

Oracle Big Data Appliance X8-2

ORACLE® BIG DATA APPLIANCE



KEY FEATURES

- Massively scalable, open infrastructure to store, analyze and manage big data
- Industry-leading security, performance and the most comprehensive big data tool set on the market all bundled in an easy to deploy appliance
- Flexible configuration and elastic hardware choices for optimizing both floor space and growth path for Hadoop leveraging Cloudera 5.x or 6.x
- Oracle Big Data SQL delivers unprecedented integration of Big Data and Oracle Database
- High performance Advanced AI and ML compute platform
- Cloudera's comprehensive software suite including Cloudera Distribution

Oracle Big Data Appliance is a flexible, high-performance, secure platform for running diverse workloads on Hadoop, Kafka and Spark. With Oracle Big Data SQL, Oracle Big Data Appliance extends Oracle's industry-leading implementation of SQL to Hadoop/NoSQL and Kafka systems. By combining the newest technologies from the Hadoop ecosystem and powerful Oracle SQL capabilities together on a single pre-configured platform, Oracle Big Data Appliance is uniquely capable to support rapid development of new Big Data applications and tight integration with existing relational data.

Oracle Big Data Appliance X8-2

Oracle Big Data Appliance is an open, multi-purpose engineered system for Hadoop and Spark workloads and streaming data processing. Big Data Appliance is designed to run diverse workloads – from Hadoop-only workloads (Yarn, Spark, Hive etc.) to interactive, all-encompassing interactive SQL queries using Oracle Big Data SQL across Apache Kafka, Hadoop and NoSQL databases. Big Data Appliance, is a Cloudera Certified platform and supports both Cloudera 5.x and Cloudera 6.x.

Big Data Appliance provides an open environment for innovation while maintaining tight integration and enterprise-level support. Organizations can deploy external software to support new functionality – such as graph analytics, natural language processing and fraud detection. Support for non-Oracle components is delivered by their respective support channels and not by Oracle.

Lower TCO and Faster Time to Value

Big Data Appliance provides unique pricing to offer both a lower initial deployment cost as well as a dramatically reduced three and four-year TCO when compared to a Do-It-Yourself Hadoop, Spark or Kafka system. Big Data Appliance bundles a highly available hardware footprint (servers, high-speed networking, power distribution units and peripherals), OS and Java support as well as subscription costs for Cloudera's Distribution including Apache Hadoop into a single price for the life of the system. Oracle delivers a single support point-of-contact that covers the hardware and the integrated software (including Cloudera CDH).

Organizations do not want to spend valuable intellectual capital assembling and tuning their Hadoop infrastructure, especially when these resources should be applied to delivering high value business solutions. Big Data Appliance delivers a pre-configured, highly tuned environment for Cloudera Enterprise and Oracle NoSQL Database. This optimized environment enables companies to focus their resources on developing

including Apache Hadoop and Apache Spark

- Oracle Enterprise Manager combined with Cloudera Manager simplifies management of the entire Big Data Appliance
- Advanced analytics with Oracle R directly interacting with data stored in HDFS
- InfiniBand connectivity between nodes and across racks as well as to Oracle Exadata and other Oracle Engineered Systems

KEY BENEFITS

- Optimized, Open and Secure Big Data Solution
- Simplified operations, updates and patch management through a single command utility of the entire stack (OS, Java, Oracle NoSQL Database and the Cloudera stack)
- Oracle Big Data SQL provides the most complete SQL solution for Big Data when integrated with Oracle Exadata
- Most comprehensive big data tool set integrated in a single appliance
- Risk-free installation and rapid time to value
- Single Management Console integrating Big Data Appliance hardware and software monitoring through Oracle Enterprise Manager
- Single-vendor support for your entire big data solution covering both hardware and software

RELATED PRODUCTS

compelling business applications – lowering the risk for the solution. Additionally, the pre-tuned environment avoids extensive ramp-up time for new applications due to performance and production issues.

Oracle continues to invest in optimizing the software stack, and incremental benefits are delivered to customers in subsequent versions of the software, without paying for expensive consulting contracts to tune Hadoop.

Comprehensive Security

Securing data is critical to all data solutions in the enterprise; Big Data Appliance provides strong authentication, authorization and auditing of data in Hadoop out of the box.

