Starburst

E-BOOK

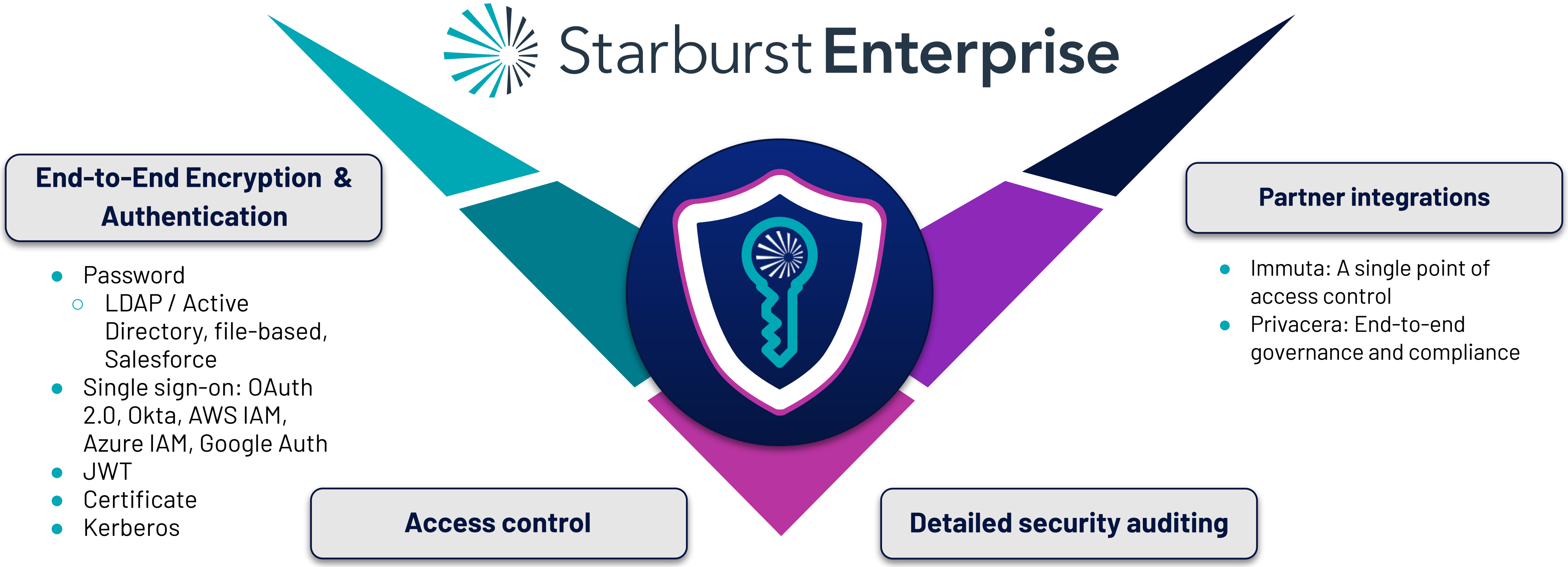
Enterprise Security Guide

Table of Contents

Introduction

Starburst Enterprise is a distributed SQL query engine that can be deployed in any infrastructure. Built on the open source **Trino** project developed at Facebook, Starburst Enterprise is used by some of the largest, well-known companies in the world such as Slack, Comcast, Zalando, and FINRA.

Starburst Enterprise is a fully supported, production-tested and enterprise grade distribution of the open source Trino query engine. It includes additional connectors for commercial database systems, query optimization, as well as management tools. One of the core reasons organizations select Starburst is the added security features that we've built into the Trino engine. These include fine-grained access control, data masking and encryption, column and row-level security, and query auditing for example. This guide provides details on all of the enterprise-grade security features available in Starburst Enterprise.



End-to-End Encryption & Authentication

- Password
  - LDAP / Active Directory, file-based, Salesforce
- Single sign-on: OAuth 2.0, Okta, AWS IAM, Azure IAM, Google Auth
- JWT
- Certificate
- Kerberos

Partner integrations

- Immuta: A single point of access control
- Privacera: End-to-end governance and compliance

Access control

- Native, integrated RBAC
- Easily configure user data and platform access
- Flexible configuration of roles and privileges
- Apache Ranger
- File-based

Detailed security auditing

- Query and usage history
- Event logging
- Cluster monitoring
- Monitoring dashboards
- Audit logs for security changes

→ This guide provides details on the enterprise security features provided by Starburst Enterprise, along with extensive event logging and robust fine-grained access control.

Starburst Enterprise Architecture

The lightweight, standalone architecture of Starburst Enterprise makes it simple to install, secure, maintain and scale. Since there is no storage of data and it can be installed in any location including cloud or on-premises, security is simple to maintain and enforce.

Starburst Enterprise's architecture consists of a coordinator node, worker nodes, and connectors for 45+ data sources. Each of these components can be secured using industry standard techniques as is best practice when deployed in a production environment. The diagram below illustrates the different components of a Starburst Enterprise cluster.

The coordinator

The coordinator is the brain behind a Trino cluster. It's responsible for:

- Accepting client connections to execute queries.
- Parsing, analyzing, planning, and optimizing query plans.
- Scheduling query data retrieval tasks on workers nodes.
- Returning the query results to the client.

Workers

The workers are responsible for the heavy lifting. Their job is to retrieve data using the Starburst and Trino connectors to filter, join, aggregate, and exchange the intermediate data before returning the results to the client via the coordinator. Starburst never stores data as part of this process.

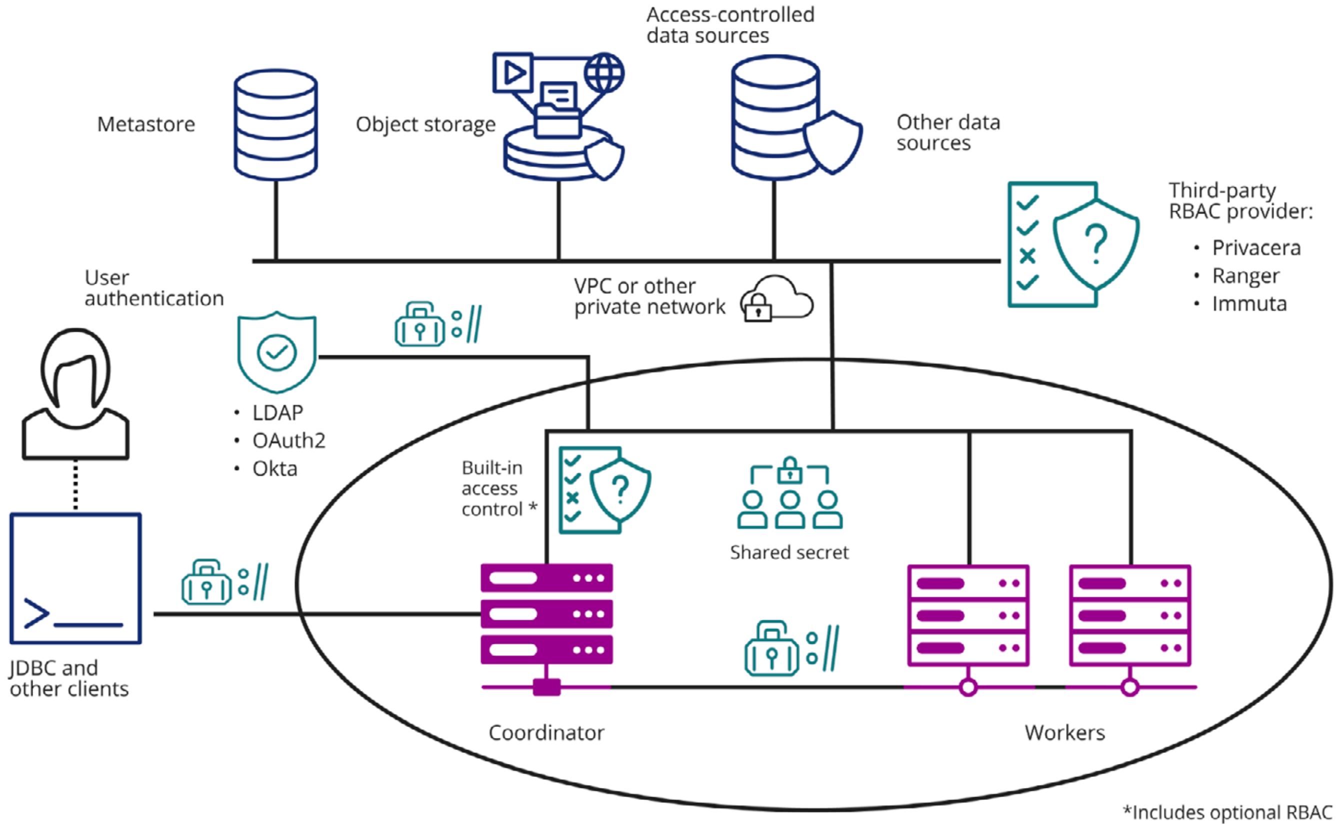
Connectors

Connectors translate data objects into something that Starburst can operate on when executing standard SQL. Connectors fall into the following categories:

- Distributed Storage - such as Amazon S3, Azure Blob and Data Lake Storage, and Google Cloud Storage using Hive compatible metadata stores such as Hive Metastore and Amazon Glue.
- Relational Database Management Systems - for example Oracle, PostgreSQL and MySQL.
- Key-Value Stores - including Cassandra, Accumulo and Redis.
- Document Stores such as MongoDB and Elasticsearch.
- Streaming Systems like Kafka and Kinesis.
- Built-in utilities, for example system, memory, and TPCH/DS.

