

Starburst Lakehouse

Accelerate data lake analytics and actionable insights

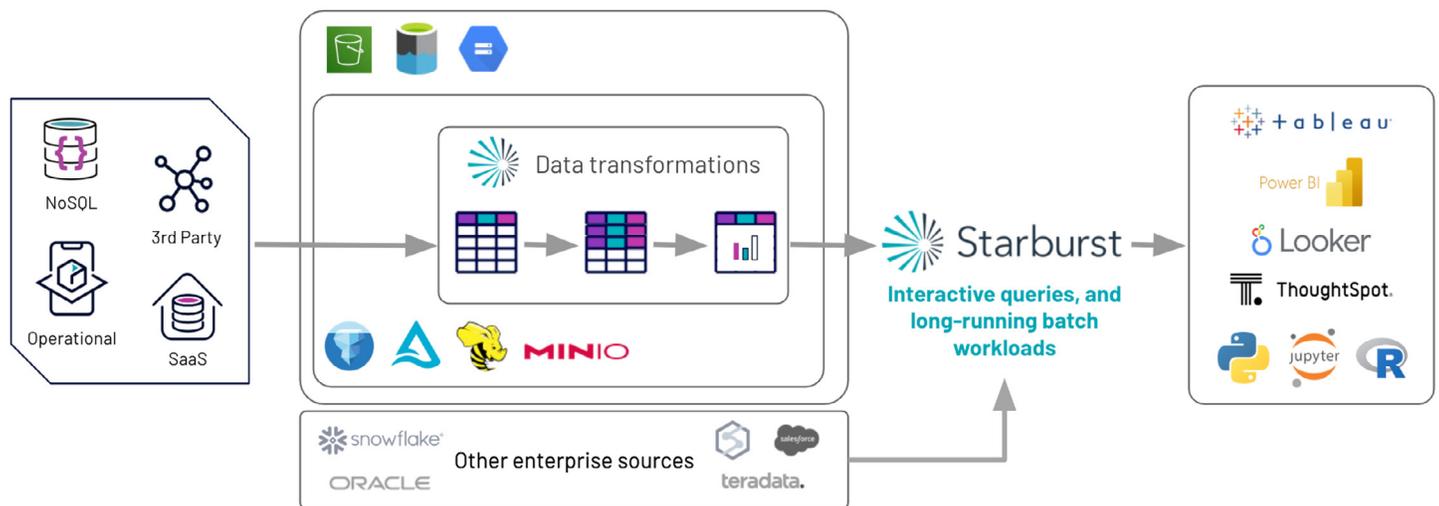
Data lakes and open file formats offer cheaper storage and lower overall costs relative to data warehouses. Although the data lake model decoupled compute and storage, it historically has failed to offer data reliability and performance modern enterprises require for mission critical analytics. Making the right data available in the right form is time-consuming, and is compounded with some slow and cumbersome existing solutions for accessing data in the lake. Data often remains isolated and unused.

To achieve typical 'data warehouse analytics' on a data lake requires a flexible, performant, and secure query engine. Starburst liberates organizations' data lakes with live, interactive queries directly on cloud data lake storage, allowing data teams to adopt a modern data lake engine approach. Starburst, based on open source Trino, is the best in class SQL-based MPP query engine for data stored in data lakes. Starburst lakehouse enables high concurrency, scalability, and performance while increasing productivity and lowering infrastructure costs. Starburst supports self-hosted, fully managed, on-premises, and hybrid deployment while also supporting Data Manipulation Language (DML) and write operations for the data lake. With Starburst you gain the ability to access and join data from disparate sources to access even more new insights.



Starburst lakehouse enables organizations to leverage the flexible storage and costs of a data lake, with the advanced analytics capabilities and data transformations required for critical workloads, powering the data pipelines, dashboards, and reports for companies large and small. The fully managed Starburst Galaxy enables live data lake access in minutes, and a self-managed solution, Starburst Enterprise, allows the customization for any enterprise infrastructure and need. Within one platform, customers can achieve fast interactive and ad-hoc queries, and the ability to run long-running queries, faster than other solutions, on relatively small clusters.

