QueryGrid and Presto

Enabling faster, more scalable, interactive querying of Hadoop and other data sources.

Integrated Processing of QueryGrid

Teradata QueryGrid™ lets your business work with a seamless data fabric across your data and analytical engines for no-hassle analytics. Users get the most value out of data by taking advantage of specialized processing engines operating as a cohesive analytic environment. By transparently harnessing the combined power of multiple analytic engines to address a business question, QueryGrid enables users to initiate a query in one platform that reaches across to others and combines data from multiple platforms in a single query.

Presto

Originally developed by Facebook, Presto is an open source distributed SQL query engine for running interactive analytic queries against data sources of all sizes ranging from gigabytes to petabytes. Presto, unlike some other SQL on Hadoop engines, leverages standard ANSI SQL and has been architected from the ground up for

high performance interactive query processing against Hadoop and other data sources. Many SQL on Hadoop engines are limited in performance by the restriction of writing processing steps to disk, while Presto's pure memory-based architecture is built for speed, allowing Presto to support large numbers of concurrent interactive queries against huge data sets, Presto's flexible architecture unique in its ability to not only query Hadoop, but numerous other data sources such as Amazon S3, Cassandra, MySQL and PostgreSQL. Presto's flexibility also extends to its support of all major Hadoop distribution, enabling companies to leverage the distribution of their choice vs. being locked into a single distribution.

Teradata's QueryGrid Connector for Presto

Teradata has worked extensively to create a low latency, high performing connector that supports high concurrency, and parallel processing between Teradata and
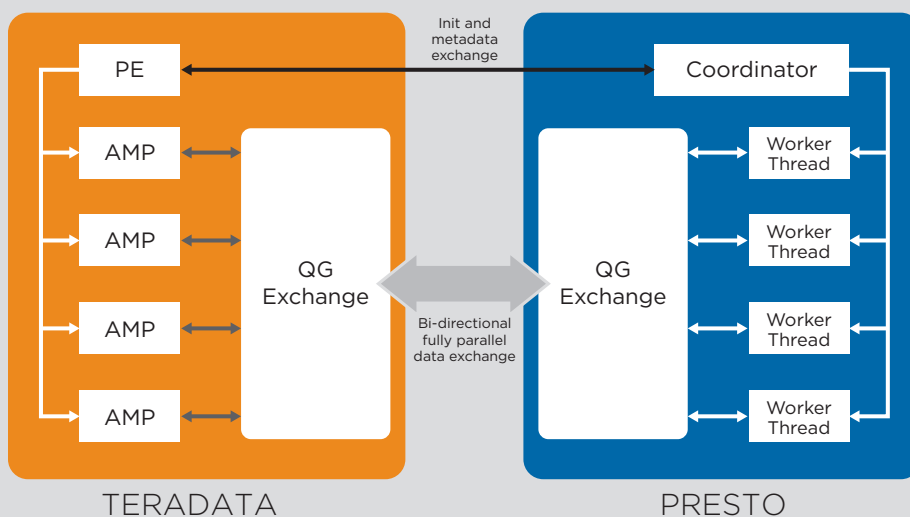

Figure 1. Presto QueryGrid Integration.

Key features

- Low latency
- High performance
- Concurrency
- SQL pushdown
- Data conversion
- Compression
- Efficient CPU usage

Presto. The Teradata QueryGrid connector for Presto enables users to execute a query within Teradata that will reach out to Presto, execute a query against one of the data platforms Presto supports, such as Hadoop, and then combine that result set with data within the Teradata database platform. The QueryGrid connector can now initiate a query from Presto to reach out to Teradata as well. The connector is architected to be as efficient as possible, leveraging SQL pushdown, auto data conversion, compression as well as optimized CPU usage.

There is no complex mapping or syntax that is needed within a QueryGrid query to reach out to Presto from Teradata or vice versa. All that is needed is an initial set up of the QueryGrid connector for Presto on both the Teradata and Presto side. Once that is in place, when initiating a query in Teradata that queries Presto you simply use the syntax **@presto** to specify the target. For example if I am querying the web logs table in Hadoop from Teradata through Presto using the QueryGrid connector and looking for the sum of time spent in a specific date range I would simply execute the following query.

> SELECT SUM(time_spent) FROM web_data.web_
> logs**@presto** WHERE visit_start BETWEEN
> '2015-12-01' AND '2015-12-31'

To query tables that exist in Teradata from Presto all that is needed is to put the syntax **teradata.** before any table name. For example if I am working in Presto and I wanted to get the average price from the sales table for the state of Massachusetts, in Teradata, I would execute the following query—

> SELECT AVG(price) FROM **teradata.sales.transactions**
> WHERE state='MA';

QueryGrid connector for Presto's syntax is simple, and easy to use enabling business user to quickly and interactively query between systems.

QueryGrid and the Unified Data Architecture

Presto is a perfect fit with Teradata's Unified Data Architecture, a vision of an analytical ecosystem. Advancing the UDA requires better SQL-on-Hadoop capabilities, whereby SQL is the interaction protocol across the UDA. Teradata views Presto as the key engine to enable interactive querying against Hadoop within the UDA. Within Teradata's QueryGrid solution Presto enables customers to more easily execute queries from Teradata to Hadoop as well as from Hadoop to Teradata With the Presto QueryGrid connector, the breath and value of the QueryGrid functionality within Teradata's Unified Data Architecture has expanded. Now with the interconnectivity of Teradata and Presto, Teradata users can not only interactively query Hadoop from their Teradata environment but also numerous other platforms that existing Presto open source connectors support.

Businesses now have the ability to expand the productivity of their business users by enabling rapid and easy access to numerous platforms data through Teradata, Presto and QueryGrid.

QueryGrid Connector for Presto requirements

- Teradata 15.00.04.04 or higher
- Teradata certified open source distribution of Presto
- Teradata enterprise support for Presto.

TERADATA