Access control

Lastly, fine-grained access control policies are enforced during query time using Starburst's built-in access controls. This includes column and row level authorization.



*Includes optional RBAC

Achieving Enterprise-Grade Security with Starburst

We know very well that data security is critical, and that sensitive information can be highly destructive and costly to an organization when it falls into the wrong hands. Starburst Enterprise contains many security features that enterprise companies expect, but are not available in open source Trino.

Ensuring data is encrypted from the sources to the end user is now a standard in enterprise environments. Controlling access down to the column and row level is usually the function of third party software, but is included in Starburst Enterprise and is constantly being improved. Providing a full data access audit trail is also essential to ensure companies are in compliance with state and federal regulations.

The following sections detail out how Starburst Enterprise implements the following enterprise features in order to secure Trino:

Starburst

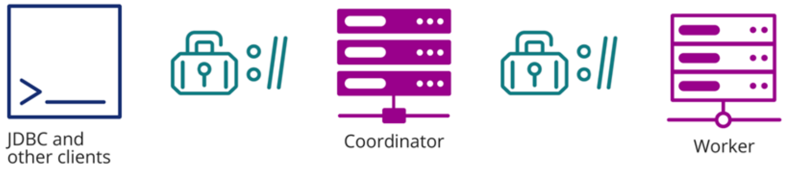
- **End-to-End Data Encryption and Authentication**
- **Fine-Grained Access Control**
- **Detailed Security Auditing**
- **Partner Integrations**

End-to-End Encryption and Authentication

The leading reason Starburst clients choose Starburst Enterprise over open source Trino is the built-in comprehensive security features and configurations. This document highlights the key elements of end-to-end security.

Encrypting User Access

When users connect to Trino to issue queries, they are connecting to the coordinator node.

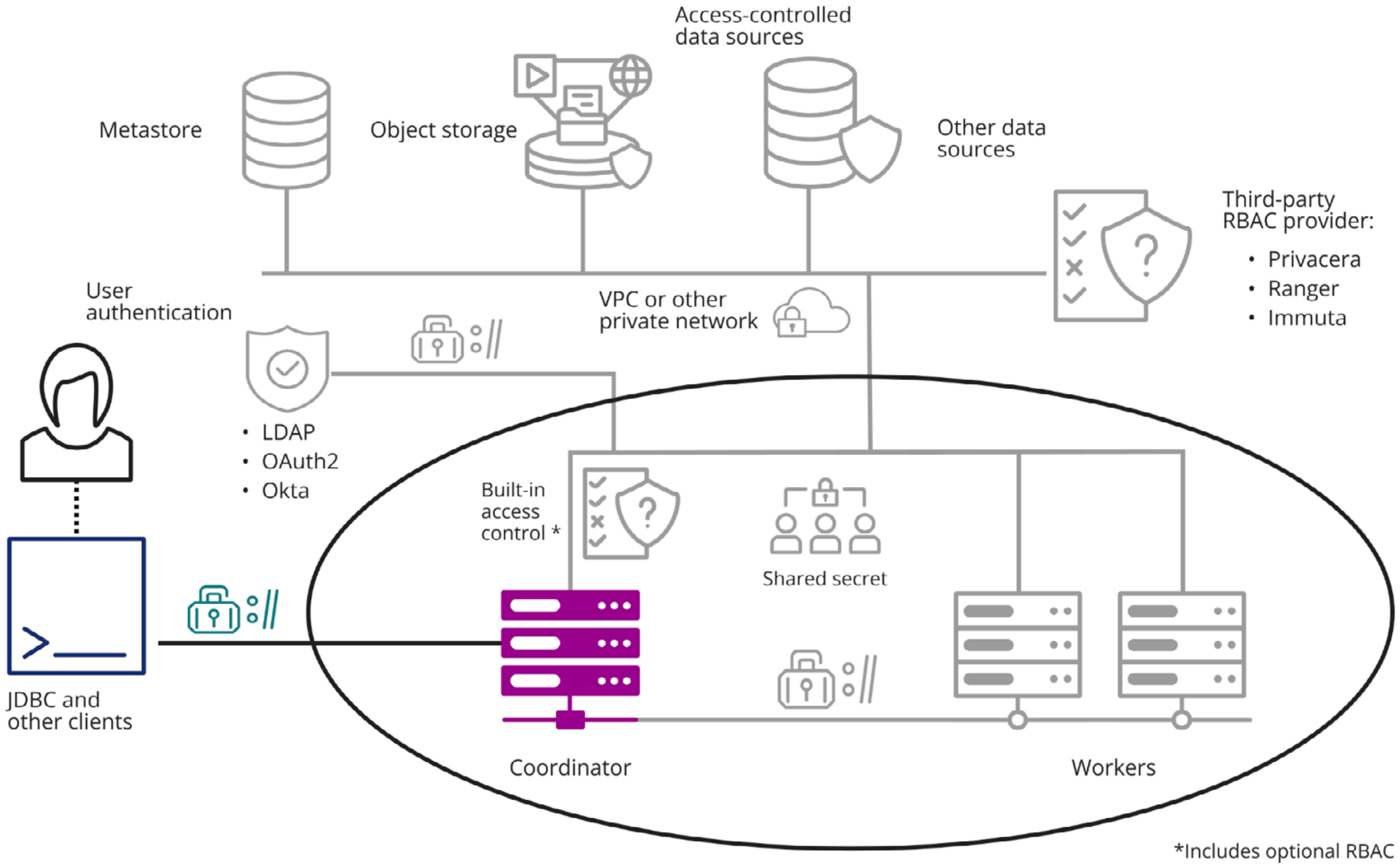


Starburst Enterprise supports both trusted CA and self-signed certificates. When using a self-signed certificate, clients must have a password-protected truststore file containing the coordinator's certificate. Most ODBC/JDBC clients that connect to Trino will not require the truststore file when using a trusted certificate.

This will encrypt traffic from the end users to the coordinator node. Traffic from the coordinator to the worker nodes will still remain unencrypted.

Configuring HTTPS between the coordinator node and the workers is similar to securing the coordinator. Workers are configured to communicate with the coordinator using standard HTTPS methods.

Configuring secure communication is covered later in this document.



Authenticating Users

Authenticating users to access Trino can be handled using a few different methods. Currently, Trino supports authenticating against external sources through LDAP, SSO (single sign-on) or Kerberos using HTTPS as illustrated in the diagram below:

LDAP

Authentication through the industry-standard LDAP protocol supports many different providers such as Active Directory and OpenLDAP. Trino supports authenticating users using Secure LDAP (LDAPS) which requires the external LDAP server to be configured with TLS, an industry standard practice.

