Starburst
ANALYTICS ANYWHERE

SOLUTION BRIEF

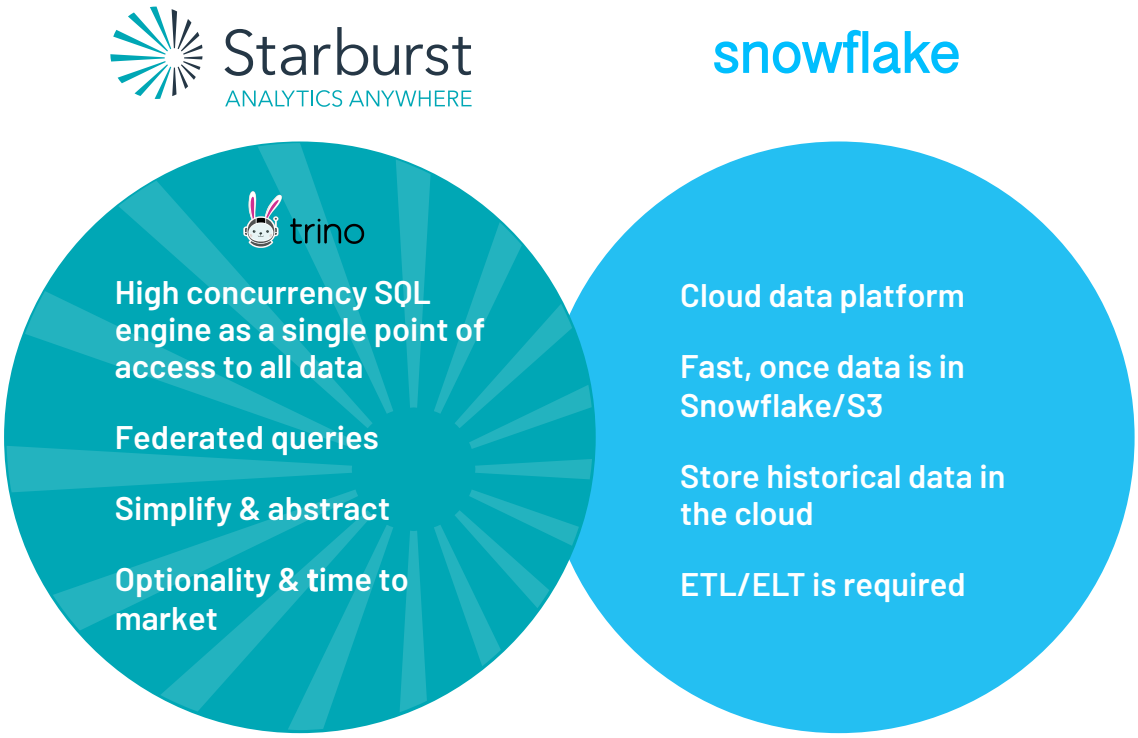
Extending your analytics with Starburst & Snowflake

Different data, different needs

Analytics, despite being business critical, is actually the last step in the long journey of the data it uses. Most of that data wasn't even generated to be used for analytics or with analytics in mind. It was created by and for applications, devices and users. As such, that data creation can span many different platforms, locations, formats and technologies. Bringing all that data together to enable timely & complete decisions then becomes the most serious challenge in analytics.

Generally, data can be generated in two basic methods. The first is the operational data, made up of highly structured data from sources like relational databases. This data is used to control inventory, store purchase orders, manage client contacts, and more. It has comprised the vast majority of data, until lately. The second method of data generation is the up and coming unstructured & semi-structured data sets like real-time streaming, IoT devices, web & mobile application output, logs, and scale-out application outputs. With scale required for these cloud-native applications and processes, this data is stored in the lowest-cost, most-scalable locations possible: object storage like HDFS, S3, GCS & ADLS. To process this type of data requires a whole new approach.

For all your analytics needs, Starburst and Snowflake offer platforms designed to complement each other and integrate with any architecture for full coverage and flexibility.



Unified analytics. Better together.

Starburst is query fabric above all your sources. If you're migrating to Snowflake or have already moved core data there, you already know the tremendous benefits and speed it delivers. Keeping existing Snowflake workloads in place, Starburst Enterprise provides optionality for other workloads. For example, mostly read only or inset only workloads are great for Starburst and can be much more cost effective. Bottom line is you get to own your own data and make the best storage and compute decisions to meet your needs. True separation of storage and compute is owning your own data doing whatever you want it without being nickel and dimed.

