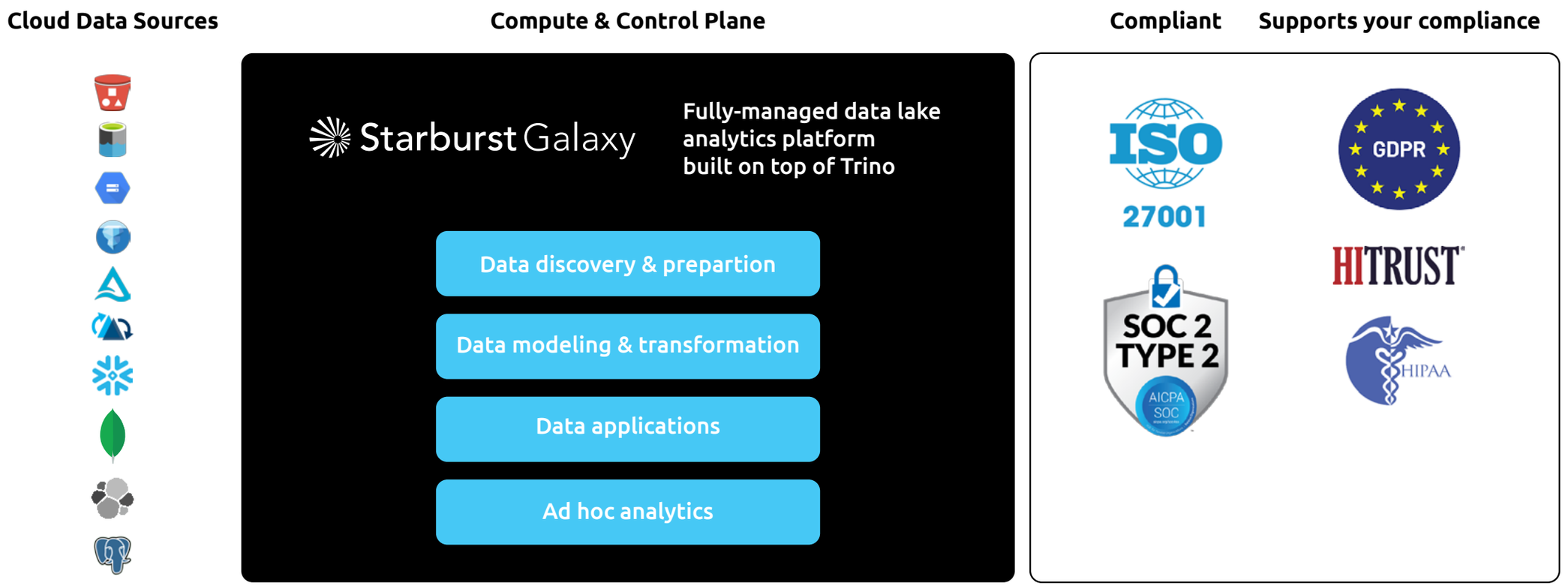
E-BOOK

Starburst Galaxy Security Guide

Table of Contents

This document provides an overview of the Starburst Galaxy security capabilities and features.
For additional details refer to the detailed documentation, here.



Enterprise-grade security for Galaxy

Galaxy delivers enterprise-grade security features that are embedded within a secure and mature cloud data platform. Galaxy controls integrate with your existing security policies and standards to support compliance with regulatory requirements.

At its core, Galaxy is a compute engine that leverages the customer's storage environment. This unique approach empowers our customers to have greater control over their data security.

Compliance & Regulation

Starburst undergoes independent verification of platform security, privacy, and compliance controls to help you meet regulatory and policy objectives, including the unique compliance needs of highly regulated industries.

Data Sources


Cloud data lake


Object storage


Cloud data warehouse


RDBMS


Streaming

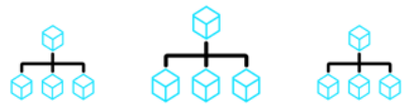
NoSQL

- Galaxy is designed to store data within the customer's environment and existing control framework.
- Data is always encrypted in transit, whether it's across a public or private network.
- Data source connectors support modern authentication and authorization protocols.

Compute Plane

Starburst Galaxy

Analytics engine for analyzing large amounts of distributed data with high concurrency.

