# IBM Db2 Big SQL with Hortonworks Data Platform improves analytics and meets enterprise data lake requirements; IBM Db2 Big SQL with Hortonworks DataFlow maximizes the value of data in motion

## Overview

IBM[R] Db2[R] Big SQL, an advanced SQL engine on Hadoop, has made strides with the fast-evolving open source ecosystem by supercharging your analytical workloads on data lakes. The core capabilities of Db2 Big SQL focus on data virtualization, SQL compatibility, scalability, performance, and, of course, enterprise security and governance. This makes it a desirable query engine to seek insights from disparate data sources, including Hadoop.

Db2 Big SQL V6.0 now supports the newly released Hadoop platform Hortonworks Data Platform V3.1. This new platform, HDP 3.1, delivers new capabilities for the enterprise to help enable:

- Agile application deployment
- New machine learning and deep learning workloads
- Real-time database
- Security and governance

For more information about Hortonworks DataFlow V3.3.1 and Hortonworks Data Platform V3.1, see Software Announcement 219-106, dated January 29, 2019.

## Key requirements

Supports RHEL and CentOS operating systems. For details, see the Hardware requirements and Software requirements sections.

## Planned availability date

February 28, 2019

## Description

### Enterprise readiness

Db2 Big SQL V6.0.0 provides enterprise readiness, enhanced performance, and ease of use capabilities. Big SQL offers:

- Integration with HDP V3.1
  - Provides tight integration with various components in HDP V3.1, especially integration with Hive 3.0 APIs and metadata for smooth interoperability between Hive and Db2 Big SQL tables
  - Integrates with YARN 3.0 (which has now absorbed the Apache Slider project for long-running processes) to manage resources centrally while sharing the node with many services
  - Integrates with other components in the platform such as HBase, Sqoop, Atlas, Ranger, and so on
- A new Zookeeper-based solution for automatic failover for HA

  Being highly available is critical for any enterprise, especially being able to automatically fail over from a primary to a secondary system. When an outage occurs, recovery needs to be seamless with no disruption or manual intervention. Db2 Big SQL introduces a new solution that is based on Zookeeper technology to be synonymous with other Hadoop components.

- The ability to create and manage tag-based policies by using Apache Atlas and Apache Ranger for Db2 Big SQL tables

  With the integration of Atlas and Ranger, users can now define and manage resource-based policies by using tags. Once a resource is tagged, the authorization for the tag is automatically enforced, eliminating the need to create or update the policies for the resource. This expands the capabilities of centralized security to not only access-based policies, but resource-based based policies as well.

- The ability to create and manage access policies in Apache Ranger for federated sources

  Apache Ranger policies for Db2 Big SQL were once only able to provide access control for Hadoop objects. But now, you can also set Ranger policies for federated sources, which makes Apache Ranger a centralized access policy manager for Hadoop.

### Performance

Every enterprise looks for high performance for its workloads and applications. Db2 Big SQL brings the advanced SQL engine to Hadoop. It enables best query execution, even when the query is very complex or tool generated, without hitting any Out-Of-Memory errors. Highlights include:

- Better reader performance on sorted tables

  When scanning ORC and Parquet tables, Db2 Big SQL readers can apply query predicates to skip reading portions of files. When tables are sorted on columns that are predicates in queries, table scan times can be significantly reduced. Testing shows a 4-6x speedup in top 10 TPC-DS queries.

- Join Range Filter Predicate (JRFP) turned ON by default

  JRFP has been available since version 4.2 but had limitations. In this release, this capability is enhanced to generate JRFP pushdowns for as many joins as possible. This capability is turned on by default and the SQL optimizer will push down when appropriate. Significant performance improvement was observed, especially for Star Schema queries as it skips reading large portions of fact table rows, thereby reducing the overall query processing times. Performance testing shows a 3-6x speedup for top 10 most improved TCP-DS queries.

- Enhanced cache management in Scheduler

  Scheduler takes care of coordinating the jobs among the various nodes that have Db2 Big SQL installed. With enhanced cache management introduced, the scheduler provides improved performance in coordinating the jobs.

- Enhanced query pushdown in LOAD

  Query pushdown is one of the core capabilities of Db2 Big SQL where the execution is broken down to smaller units and distributed efficiently for faster and distributed processing. In this release, the ability to allow "sort by" in LOAD enables pushdown to improve performance.

