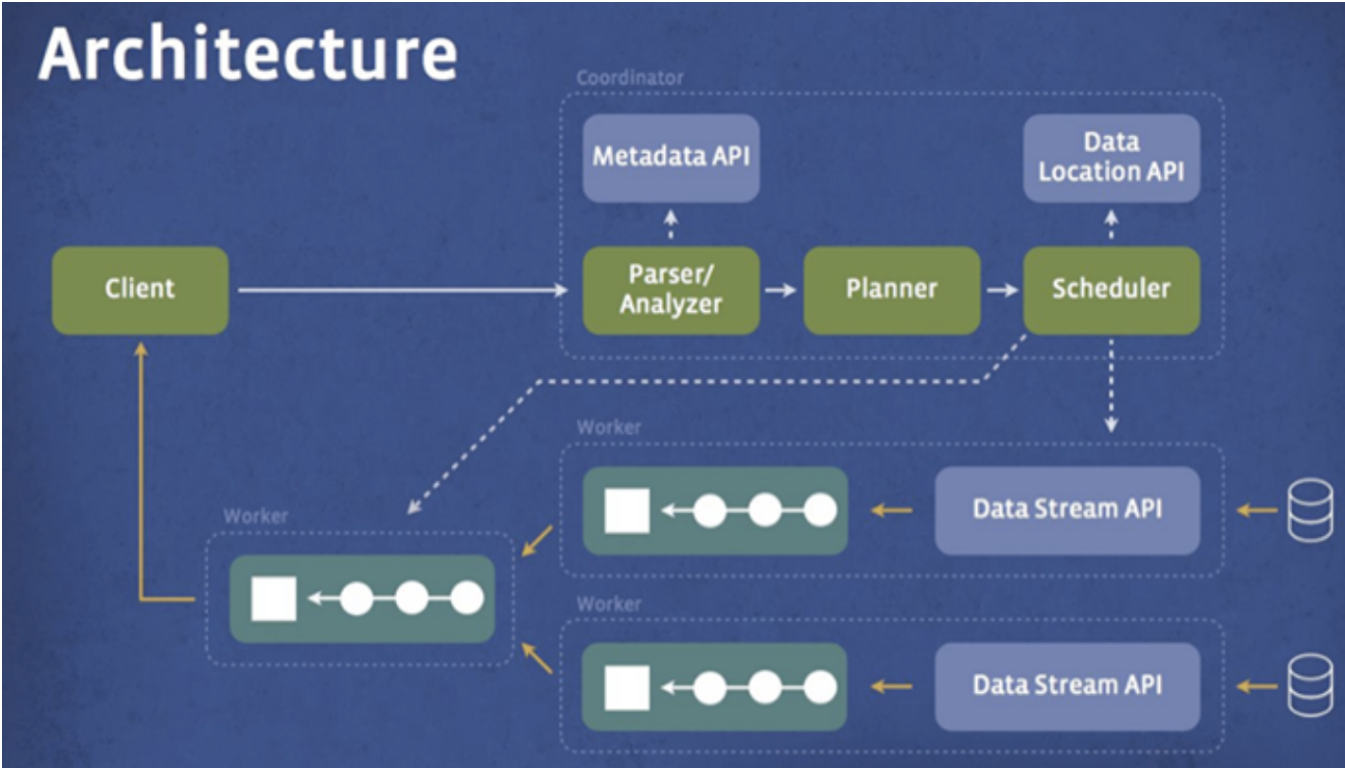


About Presto Distributed Query Engine

Presto is an open-source, distributed SQL query engine. Presto can query data where it is stored, without needing to move data into a separate analytics system. Treasure Data has customized Presto to talk directly with our distributed columnar storage layer.

How it Works

Presto uses an architecture similar to a classic massively parallel processing (MPP) database management system.



It has one coordinator node working in sync with multiple worker nodes. A user can submit their SQL query to the coordinator which uses a custom query and execution engine to parse, plan, and schedule a distributed query plan across the worker nodes. Presto is designed to support standard ANSI SQL semantics, including complex queries, aggregations, joins, left/right outer joins, sub-queries, window functions, distinct counts, and approximate percentiles.

After a query is compiled, Presto processes the request into multiple stages across the worker nodes. All processing is in-memory, and pipelined across the network between stages, to avoid any unnecessary I/O overhead.