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 Appliance leverages Apache Sentry to authorize SQL access via tools like Hive, Impala and Big Data SQL.

Both network encryption and encryption of data-at-rest are configured with Oracle Big Data Appliance. Big Data Appliance supports the latest innovations in encryption of data-at-rest by supporting HDFS Transparent Encryption with Cloudera Key Trustee. This implementation enables the tightest security on all data in HDFS. Network encryption prevents network sniffing from capturing protected data and is enabled on Big Data Appliance through a simple check box.

To ensure security and data access compliance, Big Data Appliance delivers Cloudera Navigator to track and trace all access to the data in Big Data Appliance.

In addition to securing the Hadoop system, Oracle Big Data SQL enables organizations to leverage Oracle Database security capabilities when querying Hadoop, Kafka and NoSQL data. A secure Big Data Appliance combined with Oracle Big Data SQL delivers the most comprehensive security of any big data system in the market today.

Simplified Operations

Big Data Appliance simplifies day-to-day operations by providing a simple one-command installation, update, patch and expansion utility – Mammoth – which enables rapid deployment updates (typically quarterly) to the frequently evolving Hadoop stack without incurring significant downtime. Mammoth also enables Oracle-tested, seamless upgrades between Hadoop versions and automated service management to ensure the best balance between Hadoop Master Nodes and Data Nodes.

Big Data Appliance is supported by Oracle, giving organizations a single point of support for their hardware, all integrated software (including all Cloudera software) and any additional Oracle software installed.

Elastic Configurations

Big Data Appliance is designed to expand as your data and requirements grow. Initial big data implementations may start with Big Data Appliance Starter Rack. The Starter Rack comes fully equipped with a complete set of switches and power distribution units

- Oracle Exadata
- Oracle Big Data SQL
- Oracle Big Data Spatial and Graph
- Oracle Big Data Connectors
- Oracle Big Data Discovery
- Oracle NoSQL Database
- Oracle Exalytics
- Oracle Business Intelligence Enterprise Edition
- Oracle GoldenGate for Big Data
- Oracle Data Integrator
- Oracle Stream Explorer
- Oracle Enterprise Manager

RELATED SERVICES

- Advanced Customer Services
- Product Support Services
- Consulting Services
- Oracle University Training

(PDU) required for a full rack. The Starter Rack and switching gear enables the appliance to be easily and efficiently expanded in single node hardware increments to larger configurations using Oracle Big Data Appliance X8-2 High Capacity (HC) Node plus InfiniBand Infrastructure.

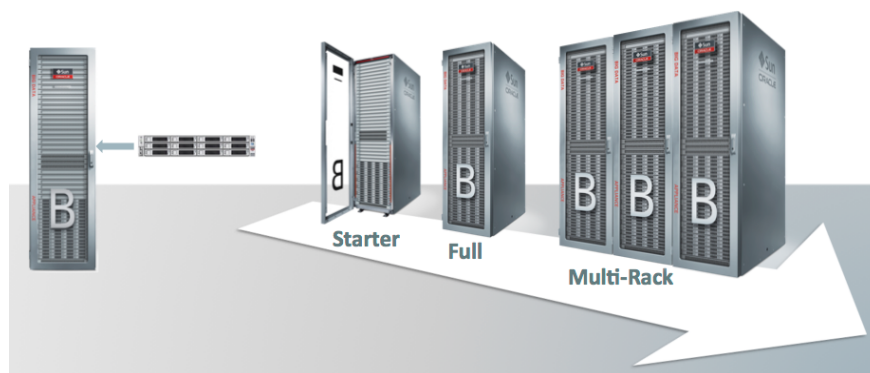


Figure 1. Modular Hardware Building Blocks

In addition to expanding the system within a rack, multiple racks can be connected using the integrated InfiniBand fabric to form larger configurations; up to 18 racks can be connected in a non-blocking manner without the need for any external switches. Larger non-blocking configurations are supported with additional external InfiniBand switches, while larger blocking network configurations can be supported without additional switches. The use of InfiniBand dramatically reduces the cost of large configurations by reducing the need for a high-bandwidth top of rack switching infrastructure.

