Better, faster data-based decisions

Starburst Enterprise on Red Hat OpenShift

Highlights

Stop moving data and start unlocking its value by accessing and analyzing data from any data source at scale.

Gain insights quickly without worrying about where data is stored, what form it is stored in, or whether it is in flight or at rest.

Modernize your data infrastructure at your own pace while answering your most critical business questions.

Deploy compute and storage resources efficiently and cost-effectively to perform rapid reporting, business intelligence analysis, and SQL queries.

Are disparate data sources slowing you down?

Making better and more timely decisions is essential to improving performance and remaining competitive. At the same time, many large organizations struggle with how best to combine, query, and analyze isolated and disparate data sets. Traditional data warehouse products approach the problem with an extract, transform, and load (ETL) process that copies data from 1 or more sources into a destination system. This outdated and monolithic technique is inefficient and keeps business analysts from running fast analytics on their data, potentially delaying critical insights.

In contrast, Starburst Enterprise gives organizations the freedom to examine diverse data sets wherever they are located without building a data warehouse. You can run multiple query clusters, scaling up or down dynamically and optimizing for speed and cost as desired. Starburst Enterprise lets you access multiple software-defined data storage platforms deployed with Red Hat® OpenShift®. For instance, data can be queried instantly and simultaneously from a warehouse, data lake or lakehouse, an SQL or NoSQL database running on Red Hat OpenShift or externally, and from data sources existing in other environments.

Starburst Enterprise on Red Hat OpenShift

Distributed cloud and hybrid cloud applications are increasingly popular, but the transition to cloud deployments cannot take place instantly. As a result, many organizations rely on a combination of traditional and modern applications to run the business and make critical decisions. Likewise, most employ a combination of both traditional and modern data sources with data scattered across datacenters and cloud and vendor environments.

For example, an analyst might need to combine data from a PostgreSQL application on a Kubernetes persistent volume (PV), a general ledger running in Microsoft SQL Server, and archived client data in an object store. Analysts need reliable, consistent user and operational experiences that let them develop applications and analyze data from diverse sources rapidly without having to worry about the underlying infrastructure.

Starburst Enterprise

Starburst Enterprise provides a modern solution built on an open source distributed SQL query engine project, Trino. Trino can perform interactive analytic queries against data sources of all sizes, ranging from gigabytes to petabytes. It approaches the speed of commercial data warehouses while scaling to the size of large organizations. Starburst Enterprise deployed on Red Hat OpenShift adds the tools and 24x7 support you need for big data access at scale.

The Starburst Enterprise platform provides distributed query support for varied data sources, including:

- NoSQL systems (MongoDB, Cassandra, Redis).
- SQL databases (Microsoft SQL Server, MySQL, PostgreSQL).
- Data warehouses (IBM Db2 Warehouse, Teradata, Oracle Exadata, Snowflake).

- Hive (HDFS, Cloudera, MapR).
- Data services (Kafka, Elasticsearch, Red Hat OpenShift Data Foundation).
- Cloud object storage (AWS S3, ADLS, Azure Blob, Ceph®, IBM COS).

Starburst on Red Hat OpenShift

Operating Trino at scale can present challenges, especially if you are attempting to size and configure environments manually. You need to achieve petabyte scale with autoscaling and be able to decommission unneeded resources gracefully. Combining Starburst Enterprise with Red Hat OpenShift addresses these needs by offering automation, high availability, elasticity, and monitoring for Trino clusters.

