Oracle Big Data Appliance is a flexible, high-performance, secure platform for running diverse workloads on Hadoop, Kafka and NoSQL. 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 X7-2

Oracle Big Data Appliance is an open, multi-purpose engineered system for Hadoop 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. These capabilities are available for on-premises deployment as well as a Public Cloud Service – Oracle Big Data Cloud Service. Big Data Appliance, Big Data Cloud Machine and Big Data Cloud Service are Cloudera Certified platforms.

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.

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
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
- Handle low-latency workloads with the pre-installed and configured Oracle NoSQL Database
- 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 though 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

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 new 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

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.

**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 to four-year TCO when compared to a DIY Hadoop system. Big Data Appliance bundles the hardware (servers, high-speed networking, power distribution units and peripherals), OS support and subscription costs for the Cloudera software into a single price for the life of the system. A single hardware support license covers the hardware as well as the integrated software.

Organizations do not want to spend valuable intellectual capital assembling and tuning an optimized Hadoop/NoSQL infrastructure, especially when these resources can 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 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.

**Comprehensive Security**

Securing data is critical to Big 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 (an open-source project of which Oracle developers are founding members) to authorize SQL access via tools like Hive and Impala. By delivering and developing Sentry, Oracle delivers Big Data Appliance with
the highest data security levels currently available for Hadoop.

Both network encryption and encryption of data-at-rest are included with Oracle Big Data Appliance and supported by Oracle. Big Data Appliance supports the latest innovations in encryption of data-at-rest by supporting native HDFS encryption with a key management facility. 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 and NoSQL data. A secure Big Data Appliance combined with Oracle Big Data SQL delivers the most comprehensive security of any big data system.

**Simplified Operations**

Oracle Enterprise Manager provides a single entry point for managing the entire system – both hardware and software – providing continuity across other Oracle products in the organization. To provide deep management capabilities for Hadoop, Enterprise Manager enables a context-aware integration with Cloudera Manager.

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 (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 X7-2 High Capacity (HC) Node plus InfiniBand Infrastructure.
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 by connecting InfiniBand cables 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 the top of rack switching fabric.

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.

**Connectivity and Performance Enhancements**

The Table Access for Hadoop feature is an Oracle Big Data Appliance feature that turns Oracle Database tables into Hadoop and Spark data sources enabling query lookup to Oracle Database

Table Access for 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).

Perfect Balance is a Big Data Appliance feature that enables MapReduce jobs on Big Data Appliance to better handle data skew. While the default Hadoop method of distributing the reduce load is appropriate for many jobs, it does not distribute the load evenly for jobs with significant data skew. Perfect Balance addresses this scenario by detecting and optimizing for data skew.

**Software Details**

<table>
<thead>
<tr>
<th>Big Data Appliance X7-2 – Included Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating System:</td>
</tr>
<tr>
<td>• Oracle Linux 5 or Oracle Linux 6; Oracle Linux 7 for Edge Nodes</td>
</tr>
</tbody>
</table>
Integrated Software:
Cloudera Enterprise 5 – 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 Perfect Balance
Oracle Table Access for 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
** Oracle NoSQL Database CE Support is not included in the Support for Big Data Appliance. Instead a separate support subscription for Oracle NoSQL DB CE is required.

Big Data Appliance X7-2 – 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 Audit Vault and Database Firewall for Hadoop Auditing

Oracle Data Integrator
Oracle GoldenGate

Oracle NoSQL Database Enterprise Edition

Oracle Big Data Spatial and Graph
Hardware Details and Specifications

BIG DATA APPLIANCE X7-2 – HARDWARE

<table>
<thead>
<tr>
<th>Full Rack</th>
<th>Starter Rack</th>
<th>HC Node plus InfiniBand Infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 x Compute / Storage Nodes</td>
<td>6 x Compute / Storage Nodes</td>
<td>1 x Compute / Storage Nodes</td>
</tr>
</tbody>
</table>

Per Node:
- 2 x 24-Core (2.1GHz) Intel® Xeon® 8160
- 8 x 32GB DDR4-2666 MHz Memory, expandable to 1.5TB
- 12 x 10 TB 7,200 RPM High Capacity SAS Drives
- 2 x 150GB 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
- 32 x InfiniBand ports
- 8 x 10Gb Ethernet ports

