

IBM Big SQL: Making your big data SQL-accessible

Big data is a big business imperative, and analysts agree that SQL is the key to making it easier to unlock value from Apache Hadoop. Hadoop introduces a wealth of new technologies, including analytics engines, scripting languages and application programming interfaces (APIs). With the expansion of new technologies in the Hadoop ecosystem, it is critical to choose a Hadoop platform that aligns with open standards such as the Open Data Platform initiative (ODPi). The Hortonworks Data Platform (HDP) is ODPi-compliant, and Big SQL is engineered to run on HDP—reinforcing the value of ODPi.

With HDP as its foundation, Big SQL enables users to query Hive and HBase data using ANSI-compliant SQL. While Hadoop is highly scalable, the Big SQL advanced cost-based

optimizer and massively parallel processing (MPP) architecture can run queries smarter, not harder, supporting more concurrent users and more complex SQL with less hardware compared to other SQL solutions for Hadoop. Big SQL is also part of the IBM big data analytics offerings such as IBM Big Replicate, IBM BigIntegrate and IBM BigQuality.

Big SQL is also a powerful platform for data warehouse offloading and consolidation, a vital use case for many Hadoop users. Big SQL is the first and only SQL-on-Hadoop solution to understand commonly used SQL syntax from other vendors and products such as Oracle, IBM Db2 and IBM Netezza. And where data can't be moved to Hadoop, Big SQL provides federated access to relational database management system (RDBMS) sources outside of Hadoop with IBM Fluid Query technology.

Key IBM Big SQL features

- **Superior SQL-on-Hadoop performance** using elastic boost technology
- **The ability to understand SQL dialects from other vendors and products** for Oracle, IBM® Db2® and IBM Netezza®, which helps make it faster and easier to offload old data from existing enterprise data warehouses or data marts to free up capacity, while preserving most of the SQL from those platforms
- **A hybrid engine for Hadoop** that exploits Hive, HBase and Apache Spark concurrently for best-in-class analytic capabilities
- **Deeper integration** with Spark than other SQL-on-Hadoop technologies, enabling new use cases
- **Advanced row and column security:** The integration of Big SQL and Spark enhances “shopping for data” security because sensitive attributes can be masked by default with no backdoors to the data, which empowers self-service access to data in a safe and governed manner

Hadoop has many next-generation analytics engines to solve big data problems. Big SQL has superior SQL-on-Hadoop performance using elastic boost technology. Big SQL also has deep Spark 2.1 integration, enabling new and varied use cases. Need high-performance scans or high-performance inserts, updates or deletes? Need machine learning or graph analytics with Spark, with a single security model? Big SQL is a SQL engine for Hadoop that concurrently exploits Hive, HBase and Spark using a single database connection—even a single query. For this reason, Big SQL is also an effective hybrid engine.

The best part? All of the data belongs to Hadoop. Big SQL tables are Hive tables, HBase tables or Spark resilient distributed data sets (RDDs) and integrated with the Hive metastore. Save costs by offloading data to Hadoop and exploit next-generation, fit-for-purpose analytics with the SQL optimizer for Hadoop.

To learn more about IBM Big SQL, contact your IBM representative or IBM Business Partner, or visit: ibm.biz/ibmbigsql