

# Oracle Big Data Cloud Service

Delivering Hadoop, Spark and Data Science with Oracle Security and Cloud Simplicity



**KEY BENEFITS**

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

- Dedicated instances
- Dedicated Networking
- Optimized Configurations
- Simple security, including encryption
- Comprehensive Software Stack

**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
- Open platform offering full control of the environment, as well as enabling 3<sup>rd</sup> party software installs into the

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 includes several high-end analytics components, ranging from Oracle Big Data Spatial and Graph to R and embedded components like Apache Spark and Impala.

**Big Data Cloud Service – Overview**

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 (ETL, Spark, Hive etc.) to interactive, all-encompassing interactive SQL queries using SQL-on-Hadoop tools and/or Oracle Big Data SQL Cloud Service.

With Oracle Public 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:

- Dedicated instances in Oracle Public Cloud, featuring 40Gb/sec InfiniBand fabric inside Big Data Cloud Service and dedicated, direct attached storage for HDFS files
- Cloudera’s comprehensive software suite including Cloudera Distribution including Apache Hadoop and Apache Spark as well as Apache Kafka
- Oracle Big Data Spatial and Graph provides cutting-edge tools for exploring and analyzing massive graphs and geo-locational data
- Fast provisioning, automatically creating Highly Available and fully secured clusters in a matter of hours
- Simplified operations, and automated lifecycle management through a single command utility of the entire stack

Big Data Cloud Service Included Software
Software Automatically Installed on Provisioned Instances
Oracle Linux 6
Oracle Java – JDK 8
Cloudera Enterprise (Data Hub Edition) 5.x <ul style="list-style-type: none"> <li>• Cloudera’s Distribution including Apache Hadoop (CDH)</li> <li>• Cloudera Impala</li> <li>• HBase (as well as support for Accumulo)</li> <li>• Cloudera Search</li> </ul>

cluster

- Flexible subscription models, offering both metered and non-metered consumption models
- Oracle Big Data Spatial and Graph for cutting-edge analytics on places and networks
- Oracle Big Data Connectors to simplify integration of data and enable Spark based R analytics
- Oracle Data Integrator to simplify big data operations through visual code-building

#### RELATED PRODUCTS

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

<ul style="list-style-type: none"><li>• Apache Spark</li></ul>
Cloudera Manager including: <ul style="list-style-type: none"><li>• Cloudera Back-up and Disaster Recovery (BDR)</li><li>• Cloudera Navigator</li></ul>
Oracle R Distribution
Oracle Big Data Connectors <ul style="list-style-type: none"><li>• Oracle SQL Connector for Hadoop</li><li>• Oracle Loader for Hadoop</li><li>• Oracle XQuery for Hadoop</li><li>• Oracle R Advanced Analytics for Hadoop</li><li>• Oracle Data Integrator Enterprise Edition</li></ul>
Oracle Big Data Spatial and Graph
<b>Additional Services (separate subscription required) and Software</b>
Oracle Big Data SQL Cloud Service*
Oracle Big Data Discovery Cloud Service

\* Requires a subscription to Oracle Database Cloud Service – Exadata Edition as well as a separate subscription to Oracle Big Data SQL Cloud Service

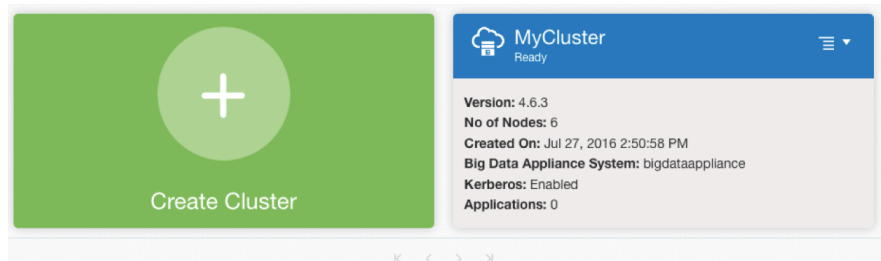
## Subscription Details

Big Data Cloud Service is available through two subscription offerings: Metered and Non-metered:

- A Metered Big Data Cloud Service subscription enables short-term pay-as-you-go service usage, with a minimum duration of one month. Pre-paid Oracle Big Data Public Cloud Services funds can be consumed. Ideal applications and use cases include short-term test and development projects, pilot projects, trials and other validation projects.
- A Non-metered Big Data Cloud Service subscription requires a minimum term of 12 months. The service is purchased directly with a service subscription. The non-metered offering is ideally suited for deploying production clusters in the Cloud on a long-term Hadoop deployment. It is also suitable for ongoing long-term sandbox and data science projects.

## Leading Hadoop Distribution Made Easy

With Big Data Cloud Service, customers have complete access to all Cloudera CDH Enterprise Data Hub features and functions as well as integration with Oracle Public Cloud components.