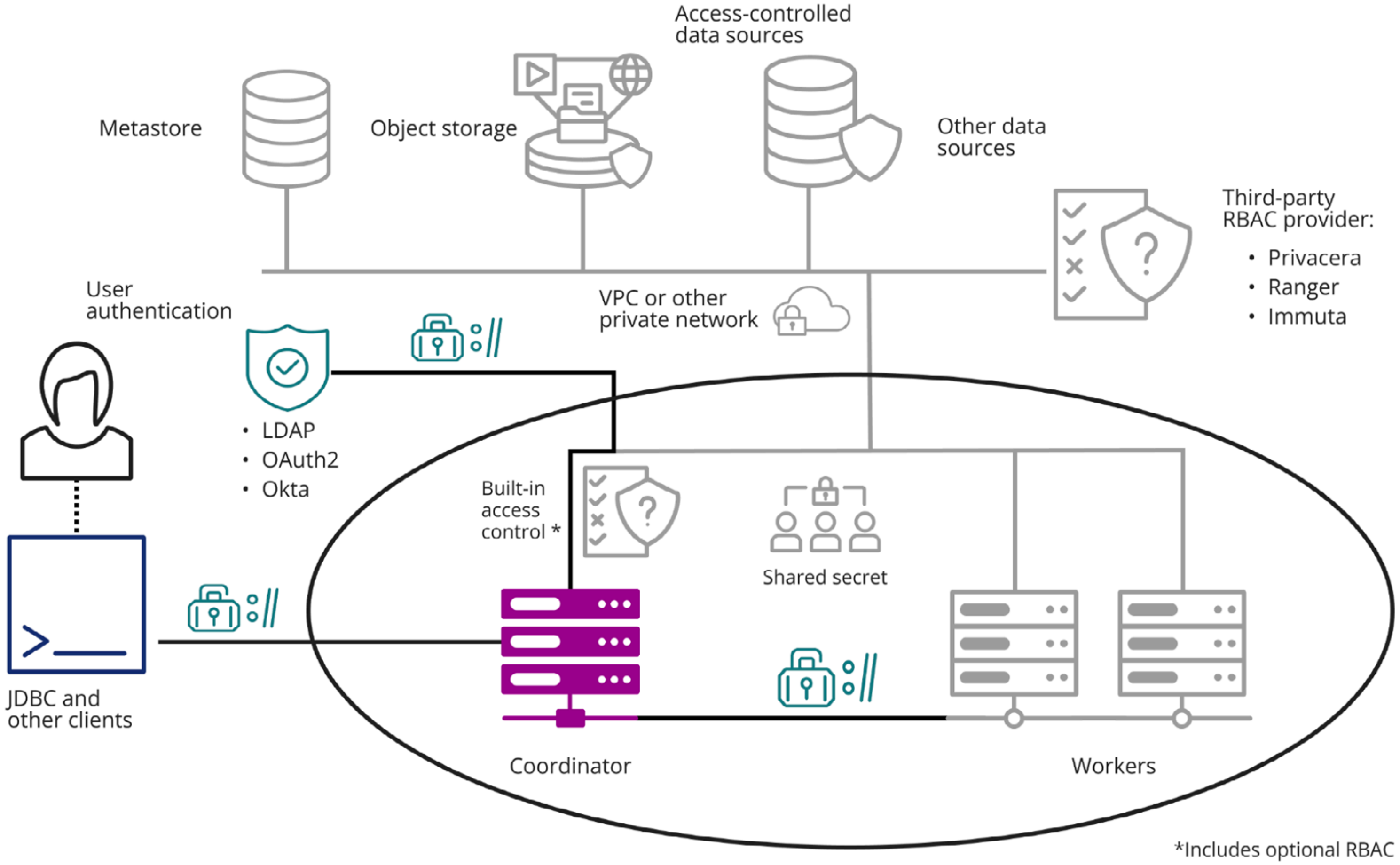
NOTE: Enabling LDAP authentication requires HTTPS to be configured on the coordinator node as described earlier in this document. Once the LDAP server's TLS certificate is imported into Trino's truststore, users will be authenticated through the coordinator.

Identity Provider (Single Sign-On)

Starburst Enterprise supports Identity Provider (IdP) through Okta as of this writing. More providers will be added in coming releases. IdPs enable end-users authentication using a provider such as Okta or Ping.

