IBM Db2 Big SQL

Make better decisions fast with powerful SQL-on-Hadoop for big data analytics

The Internet of Things (IoT), artificial intelligence (AI), social media and mobile applications are driving an increase in data volume, velocity and variety. To capitalize on this trend and obtain faster actionable insights, organizations are deploying Apache Hadoop.

IBM has partnered with Cloudera to offer an end-to-end, ODPi-compliant, enterprise grade, SQL-on-Hadoop solution. Together we are improving data exploration, discovery, testing and advanced querying, helping you to gain a 360-degree view of your customers, processes and systems. This solution works on:

The Cloudera Enterprise Data Hub (EDH) and Hortonworks Data Platform (HDP) - are enterprise open-source-based Hadoop distributions offering highly scalable platforms that:
- Optimizes data ingestion with a centralized architecture (YARN)
- Supports diverse use cases while offering enhanced governance and security
- Minimizes expense and effort required to set up a customer’s IT infrastructure by using HDP’s data and processing capabilities

IBM Db2 Big SQL - a hybrid ANSI-compliant data virtualization tool for accessing, querying and summarizing data across the enterprise which:
- Provides a massively parallel processing (MPP) architecture
- Exploits Hive, HBase and Apache Spark concurrently for best-in-class analytic capabilities
- Requires only a single database connection or query to connect disparate sources such as HDFS, RDMS, NoSQL databases, object stores and WebHDFS
- Provides low latency support for ad-hoc and complex queries, high performance, and federation capabilities
- Understands dialects from other vendors and various products from Oracle, IBM Db2 and IBM Netezza
- Enables advanced row and column security
Features of the IBM Db2 Big SQL solution include:

High performance
IBM Db2 Big SQL is built with an advanced SQL compiler and cost-based optimizer, so that complex and analytical SQL workloads maintain high performance, concurrency, and scalability. It provides an ANSI-compliant SQL parser that can run 99 TPC-DS queries and structured streaming with new APIs.

Integration with Apache Spark
Db2 Big SQL integrates with Spark for easier insight delivery and faster processing. This integration will also enable operationalizing the machine learning models with fresh data, not only from Hadoop but also from disparate sources.

Federation capabilities
Db2 Big SQL uses a single database connection, enabling you to access data across Hadoop and relational databases, whether they are on the cloud, on premises, or both. It also includes Fluid Query capabilities to enhance virtualization with various data warehouses. You can also federate with S3 Object Storage and WebHDFS (technical preview only).

Compatibility with multiple SQL dialects
Being compatible with a number of SQL dialects makes the platform well-suited for RDBMS (including Db2, Netezza and Oracle) offload as well as fast and easy consolidation. Data can be offloaded from existing enterprise data warehouses or data marts to free up capacity while preserving most of the familiar SQL.

Integration with BLU Acceleration
As a technical preview, Db2 Big SQL integrates with BLU Acceleration®, a set of advanced query processing technologies including column-organized tables, actionable compression, data skipping, CPU optimization and scan-friendly memory caching.

Integration with YARN through Slider
Improved integration with YARN is achieved through Apache Slider, enabling non-YARN services to be deployed on YARN, and control over YARN resource allocation and scheduling.

Elastic scalability
Db2 Big SQL offers the only SQL engine able to successfully run all 99 TPCDS queries up to 100TB with numerous concurrent users. It also has the ability to run multiple workers per node for efficient CPU and memory utilization.

Security-rich SQL
Robust role-based access control (RBAC), row-based dynamic filtering, column-based dynamic masking, and Apache Ranger integration are included with Db2 Big SQL to provide centralized security administration and auditing for data lakes.

Standards-compliant Open Database Connectivity and Java Database Connectivity
For developers, the usage pattern allows you to access the database with specific products or tooling that allow only Open Database Connectivity (ODBC) or Java™ Database Connectivity (JDBC).

Point querying capabilities
Db2 Big SQL enables short rapid queries to be used that search by key words or key word ranges. It uses HBase when random, real-time read/write access to your
For more information
To learn more about IBM Db2 Big SQL and its ability to help you work with SQL on Hadoop and in other locations, visit the IBM Db2 Big SQL website or start a trial at no cost.