IBM Db2 Big SQL FAQs

We currently use Apache Hive. Why would we want to use Db2 Big SQL?

Apache Hive and IBM® Db2® Big SQL can coexist and complement each other in a cluster.

Hive enables SQL developers to write Hive query language statements that are similar to standard SQL statements. It is primarily used for geospatial analytics and is known for ACID capabilities and fast ingestion.

Db2 Big SQL uses Hive database catalog (HCatalog) for table definitions, location, storage format and encoding of input files. It builds on the solution by adding federation for complex queries, application portability, high concurrency and enterprise-grade security.

Cloudera's Enterprise Data Hub comes with Hive and Impala. Why would we want to use Db2 Big SQL?

There are many open source SQL engines for Hadoop. Hive and Impala can handle SQL for single streams, but when the SQL gets complex and needs to handle multiple concurrent users with good performance, then a robust SQL engine such as Db2 Big SQL will add value to the environment. If you are offloading from Oracle, IBM Db2 or IBM Netezza®, you probably have many applications already created. Db2 Big SQL allows you to reuse those application without many rewrites. Hive, Big SQL and Impala can coexist and complement each other.

What is the advantage of having Db2 Big SQL integrate with Apache Spark?

The Db2 Big SQL bidirectional integration with Spark drives machine learning (ML), complex analytics and model building. ML is the basis for growing technologies such as artificial intelligence (AI) and the Internet of Things (IoT). Unlike other open source connectors, Big SQL integrates with Spark natively.

Does Db2 Big SQL support federation?

Db2 Big SQL supports federation to many data sources, meaning you can send distributed requests to those data sources within a single SQL statement. Integrate with Oracle, the IBM Db2 product family (including IBM Db2, the AI database), and IBM Netezza. Provide federated access to relational database management system (RDBMS) sources outside of Hadoop with IBM Fluid Query. Connect to HDFS, RDBMS, NoSQL databases, object stores and Web HDFS.

How does Big SQL integrate with other open source components?

Big SQL uses the compute power of Db2 to process ANSI SQL commands in conjunction with a cost-optimized query engine to provide optimal performance for open source file formats such as ORC, Parquet, Avro, Text, and more. It also shares metadata with a Hive metastore.

How will Db2 Big SQL integrate with my other Db2 products?

Db2 Big SQL includes the IBM Common SQL Engine (PDF, 150 KB), an integral feature of each IBM Db2 product. The Common SQL Engine is part of a comprehensive IBM strategy for flexibility and portability—one that includes application compatibility, strong data integration and flexible licensing. The Common SQL Engine includes an Oracle Application Compatibility layer, allowing Oracle applications to integrate with the IBM Db2 family of offerings.

IBM®