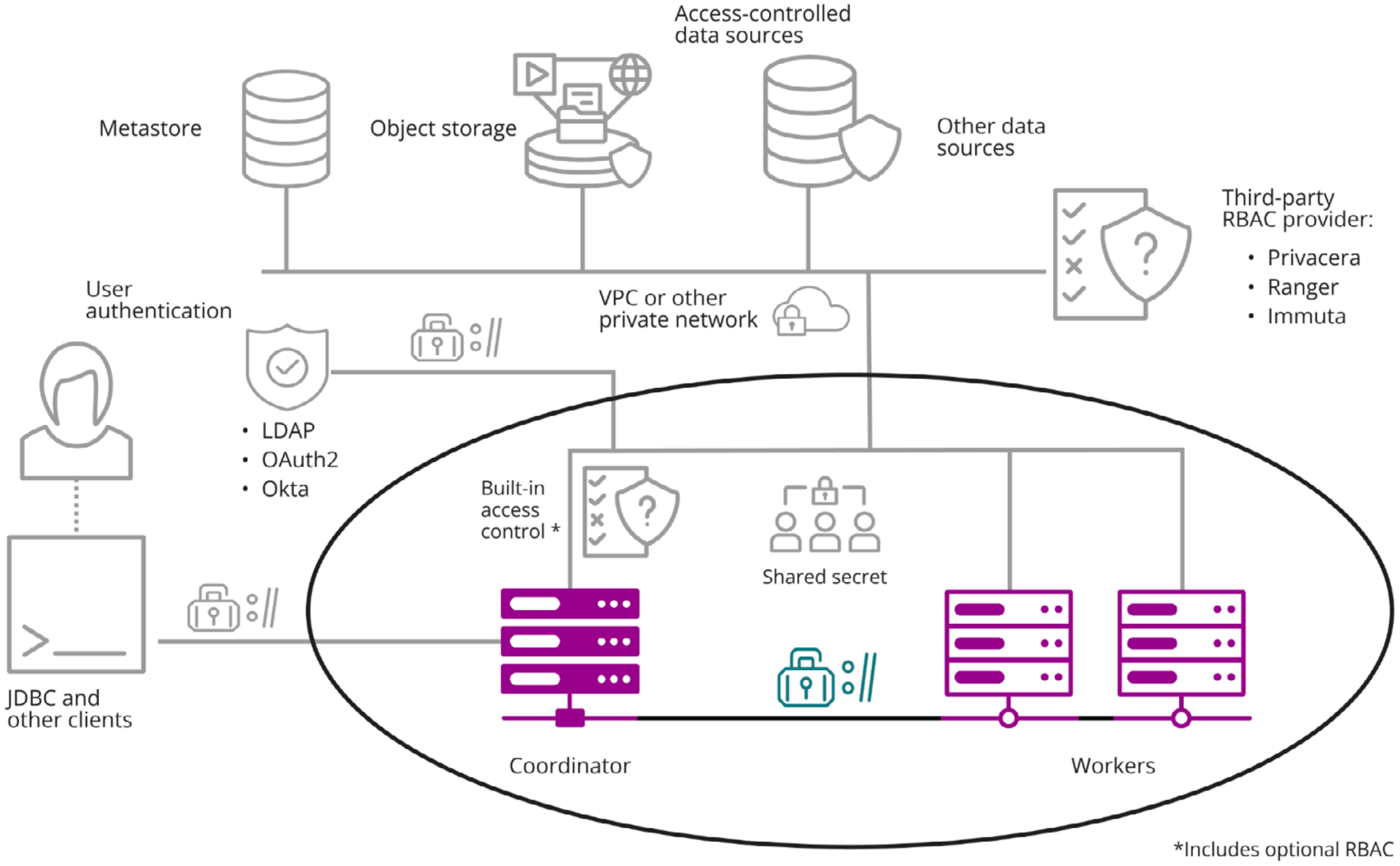
Kerberos

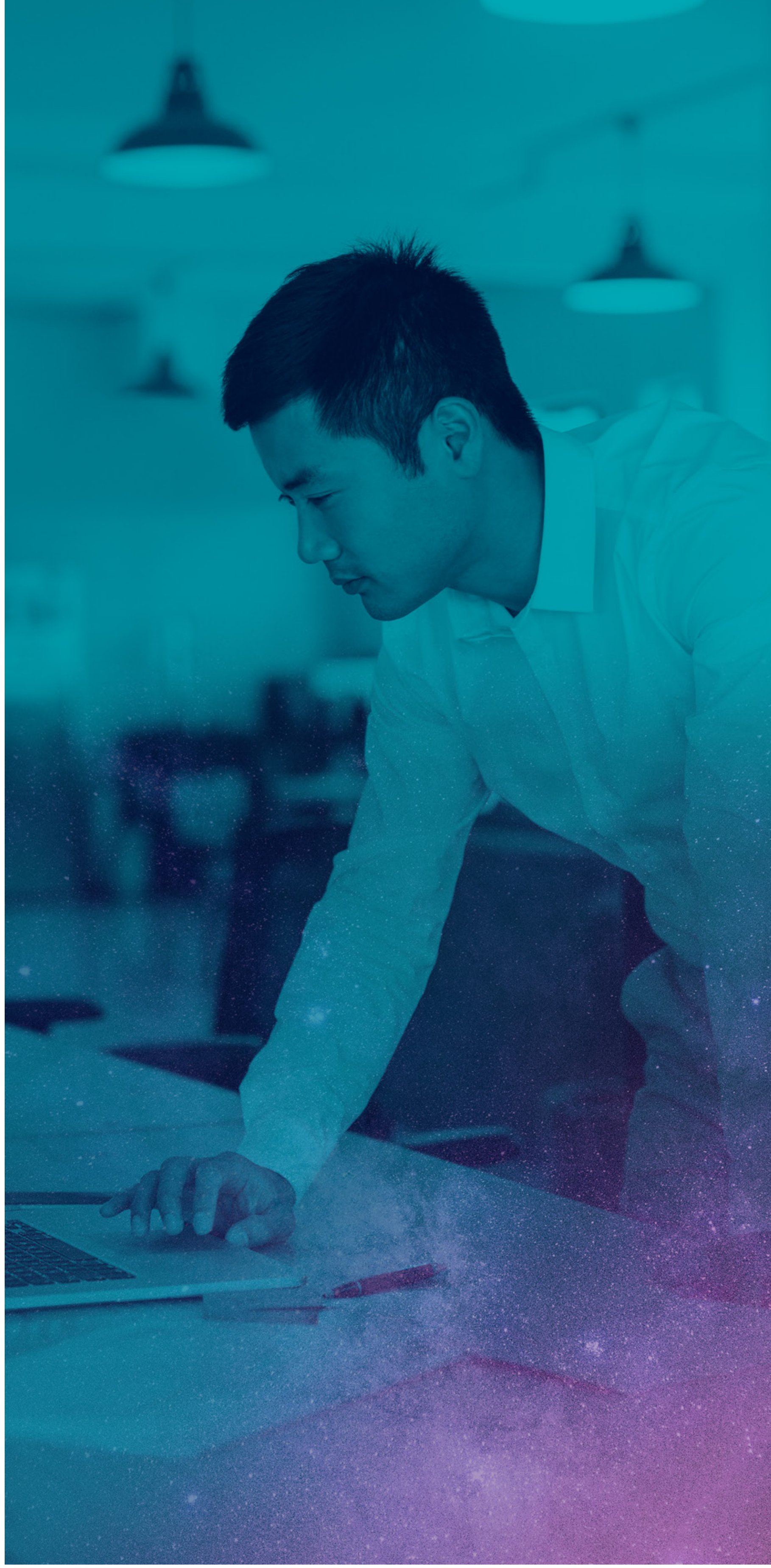
Alternatively, user authentication can be achieved via Kerberos. A Kerberos KDC service must be accessible by the Trino coordinator server. HTTPS must be enabled on the coordinator in order to secure user authentication requests. More detailed information can be found in the Kerberos documentation.



Securing Internal Communication

Full end-to-end encryption of internal traffic between Trino nodes is possible for highly secure environments. Securing internal communication ensures that the intermediate data exchanged between workers and the coordinator during query processing is encrypted in transit.

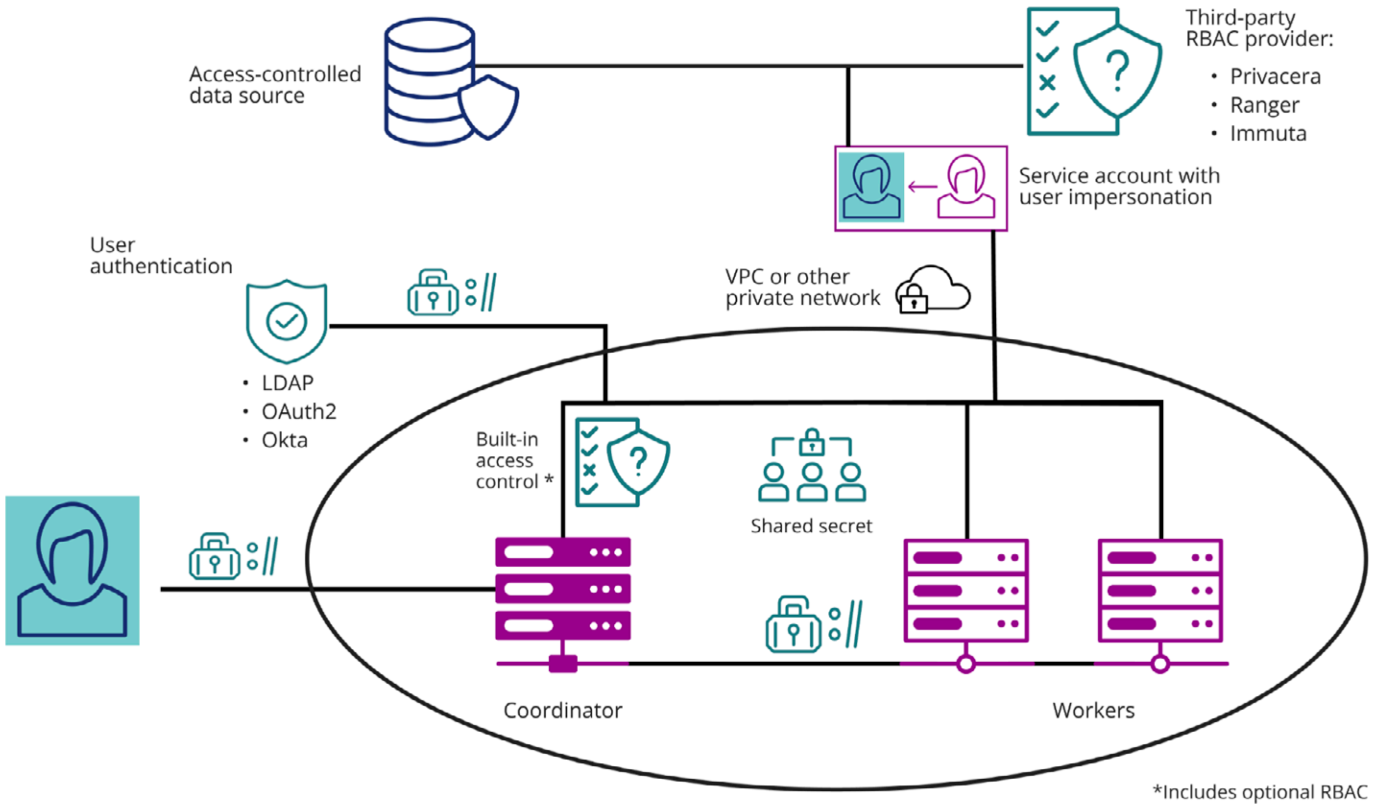




Securing connectors

Securing access to your data source depends on the particular connector and the source system capabilities. Often, several different methods are supported to fit various customer setups.

In most cases, the connector configuration file contains security credentials that allow Trino to access the data source. Often the access to data mediated with service accounts that channel all Trino end-users data access requests. We recommend using service accounts with read-only access to objects in the source system.



Access-controlled data source
Third-party RBAC provider:
• Privacera
• Ranger
• Immuta
Service account with user impersonation
User authentication
• LDAP
• OAuth2
• Okta
VPC or other private network
Built-in access control *
Shared secret
Coordinator
Workers

*Includes optional RBAC

End-user impersonation

Starburst Enterprise includes a feature in some of the connectors called end-user impersonation. Typically, a service account is used to access the data source. For auditing purposes however, this may not be desirable as auditing shows only a single user accessing the data. By implementing end-user impersonation, queries run as the user who initially executed the query. This means the privileges for that user are applied and also audited appropriately.

Credential passthrough

Starburst Enterprise provides credential passthrough through credentials for IdPs, Kerberos, and passwords. Users are required to supply their credentials to Trino. These credentials are then used to connect to the underlying data source. As a result, any data access via Trino is subject to the access controls applicable to that user. The password credential pass-through feature guarantees that Starburst Enterprise uses the same credentials as a user accessing a data source directly. This allows you to authenticate using the CLI or client application via the JDBC or ODBC driver bypassing credentials through to Starburst Enterprise and from there to the underlying data source.

