

Oracle Big Data SQL Cloud Service

High Performance Data Virtualization with Cloud Simplicity and Security

ORACLE®

**BIG DATA
CLOUD SERVICE**

KEY BENEFITS

Oracle Big Data SQL Cloud Service enables the power of all your Oracle SQL queries across a spectrum of data stores. Big Data SQL Cloud Service provides a consistently high-performance environment for big data analytics and data management through:

- Dedicated instances for both the Hadoop cluster and the Exadata Database
- Dedicated, High-Speed Networking between the environments
- Optimized Configurations
- Simple security, including encryption

KEY FEATURES

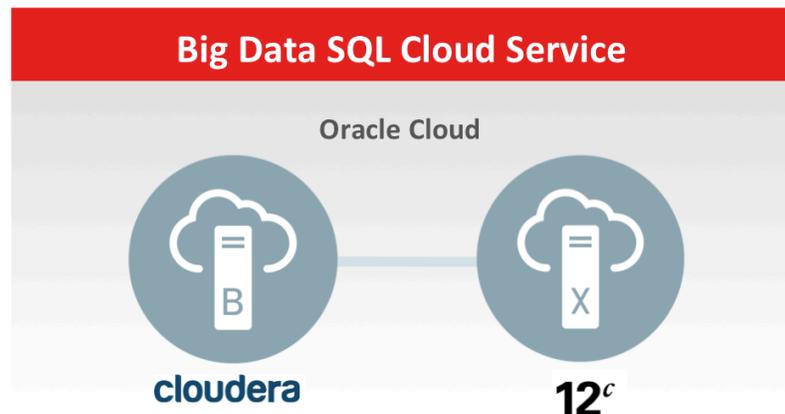
Oracle Big Data SQL Cloud Service provides a comprehensive environment for accelerating big data analytics. The service offers:

- Unprecedented networking speeds in a cloud environment for Analytics and Data Virtualization
- Oracle Database Security across both Oracle Database data and data stored in Hadoop

Oracle Big Data SQL Cloud Service enables organizations to immediately analyze data across Apache Hadoop, NoSQL and Oracle Database leveraging their existing SQL skills, security policies and applications with extreme performance. From simplifying data science efforts to unlocking data lakes, Big Data SQL makes the benefits of Big Data available to the largest group of end users possible, through an easy to configure cloud service.

Big Data SQL Cloud Service – Overview

Oracle Big Data Cloud Service delivers a data virtualization layer across Oracle Database on Exadata Cloud Service and Hadoop on Big Data Cloud Service delivering all of Oracle SQL on a wide range of data sources and formats while applying Oracle Security features to all your data.



Oracle Big Data SQL is a data virtualization innovation from Oracle. It is a new architecture and solution for SQL and other data APIs (such as REST and Node.js) on disparate data sets, seamlessly integrating data in Apache Hadoop and a number of NoSQL databases with data stored in Oracle Database. Using Oracle Big Data SQL, organizations can:

- Combine data from Oracle Database, Apache Hadoop and NoSQL in a single SQL query
- Query and analyze data in Apache Hadoop and NoSQL
- Maximize query performance on all data using advanced techniques like

- Cost effective economics model
- Unique architecture enabling fast SQL across tiers through technical innovations only available with Big Data SQL

RELATED PRODUCTS

- Oracle Database Cloud Service – Exadata Edition
- Oracle Big Data Cloud Service
- Oracle IoT Cloud Service
- Oracle Analytics Cloud Service
- Oracle Storage Cloud Service

Smart Scan, Partition Pruning, Storage Indexes, Bloom Filters and Predicate Push-Down in a distributed architecture

- Integrate big data analyses into existing applications and architectures
- Extend security and access policies from Oracle Database to data in Apache Hadoop and NoSQL
- Leverage Oracle’s high bandwidth network between Exadata Cloud Service and Big Data Cloud Service to deliver unprecedented analytics performance

Subscription Details

Big Data SQL Cloud Service is available as a Non-metered Cloud Service subscription with a minimum term of 12 months. The service is purchased directly with a service subscription and requires two underlying Oracle Public Cloud services: Big Data Cloud Service and Database Exadata Cloud Service.

The non-metered offering is ideally suited for deploying production clusters in the Cloud on a long-term Big Data Management System deployment. It is also suitable for ongoing long-term sandbox and data science projects leveraging the power of dedicated systems to ensure performance and production level application stability.

