Starburst

# Discover. Govern. Analyze. Fast.

Data lake scalability, data warehouse functionality, and the reach of a semantic layer in a single, unified platform

Starburst Galaxy is a fully-managed lakehouse platform, powered by Trino, that combines the best of the data warehouse, the data lake, and data virtualization in a single platform.

| The Challenge | The Solution |
| --- | --- |
| Data lakes are hard to build and maintain | Add warehouse-like functionality to your data lake with modern table formats |
| Data warehouses are expensive at scale | Offload your costly transformation workloads to a data lake architecture with Starburst |
| Data virtualization often fails to meet performance requirements | Leverage industry-leading query acceleration supported by query pushdown and dynamic filtering |

## What are common use cases?

### Scale your data lake engine

Easily speed up, scale, and secure your data lake workloads with a fully-managed platform optimized for Trino, the leading data lake engine.

### Reduce your data warehouse costs

Reduce your cloud data warehouse costs and improve your operational efficiencies by handling your semi-structured workloads in a data lake – while getting instant access to all of your other data sources.

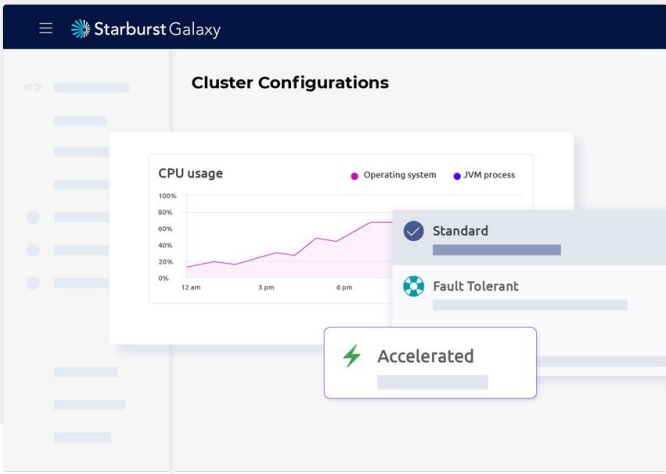
### Deliver dynamic data applications

Search and analyze semi-structured data in AWS S3, GCS, and ADLS in near real-time with automatic schema discovery and smart indexing and caching.

# Interactive data lake analytics powered by Trino

Starburst Galaxy is powered by open source Trino and is designed for analyzing large and complex data sets in and around your cloud data lake – from gigabyte to petabyte scale.

Easily tailor your compute to the individual needs of your workload with enhanced cluster execution modes like Warp Speed and query-level fault tolerance.




# Gravity

*A universal data discovery, governance, and sharing layer in Starburst Galaxy*

## Data catalog

Simply connect your data with a data catalog to automatically pull in and index relevant objects from the source.

Easily search across all of your data – across sources, regions, and clouds.

## Data governance

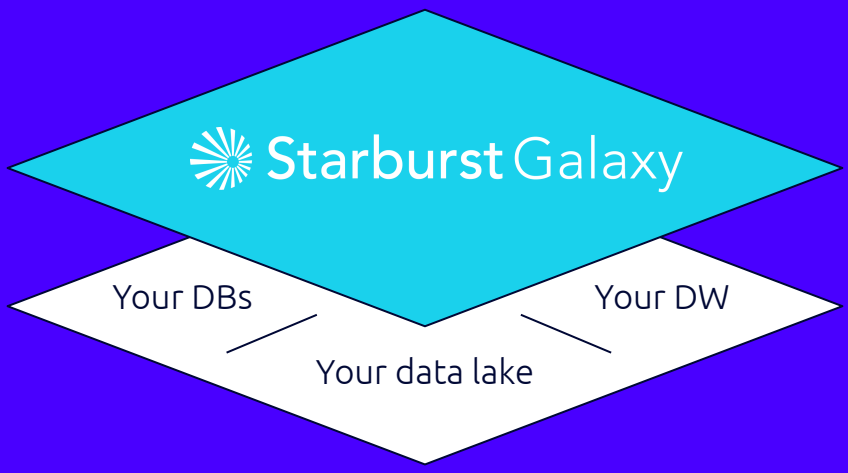
Elevate the security of your data with centralized governance across disparate sources.

Fine-grained access controls allow permissioning down to the row-level and include support for column masks.

## Data products

Natively curate, share, and govern data products.