1 x 36 Port QDR InfiniBand Spine Switch
- 36 x InfiniBand Ports

Additional Hardware Components included:
- Ethernet Administration Switch
- 2 x Redundant Power Distributions Units (PDUs)
- 42U rack packaging

Spares Kit Included:
- 1 x 10 TB High Capacity SAS disk
- InfiniBand cables

Leverages the leaf switches from the Starter Rack
Leverages the spine switch from the Starter Rack
Leverages the administration switch, PDUs and base rack from the Starter Rack
Leverages the spares kit from the Starter Rack

BIG DATA APPLIANCE X7-2 COMPONENT ENVIRONMENTAL SPECIFICATIONS

X7-2 High Capacity Node plus InfiniBand Infrastructure

Physical Dimensions

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>3.5 in. (87.6 mm)</td>
</tr>
<tr>
<td>Width</td>
<td>17.5 in. (445.0 mm)</td>
</tr>
<tr>
<td>Depth</td>
<td>29.0 in. (737.0 mm)</td>
</tr>
<tr>
<td>Weight</td>
<td>33.1 kg (73.0 lbs)</td>
</tr>
</tbody>
</table>

Power

<table>
<thead>
<tr>
<th></th>
<th>Maximum: 0.7 kW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Typical: 0.5 kW</td>
</tr>
</tbody>
</table>

Cooling

<table>
<thead>
<tr>
<th></th>
<th>Maximum: 2,481 BTU/hour</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Typical: 1,736 BTU/hour</td>
</tr>
</tbody>
</table>

Airflow

<table>
<thead>
<tr>
<th></th>
<th>Maximum: 115 CFM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Typical: 80 CFM</td>
</tr>
</tbody>
</table>
Operating temperature/humidity: 5 ºC to 32 ºC (41 ºF to 89.6 ºF), 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

<table>
<thead>
<tr>
<th>Physical Dimensions Full Rack</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
</tr>
<tr>
<td>Width</td>
</tr>
<tr>
<td>Depth</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Weight</th>
<th>Rack</th>
<th>Shipping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starter Rack</td>
<td>415kg – 915lbs</td>
<td>546kg – 1203lbs</td>
</tr>
<tr>
<td>Full Rack</td>
<td>836kg – 1843lbs</td>
<td>979kg – 2158lbs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starter Rack</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Full Rack</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cooling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starter Rack</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Full Rack</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Airflow²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starter Rack</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Full Rack</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Big Data Appliance X7-2 – Further Environmental Specifications**

Operating temperature/humidity: 5 ºC to 32 ºC (41 ºF to 89.6 ºF), 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 X7-2 – Regulations and Certifications**

<table>
<thead>
<tr>
<th>Regulations²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safety: UL/CSA 60950-1, EN 60950-1, IEC 60950-1 CB Scheme with all country differences</td>
</tr>
<tr>
<td>RFI/EMI: EN55022, EN61000-3-11, EN61000-3-12</td>
</tr>
<tr>
<td>Immunity: EN 55024</td>
</tr>
</tbody>
</table>
Emissions and Immunity: EN300 386

Certifications
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 Directives
- 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 X7-2 – EXPANSIONS

<table>
<thead>
<tr>
<th>In-Rack</th>
<th>Multi-Rack</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expansion:</td>
<td>Up to 18 racks can be connected without requiring additional InfiniBand switches.</td>
</tr>
<tr>
<td>Field upgrade leveraging up to 12 HC Nodes plus InfiniBand Infrastructure per Starter Rack.</td>
<td>InfiniBand cables to connect 3 racks are included in the rack Spares Kits.</td>
</tr>
<tr>
<td>Additional hardware included with each HC Node plus InfiniBand Infrastructure:</td>
<td>Additional optical InfiniBand cables required when connecting 4 or more racks.</td>
</tr>
<tr>
<td>• 1 x Node with direct attached storage as shown earlier</td>
<td></td>
</tr>
<tr>
<td>• InfiniBand and Ethernet cables to connect the components</td>
<td></td>
</tr>
<tr>
<td>Expansion supports multiple generations of hardware.</td>
<td></td>
</tr>
<tr>
<td>Memory</td>
<td></td>
</tr>
<tr>
<td>• Expand or update the memory in any individual node or any number of nodes from 256GB per node to 1.5TB per node</td>
<td></td>
</tr>
</tbody>
</table>

BIG DATA APPLIANCE X7-2 – 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)