With Starburst and Snowflake organizations can achieve fast access to core data to operate your business, and fast access to all other data for ad hoc analytics. Starburst enables a flexible data management approach that provides an abstracted query layer; allowing access to data wherever it lives today, and tomorrow. This allows for an optimized price performance with the flexibility to use lower cost query execution via Starburst and an improved time-to-value for overall business performance.

While Snowflake provides processing storage offload and transformation from the primary data source, Starburst enables federation over all data sources, including data lake processing.

Starburst Enterprise empowers its customers to:

Access to all data

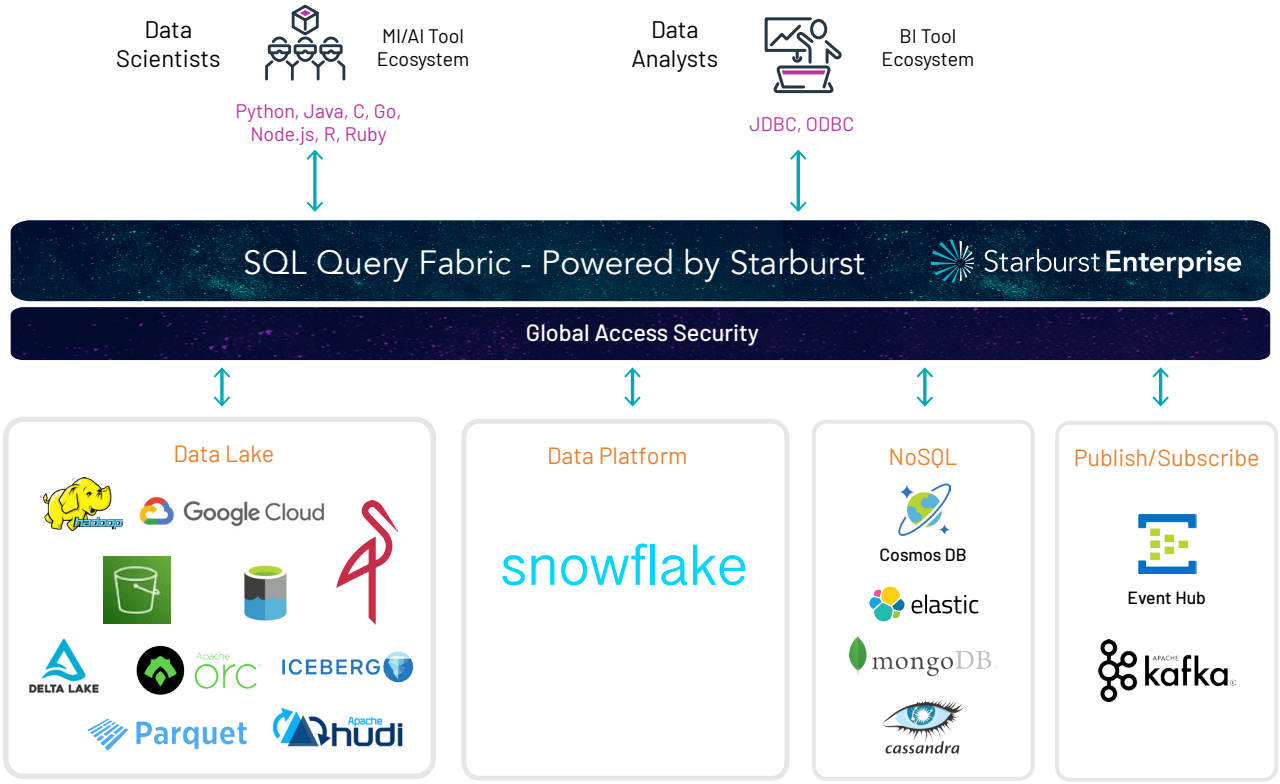
- Solves for fast access to distributed data not only in Snowflake
- MPP query engine that gives you the ability to access all data where it lives, with high concurrency
- Purpose-built for data discovery & ad hoc analytics

Work for today & tomorrow

- Separates data access from where data is stored
- Uses a language your team already knows: SQL
- Consistency of data access when migrating data to new forms of storage

Optimize price/performance

- Mitigate rising costs by offloading workloads to Starburst
- Give your data engineering team optionality in how they execute queries



Different, but complementary

In the world of analytics it would appear as if data lakes and data warehouses are competing for the same position as the single repository for data. However, there is actually not only space for both, but a need for both. This is due to the diverse nature of both the ways data is created and how it is used.