Enhanced Metadata

When dealing with large data sets stored in disparate systems, it can be difficult to know where your data is, let alone understand how the data is structured. Big Data SQL adds new external table types to Oracle Database 12c, which give users a single location to catalog and secure data in Hadoop and NoSQL systems: the Oracle Database. Big Data SQL keeps track of the metadata about external data sources – both clusters and the tables within them – without moving or copying data. External tables for Big Data SQL provide:

- Seamless metadata integration and queries which join data from Oracle Database with data from Hadoop and NoSQL databases
- Automatic mappings from metadata stored in HCatalog (or the Hive Metastore) to Oracle Tables
- Multiple cluster support to allow one Oracle Database to query multiple Hadoop clusters
- Enhanced access parameters to give database administrators the flexibility to control column mapping and data access behavior

Smart Scan – Data Driven Parallel Processing

Finding insights from Big Data can mean sifting through an extraordinary amount of data. With the massive increase in data volumes that Big Data brings, analytical performance can only be achieved by moving the analytics to the data, not the other way around. Big Data SQL applies the power of Smart Scan, first introduced in Oracle’s best-in-class Exadata Database Machine, to Hadoop and NoSQL systems. Smart Scan enables Oracle SQL operations to be pushed down to the storage tiers of the Big Data system. Paired with the horizontal scalability of these storage systems, Smart Scan automatically provides parallel processing equal to your biggest data set, enabling:

- Locally filtered data, so that only the rows and columns relevant to your query are transmitted to Oracle Database

- Join optimization via Bloom filters, speeding up joins between data in Oracle Database and massive data in Hadoop
- Scoring for data mining models and enhanced processing for querying document data sets in for example JSON or XML
- Oracle-native operators providing complete fidelity between queries run with Big Data SQL and Oracle Database alone

Storage Indexes – More effective I/O

In addition to the set of Smart Scan features, Oracle Big Data SQL provides Storage Index technology to speed up processing before any I/O occurs. As data is accessed, Oracle Big Data SQL automatically builds local, in-memory indexes that capture where relevant data is stored. On subsequent queries of the same data, Storage Index technology ensures that data blocks that are not relevant to the query are not read. Because data blocks in Big Data systems can be very large (up to hundreds of megabytes), this “I/O skipping” strategy can improve performance on some queries by orders of magnitude.

Predicate Pushdown – Harness External Storage Formats

Oracle Big Data SQL not only enables easy integration of data from Hadoop and NoSQL sources, Big Data SQL also leverages the underlying storage mechanisms to provide the best possible performance. Big Data SQL’s *predicate push down* technology allows predicates in queries issued in Oracle Database to be executed by remote systems, and to be pushed into certain file formats. Using predicate push down, Big Data SQL enables you to:

- Prune partitions from tables managed by Apache Hive
- Minimize I/O on files stored in Apache Parquet and Apache ORC formats
- Enable remote reads on data stored in Oracle NoSQL Database or Apache HBase

Oracle Database Security on Hadoop Data

Oracle Big Data SQL’s unique approach to integrating data enables Oracle Database Security features on Hadoop and NoSQL data. Using standard Oracle security mechanisms, you can secure Big Data using:

- Standard Oracle Database roles and privileges to govern access to data
- Data redaction, to ensure that sensitive information is obscured when accessed by unauthorized users
- Virtual Private Databases to better enforce governance policies

Extended Information Life-Cycle Management

For many years, Oracle Database has provided rich support for Information Lifecycle Management (ILM). Numerous capabilities are available for data tiering – or storing data in different media based on access requirements and storage cost considerations.

Copy to Hadoop

Copying data from Oracle Database to Hadoop can be complicated. Oracle Big Data SQL includes the Oracle Copy to Hadoop utility. This utility simplifies copying Oracle data to the Hadoop Distributed File System (HDFS). Data copied to the Hadoop cluster

by Copy to Hadoop is stored in Oracle Data Pump format. This format optimizes queries thru Big Data SQL: 1) the data is stored as Oracle data types – eliminating data type conversions and 2) the data is queried directly – without requiring the overhead associated with Java SerDes. Native Hadoop tools like Hive can easily access these same Oracle Data Pump export files using optimized input format classes.

Smart Scan on Tablespaces in HDFS

Oracle Partitioning is the enabling technology that allows a single table's data partitions to be stored on the various tiers. This enables immutable archive data within a table to reside in Hadoop in highly optimized Oracle Database storage formats. Database queries seamlessly access this archive data as they would any other data – exploiting the optimized access and storage structures (like indexes) for fast query performance.

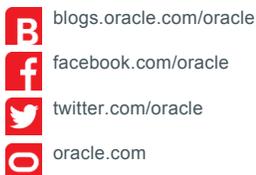
In addition, Big Data SQL's Smart Scan capabilities enable compound performance benefits. Big Data SQL Smart Scan utilizes the massively parallel processing power of the Hadoop cluster to filter data at its source – greatly reducing data movement and network traffic between the cluster and the database.



CONTACT US

For more information about Big Data Cloud Service, visit cloud.oracle.com/bigdata or call +1.800.ORACLE1 to speak to an Oracle representative.

CONNECT WITH US



Integrated Cloud Applications & Platform Services

Copyright © 2017, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. UNIX is a registered trademark of The Open Group.