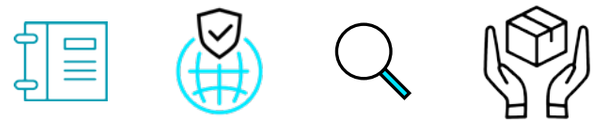

Trino Clusters

- Each customer has a dedicated compute plane with the ability to manage what specific users can do.
- At the customer's discretion, data may be cached for performance, it is encrypted in memory and deleted after use.
- The compute plane securely connects to customer data sources based on the configured access privileges.

Control Plane

Starburst Galaxy

Manages the overall application and it is deployed and managed by Starburst in our cloud environments



- Access to the control plane, compute plane and data sources can be managed with RBAC and ABAC.
- Row level filters and column masking can be configured by the customer for each data asset.
- The customer can specify where their data will be processed.

Starburst Galaxy Architecture

Data Sources

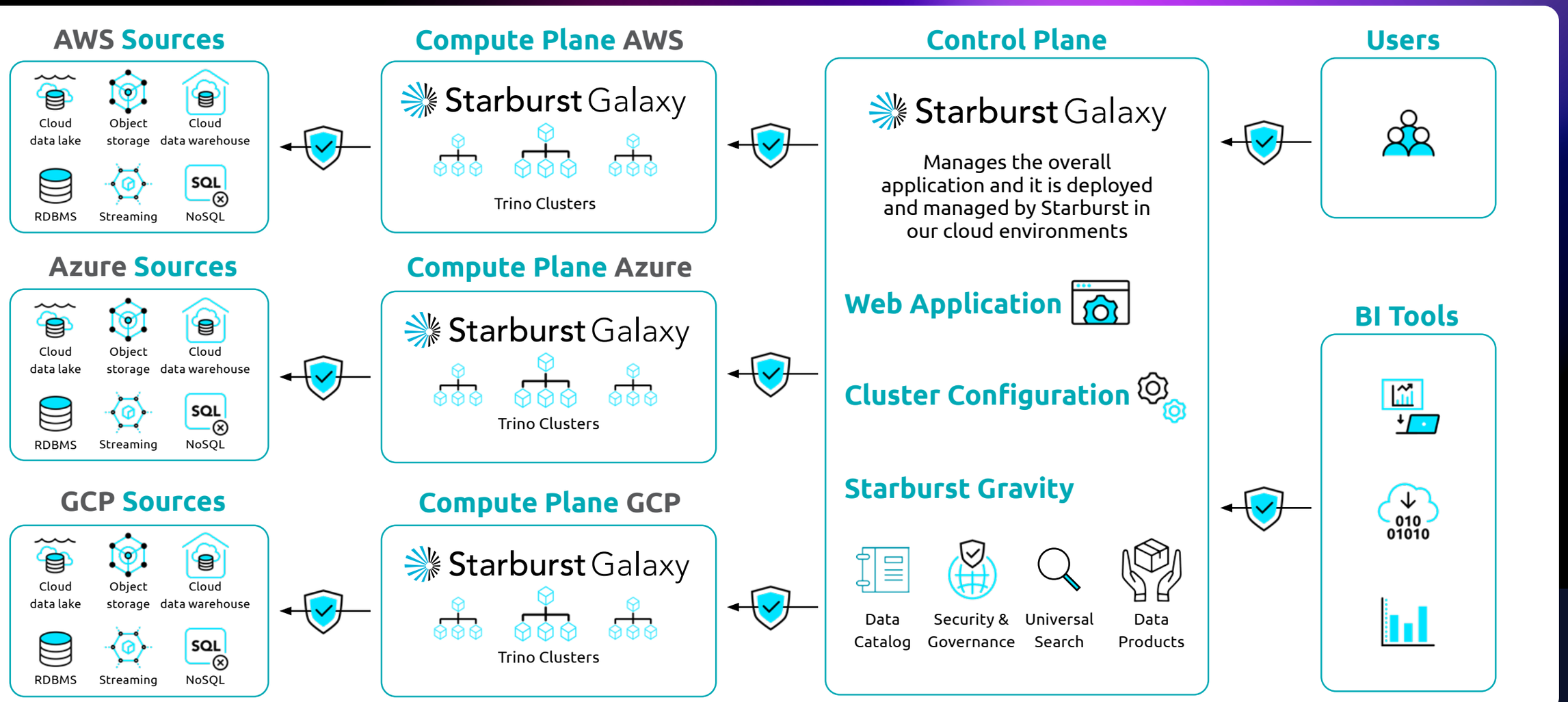
The customer's cloud architecture and data sources are unchanged. Galaxy connects to the data where it sits today. Data is always stored within the customer's environment.