Data warehouses and dake lakes can serve their designed purposes individually as well as combine to produce the holistic view businesses need for thorough, accurate and timely decision making.

Data Warehouses

Operational data is the traditional variety that is behind common applications like ERPs and CRMs. Basically, data behind the common tools and applications that organizations use to run their business. This data has been around for a long while and will continue to be. It is vital for decision making and exists in every business.

Data warehouses are built to pull these data sets away from their primary location, transform them with business rules and combine the required pieces together to form the backbone of traditional analytics.

Purpose: Small to large size, specialized, tightly integrated, highly formatted, business rules applied, and pre-aggregated to fit needs.

Starburst & Data Lakes

More recently, cloud native technologies, applications and devices have created a separate but growing type of data. This type of data is generated on a much larger scale, is typically formed of files, and doesn't always have a defined or rigid structure. Due to its scale, a new form of storage was needed. Enter the data lakes.

Data lakes store unformatted, or raw, data on a very large scale at very low cost. This can be anything from fresh real-time data to large bulks of historical data. To process these types of data requires a special & new type of analytics engine. Starburst was designed to process this type of data.

Purpose: Medium to limitless size, lowest cost, unrestricted repository, raw or unstructured.

Values and use cases

To solve all an organization's data needs, both warehouses and lakes (with Starburst) are required.

Data Warehouses

- Highly transformed data (business rules)
- Remove processing from primary source
- Transform format for optimal performance
- Move data once (traditional)

Starburst & Data Lakes

- Lowest cost, for scale
- Immediate access
- Real-time & historical
- Federate over all sources
- Move data once (caching)