Data teams can easily create new data products by seamlessly joining data from your data lake and surrounding sources without the hassle of data movement.




# Your storage, our platform.

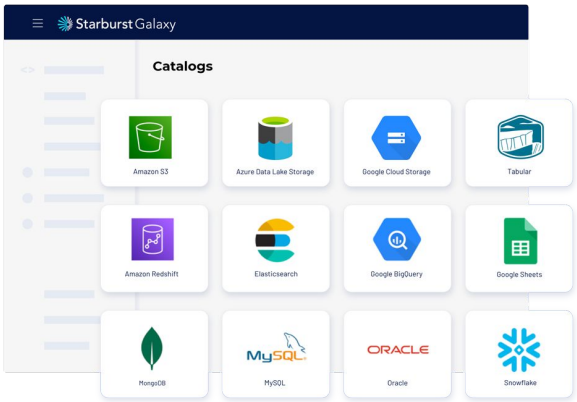
Optimized for Iceberg with support for all modern table formats, including Delta Lake, Hudi, and Hive

Runs on AWS, Azure, and GCP with ability to query across clouds and even your on-premise data

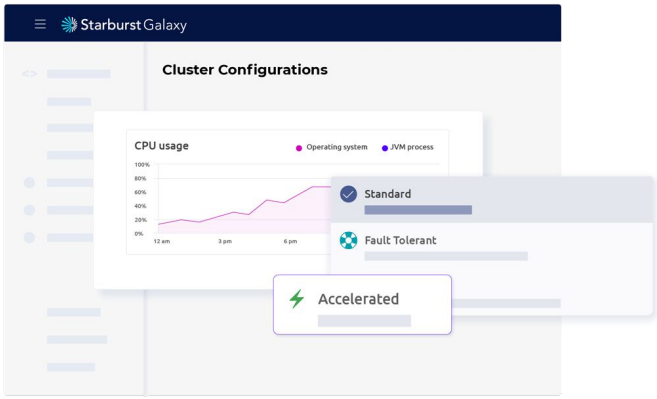
Secure from the start with built-in security controls for all your data