Each Big Data Cloud Service instance is configured such that there is a single Virtual Machine (VM), which is dedicated to the customer, in each Hadoop server of the Big Data system. Customers have root privileges for the Hadoop server VMs and they can configure the system as they like, and load additional software on the nodes to conform to business standards or analytics requirements.

Customers perform familiar Hadoop administration and OS administration tasks aided by cloud automation for patching and upgrades. Upgrades are pre-tested by Oracle and are implemented in a standard way.

All supporting infrastructure for Big Data Cloud Service is deployed, maintained and managed by Oracle, including datacenter networking, private Big Data Cloud Service InfiniBand network, physical compute and storage servers, firmware, etc. This enables customers to focus on their business needs and application requirements, and not IT infrastructure management.

Customers on their preferred schedule initiate Hadoop, Spark and OS patching while infrastructure patching is performed automatically by Oracle, with little/no disruption to the service.

## Elastic Scaling and Compute Bursting

Big Data Cloud Service enables elastic scaling and hourly compute bursting on a fully dedicated platform delivering both the performance of a dedicated and optimized Hadoop and Spark environment as well as flexibility to adjust dynamically to peak work load requirements.

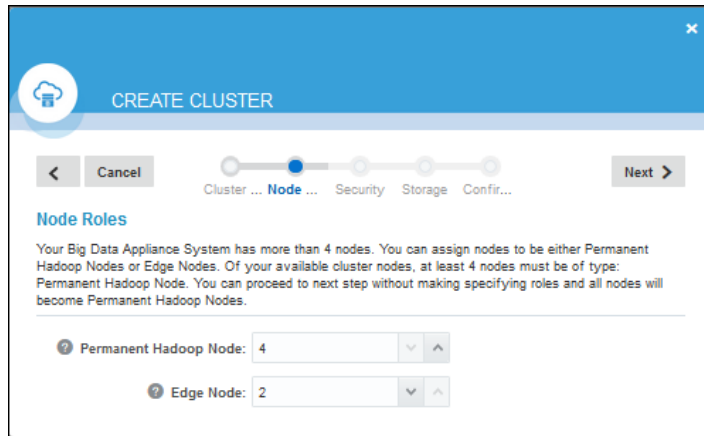
Big Data Cloud Service features no over-provisioning of hardware to ensure response times and throughput are predictable delivering analytics for mission critical big data applications. This contrasts with other cloud providers that silently overprovision hardware and may not be able to deliver the expected resources during busy periods such as quarter-ending time.

In addition, Big Data Cloud Service enables Compute Bursting, enabling customers to grow, and then later shrink, their cluster CPU capacity over their base subscription level to meet their peak or seasonal demands. With this feature, non-metered subscription customers can add up to 192 additional cluster OCPUs.

The screenshot shows the 'Instance Details' configuration interface. It includes the following fields and values:

- Name:** bigdataappliance
- Additional Nodes:** 2 Current, 2 Current (with an information icon)
- Additional Number of OCPU (Cores):** -32, 32 Current, 32 Current (with an information icon)
- Cloudera Manager admin password:** (masked with asterisks)
- Confirm Cloudera Manager admin password:** (masked with asterisks)

Compute Bursting provides Big Data Cloud Service customers with the best of both worlds: low-cost non-metered pricing for their normal needs, plus the flexibility to rapidly adjust processor capacity as their business conditions change. This avoids the costly practice of sizing for the highest possible peak workload, which is required for on-premises systems and reserved cloud capacity on other cloud providers. Compute Bursting with Big Data Cloud Service effectively lowers the costs of cloud deployment by allowing to provision only for steady state and enabling a seamless scale-up right when it is needed.



To facilitate applications to be installed according to Hadoop best practices, Big Data Cloud Service enables Edge Nodes. Edge nodes are directly attached to the InfiniBand network and can be configured via the Service Console. Connecting over InfiniBand enables the edge node to serve as a bastion as well as a staging area to land data.

### Open for Innovation

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 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.

### Simple to Configure End-to-End 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.

### Big Data Cloud Service – Customer Benefits

Driving business benefits from Big Data infrastructure is critical. Big Data Cloud Service delivers. Specific business benefits include:

- Secure clusters, ensuring that data infrastructure remains secure. In Big Data Cloud Service, security is simple to enable, is tested and remains in place across upgrades.
- Highly Available, ensuring your environment remains functional even during certain infrastructure failures. Big Data Cloud Service features a tested and automated High

Availability architecture.

- Full Control, enables your users to configure the system for their needs. This enables the users to use their favorite tools, both from partners and from Oracle directly on Big Data Cloud Service.

**ORACLE®**

CONTACT US

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

CONNECT WITH US



[blogs.oracle.com/oracle](http://blogs.oracle.com/oracle)



[facebook.com/oracle](https://facebook.com/oracle)



[twitter.com/oracle](https://twitter.com/oracle)



[oracle.com](http://oracle.com)

**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.



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