Compute Plane

The Galaxy compute plane is where both ad hoc and batch workloads are executed on the only SaaS-based Trino MPP engine. Each customer has dedicated, fully scalable Trino clusters, within their private compute plane.

Control Plane

The Galaxy control plane is the central portal for managing all aspects of the application and configuring how data is accessed and transformed. The control plane is managed by Starburst and deployed in our cloud environment. Galaxy customer's use the control plane to securely manage their environment. Within Starburst, only a dedicated secure admin team has access to perform platform management activities.



Starburst Galaxy Architecture (continued)

Deployment Model

Galaxy is a fully managed, secure SaaS solution, there is no on-premise infrastructure required.

With Galaxy, the compute layer exists in each customer's private, dedicated Galaxy account.

The data sources are managed by the customer within their cloud infrastructure. The data sources remain under the customer's complete control.

Data Movement

There is no data migration required. Galaxy can federate across your existing data sources, pulling only the granular data assets that are requested by the query.

Data from multiple sources can then be integrated and transformed, before it is delivered to the customer. The result of the query is not persistent within the Galaxy infrastructure, it is only delivered to the consumer.

Data Security

Data is always encrypted at rest or in transit, whether it's across a public or private network. Cloudflare integration provides robust network protection for Galaxy, ensuring speed, availability and security.

- End-to-end encryption
- Fine-grained data security and masking
- Connection via AWS PrivateLink (Azure and GCP private connect coming soon)

Data Access

Data access is configured via RBAC and ABAC by the customer. Galaxy can integrate with third-party authentication services to increase security and centralize security controls.

- SSO with identity provider integration
- SCIM provisioning to manage user identities and groups
- Service principals or service accounts to manage application identities

Starburst Galaxy Identity & Access

Galaxy offers best in class access management capabilities.

Our rapidly evolving ABAC capabilities combined with our insightful end-to-end activity monitoring capabilities, provide the infosecurity teams with a modern toolset for protecting even their higher risk data assets.

1

User authenticates with their email and password or via Single Sign-on.

Galaxy can integrate with IdPs that support SAML, like Okta, AzureAD, Google Workspace, PingID, Auth0 or a custom IdP. Galaxy also support the SCIM protocol, which allows changes in the IdP's user and group database to be transmitted to Galaxy.

Credentials are encrypted and no Starburst personnel have access to a user's plain text credentials. All other customer-related data such as catalog configuration or cluster setup are encrypted in motion and at rest the moment it enters Starburst systems.

2

Starburst Galaxy includes a role-based access control (RBAC) system to support Starburst Galaxy, the clusters, and the configured catalogs with the data from the data sources for every user.

Users are assigned one or more roles. A role has a name and an optional description, and can be assigned privileges on entities, such as cluster management, user creation, audit log viewing, and others. You can manage users, roles, and privileges in the Starburst Galaxy user interface.

Starburst Galaxy includes an attribute-based access control (ABAC) system that uses policies and attributes, such as tags, to help further manage role access to entities like catalogs, schemas, tables, and views. You can manage policies and tags in the Starburst Galaxy user interface.

3

Galaxy access to the customer's data sources is managed by the customer through their dedicated control plane.

The primary controls are established at the data sources with their cloud provider. Galaxy is configured to only access the specifically approved, granular data assets. For added security, columns and rows may be filtered depending on each user's unique access rights.

Data authorization is decided before the query is executed. There is no need to attach data to a query in order to make an access decision for the data. If you don't have access to the data, you can't run the query.

Starburst Galaxy Data Protection

Starburst maintains a high standard for our internal security posture, as evidenced by the independent certifications.

Security Testing

Starburst undergoes annual penetration tests, which includes assessments against the OWASP framework. In addition, our vulnerability assessment program (including our consistent code scanning practice, using Veracode) includes regularly confirming what OWASP risks are identified.

Data Storage & Encryption

Galaxy stores data in the customer's environment, except when clusters with Warp Speed acceleration use caches that reside on Galaxy SSD storage.

These caches are fully encrypted at rest and there is no means of direct access for end users.

