

REPORT REPRINT

Starburst previews Galaxy for analytics acceleration across multicloud and hybrid IT

MARCH 25 2021

By Matt Aslett

Distributed query specialist Starburst has announced its plans to deliver a managed cloud offering, Starburst Galaxy, that will enable users to accelerate the analysis of data in data lakes and databases across multiple clouds, as well as on-premises.

THIS REPORT, LICENSED TO STARBURST DATA, DEVELOPED AND AS PROVIDED BY 451 RESEARCH, LLC, WAS PUBLISHED AS PART OF OUR SYNDICATED MARKET INSIGHT SUBSCRIPTION SERVICE. IT SHALL BE OWNED IN ITS ENTIRETY BY 451 RESEARCH, LLC. THIS REPORT IS SOLELY INTENDED FOR USE BY THE RECIPIENT AND MAY NOT BE REPRODUCED OR RE-POSTED, IN WHOLE OR IN PART, BY THE RECIPIENT WITHOUT EXPRESS PERMISSION FROM 451 RESEARCH.

451 Research

S&P Global

Market Intelligence

Introduction

Starburst has established itself as one of the key vendors in the data abstraction and analytics acceleration space thanks to its Starburst Enterprise offering, based on the open source Trino project (formerly PrestoSQL). Having recently raised a \$100m series C funding round, valuing the company at \$1.2bn, Starburst has announced plans to deliver a managed cloud offering, Starburst Galaxy, that will enable users to accelerate the analysis of data in data lakes and databases across multiple clouds, as well as on-premises.

451 TAKE

Starburst has evolved rapidly in recent months, raising \$164m and gaining unicorn status in a little over a year after it started raising outside funding. The company has also grown a healthy number of customers, an increasing proportion of which are coming directly to Starburst, rather than starting with Trino/Presto and then looking for assistance. Distancing itself from the successful Presto brand is a potential banana skin, but one that can be counteracted by driving developer and user awareness of Trino. In the meantime, Starburst is focusing customer attention on its broader plans for Starburst Galaxy, which will enable customers to take a more strategic view of data abstraction and analytics acceleration across their data estates, whether on-premises, in the cloud or (probably) both.

Details

Starburst was spun off from Teradata in 2017 to build a business around what was then the PrestoSQL open source distributed query engine project. Much has changed since then – not least PrestoSQL rebranding as Trino (see below) and Starburst establishing itself as a key player in the data abstraction and analytics acceleration space with 250 employees and over 100 paying customers for the Starburst Enterprise product.

While the company was initially bootstrapped, it has also raised \$164m in three quick-fire funding rounds since November 2019. The latest \$100m series C round was announced in January and led by Andreessen Horowitz, with participation from Salesforce Ventures, Index Ventures and Coatue, valuing the company at \$1.2bn.

That valuation is based on the opportunity for Starburst's products, which enable users to create clusters of the Trino distributed SQL query engine for the purpose of accelerating the analysis of data stored in cloud storage services, as well as Apache Hadoop-based products and services, and relational and nonrelational databases, as part of what we have called an abstracted data architecture.

In addition to enabling querying of data in each of those repositories and locations, Trino enables users to federate queries across multiple data repositories and locations, while Starburst Enterprise provides additional security, performance and management capabilities not available with open source Trino.

Starburst recently announced the expansion of its product portfolio by launching the public beta release of its Starburst Galaxy managed cloud service. While Starburst Enterprise can be deployed on-premises and in multiple clouds (Amazon Web Services, Microsoft Azure and Google Cloud), Starburst Galaxy is a hosted cloud managed service.

REPORT REPRINT

Initially available on AWS, with Azure and Google Cloud support in development, Starburst Galaxy delivers a control plane that is managed by Starburst in its own virtual private cloud, through which customers are able to manage one or more Trino compute clusters, as well as the associated data sources, in their own virtual private cloud environment or with on-premises infrastructure. Starburst Galaxy provides cluster monitoring, as well as a notebooks development environment through which data engineers can explore, query and visualize data sources, as well as create notebooks for sharing with data analysts.

While Starburst Galaxy provides some basic visualization capabilities, these are aimed at enabling data engineers to perform ad hoc queries and visually validate queries – as with Starburst Enterprise, the actual analysis is performed by data analysts using their chosen business intelligence and visualization tools, with Trino providing query acceleration and federation behind the scenes.

In addition to being consumed as a service, Starburst Galaxy has also been designed to address multicloud and hybrid IT environments, with the aim of enabling analysts to query data across an enterprise's data estate regardless of where it resides. While full support for accessing and querying data across multiple clouds and datacenters is a work in progress, it will be enabled by Starburst Remote Connector, which was a new addition in the latest version of Starburst Enterprise and provides connections between geographically distributed Trino clusters, as well as access to remote data sets for federated query and pushdown query processing.

Another point of interest on the Starburst roadmap is support for data manipulation language capabilities, to enable the modification of data (e.g., insert, update, delete and merge). Capabilities such as these are rapidly being introduced into data lake environments via projects such as Apache Iceberg, which is already supported by the open source Trino project.

Trino originated at Facebook in 2012 with the creation of the Presto distributed query engine. While Facebook open-sourced Presto via the PrestoDB project in 2013, its lead engineers left the company and launched their own variant in 2019 (known as PrestoSQL), with the aim of guiding the ongoing development of the project outside of Facebook. Those lead engineers later joined Starburst as cofounders.

The PrestoSQL project has now been rebranded to Trino and is developed by the Trino Software Foundation. Meanwhile, the PrestoDB project is now developed by the Presto Foundation, under the umbrella of the Linux Foundation. Ahana is building a business around PrestoDB and is therefore a direct competitor to Starburst. Another key competitor is Amazon Web Services with its Athena service, which is based on an older version of PrestoDB, while Qubole's managed Presto service is based on PrestoSQL (now Trino).