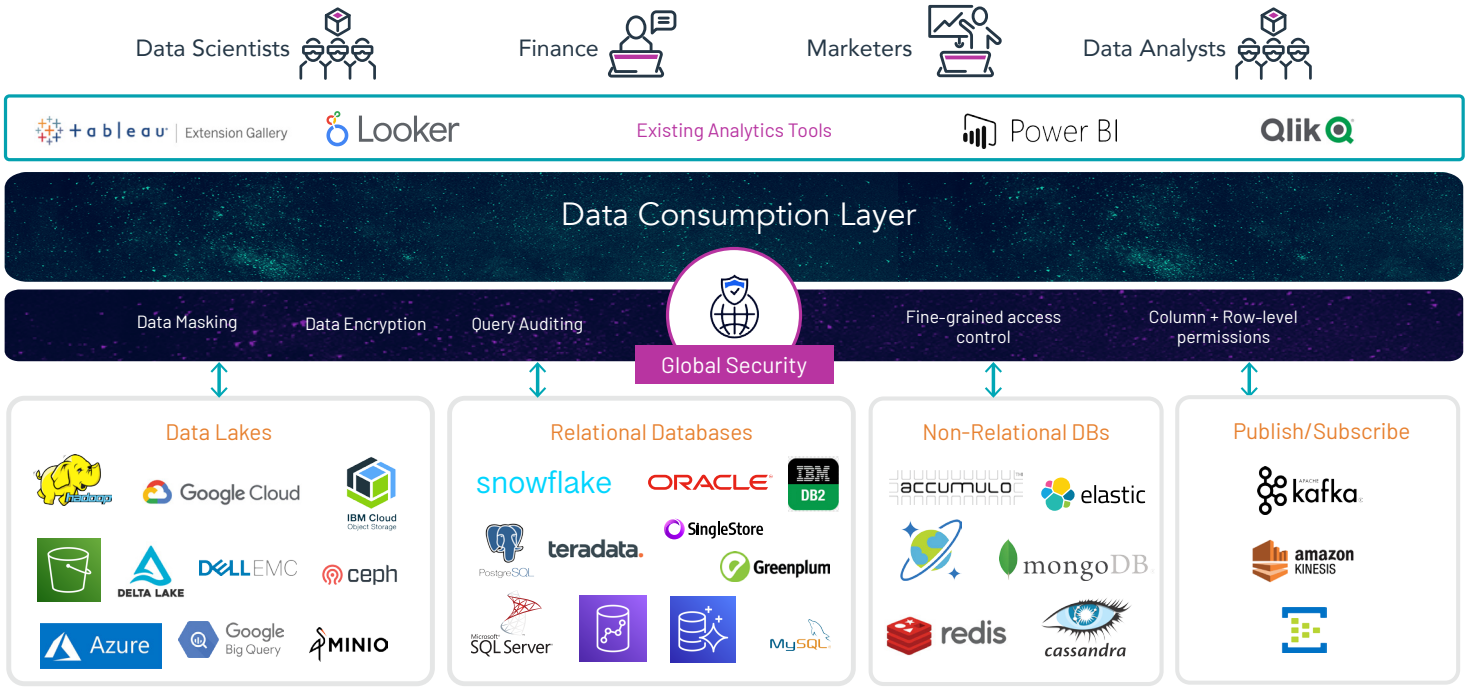
Working together

Federation: Single point of access

Data warehouses are able to run analytics only on what has been imported into their system. ETL is the process used to migrate data into the warehouse and transform it into a usable form. Once inside, all data appears to be the same and can be used as one. But, this leaves out any data that has yet to be brought in. This could be new data, data too large to migrate in or data that has yet to be integrated.

Once data has been brought into the warehouse, no further load is placed on the data source. For workloads like traditional transaction processing in a RDBMS, it is very common to exclude analytics processing so as to not overwork the source database. The value of Snowflake is offloading work from primary data source when needed and to store active and historical data in one place.

Starburst can establish a new link to almost any data source in minutes as well as local and remote data sources. This allows for both very comprehensive analytics as well as very agile time to market for producing new reports. By not requiring an ETL (data migration) the analytics teams can get to work right away. Also, by not storing the data twice, the costs can be reduced against a data warehouse type solution.

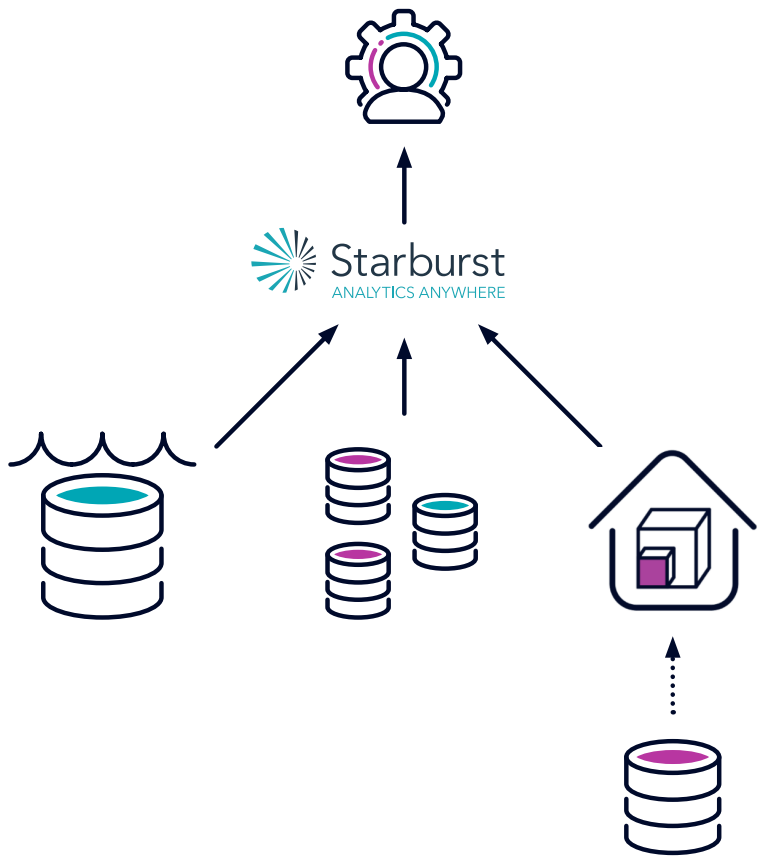


There are times where it makes sense to federate with Starburst, to enable a single point of access, but yet also migrate some data into a warehouse. A few examples would be:

- When analytics workloads need to be separated from the primary data source so as to not harm their processing,
- When hybrid-cloud / on-prem architectures where a slow connection or egress charges would cause slow or costly analytics, it would be better to migrate the data once.
- When data needs significant transformation prior to running analytics, it can be better to run a transformation into a new location (warehouse) and have Starburst use that as a source.

Complete solution

When used together, Starburst and data warehouses can produce an optimal solution that achieves the lowest cost, highest performance, most comprehensive access and quickest time to market.



Starburst Enterprise and Snowflake promote better and more timely insights by letting organizations rapidly analyze data across multiple disparate and distributed data platforms. Companies using the Starburst Enterprise with Snowflake discover and extract value from data faster — and turn these insights into actionable business initiatives that drive significant revenue gains.

Starburst
ANALYTICS ANYWHERE

Unlock the value of your data with Starburst and Snowflake today.
For more information, contact us at starburst.io