including long-running ELT/ETL workloads. Developments to open source Trino have enhanced the engine's ability to handle these workloads by enabling query fault tolerance (query and task level retries) for long-running queries.

Starburst completes Trino, with enterprise-ready enhancements to performance, connectivity, security, and data discovery. The combination of enterprise enhancements made by Starburst with open source Trino enhancements makes Starburst the data processing engine of choice for data teams. Starburst provides a reliable and consistent engine for processing long-running queries in a cost effective and highly performant manner.

With query fault tolerance enabled on a Starburst cluster, if a long-running query experiences a failure it does not have to restart. Instead, a cluster will retry the query at the task or query-level. This provides data engineers with the reliability necessary for long-running data processing jobs like ELT. The marriage of query fault tolerance and the engine's best-in-class performance improves outcomes by providing consistent landing times for long-running data processing workloads.

Flexibility and performance no matter the workload
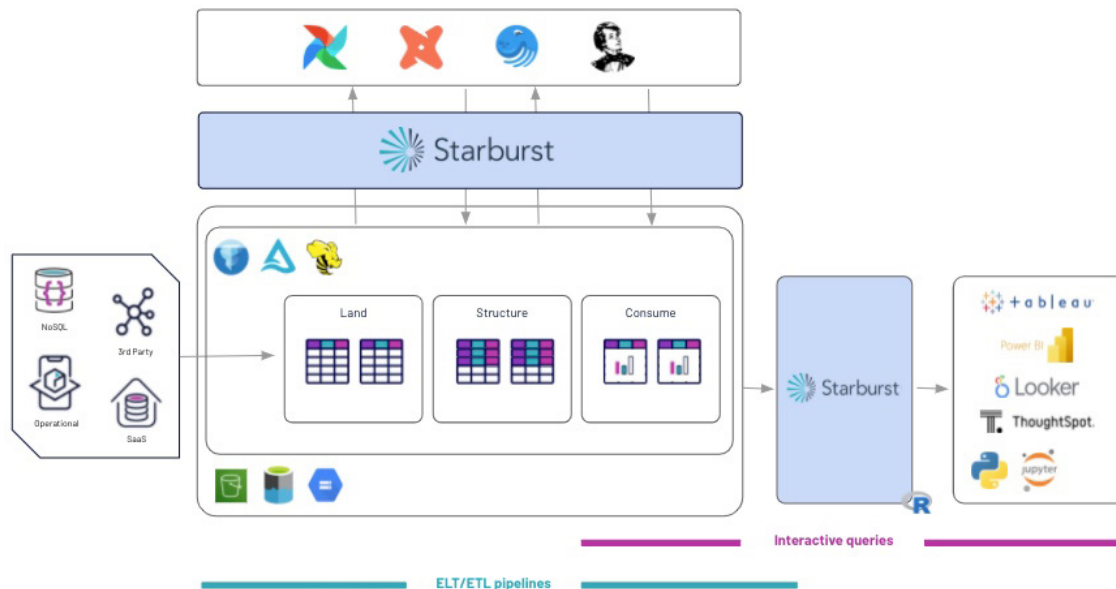
One engine for data processing and interactive workloads

Starburst's engine is designed for interactive and batch workloads simplifying your data architecture and reducing the number of analytics tools required in your workflows. Starburst for ELT data processing provides cost savings, accelerates time to insight, and streamlines your data infrastructure.

Leverage connectivity to simplify your architecture

Starburst includes over 40 enterprise connectors providing access to a wide array of data sources including data lakes, data warehouses, relational database services, and much more. Starburst engine's connector-based architecture allows you to connect directly to multiple data sources and use one SQL statement to bring the data together in the target system. This data federation capability helps organizations achieve a greater degree of flexibility by simplifying data access.

Improve discovery and consumption

Starburst improves data access, discovery, and consumption by providing a simple framework to query data using SQL and a rich ecosystem of integrations. Starburst's query engine is ANSI-SQL compliant, enabling your analyst team to create new data pipelines and reduce organizational bottlenecks. Starburst has integrations with tools like Airflow, dbt, Great Expectations, Dagster, and more to help manage your ELT workflows.

Scales to meet your organization's needs

Organizations leverage Starburst's engine to provide a true separation between the storage and compute layer of their data architecture. This separation allows organizations to scale as their data needs grow, allowing for reduced infrastructure costs via elastic scaling. This decoupling also enables the use of inexpensive data storage options while still reaching target performance benchmarks.

In the cloud, or on-premises, we're there for you

Starburst is adaptable to all environments whether your organization's data is on-premises, or in the cloud. Traditional approaches to data architecture rely on a monolithic design which promises a single source of truth, but also comes with heavy risks associated with having your data on one platform. Starburst provides the optionality, and flexibility, by giving you access to data directly where it lives, regardless of data source, cloud, region, and federates across all of them. These deployment options, with the ability to handle ad-hoc, and long-running batch workloads, gives you the freedom to access data, and transform data, no matter the size and complexity. It provides the flexibility to create more efficient architectural patterns in your data ecosystem.

Driving better performance with ELT data processing features

Data lake analytics

Starburst includes all the necessary components to enable a highly performant data lake analytics engine. Users can easily transform, join, and enrich data, without reducing performance unlike traditional data processing engines. Starburst as a data lake engine includes support for data manipulation language and write operations on the most popular object storage and open table formats, to empower data analysts, ,and engineers, with data transformation capabilities to empower mission critical analytics. These powerful tools help in each step of the ELT process on the data lake.

Cross-cloud analytics

Starburst Stargate unlocks organizational value by linking catalogs and data sources supported by one Starburst cluster to other catalogs and data sources in remote Starburst clusters. Starburst Stargate is a gateway for providing data access across geographic regions while ensuring access controls are in place and data residency requirements are honored. Starburst provides an even greater degree of flexibility, across clouds and regions, to support new dynamic ELT data processing patterns.

Data products

Easily create, publish and share a curated, high-quality dataset and relevant data. Data products enables data producers, data consumers, and domain experts to generate insights without the burden of a centralized data strategy. Data products allows engineers using ELT to create and share insights across the broader data team more efficiently. This shortens the path to insights on the most useful data sets, providing organizations with the data needed to continue to innovate and outpace the competition.

Great Lakes connectivity

Great Lakes connectivity provides unified access to open t able formats like Hive, Delta Lake and Iceberg in Starburst Galaxy. Choosing the file format to create is handled with simple SQL and querying is transparent to the end users. Open table formats allow for even greater performance gains when running ELT workloads.

For companies large and small, Starburst is the ELT engine of choice.

Leading messaging platform

This leading messaging platform manages 270M messages, 215B logs, and 700B records daily. After choosing Starburst, deployed to access Amazon S3, they achieved cluster optimization for fast, simple queries, while still having availability for more complex queries and more concurrent users. Starburst improved the user experience, strengthened security, and allowed the messaging leader to extract more value providing a highly performant engine for data processing workloads. As a result, the platform achieved 80% improvement in cluster stability to reduce query failures, improved query success rate from 50% to 99%, and 20% reduction in EC2 costs.

To learn more about Starburst for ELT data processing, visit starburst.io

Starburst
ANALYTICS ANYWHERE