### Ease of use

Ease of use provides efficient, effective, engaging, error-tolerant, and easy-to-learn capabilities in the product. One capability focused on in this release is simplification of YARN configuration for Db2 Big SQL. Highlights include:

- YARN is widely used to centrally manage resources for various services that run in a Hadoop cluster. But it is always challenging to have the right configuration. Enhancements made to have better control over YARN include:
  - Notify users through Ambari UI when disk failures are detected for all nodes, head and worker nodes
  - Prevent YARN from starting that node which will block any queries to be sent to that node, when there is a damaged DB or disk
- HBase is no longer mandatory when installing Db2 Big SQL.
- Db2 Big SQL can be installed on heterogeneous clusters, that is, a cluster with mixed hardware and OS versions.
- Db2 Big SQL can now support huge tables. The maximum number of columns is increased to 2048, which enables offloading from Teradata and reusing the same schema definitions.
- An all-new look and feel of product documentation can help you find the details that you are looking for or find solutions to the problems you want to solve. Developer-written Best Practices are also now part of the documentation for some quick tips and tricks.
- Enhanced compatibility with Hive 3.0 supports Decimal (31,0) schema evolution and synchronization between Hive metadata and Db2 Big SQL.
- A new GUI tool, Unified Console (previously DSM), allows historical monitoring of queries and workloads, along with real-time monitoring of how the cluster is performing and where the bottlenecks are.

### Accessibility by people with disabilities

A US Section 508 Accessibility Compliance Report containing details on accessibility compliance can be found on the Product accessibility information website.

## Reference information

For additional information about IBM Db2 Big SQL, see Software Announcement 217-503, dated October 24, 2017.

## Program number

| Program number | VRM | Program name |
| --- | --- | --- |
| 5737-E74 | 6.0.0 | IBM Db2 Big SQL |
| 5737-F94 | 6.0.0 | IBM Db2 Big SQL with Hortonworks DataFlow |

5737-E74 includes IBM Db2 Big SQL with Hortonworks Data Platform.

## Offering Information

Product information is available on the  IBM Offering Information website.

More information is also available on the  Passport Advantage(R) and Passport Advantage Express(R)  website.

## Publications

None.

## Services

### Software Services

IBM Software Services has the breadth, depth, and reach to manage your services needs. You can leverage the deep technical skills of our lab-based, software services team and the business consulting, project management, and infrastructure expertise of our IBM Global Services team. Also, we extend our IBM Software Services reach through IBM Business Partners to provide an extensive portfolio of capabilities. Together, we provide the global reach, intellectual capital, industry insight, and technology leadership to support a wide range of critical business needs.

To learn more about IBM Software Services, contact your Lab Services Sales or Delivery Leader.

## Technical information

### Specified operating environment

#### *Hardware requirements*
- Minimum of 3 SKUs with each SKU of 4 nodes each (for example, 12 nodes, in cluster installation)
- Processor: x86 or IBM POWER(R) 64-bit systems
- Memory: Minimum 24 GB of memory
- Disk space: Minimum 80 GB of disk storage
    - Minimum 32 GB of disk storage on the management node
    - Minimum 20 GB of disk storage on other nodes

#### *Software requirements*
Operating systems:

Intel™ x86-64:

- RHEL 7.2, 7.3, 7.4, 7.5, and 7.6
- CentOS 7.2, 7.3, 7.4, 7.5, and 7.6

Processors:

- POWER8: RHEL 7.2, 7.3, and 7.4
- POWER9 (ALT): RHEL 7.6

For more information, see System requirements for IBM Big SQL.

## Planning information

### *Packaging*

This offering is delivered through the internet as an electronic download. There is no physical media.

This program, when downloaded from a website, contains the applicable IBM license agreement and License Information, if appropriate, which will be presented for acceptance at the time of installation of the program. For future reference, the license and License Information will be stored in a directory such as LICENSE.TXT.

## Ordering information

For ordering information, consult your IBM representative or authorized IBM Business Partner, or go to the  Passport Advantage  website.

This product is available only through Passport Advantage. It is not available as shrinkwrap.

This product may only be sold directly by IBM or by authorized IBM Business Partners for Channel Value Rewards.

More information can be found on the  IBM Channel Value Rewards website.

To locate IBM Business Partners for Channel Value Rewards in your geography for a specific Channel Value Rewards portfolio, go to the Find a Business Partner page.

### Passport Advantage

No new parts.