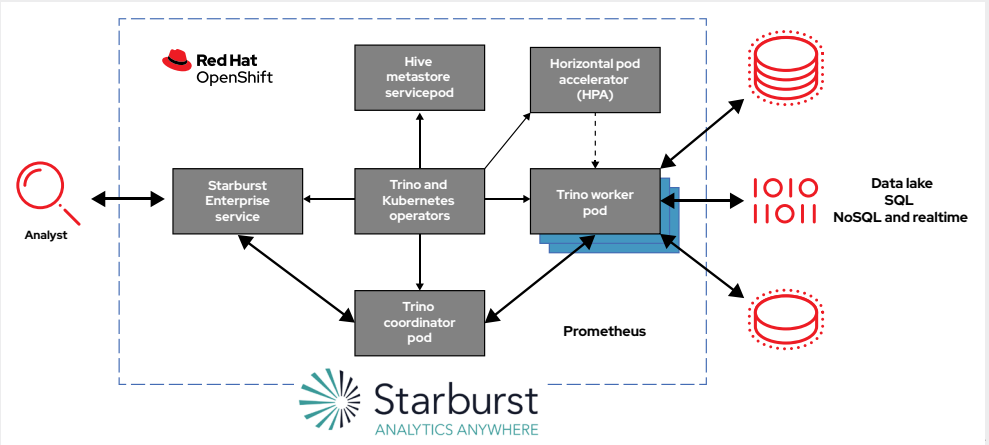
Kubernetes Operators delivered with Red Hat OpenShift automate installation, upgrades, and life cycle management throughout the container stack. A Kubernetes Operator for Starburst Enterprise lets Red Hat OpenShift greatly simplify the administration of a Starburst Enterprise cluster. Together, these operators offer benefits that include:

Red Hat OpenShift automates the provisioning, management, and scaling of applications so that you can focus on writing the code for your next big idea.

- **Automation.** Red Hat OpenShift and Starburst Enterprise operators provide automatic configuration, tuning, and management of Starburst Enterprise clusters. Red Hat OpenShift operators determine what to deploy, including identifying hardware and provisioning new instances. Starburst Enterprise operators manage updates to the environment.
- **High availability.** Continuous operation of the Trino coordinator is essential. Using liveness probes, the Red Hat OpenShift load balancer can keep services like the Trino coordinator in an always-on state.
- **Elastic scalability.** Red Hat OpenShift simplifies deploying, operating, and scaling workloads in Kubernetes infrastructure. It can automatically scale the Trino worker pods based on query load. Using the Kubernetes Horizontal Pod Autoscaler (HPA), you can specify thresholds for Trino worker pods. As the number of queries increases, the HPA will automatically spin up additional worker pods based on specified system constraints.
- **Graceful scale-down and decommissioning.** With Red Hat OpenShift, reduced load does not imply system downtime or killed queries. The Kubernetes HPA will gracefully decommission unused Trino worker pods and free system resources for other tasks without service interruptions.
- **Monitoring for all hardware and software layers.** Prometheus, the cluster monitoring service for Red Hat OpenShift, delivers metrics and alerts that inform Kubernetes orchestration and populate the Red Hat OpenShift dashboard. Prometheus informs Kubernetes if a pod is offline and provides metrics to the HPA to let it know whether to commission or decommission additional Trino pods.
- **Red Hat partner ecosystem support.** Starburst Enterprise lets you use data storage platforms associated with Red Hat OpenShift. Application developers can make use of SQL and NoSQL databases backed by popular cloud-native storage solutions, such as Red Hat OpenShift Data Foundation, Portworx by Pure Storage, or NetApp. Business users can pull data from data lakes running on AWS S3. In addition, several Red Hat ecosystem partners, such as Aerospike and Dell Technologies, integrate Starburst into their products.

Figure 1 illustrates how Starburst Enterprise works with Red Hat OpenShift and related Red Hat storage platforms.

Starburst Enterprise lets you run analytics anywhere to make better business decisions.



Red Hat OpenShift is available as self-managed on-site or extended to the cloud with Azure Red Hat OpenShift and Red Hat OpenShift on AWS.

Figure 1: Starburst Enterprise on Red Hat OpenShift.

