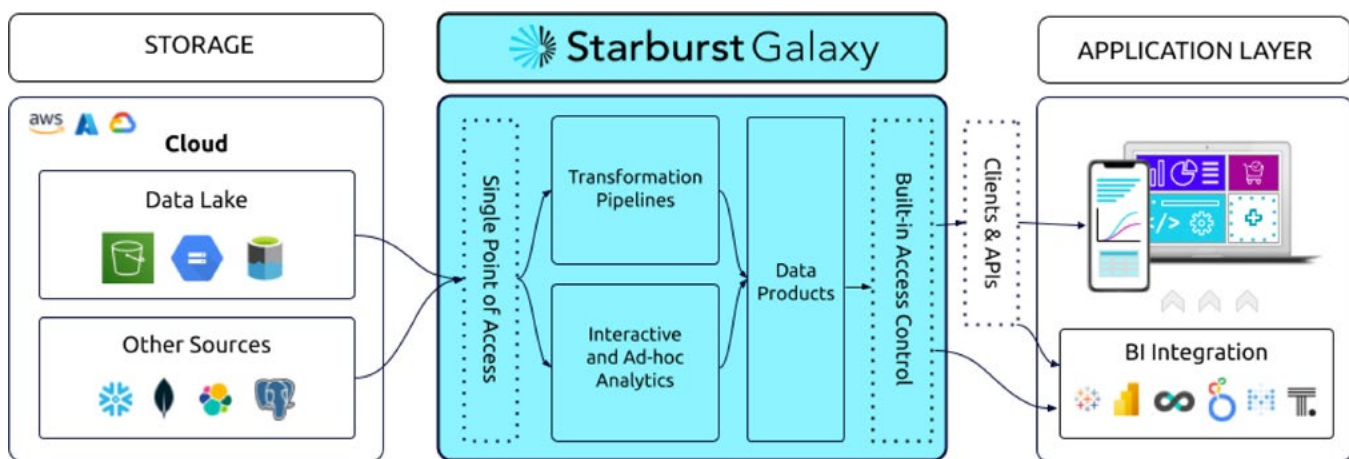




Starburst-powered data applications

Modern data applications require a modern approach. Build and scale data-intensive applications with Starburst, never compromising performance or cost.



Introduction

This journey to building and scaling data applications is not without its challenges. Developers often struggle with data silos, complex data pipelines, provisioning, and security bottlenecks. Additionally, they face challenges in performance tuning and ensuring an exceptional user experience, all while dealing with the operational overhead of scaling applications and related processes. These obstacles ultimately inhibit productivity and delay progress, stealing valuable time and resources from data teams.

Whether you're building the next best-in-class application for threat detection, global supply chain visibility, a machine learning engine for customer 360, or even clickstream analytics platform, you need a fully-managed platform that scales without compromise. **Many technologies today claim to scale your applications with ease, but regret mention the performance, cost, and maintenance implications that follow** - whether it be growth in data size, data complexity, or user concurrency.

Starburst is designed with price-performance at its core, enabling teams to scale from hundreds to thousands of users without sacrificing the bottom line. Starburst's data lake analytics platform robust capabilities enable these teams to effortlessly create and expand custom data applications, providing seamless, high-performance access to data from any location.

Starburst is designed with price-performance at its core, enabling teams to scale from hundreds to thousands of users without sacrificing the bottom line. Starburst's data lake analytics platform robust capabilities enable these teams to effortlessly create and expand custom data applications, providing seamless, high-performance access to data from any location.

No matter the data size, complexity, or user concurrency, Starburst effectively addresses data application challenges at scale on multiple fronts.

Data Volume: With a modern approach, Starburst directly connects to data in and around your lake, streamlining access to handle extensive data volumes efficiently.

Data Complexity: Leveraging Warp Speed its proprietary smart indexing and caching technology, Starburst accelerates query performance, particularly for intricate data structures.

User Concurrency: By utilizing Trino's MPP capabilities, Starburst scales to internet-level user concurrency, ensuring smooth interactions even with numerous simultaneous users. This holistic approach optimizes data access, query speed, and user experience, enabling seamless data application management. unnecessary table scanning, saving valuable cloud compute costs.

Key pillars

Scale

Scale applications without compromise

Starburst enables seamless scalability of data applications by leveraging distributed processing and parallelization, smart indexing and caching through Warp Speed, ensuring optimal performance even with exponentially growing data volumes.

Development

Focus on innovation, not infrastructure management

Starburst Galaxy empowers teams to focus on building unique features by minimizing the time spent on infrastructure management. With streamlined data processing and integration, Galaxy accelerates product development, driving business growth and agility while reducing operational overhead.

Security

Robust, fine-grained security made easy

Starburst Galaxy provides comprehensive built-in access controls via row level filtering, dynamic column masking, and encryption capabilities within a universal governance platform. No third-party integrations needed.

Cost

Grow beyond limits, not budgets

No storage costs, no hidden fees, just compute. Galaxy provides a number of premium features, across dynamic autoscaling, query optimization, smart indexing, and more, all optimized for price-performance, at no additional cost.

Use cases

These applications span industries, from dashboards to chatbots, demanding strategic alignment for efficient scalability. New ventures require thorough research, planning, and technology deployment for lasting success.

Today, Starburst-fueled data apps drive real-time insights across vital sectors like log analysis, cyber threats, fraud detection, supply chains, machine data, clickstreams, and customer 360.



Let's take a deeper look into how Starburst future-proofs cutting edge data applications.