### Charge metric

The charge metrics for this licensed product can be found in the following License Information document:

| Program identifier | License Information document title | License Information document number |
|---|---|---|
| 5737-E74 | IBM Db2 Big SQL | L-MLOY-AZHSS7 |

Select your language of choice and scroll down to the "Charge Metrics" section. Follow-on releases, if any, may have updated terms. See the License Information documents website for more information.

## Terms and conditions

The information provided in this announcement letter is for reference and convenience purposes only. The terms and conditions that govern any transaction with IBM are contained in the applicable contract documents such as the IBM International Program License Agreement, IBM International Passport Advantage Agreement, and IBM Agreement for Acquisition of Software Maintenance.

This product is available only through Passport Advantage.

### Licensing

IBM International Program License Agreement including the License Information document and Proof of Entitlement (PoE) govern your use of the program. PoEs are required for all authorized use. Part number products only, offered outside of Passport Advantage, where applicable, are license only and do not include Software Maintenance.

This software license includes Software Subscription and Support (also referred to as Software Maintenance).

## Software Maintenance

Licenses under the IBM International Program License Agreement (IPLA) provide for support with ongoing access to releases and versions of the program. IBM includes one year of Software Subscription and Support (also referred to as Software Maintenance) with the initial license acquisition of each program acquired. The initial period of Software Subscription and Support can be extended by the purchase of a renewal option, if available. Two charges apply: a one-time license charge for use of the program and an annual renewable charge for the enhanced support that includes telephone assistance (voice support for defects during normal business hours), as well as access to updates, releases, and versions of the program as long as support is in effect.

## License Information number

The following License Information document applies to the offering in this announcement:

| Program identifier | License Information document title | License Information document number |
|---|---|---|
| 5737-E74 | IBM Db2 Big SQL | L-MLOY-AZHSS7 |

Follow-on releases, if any, may have updated terms. See the License Information documents website for more information.

## Limited warranty applies

Yes.

## Limited warranty

IBM warrants that when the program is used in the specified operating environment, it will conform to its specifications. The warranty applies only to the unmodified portion of the program. IBM does not warrant uninterrupted or error-free operation of the program or that IBM will correct all program defects. You are responsible for the results obtained from the use of the program.

IBM provides you with access to IBM databases containing information on known program defects, defect corrections, restrictions, and bypasses at no additional charge. For further information, see the IBM Software Support Handbook.

IBM will maintain this information for at least one year after the original licensee acquires the program (warranty period).

## Program technical support

Technical support of a program product version or release will be available for a minimum of two years from the general availability date, as long as your Software Subscription and Support (also referred to as Software Maintenance) is in effect.

For more information on Elite support, see IBM Elite Support Definition.

For software support lifecycle details, see the Software Support Lifecycle.

This technical support allows you to obtain assistance (by telephone or electronic means) from IBM for product-specific, task-oriented questions regarding the installation and operation of the program product. Software Subscription and Support (Software Maintenance) also provides you with access to versions, releases, and updates (CD releases, Long Term Support Releases or fixes) of the program.

You will be notified, through an announcement letter, of discontinuance of support with six months' notice. If you require additional technical support from IBM, including an extension of support beyond the discontinuance date, contact your IBM representative or IBM Business Partner. This extension may be available for a fee.

For additional information on the IBM Software Support Lifecycle Policy, see the IBM Software Support Lifecycle Policy website.

## Money-back guarantee

If for any reason you are dissatisfied with the program and you are the original licensee, you may obtain a refund of the amount you paid for it, if within 30 days of your invoice date you return the program and its PoE to the party from whom you obtained it. If you downloaded the program, you may contact the party from whom you acquired it for instructions on how to obtain the refund.

For clarification, note that (1) for programs acquired under the IBM International Passport Advantage offering, this term applies only to your first acquisition of the program and (2) for programs acquired under any of IBM's On/Off Capacity on Demand (On/Off CoD) software offerings, this term does not apply since these offerings apply to programs already acquired and in use by you.

## Volume orders (IVO)

No.

## Passport Advantage applies

Yes. Information is available on the Passport Advantage and Passport Advantage Express website.

## Software Subscription and Support applies

Yes. Software Subscription and Support, also referred to as Software Maintenance, is included with licenses purchased through Passport Advantage and Passport Advantage Express. Product upgrades and Technical Support are provided by the Software Subscription and Support offering as described in the Agreements. Product upgrades provide the latest versions and releases to entitled software, and Technical Support provides voice and electronic access to IBM support organizations, worldwide.