Use cases

Starburst Enterprise accommodates a wide range of use cases when running on Red Hat OpenShift, including:

- **Data modernization.** Starburst Enterprise lets you modernize data at your own pace while you work with the environment you have. You can update, migrate, and move data as it makes sense for the business—without forced data migrations.
- **ETL workloads.** Starburst Enterprise is ANSI SQL-compliant for support of create table and insert statements. It can act as the SQL engine for ETL jobs, providing a single platform for both query and migration needs. For example, archived data from an Apache Hadoop cluster could be moved to a data lake on an object store, allowing federated Trino queries against that data along with data from other sources that are not ready to migrate.
- **Interactive data investigation.** Starburst Enterprise allows rapid ad hoc interactive queries from a range of data sources—including traditional, real-time, and object stores. Database administrators can query underlying sources from their SQL or business intelligence tools of choice. Data can be queried rapidly from a single source or combined through federated joins.
- **Business intelligence (BI) dashboarding and reporting.** Data consumers can work with their favorite dashboarding and reporting BI tools, such as Tableau, MicroStrategy, or Metabase. Because Starburst Enterprise separates compute and storage resources, it provides the interactive responsiveness these tools require.
- **Data science.** Data scientists need access to data for model development and machine learning (ML). Starburst Enterprise, used in conjunction with Red Hat OpenShift AI, fulfills these requirements. This solution allows data scientists to rapidly perform data discovery in place and ingest appropriate source data into their machine learning scratch space through a standard open or Java™ database connectivity (ODBC/JDBC) package interface. Red Hat OpenShift AI also offers tools across the full life cycle of AI/ML experiments.

Extending data discovery for analytics

By providing a standardized foundation for creating production AI/ML models and running the resulting applications, Red Hat OpenShift AI delivers the consistency, ease of use, and cloud-to-edge deployment options of Red Hat OpenShift, allowing platform engineers to create scalable configurations that are specific to the needs of their data scientists and developers.

Red Hat OpenShift Data Science is a product in the Red Hat OpenShift AI family and provides tools to build, deploy, and monitor models. Starburst is integrated into the OpenShift Data Science dashboard, providing tools for data acquisition and preparation for ML models.

With Starburst Enterprise, you can perform federated data queries across your different data sources—whether structured, semi-structured, or unstructured—even using different protocols. You can also perform in situ data analysis across various file systems, databases, and object stores delivered in storage platforms. This can significantly impact time to value for data analysts and data scientists who are performing discovery across many data sources to support their models.

As a result, *data scientists* can:

- Discover the data they need to support their ML models through federated queries, getting them to the data they need, regardless of where it rests.
- More easily join datasets via parallel connections, connect multiple sources, and analyze data more quickly.
- Pull only the data necessary for model training, once they have a clear vision of the data and their model. This can accelerate data acquisition and preparation while reducing data egress costs.

Data directors and *data architects* can:

- Execute high-speed queries to give business analysts the information they need now.
- Optimize their cloud spend by separating compute and storage while eliminating the time spent on complex ETL processes and data preparation.
- Build modern data architectures that are open, flexible, and capable of existing within any stack.

Conclusion

Starburst Enterprise and Red Hat OpenShift promote better and more timely insights by letting you rapidly analyze data across multiple disparate and distributed data platforms. The combination provides critical automation, high availability, elasticity, and monitoring that meets the demands of enterprise organizations. When developing AI-enabled applications with Red Hat OpenShift Data Science, Starburst provides essential data services for IT operations leaders, data scientists, and developers—facilitating their ability to train, serve, monitor, and manage the life cycle of AI/ML models and applications from experiments to production. Learn more about Starburst Enterprise in the Red Hat Ecosystem Catalog.

Find out more

Make better decisions with fast, comprehensive data access (brief)

What can you accomplish with fast access to all data? (checklist)

About Red Hat

Red Hat helps customers standardize across environments, develop cloud-native applications, and integrate, automate, secure, and manage complex environments with award-winning support, training, and consulting services.

facebook.com/redhatinc
@RedHat
linkedin.com/company/red-hat

redhat.com
#454171_0823

North America
1 888 REDHAT1
www.redhat.com

Europe, Middle East, and Africa
00800 7334 2835
europe@redhat.com

Asia Pacific
+65 6490 4200
apac@redhat.com

Latin America
+54 11 4329 7300
info-latam@redhat.com