Empowering a new era of data lake analytics



Engine built for performance and flexibility

Starburst's MPP query engine is built for speed and performance at scale and is adaptable to all customer environments and technology investments, today and tomorrow. Performance is enhanced by advanced workload execution optimization - cost-based optimizations, pushdown, dynamic filtering, materialized views, parallel connectors, caching, and multi-regional metastore. We give organizations control over query response time and cost, all with an engine flexible to do ad-hoc and interactive queries, in addition, or long-running scheduled batch workloads.

Data lakehouse for start-ups and small businesses

First created to overcome the limitations of traditional data warehouses, data lakes offer the scalability, speed, and cost-effectiveness to help you manage large volumes and multiple types of data across your various analytics initiatives. This makes a great option for start-ups and small businesses without a mature analytics strategy. Starburst Galaxy, a fully-managed distributed SQL engine data platform enables organizations the speed and ability to deploy and access their data lake in minutes.

Joining data lakes with other enterprise sources

Starburst offers live access to 50+ enterprise data sources including modern and legacy enterprise sources. We deliver data warehouse functionality to all the major data lakes - Amazon S3, ADLS, GCS, and table formats such as Delta, Iceberg, MinIO, and more. With additional enterprise connectors like Snowflake, Oracle,

and Teradata, you can easily unlock new insights by combining data in the lake with other sources. Additionally, with dbt Labs, and Airflow integrations, data teams can continue to use tools (BI, ML/AI, workflow management, and governance) you know and love while getting more value out of your data.

Bringing the data lake to BI

Starburst also enables easy discovery and consumption of high-quality data through the creation and management of data products, data catalogs, and materialized views. Queried results can be consumed with client connections and integrations with your favorite BI platforms. We dramatically improve the speed and concurrency for ad-hoc, batch, and reporting queries live against your data lake storage, there's less dependence on IT or data engineering. Starburst enables you to do all your BI tasks at immense scale, without sacrificing performance, directly against cloud data lake storage.

Delivering security and compliance to data lake

All this speed, with the governance and security required for enterprises. Starburst offers centralized and fine-grained control over access to all your data- query auditing, RBAC, ABAC, data masking, encryption in-flight, and more. Customers can comply with various regulatory compliance standards, as Starburst allows you to safely analyze global data across clouds and regions.

Enterprise-ready features for your data lake analytics



Lakehouse analytics

Starburst includes all of the tools and features needed to build the modern lake house. Customers can easily transform, join and enrich data using standard SQL, without being strained like traditional engines. Starburst's data lake connectivity includes support for DML to empower data analysts, platform administrators, and engineers with advanced analytics and data transformation to support mission critical analytics.



Data products

Easily create, publish, and share a curated, high-quality dataset and relevant metadata. Data products enable data producers, data consumers, and domain experts to generate insights without the burden of a centralized data strategy. Data products can benefit any analytics journey, large or small, on-premises or multiple clouds, in a warehouse or a lake. And federates across all of them.



Accelerated Parquet reader

Trino has long served organizations with the ability to read Parquet files. And while already fast, Starburst delivers additional value with an accelerated Parquet reader which improves read performance on Parquet on average by 20%.



Great Lakes connectivity

Working with different file and table formats shouldn't be a hassle. Starburst Galaxy's Great Lakes connectivity provides unified access to Hive, Delta Lake, and Iceberg in only a few clicks. Choosing which file format to create is handled with simple SQL and querying is transparent to the end users so they don't need to worry about what's "under the hood".

Starburst is the data lake engine of choice for companies, large and small, across the globe.



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 from its data lake. 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.



Global cloud computing technology company

Frustrated with sub-optimal dashboard response times negatively impacting customer experience, this customer deployed Starburst in AWS federating across Amazon S3 and Elastic. As a result, dashboard responsiveness and overall portal experience improved to optimal levels, while meeting strict security requirements. The customer successfully decreased query response time from hours/minutes to seconds and achieved PCI/PII and GDPR compliance. The customer also has a 50% reduction in TCO, 30% reduction in EC2 spend due from autoscaling, 99.9% query success rate, and reduced time to insights by more than 100%.



To learn more about Starburst lakehouse, visit starburst.io