IBM includes one year of Software Subscription and Support with each program license acquired. The initial period of Software Subscription and Support can be extended by the purchase of a renewal option, if available.

While your Software Subscription and Support is in effect, IBM provides you assistance for your routine, short-duration installation and usage (how-to) questions, and code-related questions. IBM provides assistance by telephone and, if available, electronic access, only to your information systems (IS) technical support personnel during the normal business hours (published prime shift hours) of your IBM support center. (This assistance is not available to your users.) IBM provides Severity 1 assistance 24 hours a day, 7 days a week. For additional details, see the IBM Software Support Handbook. Software Subscription and Support does not include assistance for the design and development of applications, your use of programs in other than their specified operating environment, or failures caused by products for which IBM is not responsible under the applicable agreements.

Unless specified otherwise in a written agreement with you, IBM does not provide support for third-party products that were not provided by IBM. Ensure that when contacting IBM for covered support, you follow problem determination and other instructions that IBM provides, including in the IBM Software Support Handbook.

For additional information about the International Passport Advantage Agreement and the IBM International Passport Advantage Express Agreement, go to the Passport Advantage and Passport Advantage Express website.

For more information on Elite support, see IBM Elite Support Definition.

For software support lifecycle details, see the Software Support Lifecycle.

### Variable charges apply

No.

### Educational allowance available

Not applicable.

## Statement of good security practices

IT system security involves protecting systems and information through intrusion prevention, detection, and response to improper access from within and outside your enterprise. Improper access can result in information being altered, destroyed, or misappropriated or can result in misuse of your systems to attack others. Without a comprehensive approach to security, no IT system or product should be considered completely secure and no single product or security measure can be completely effective in preventing improper access. IBM systems and products are designed to be part of a regulatory compliant, comprehensive security approach, which will necessarily involve additional operational procedures, and may require other systems, products, or services to be most effective.

**Important:** 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.

## Prices

### Business Partner information

If you are an IBM Business Partner acquiring products from IBM, you may link to Passport Advantage Online for resellers where you can obtain Business Partner pricing information. An IBMid and password are required to access the IBM Passport Advantage website.

### Passport Advantage

For Passport Advantage information and charges, contact your IBM representative or authorized IBM Business Partner for Channel Value Rewards. Additional information is also available on the Passport Advantage and Passport Advantage Express website.

## Order now

To order, contact the IBM Digital Sales Center, your local IBM representative, or your IBM Business Partner. To identify your local IBM representative or IBM Business Partner, call 800-IBM-4YOU (426-4968). For more information, contact the IBM Digital Sales Center.

Phone: 800-IBM-CALL (426-2255)

Fax: 800-2IBM-FAX (242-6329)

For IBM representative: askibm@ca.ibm.com

For IBM Business Partner: pwcs@us.ibm.com

IBM Digital Sales Offices
1177 S Belt Line Rd
Coppell, TX 75019-4642, US

The IBM Digital Sales Center, our national direct marketing organization, can add your name to the mailing list for catalogs of IBM products.

**Note:** Shipments will begin after the planned availability date.

**IBM Channel Value Rewards**

This product is available under Channel Value Rewards (CVR), either directly from IBM or through authorized Business Partners who invest in skills and high-value solutions. IBM clients may benefit from the industry-specific or horizontal solutions, skills, and expertise provided by these Business Partners.

Additions to CVR will be communicated through standard product announcements. To determine what IBM software is available under CVR, see the  IBM Passport Advantage Online for IBM Business Partners website.

For questions regarding CVR, see the  IBM Channel Value Rewards website.

*Trademarks*

IBM, Db2, PartnerWorld, Passport Advantage, Express and POWER are registered trademarks of IBM Corporation in the United States, other countries, or both.
Intel is a trademark of Intel Corporation or its subsidiaries in the United States and other countries.
Other company, product, and service names may be trademarks or service marks of others.

*Terms of use*

IBM products and services which are announced and available in your country can be ordered under the applicable standard agreements, terms, conditions, and prices in effect at the time. IBM reserves the right to modify or withdraw this announcement at any time without notice. This announcement is provided for your information only.Additional terms of use are located at

   Terms of use

For the most current information regarding IBM products, consult your IBM representative or reseller, or go to the IBM worldwide contacts page

   IBM United States