# How Starburst Galaxy works

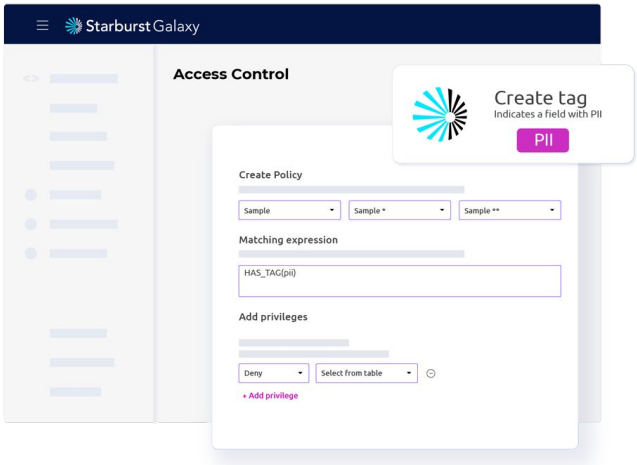
## 01 Connect and catalog your data



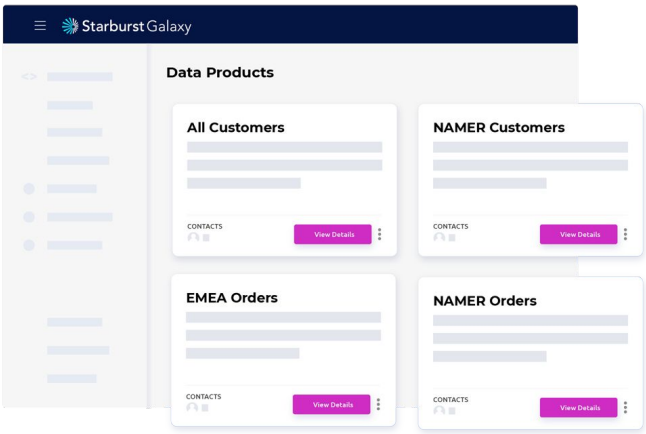

## 02 Tailor compute to the needs of your workload



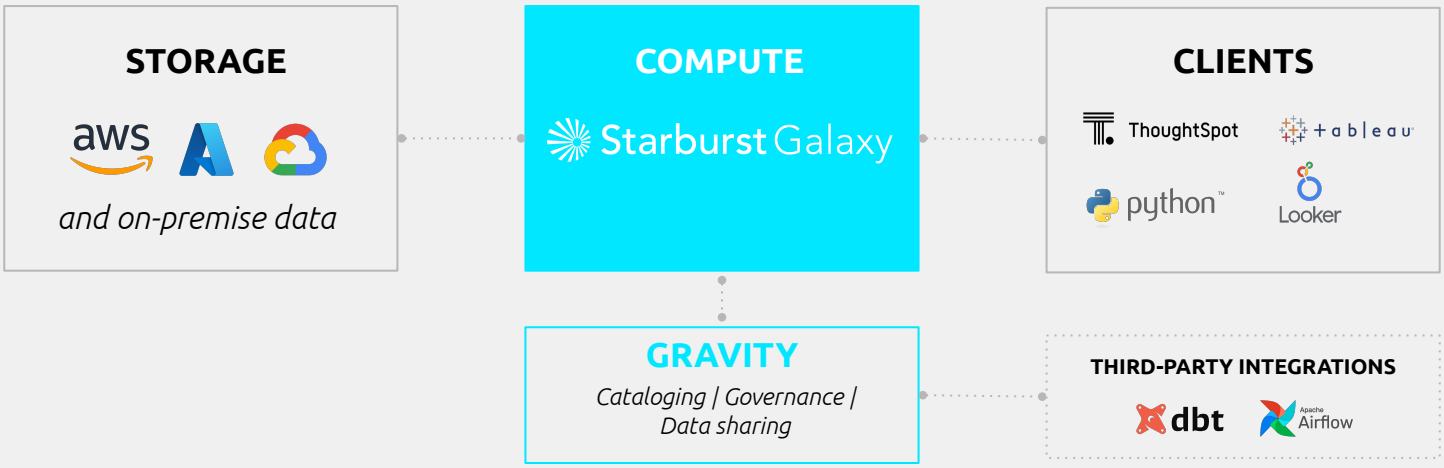

## 03 Govern your data using tags and roles




## 04 Create and share data with data products




# Architecture



# Starburst Galaxy Capabilities

| | CAPABILITY | DESCRIPTION | STATUS |
|---|---|---|---|
| DATA ENGINEERING | **Streaming ingest** | Ingest data from Kafka streams into your data lake for near real time analytics | Preview |
| | **Automated data lake maintenance** | Improve performance and usability of your data lake with automated compaction, table optimization, and data cleanup | Generally available |
| | **Job orchestration** | Orchestrate jobs directly in the Galaxy UI or integrate third-party tools like dbt and Airflow | Generally available |
| | **PyStarburst** | Build complex data pipelines in Python while leveraging the power of the leading SQL query engine, Trino | Generally available |
| GRAVITY | **Data catalog** | Simply connect your data sources and Gravity will index all the relevant objects for you, increasing discoverability | Generally available |
| | **Fine-grained data governance** | Use data attributes to write and scale ABAC policies across hundreds of roles down to the row-level | Generally available |
| | **Data observability** | Resolve upstream and downstream data quality issues fast with built-in lineage and profiling capabilities | Preview |
| | **Data sharing via data products** | Create and share curated data sets with dedicated compute via data products | Generally available |
| ENGINE | **Enhanced Trino** | Leverage additional, proprietary optimizations that build-on and extend the power of OS Trino | Generally available |
| | **Support for large transformations** | Run complex, long-running, and memory-intensive workloads without the fear of query failure | Generally available |
| | **Interactive data lake analytics** | Automatically accelerate data lake queries with Warp Speed – a smart indexing and caching technology | Generally available |
| PLATFORM | **Fully-managed, multi-cloud platform** | Spin up AWS, Azure, and GCP clusters in just a few clicks | Generally available |
| | **Open table formats** | Get first-class support for all modern table formats (Apache Iceberg, Delta Lake, and Hudi) | Generally available |
| | **Cross-cloud and on-prem data access** | Query data across regions, clouds, or even on premise while maintaining the data lake as your center of gravity | Coming soon |
| | **Enterprise-grade security and audit controls** | Protect your data and meet stringent requirements with robust operational and security controls | Generally available |
| | **24/7 x 365 support** | 24x7 support from the Trino experts | Generally available |

Generally available · Preview · Coming soon