Big Data Appliance is multitenant; it can be configured as a single cluster or as a set of clusters. This provides the flexibility customers need when deploying development, test and production clusters.

Connecting Hadoop Jobs to Oracle Database

Oracle DataSource for Apache Hadoop is an Oracle Big Data Appliance feature that turns Oracle Database tables into Hadoop and Spark data sources enabling query lookup to Oracle Database from Big Data Appliance leveraging optimal connectivity. DataSource for Apache Hadoop enables direct and consistent access to data in Oracle Database using Hive SQL, Spark SQL, as well as Hadoop and Spark APIs, which support HCatalog, InputFormat, SerDes, and StorageHandler (external tables). Data in Oracle Database is accessed in parallel using a secure connection (Kerberos, SSL, Oracle Wallet).

Oracle Big Data SQL

Oracle Big Data SQL is a data virtualization innovation from Oracle. It is a new architecture for SQL on Hadoop, seamlessly integrating data in Hadoop, Kafka and NoSQL with data in Oracle Database. Big Data SQL is available on both Oracle Big Data Appliance and other supported Apache Hadoop platforms. Using Oracle Big Data SQL, organizations can:

- Combine data from Oracle Database, Hadoop / NoSQL and Kafka in a single SQL query
- Query and analyze data in Hadoop / NoSQL, Kafka and more
- Integrate big data analysis into existing applications and architectures
- Extend security and access policies from Oracle Database to data in Hadoop and NoSQL
- Maximize query performance on all data using Smart Scan

Oracle Big Data SQL radically simplifies integrating and operating in the big data domain through two powerful features: newly expanded External Tables and Smart Scan functionality on Hadoop.

Using Big Data SQL-enabled external table types, data in Hadoop and NoSQL is exposed to Oracle Database users. These tables, once defined, automatically discover Hive metadata including data location and data parsing requirements (i.e. SerDes and StorageHandlers). This enables SQL queries to access the data in its existing format leveraging native parsing constructs.

Oracle's unique Smart Scan capability brings the proven storage processing innovations of Oracle Exadata to Oracle Big Data Appliance. The biggest performance penalties in data processing are 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
 - Hadoop data is read using native operators local to the node
- Column projection
 - Only relevant columns are returned from the source to the database engine
- Predicate evaluation and push down
 - Only relevant rows are returned from the source
 - Leverage underlying storage formats (Parquet, ORC etc.) for high selectivity on queries
- Storage Indexes
 - IO avoidance for Hadoop scans delivering massive query speed ups
- Complex function evaluation
 - SQL operators on JSON and XML types applied at the source
 - Model scoring and analytical operators evaluated at the source
- Aggregation
 - Aggregate data and return the summarized result to the database

Smart Scan and Storage Indexes coexist with Hadoop services and does not require any changes to Hadoop itself, thus staying in line with the open environment Oracle Big Data Appliance provides.

Oracle Big Data Spatial and Graph

Oracle Big Data Spatial and Graph provides advanced spatial analytic capabilities and a graph database on both Oracle Big Data Appliance and other supported Apache Hadoop and NoSQL database platforms.

The property graph component gives users a scalable graph database with industry-leading in-memory analytics. It includes 35 pre-built graph analytics enabling users to easily discover relationships, communities, influencers, and other graph patterns. The graph database is hosted on either Apache HBase or Oracle NoSQL Database and supports popular scripting languages like Python, Groovy, the open source Tinkerpop stack as well as a Java API.

The spatial analytics and services include a data enrichment service to harmonize data based on locations and place names and a wide range of 2D, 3D and raster algorithms to analyze location relationships among persons and assets in for example social media or log data. It can apply city, state, and country categorization and process and visualize geospatial map data and satellite imagery.

Oracle Big Data Connectors

In addition to providing Oracle Big Data SQL and the full Cloudera software platform, Big Data Appliance utilizes Oracle Big Data Connectors to simplify data integration and analytics. Big Data Connectors provide high-speed access to data in Hadoop from Oracle Exadata and Oracle Database – with data transfer rates on the order of 15 TB/hour. Big Data Connectors also enable integrated, highly scalable analytics – providing native access to Hadoop data and parallel processing using Oracle R Distribution. Finally, Oracle XQuery for Hadoop facilitates standard XQuery operations to process and transform documents in various formats (JSON, XML, Avro and others), executing in parallel across the Hadoop cluster.

