Oracle Big Data Cloud Service

ORACLE ®

**BIG DATA
CLOUD SERVICE**

Oracle Big Data Cloud Service is an automated service that provides a high-powered environment tailor-made for advancing businesses' analytical capabilities. With automated lifecycle management and one-click security, Big Data Cloud Service is designed to optimally and securely run a wide variety of big data workloads and technologies while simplifying operations. Big Data Cloud Service with Oracle Big Data SQL Cloud Service enables organizations to analyze data across Hadoop, NoSQL and Oracle Database Service – Exadata Edition while leveraging their existing SQL skills.

Enterprise-Grade Big Data as a Service

Oracle Big Data Cloud Service is an automated cloud service for Big Data processing. It is optimized to run a diverse set of workloads – from Hadoop-only workloads (MapReduce 2, Spark, Hive etc.) to interactive, all-encompassing interactive SQL queries using Oracle Big Data SQL Cloud Service.

Big Data Cloud Service embraces the innovations in the big data domain by providing an open environment for innovation, while automated lifecycle management and simple security ensure organizations do not compromise enterprise-level stability and safety. Organizations are free to deploy external software to support new functionality – such as graph analytics, natural language processing and fraud detection –to meet the needs of the application. Support for non-Oracle components is delivered by their respective support channels and not by Oracle.

With the Oracle Cloud, organizations get the best possible cloud environment for Big Data workloads. As part of this environment, Big Data Cloud Service provides a massively-scalable Big Data environment featuring:

- Cloudera's comprehensive software suite including Cloudera Distribution including Apache Hadoop and Apache Spark
- Big Data Connectors delivers load rates of up to 15TB per hour between Big Data Cloud Service and Oracle Exadata Cloud Service
- Big Data Spatial and Graph provides cutting-edge tools for exploring and analyzing massive graphs and geo-locational data
- Dedicated instances in the Oracle Cloud, featuring 40Gb/sec InfiniBand fabric inside Big Data Cloud Service and Database Cloud Service -- Exadata Edition.
- Simplified operations, and automated lifecycle management though a single command utility of the entire stack

KEY BENEFITS

Oracle Big Data Cloud Service speeds time to value by providing a cloud-based automated service for big data processing on Apache Hadoop and Apache Spark. Big Data Cloud Service provides a consistently high-performance environment for big data management via

- Dedicated instances
- Simple security
- Comprehensive Software

KEY FEATURES

Oracle Big Data Cloud Service provides a comprehensive environment for accelerating big data analytics. Automatically included in the software are

- Cloudera's Distribution 5.x including Apache Hadoop and Apache Spark
- Oracle Big Data Spatial and Graph for cutting-edge analytics on places and networks
- Oracle Data Integrator and ODI Big Data Module to simplify big data

ORACLE ®

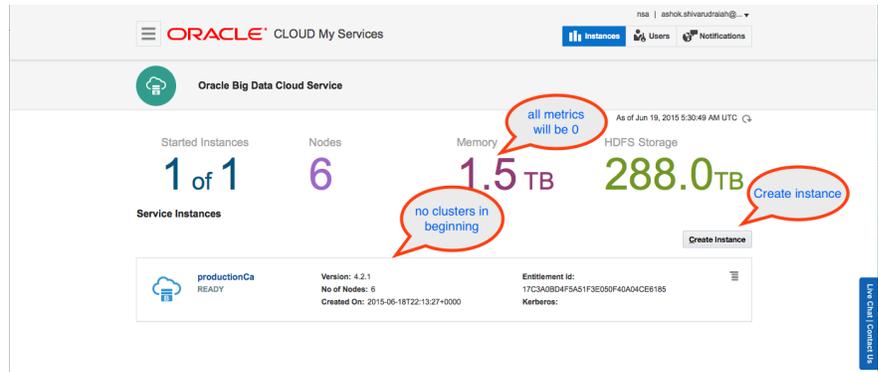operations through visual code-building

Figure 1. Managing Big Data is simple in the Oracle Cloud.