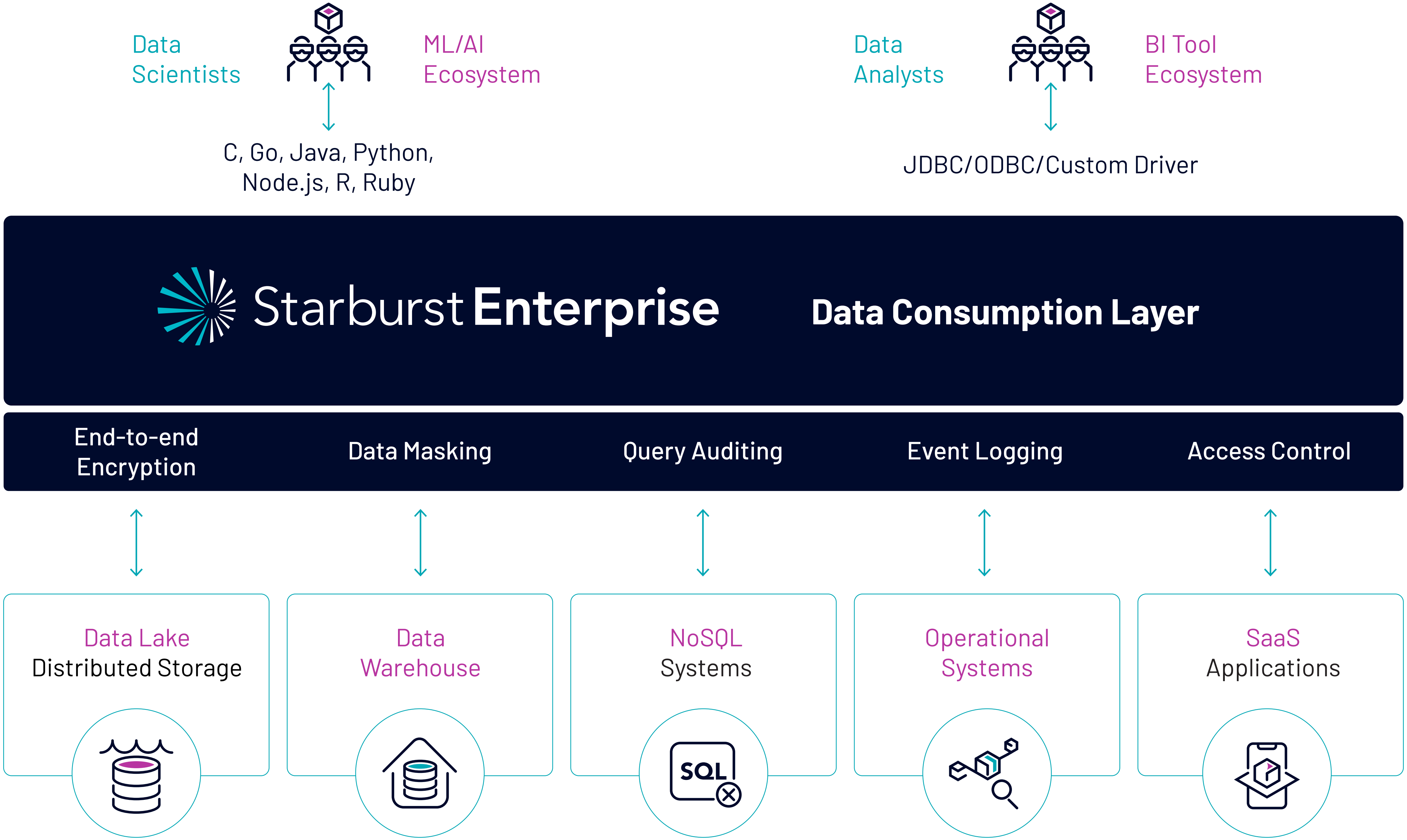
Securing sensitive data

Trino's configuration files may include sensitive information, such as plaintext passwords and usernames needed by the service account to authenticate to the data sources. In many enterprise organizations, this violates security policies. Starburst Enterprise provides the ability to securely authenticate to these data sources without having to store the plaintext passwords and usernames in the configuration files. Trino manages configuration details in static properties files called secrets. This configuration needs to include values such as usernames, passwords and other strings, that are often required to be kept secret. Only select administrators or the provisioning system has access to the actual value. The secrets support in Trino allows you to use environment variables as values for any configuration property. When loading the properties, Trino replaces the reference to the environment variable with the value of the environment variable.

Starburst Fine-Grained Access Control

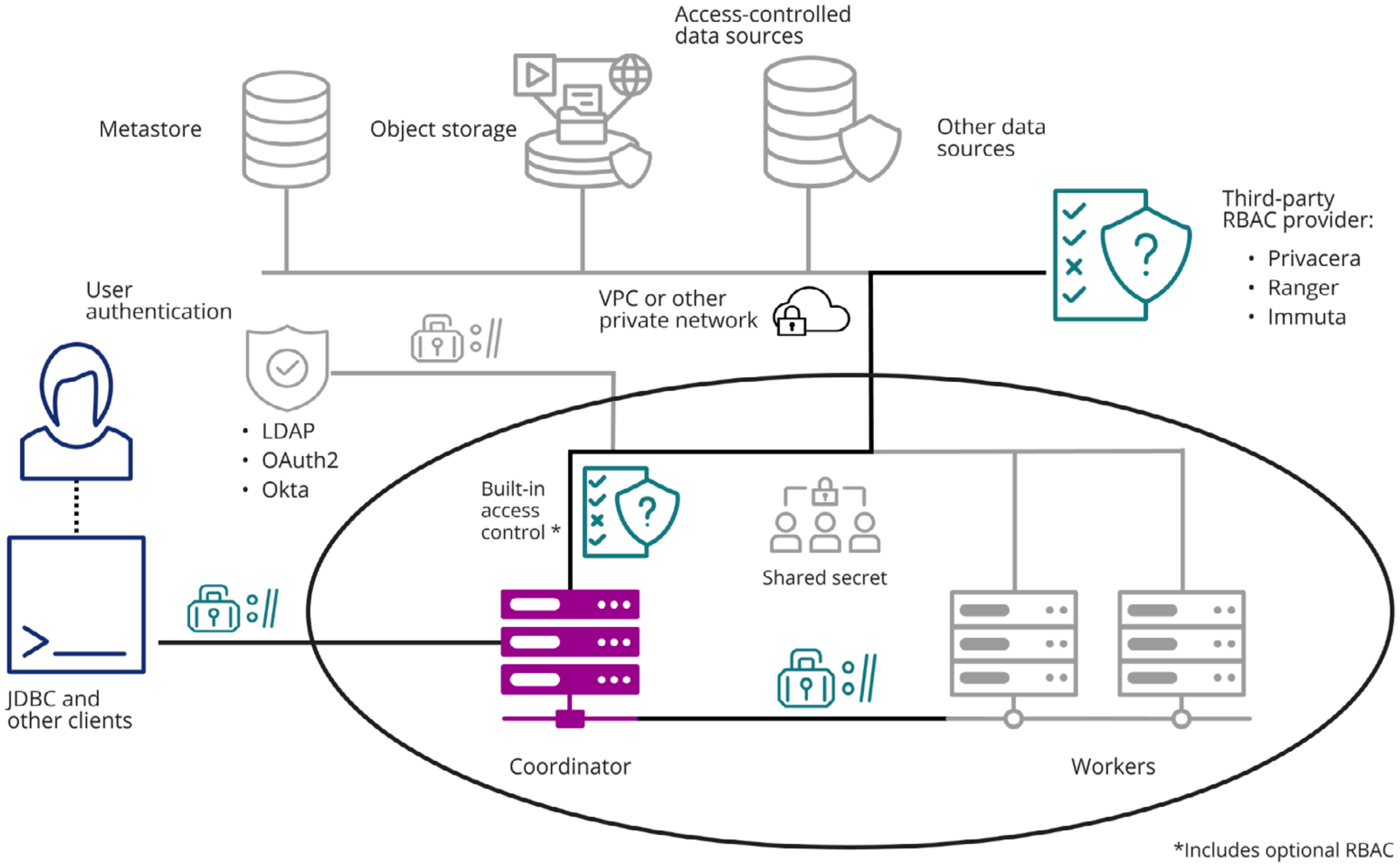
With the amount of data in a data lake and other source systems, limiting certain users and groups to sensitive data is a requirement for many organizations. Fine-grained access control allows column and row-level control against data sources.

Starburst Enterprise provides access control for all data sources. This allows fine-grained access control including column- and row-level policy enforcement as well as column-level data masking.



Built-in access control

Starburst Enterprise comes with built-in access control, and also integrates with open source Apache Ranger to provide a greater level of access control as well as offering column-level data masking. Groups and users can be synchronized to a central LDAP repository, and policies can be set to only allow approved users and groups access to data within the source with Apache Ranger.