Log analytics

Starburst's exceptional capacity to process and analyze extensive log data, combined with its distributed query framework, uniquely equips it to power log analytics applications. Its optimization techniques enhance query efficiency, enabling rapid issue identification, while compatibility with standard SQL simplifies integration. Coupled with direct data lake connectivity and scalability, Starburst empowers log analytics applications with the speed, efficiency, and responsiveness required to extract valuable insights from log data effectively.

Fraud detection

Starburst's query optimization and scalability enable seamless access to diverse real-time datasets, while parallel processing ensures efficient handling of high-volume data streams. This unique combination empowers the application to swiftly identify anomalies and fraudulent activities, enhancing accuracy and responsiveness in combating evolving threats.

Clickstream analytics

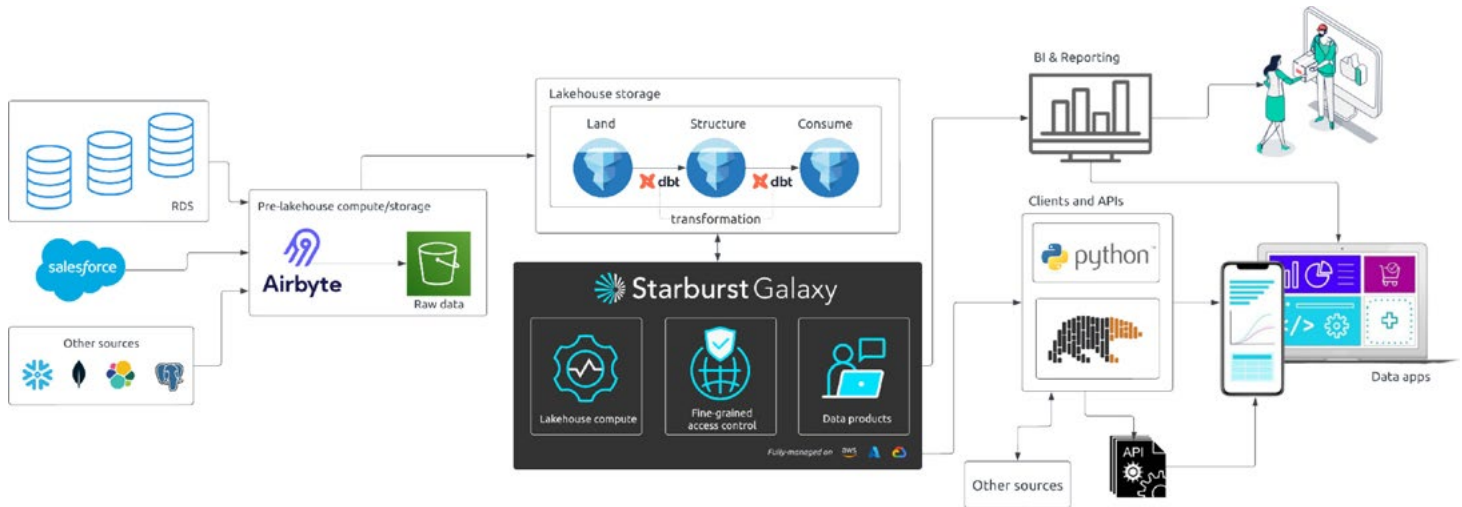
By seamlessly accessing and integrating diverse clickstream data sources in real-time, Starburst enables comprehensive insights into user behavior. Its parallel processing and scalability ensure efficient handling of large clickstream datasets, allowing for rapid analysis and actionable insights. With Starburst, the clickstream analytics application gains a unique edge, driving enhanced decision-making and customer engagement strategies.

“Starburst Galaxy allows us to take a load off of engineering. We don't have to become experts in Trino or experts in managing Trino. We can focus on business needs – whether it's performance work, feature work, or the one-off requests you get from large customers.”

SVP Engineering and Data Science, Leading Cybersecurity Company

Reference architecture

The below is a reference architecture based on discussions with customers leveraging Starburst for data applications. This visual representation highlights the integration of data sources, transformations, and storage optimization. Central to this framework is Starburst Galaxy, the fully-managed version of Starburst that runs in all three clouds. In this diagram, Galaxy is the powerful hub that facilitates data virtualization, manages data lake compute, enforces access controls, and supports data product development for downstream applications. Exploring this diagram, we witness the evolution of data from its raw state to actionable insights



1. *Data Sources Ingestion: Extract data from various sources - data lakes (Hadoop, Delta Lake, ADLS, S3, etc.), RDS, Salesforce, and other cloud data sources (Snowflake, MongoDB, Elasticsearch, PostgreSQL, etc.)*
2. *Data Integration with Airbyte: Ingest and consolidate data using Airbyte, a data integration platform.*
3. *Raw Data Storage: Store ingested data in its raw, unprocessed format for future use.*
4. *Iceberg Storage: Use an Iceberg-based storage layer to organize and manage structured data (Starburst supports multiple modern table formats like Apache Iceberg (pictured above), Delta Lake, and Hudi)*
5. *Data Consumption with dbt: Utilize dbt to perform transformations, aggregations, and business logic on structured data, preparing it for analytics.*
6. *Data Virtualization with Starburst Galaxy: Use Starburst Galaxy as a data virtualization layer to create a unified view of data across different sources.*
7. *Outputs: Deliver insights to clients through various output methods, such as visualization tools or direct API access. Power data apps and applications by exposing necessary data through APIs, enabling real-time data access and interaction.*

How to get started