Starburst Access to Data

Starburst personnel, by default, do not have access to customer clusters or the control plane. Access to customer data is granted on a need-to-know basis to provide customer support, fulfill legal requirements, or for other legitimate business purposes.

Starburst Galaxy Data Governance

Gravity provides a unified, searchable and taggable governance layer that lets users manage, govern, and secure all of their data assets in a comprehensive data lake analytics platform. With Gravity, users can create and maintain trust as organizations is maturing in their data journey.

Starburst Gravity

Universal search

Universal search provides a seamless and efficient experience, enabling data teams to make faster and more informed decisions.

Easily search for the names of catalogs, schemas, tables, views, columns, or even your data products.

Data catalogs

An automatic data cataloging solution – simply connect your data, and Gravity pulls in all the relevant metadata directly from the source.

Catalog explorer is a centralized hub for all your data assets, enabling you to discover and explore your organization's data universe effortlessly.

Access controls

Elevate the security of your data with with centralized governance across disparate sources. Fine-grained access control allows for precise permissions based on catalogs, schema, tables, columns, attributes, or user roles, ensuring authorized access to sensitive information.

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.

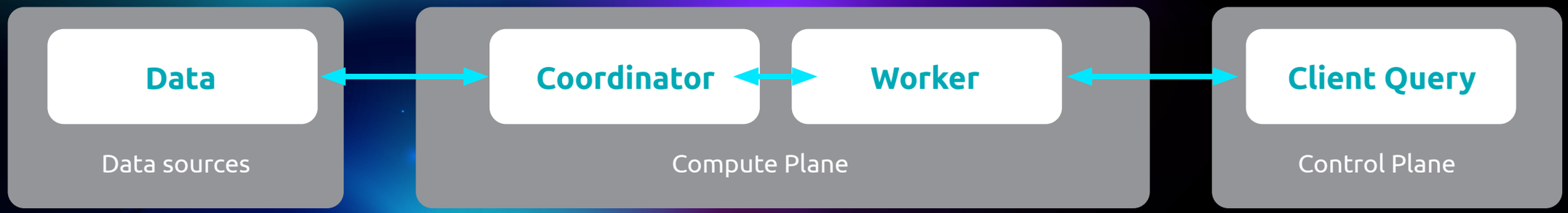
Observability

Provides visibility into the data flow from upstream data sources of data products, enabling data consumers to more confidently determine the truthfulness of data flowing into a data product.

Starburst Galaxy Network Controls

All communication is encrypted end-to-end, from connecting to the control plane, to cluster communication, to connecting with source data. The customer controls all network access to their data sources.

Secure private connectivity may be required for some cloud resources. Connection via AWS PrivateLink is available now, (Azure and GCP private connect coming soon).



Our workloads are co-located on network segments that are isolated via a Container Network Interface (CNI).

You can configure allow lists to define access for specific IPs.

Multiple secure authentication options are available for our catalog connectors.

Starburst Galaxy Audit & Monitoring

Starburst Galaxy offers market leading audit features, including comprehensive logging of events and end-to-end user activities. Galaxy also automates health and performance monitoring to provide observability to ensure services are functioning optimally.

Data Sources

Log Every Connection

Every data source query is logged. Information like user, user group, client, among others can be used for further audit and investigation purpose.

External Logs

Audit logs from external systems can be queried to support audit focused analytics.

Compute Plane

Change Management

The following actions are logged:

- **Operation:** The configuration change that was performed. Possible values include update, delete, suspend, and others depending on the object.
- **Object:** The type of object, such as cluster or catalog, for which the configuration was changed.
- **Object name:** The name of the specific object.
- **What changed:** A description of the changed attributes including values.
- **User:** The user who performed the change.
- **Time of change:** The date and time the change was performed.

Query History

The Query History section and the Telemetry Catalog tables enables you you analyze and audit at the query level. (without any performance degradation on the cluster). Information like user, user group, client, type of query (select, insert, delete, update), objects (tables/views, columns) among others can be used for further audit and investigation purpose.

Control Plane

Audit Logs

The control plane enables access to all administrative actions and compute activities performed by all users in Starburst Galaxy.

All Starburst audit logs are kept in perpetuity for security and audit purposes. They do not expire and are never purged. Audit logs are available to account administrators.

Audit Data Products

Leverage the power of Starburst's federation engine to integrate Starburst audit logs with logs across multiple systems. Auditors can integrate multiple data logs and develop their own audit focused data products and dashboards.