*Includes optional RBAC

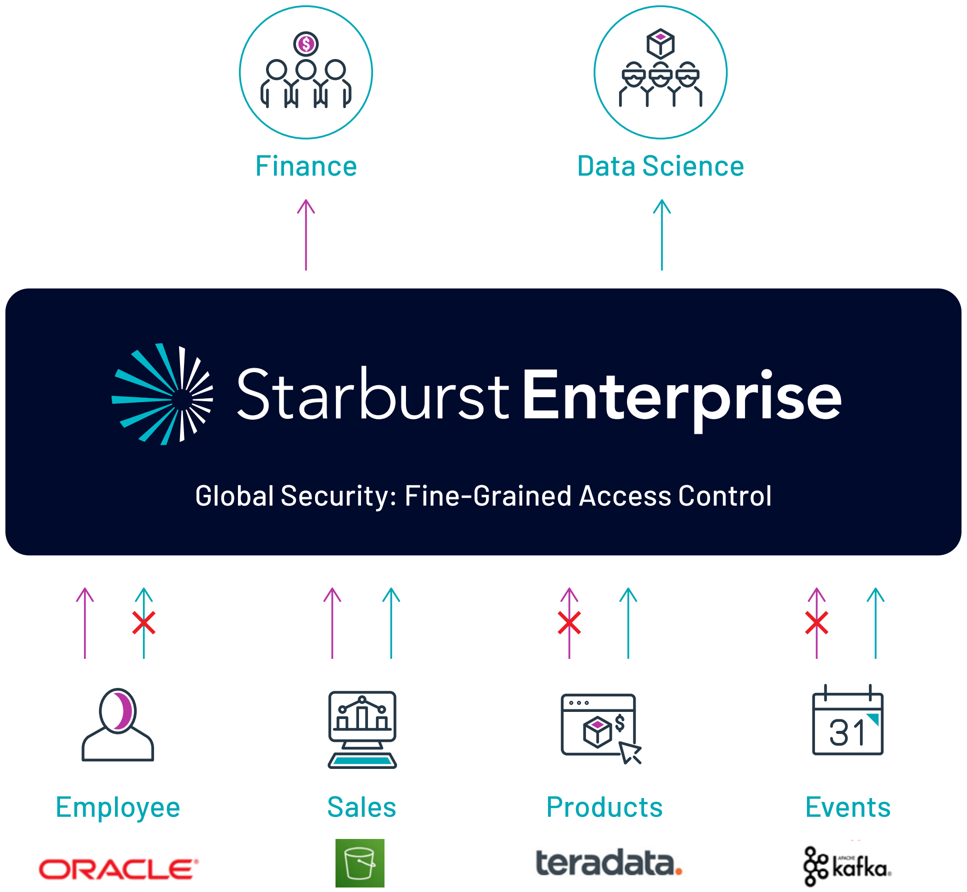
Groups and users can be synchronized to a central LDAP repository and policies can be set to only allow approved users and groups access to data within the source.

Catalog, schema and table level control

When providing access to many different data sources, it's crucial to an organization to be able to control access to this data. This becomes even more of a challenge when it comes to data residing in a data lake. These are often object stores with policy enforcement limited to the folder or bucket level.

OAuth

Starburst Enterprise includes support for OAuth 2.0 authentication. Supported identity providers include Okta, Azure Active Directory, Active Directory Federation Services, Keycloak and Google Cloud Platform. Starburst Enterprise can be configured to enable OAuth 2.0 authentication over HTTPS for the Web UI and the JDBC driver. Starburst Enterprise uses the Authorization Code flow which exchanges an Authorization Code for a token. The OAuth 2.0 token pass-through feature guarantees that Starburst Enterprise uses the same token as a user accessing a data source directly. This allows you to authenticate to Starburst Enterprise using OAuth 2.0, and the received token is passed through Starburst Enterprise and the connector to the underlying data source.



Starburst Enterprise provides the ability to limit groups and users by the catalog, schema and table. The diagram on this page illustrates a common scenario where one group of users do not have access to certain tables.

Column-level control

Tables in data lakes are often denormalized for performance and include many columns. In the case of many data lakes, this could be hundreds of columns. Providing access to a large group of users can be problematic if there is no functionality to limit access at the column level. Starburst Enterprise provides column-level policy enforcement at the group or user level. If the user submitting the query doesn't have access to a column or set of columns in the table, they will receive a message that based on an existing policy, they do not have access.

Row-level security

In a data lake or relational data source, there may be sensitive data that encompasses the entire row. Limiting access to this data for certain groups or users is a challenge in most systems. Starburst Enterprise provides the ability to limit rows of data within a table by enforcing row-level based policies.

In the diagram below, the East Manager would not be able to view rows in the sales table with a region = "West". Likewise, the West Manager wouldn't be able view rows that belong to the East region.

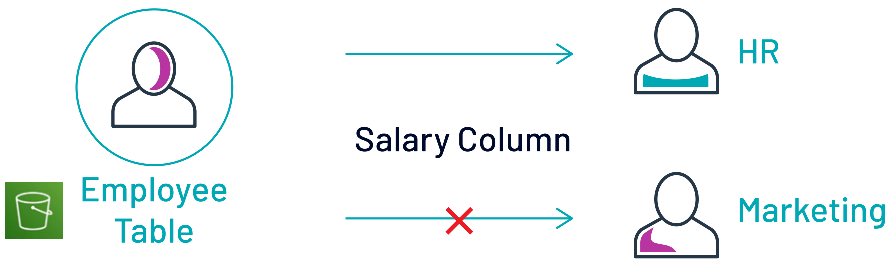
Data masking

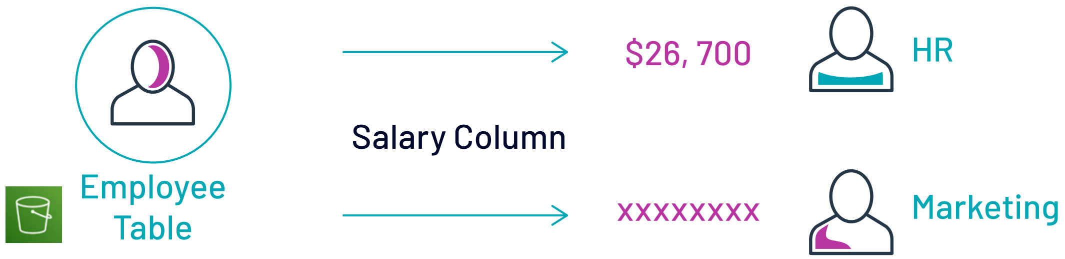
Column data can additionally be masked using Starburst Enterprise. Based on a user or group, columns of any tables can be masked using popular masking techniques such as redaction, partial masking and blanking-out sensitive data.

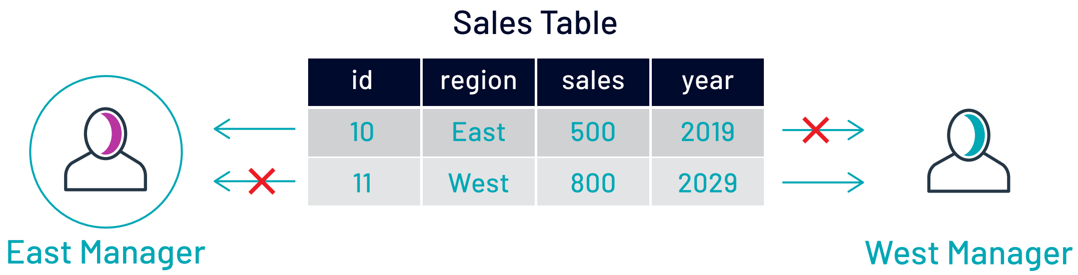
In the diagram below, when the HR and marketing users query the employee table, their results will vary based on their privileges. The HR user will be able to access the salary column but the marketing user's results will be masked.

Data products access control

With built-in access control enabled you can control which users are able to see, edit, create, delete, and publish data products. You can set privileges that apply globally (across all domains), privileges that apply to a specific domain, and privileges that apply to a specific data product.