Software Details

Big Data Appliance X8-2 – Included Software

Operating System:

- Oracle Linux 6 or Oracle Linux 7
-

Integrated Software:

Cloudera Enterprise 5 or Cloudera Enterprise 6 – Data Hub Edition with support for:

- Cloudera's Distribution including Apache Hadoop (CDH)
- Cloudera Impala
- Cloudera Search
- Apache HBase and Apache Accumulo
- Apache Spark
- Apache Kafka
- Cloudera Manager with support for:
 - Cloudera Navigator
 - Cloudera Back-up and Disaster Recovery (BDR)

Oracle DataSource for Apache Hadoop

Other:

Oracle Java JDK 8

MySQL Database Enterprise Server - Advanced Edition*

Oracle Big Data Appliance Enterprise Manager Plug-In

Oracle R Distribution

Oracle NoSQL Database Community Edition (CE)**

* Restricted Use License

** No support for Oracle NoSQL Database CE available

Big Data Appliance X8-2 – Optional Oracle Software

Oracle Big Data SQL

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
-

Oracle Data Integrator

Oracle GoldenGate for Big Data

Oracle NoSQL Database Enterprise Edition

Oracle Big Data Spatial and Graph

Hardware Details and Specifications

BIG DATA APPLIANCE X8-2 – HARDWARE

Full Rack	Starter Rack	HC Node plus InfiniBand Infrastructure
18 x Compute / Storage Nodes	6 x Compute / Storage Nodes	1 x Compute / Storage Nodes

Per Node:

- 2 x 24-Core (2.4GHz) Intel® Xeon® P8260)
- 8 x 32GB DDR4-2666 MHz Memory, expandable to 1.5TB
- 12 x 14 TB 7,200 RPM High Capacity SAS Drives
- 2 x 240GB M.2 SATA SSD Drives
- 2 x QDR 40Gb/sec InfiniBand Ports
- 1 x Dual-port InfiniBand QDR CX3 (40 Gb/sec) PCIe HCA
- 1 x Built-in RJ45 1 Gigabit Ethernet port

2 x 32 Port QDR InfiniBand Leaf Switch <ul style="list-style-type: none"> • 32 x InfiniBand ports • 8 x 10Gb Ethernet ports 	Leverages the leaf switches from the Starter Rack
1 x 36 Port QDR InfiniBand Spine Switch <ul style="list-style-type: none"> • 36 x InfiniBand Ports 	Leverages the spine switch from the Starter Rack
Additional Hardware Components included: <ul style="list-style-type: none"> • Ethernet Administration Switch • 2 x Redundant Power Distributions Units (PDUs) • 42U rack packaging 	Leverages the administration switch, PDUs and base rack from the Starter Rack
Spares Kit Included: <ul style="list-style-type: none"> • 1 x 14 TB High Capacity SAS disk • InfiniBand cables 	Leverages the spares kit from the Starter Rack

BIG DATA APPLIANCE X8-2 COMPONENT ENVIRONMENTAL SPECIFICATIONS

X8-2 High Capacity Node plus InfiniBand Infrastructure

Physical Dimensions

Height	3.5 in. (87.6 mm)
Width	17.5 in. (445.0 mm)
Depth	29.0 in. (737.0 mm)
Weight	33.1 kg (73.0 lbs)
Power	Maximum: 0.7 kW
	Typical ¹ : 0.5 kW
Cooling	Maximum: 2,481 BTU/hour
	Typical ¹ : 1,736 BTU/hour
Airflow ²	Maximum: 115 CFM
	Typical ¹ : 80 CFM

Operating temperature/humidity: 5 °C to 32 °C (41 °F to 89.6 °F), 10% to 90% relative humidity, non-condensing

Altitude Operating: Up to 3,048 m, max. ambient temperature is de-rated by 1° C per 300m above 900m

¹ Typical power usage varies by application workload² Airflow must be front to back