Big Data SQL Cloud Service

Oracle Big Data SQL, an innovation from Oracle only available on the Oracle Cloud is a new architecture for SQL on Hadoop seamlessly integrating data in Hadoop and NoSQL with data in Oracle Database Cloud Service. With Oracle Big Data SQL Cloud Service, organizations can:

- Leverage existing SQL skills and tools to
  o Combine data from Oracle Database Cloud Service, Hadoop and NoSQL in a single SQL query
  o Query and analyze data in Hadoop and NoSQL
- Quickly integrate big data analysis into existing applications and architectures
- Easily extend security and access policies from Oracle database to data in Hadoop and NoSQL
- Get maximum query performance on all data using Smart Scan

Oracle Big Data SQL Cloud Service radically simplifies integrating and operating in the big data domain through two powerful features: newly expanded External Tables and Smart Scan functionality on Hadoop. Oracle's unique Smart Scan capability for Hadoop brings the proven storage processing innovations of Oracle Exadata to the Oracle Big Data Cloud.

Even in the cloud, the biggest performance penalties in data processing are a typically the result of excess data movement. Instead of sending all scanned data to the compute resources, Smart Scan on Hadoop radically minimizes data movement to the compute nodes by applying the following techniques at the storage level:

- Data-local scans:
  o Hadoop data is read using native operators at the source
- Column projection
  o Only relevant columns are returned from the source
- Predicate evaluation
  o Only relevant rows are returned from the source
- Complex function evaluation
  o SQL operators on JSON and XML types applied at the source
  o Model scoring and analytical operators evaluated at the source

Simple Security

Securing data is critical to Big Data solutions in the enterprise; Big Data Cloud Service provides strong authentication, authorization and auditing of data in Hadoop with just a single click.
Strong authentication is provided using Kerberos. This ensures that all users are who they claim to be – and that rogue services are not added to the system.

Big Data Cloud Service leverages Apache Sentry (an open-source project of which

Oracle is a founding member) to authorize SQL access via tools like Hive and Impala. By delivering and developing Sentry, Oracle delivers Big Data Cloud Security with the highest data security levels currently available for Hadoop.

Both encryption of data-at-rest and network encryption are capabilities included with Oracle Big Data Cloud Service and supported by Oracle. Network encryption prevents network sniffing from capturing protected data.

In addition to securing the Hadoop system, Oracle Big Data SQL Cloud Service enables organizations to leverage Oracle's security capabilities when querying data by utilizing Oracle Database Cloud Service – Exadata Edition. Big Data Cloud Service with Hadoop security in place, in combination with Oracle Big Data SQL Cloud Service delivers the most comprehensive security of any big data system.

Figure 2. Cluster tagging and security administration are handled via a simple web interface.

Big Data Cloud Service Included Software
Software Automatically Installed on Provisioned Instances
Oracle Linux 6 with Unbreakable Enterprise Kernel
Oracle Java – JDK 8
Cloudera Enterprise (Data Hub Edition) 5.x • Cloudera's Distribution including Apache Hadoop (CDH) with support for YARN and MR2 • Cloudera Impala • HBase (as well as support for Accumulo) • Cloudera Search • Apache Spark
Cloudera Manager including: • Cloudera Back-up and Disaster Recovery (BDR) • Cloudera Navigator
Oracle R Distribution
Oracle Big Data Connectors • Oracle SQL Connector for Hadoop • Oracle Loader for Hadoop • Oracle XQuery for Hadoop

"[

• Oracle R Advanced Analytics for Hadoop • Oracle Data Integrator Application Adapter for Hadoop
Oracle Big Data Spatial and Graph
Oracle Data Integrator and Big Data Extensions
Optional Services (separately subscribed)
Oracle Big Data SQL Cloud Service*

* Requires a subscription to Oracle Database Cloud Service – Exadata Edition

CONNECT WITH US

blogs.oracle.com/bigdata

facebook.com/oraclebigdata

twitter.com/oraclebigdata

oracle.com/bigdata

Hardware and Software, Engineered to Work Together

Oracle is committed to developing practices and products that help protect the environment