Sales Table

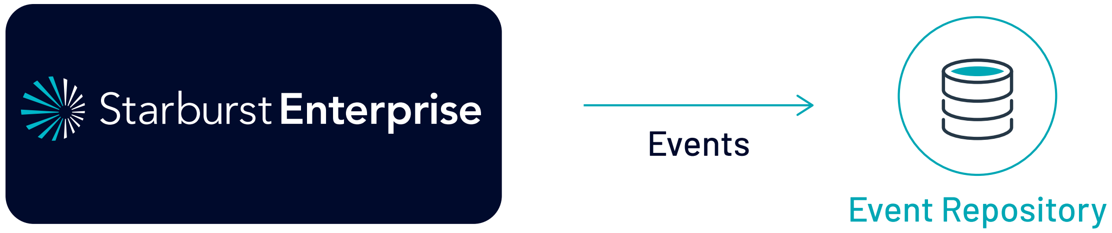
id	region	sales	year
10	East	500	2019
11	West	800	2029

Detailed Security Auditing

Logging information about queries is a requirement for many organizations. Starburst Enterprise includes logging capabilities.

Event Logging

Starburst Enterprise logs each query in great detail to a remote database. From here, this data can be queried with Starburst Enterprise or pulled into an existing event logging system.



Example fields that are available:

Column Name	Description
`query_id`	Randomly generated id of the query
`execution_time`	How long the query took to complete
`user`	The user that executed the query
`query`	The text of the query
`total_rows`	How many rows the query produced
`written_rows`	How many rows the query wrote to the target
`cpu_time`	Total cpu consumed on the cluster
`client_info`	Detailed information about the client. (JDBC,etc..)
`query_plan`	Plan the cost based optimized produced

Compliance

The ability to provide real-time query logging has been standard practice in the database industry for years.

As data volume constantly increases, companies are under more and more pressure to monitor data access within their organization. With Starburst Enterprise's event logging functionality, a full, GDPR-level audit trail is available in real-time. This allows tracking access to all data sources used in queries submitted to Starburst Enterprise.

Chargeback

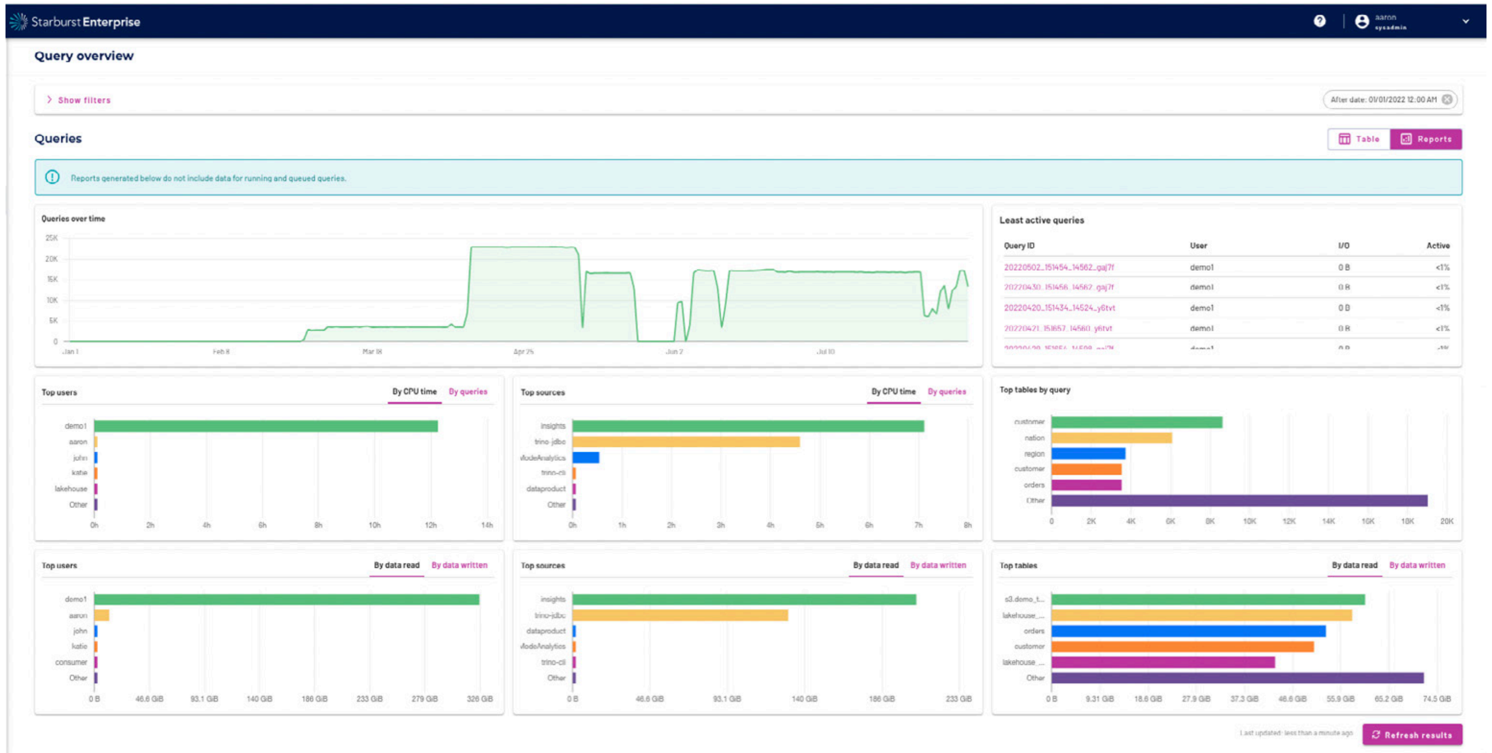
Starburst Enterprise is used by many different departments and user groups. It can be difficult for a centralized IT organization to determine the resource usage for these different users. With event logging, each query is logged with the user that executed the query, the text of the query and the elapsed time, as well as other metrics such as RAM and CPU to provide a more granular level of detail of actual usage.

Performance tuning

The data collected for each query includes resource utilization. This data enables resource usage per query and can be used to determine queries, users and data sources where performance tuning might be considered. Reporting can easily be created to monitor Starburst Enterprise usage based on users, connectors and tables.

Insights

Insights provides a graphical interface to view query history. Users can easily search and view queries that were executed on the cluster. In addition, detailed statistics are available for each query. Insights provides a visual overview of important metrics about your Starburst Enterprise cluster for all types of users, from platform administrators to data consumers. From the Insights interface, you can access detailed query history, including single-query statistics and query plans, and cluster performance information from a selected date range.

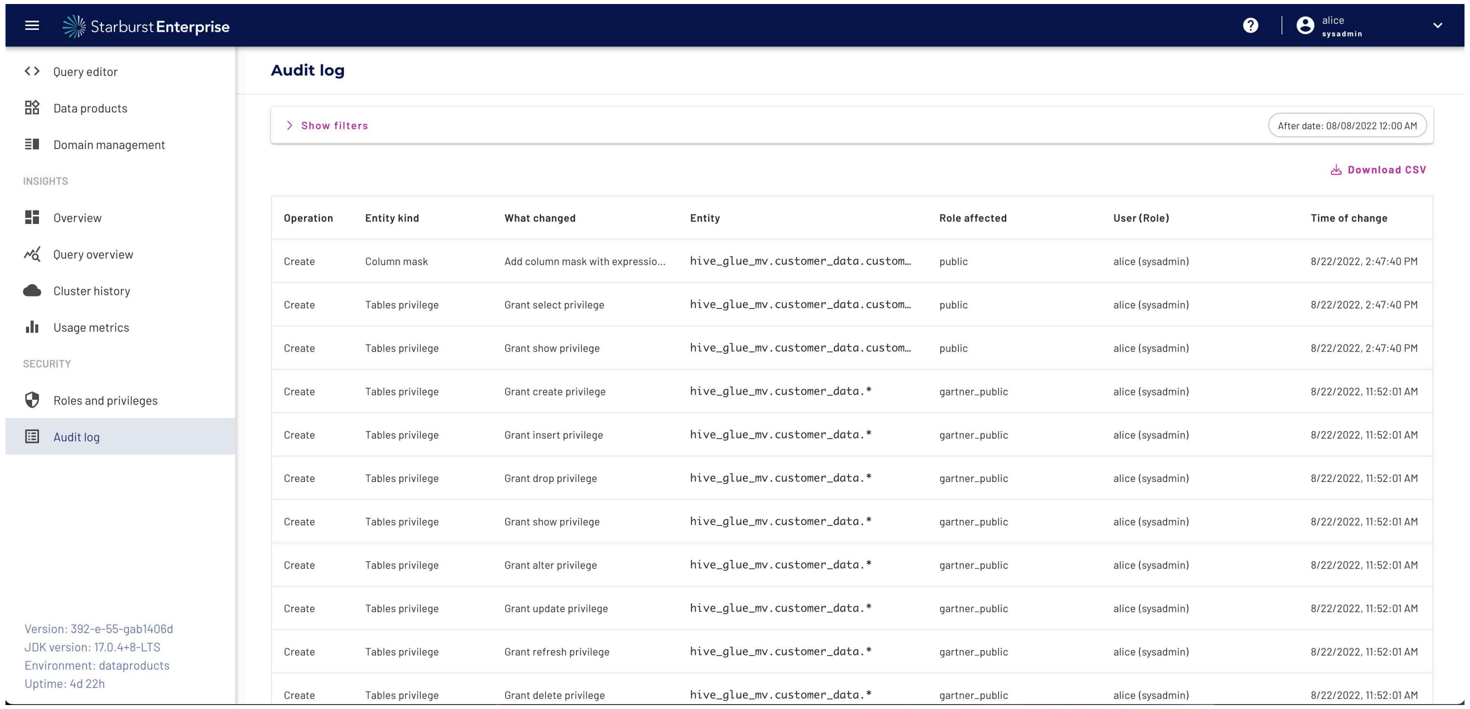


Reports generated below do not include data for running and queued queries.