BIG DATA APPLIANCE X8-2 – ENVIRONMENTAL SPECIFICATIONS

Physical Dimensions Full Rack

Height	42U, 78.66" - 1998 mm	
Width	23.62" - 600mm	
Depth	47.24" - 1200 mm	
Weight	Rack	Shipping
Starter Rack	415kg – 915lbs	546kg – 1203lbs
Full Rack	836kg – 1843lbs	979kg – 2158lbs

Power

Starter Rack	Maximum: 5.3KW
	Typical ¹ : 3.7KW
Full Rack	Maximum: 13.4KW
	Typical ¹ : 9.4KW

Cooling

Starter Rack	Maximum: 18,087 BTU/hour
	Typical ¹ : 12,661 BTU/hour
Full Rack	Maximum: 45,659 BTU/hour
	Typical ¹ : 31,961 BTU/hour

Airflow²

Starter Rack	Maximum: 837 CFM
	Typical ¹ : 586 CFM
Full Rack	Maximum: 2114 CFM
	Typical ¹ : 1480 CFM

Big Data Appliance X8-2 – Further Environmental Specifications

Operating temperature/humidity: 5 °C to 32 °C (41 °F to 89.6 °F), 10% to 90% relative humidity, non-condensing

Altitude Operating: Up to 3,048 m, max. ambient temperature is de-rated by 1° C per 300m above 900m

Big Data Appliance X8-2 – Regulations and CertificationsRegulations³

Safety: UL/CSA 60950-1, EN 60950-1, IEC 60950-1 CB Scheme with all country differences
 RF/EMI: EN55022, EN61000-3-11, EN61000-3-12
 Immunity: EN 55024
 Emissions and Immunity: EN300 386

Certifications³

North America (NRTL), European Union (EU), International CB Scheme, BSMI (Taiwan), C-Tick (Australia), CCC (PRC), MSIP (Korea), CU EAC (Customs Union), VCCI (Japan)

European Union Directives³

2006/95/EC Low Voltage Directive, 2004/108/EC EMC Directive, 2011/65/EU RoHS Directive, 2012/19/EU WEEE Directive

¹ Typical power usage varies by application workload

² Airflow must be front to back

³ All standards and certifications referenced are to the latest official version at the time the data sheet was written. Other country regulations/certifications may apply. In some cases, as applicable, regulatory and certification compliance were obtained at the component level.

BIG DATA APPLIANCE X8-2 – EXPANSIONS

In-Rack	Multi-Rack
<p>Expansion:</p> <p>Field upgrade leveraging up to 12 HC Nodes plus InfiniBand Infrastructure per Starter Rack.</p> <p>Additional hardware included with each HC Node plus InfiniBand Infrastructure:</p> <ul style="list-style-type: none"> · 1 x Node with direct attached storage as shown earlier · InfiniBand and Ethernet cables to connect the components <p>Expansion supports multiple generations of hardware.</p>	<p>Up to 18 racks can be connected without requiring additional InfiniBand switches.</p> <p>InfiniBand cables to connect 3 racks are included in the rack Spares Kits.</p> <p>Additional optical InfiniBand cables required when connecting 4 or more racks.</p>
<p>Memory</p> <ul style="list-style-type: none"> • Expand or update the memory in any individual node or any number of nodes from 256GB per node to 1.5TB per node 	

BIG DATA APPLIANCE X8-2 – SUPPORT SERVICES

Hardware Warranty: 1 year with a 4 hour web/phone response during normal business hours (Mon-Fri 8AM-5PM), with 2 business day on-site response/Parts Exchange

Oracle Premier Support for Systems: Oracle Linux and integrated software support and 24x7 with 2 hour on-site hardware service response (subject to proximity to service center)

Oracle Premier Support for Operating Systems

Oracle Customer Data and Device Retention

System Installation Services

Software Configuration Services

System Expansion Support Services including hardware installation and software configuration

Quarterly on-site patch deployment service

Oracle Automatic Service Request (ASR)



CONTACT US

For more information about Big Data Appliance X8-2, visit oracle.com or call +1.800.ORACLE1 to speak to an Oracle representative.

CONNECT WITH US



blogs.oracle.com/oracle



facebook.com/oracle



twitter.com/oracle



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. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group. Cloudera, Cloudera CDH, and Cloudera Manager, Cloudera BDR and Cloudera Navigator are registered and unregistered trademarks of Cloudera, Inc. 1219



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