



---

## Highlights

- Enterprise-grade ANSI SQL engine for data on Hadoop or object store with high concurrency and performance
  - Low latency, support for ad-hoc and complex queries, security and federation capabilities
  - Single database connection federating to disparate sources; HDFS, RDMS, NoSQL databases, object stores and WebHDFS
  - Get sub-second latency for star-schema queries by enabling MQTs on aggregated data
  - A hybrid Hadoop engine that exploits Hive, HBase and Apache Spark concurrently for best-in-class analytics
- 

# IBM Db2 Big SQL

Make better decisions fast with powerful Db2 engine 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.

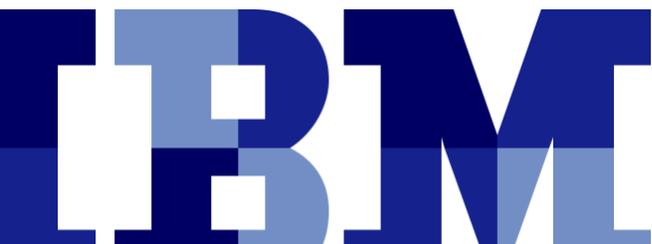
IBM has partnered with Cloudera to offer an end-to-end, ODPi-compliant, enterprise grade, SQL-on-Hadoop solution. Together we are improving data exploration, discovery, testing and advanced querying, helping you to gain a 360-degree view of your customers, processes and systems. This solution works on:

**The Cloudera Data Platform Data Center (CDP-DC), along with legacy platforms 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
- Supports diverse use cases while offering enhanced governance and security
- Minimizes expense and effort required for an IT infrastructure by using platform'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



## IBM Data and AI

### 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.

#### Enhance performance using MQTs

Db2 Big SQL offers an array of performance features to help improve performance of queries in the Hadoop ecosystem. One of these features is the ability to create and automatically match materialized query tables (MQTs) to incoming queries. MQTs can simplify query processing and improve performance because expensive join or aggregation operations can be calculated and cached in MQTs. Once these MQTs are created, Big SQL will automatically rewrite queries to take advantage of the newly created MQTs when possible to improve performance.

#### Advanced Workload Management

Workload Management (WLM) offers the capability to monitor and control statements executing on the cluster to make efficient use of cluster resources, ensuring that the cluster is not over or underutilized. At a high level, a certain number of concurrent queries can be executed so as to not over saturate the cluster resources. When this threshold is exceeded, incoming work is queued until some of the earlier queries have completed.

#### 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).

#### Accessing JSON and XML data using built-in SQL JSON functions

One of the challenges when accessing JSON data is sometimes referencing the fields in the JSON itself. The acknowledged strengths of JSON can also contribute to its complexity particularly in the area of parsing. In addition to built-in support using SerDes, IBM Db2 Big SQL now has built-in JSON SQL functions that will allow the user to store and query JSON data in the same manner as XML data can be handled using the set of XML functions provided in IBM Db2 Big SQL.

## IBM Data and AI

### 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.



---

© Copyright IBM Corporation 2020

IBM Corporation  
IBM Analytics  
Route 100  
Somers, NY 10589

Produced in the United States of America  
February 2020

IBM, the IBM logo, ibm.com, BLU Acceleration, Db2 and Netezza are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

Statement of Good Security Practices: IT system security involves protecting systems and information through prevention, detection and response to improper access from within and outside your enterprise. Improper access can result in information being altered, destroyed, misappropriated or misused or can result in damage to or misuse of your systems, including for use in attacks on others. No IT system or product should be considered completely secure and no single product, service or security measure can be completely effective in preventing improper use or access. IBM systems, products and services are designed to be part of a lawful, comprehensive security approach, which will necessarily involve additional operational procedures, and may require other systems, products or services to be most effective. IBM DOES NOT WARRANT THAT ANY SYSTEMS, PRODUCTS OR SERVICES ARE IMMUNE FROM, OR WILL MAKE YOUR ENTERPRISE IMMUNE FROM, THE MALICIOUS OR ILLEGAL CONDUCT OF ANY PARTY.



Please Recycle

---