Query ID	User	I/O	Active
20220502_151454_14562_gaj7f	demo1	0 B	<1%
20220430_151456_14562_gaj7f	demo1	0 B	<1%
20220420_151434_14524_y6tvt	demo1	0 B	<1%
20220421_151657_14560_y6tvt	demo1	0 B	<1%

Built-in access control audit log

Starburst Enterprise maintains logs of all access control changes performed through its built-in access control functionality. When enabled, the Audit log feature appears for members of the sysadmin role.



Version: 392-e-55-gab1406d
JDK version: 17.0.4+8-LTS
Environment: dataproducts
Uptime: 4d 22h

Audit log

Operation	Entity kind	What changed	Entity	Role affected	User (Role)	Time of change
Create	Column mask	Add column mask with expressio...	hive_glue_mv.customer_data.custom...	public	alice (sysadmin)	8/22/2022, 2:47:40 PM
Create	Tables privilege	Grant select privilege	hive_glue_mv.customer_data.custom...	public	alice (sysadmin)	8/22/2022, 2:47:40 PM
Create	Tables privilege	Grant show privilege	hive_glue_mv.customer_data.custom...	public	alice (sysadmin)	8/22/2022, 2:47:40 PM
Create	Tables privilege	Grant create privilege	hive_glue_mv.customer_data.*	gartner_public	alice (sysadmin)	8/22/2022, 11:52:01 AM
Create	Tables privilege	Grant insert privilege	hive_glue_mv.customer_data.*	gartner_public	alice (sysadmin)	8/22/2022, 11:52:01 AM
Create	Tables privilege	Grant drop privilege	hive_glue_mv.customer_data.*	gartner_public	alice (sysadmin)	8/22/2022, 11:52:01 AM
Create	Tables privilege	Grant show privilege	hive_glue_mv.customer_data.*	gartner_public	alice (sysadmin)	8/22/2022, 11:52:01 AM
Create	Tables privilege	Grant alter privilege	hive_glue_mv.customer_data.*	gartner_public	alice (sysadmin)	8/22/2022, 11:52:01 AM
Create	Tables privilege	Grant update privilege	hive_glue_mv.customer_data.*	gartner_public	alice (sysadmin)	8/22/2022, 11:52:01 AM
Create	Tables privilege	Grant refresh privilege	hive_glue_mv.customer_data.*	gartner_public	alice (sysadmin)	8/22/2022, 11:52:01 AM
Create	Tables privilege	Grant delete privilege	hive_glue_mv.customer_data.*	gartner_public	alice (sysadmin)	8/22/2022, 11:52:01 AM

Partner Integrations

Immuta

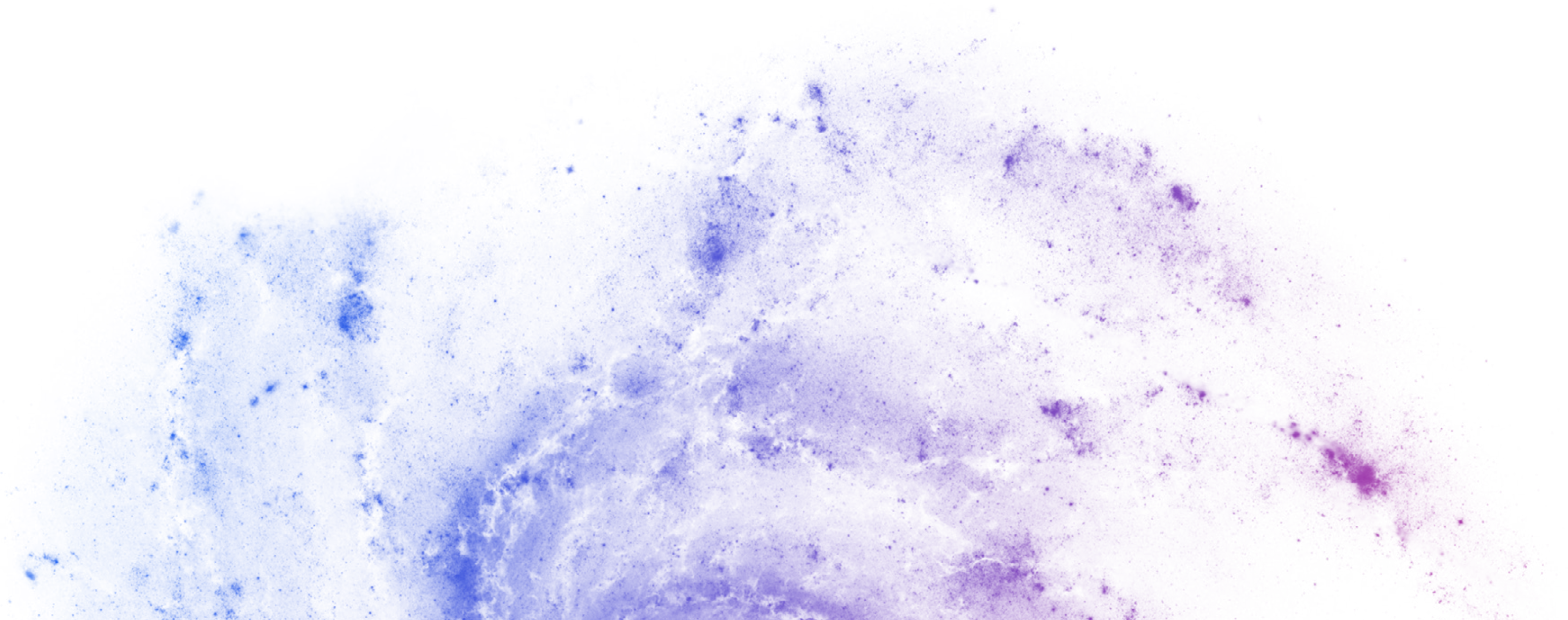
Immuta's native integration with Starburst enables data engineering and operations teams to scale secure data access so that any user can access and share data — even the most sensitive data — across all compute environments in real-time. With Starburst as a single point of access and Immuta as a single point of access control, data teams can optimize performance and streamline self-service data access from a centralized access control plane.

To learn more about the Immuta integration, visit our documentation: **docs.starburst.io/latest/security/immuta-overview.html**

Privacera

Privacera combined with Starburst provides enterprises with centralized, secure access to data that includes end-to-end governance and compliance. Privacera integrates with Starburst Enterprise enabling users to seamlessly integrate with a secure infrastructure to run federated queries across multiple databases without sacrificing privacy, governance, or compliance. Regardless of where data resides, or where users are querying from, combination of Starburst and Privacera provides a secure infrastructure to run federated analytics across multi-cloud or hybrid cloud infrastructures breaking down data silos and accelerating time-to-insight while ensuring consistent data governance and compliance.

To learn more about the Privacera integration, visit our documentation: **docs.starburst.io/latest/security/global-privacera.html**



Summary

Trino is adopted by some of the world's most innovative companies such as Pinterest, Lyft, Netflix, LinkedIn and many more. As the awareness and adoption of Trino has grown, Starburst has worked to deliver enterprise-grade features and support to ensure organizations are successful with Trino.

A critical part of ensuring Trino is successful at scale is providing a centralized security framework that meets enterprise standards, and Starburst Enterprise is designed for that. Controlling access to data wherever it lies within a data-intensive organization conforms to strict new data governance and security regulations.

Visit **docs.starburst.io/latest/security.html** to learn more about our robust security features

Starburst **Enterprise**