HPE Reference Configuration for Hortonworks Data Platform (HDP) on HPE Elastic Platform for Big Data Analytics (EPA)

HPE EPA Balanced and Density Optimized (BDO) traditional cluster design for HDP 2.6
Contents

Executive summary ..........................................................................................................................3
HPE Pointnext Services ...................................................................................................................4
  Hortonworks Data Platform Overview .........................................................................................4
Solution overview ..........................................................................................................................5
  HPE Balanced and Density Optimized (BDO) solution (Standard) ..................................................5
Solution components and configuration guide .............................................................................6
  Single-rack Reference Configuration .........................................................................................6
Pre-deployment considerations ......................................................................................................7
High Availability considerations ....................................................................................................8
Software components for control blocks .......................................................................................8
Software components for compute/storage blocks .......................................................................9
Capacity and sizing ........................................................................................................................10
  Hadoop Best practices and tuning guidelines .............................................................................10
HPE Sizer for the Elastic Platform for Big Data Analytics ................................................................11
HPE Insight Cluster Management Utility ......................................................................................11
Summary ......................................................................................................................................13
 Appendix A: Hadoop cluster tuning/optimization .......................................................................13
 Appendix B: HPE Pointnext value-added services and support .....................................................16
Resources and additional links ....................................................................................................18
Executive summary

Hewlett Packard Enterprise and Apache Hadoop allow you to derive new business insights from all of your data by providing a platform to store, manage and process data at scale. As organizations start to capture and collect these big datasets, increased storage becomes a necessity. With a centralized repository, increased compute power and storage are requirements to respond to the sheer scale of Big Data and its projected growth. Enterprises need Big Data platforms that are purpose-built to support their Big Data strategies. This white paper provides several performance optimized configurations for deploying Hortonworks Data Platform (HDP) clusters of varying sizes on HPE infrastructure that provide a significant reduction in complexity and a recognized increase in value and performance.

The configurations are based on Hortonworks Data Platform (HDP), 100% open source distribution of Apache Hadoop, specifically HDP 2.6 and the HPE ProLiant DL380 Gen10 server platform. The configurations reflected in this document have been jointly designed and developed by HPE and Hortonworks to provide optimum computational performance for Hadoop and are also compatible with other HDP 2.x releases.

HPE Elastic Platform for Big Data Analytics (EPA) is designed as a modular infrastructure foundation to address the need for a scalable multi-tenant platform, by enabling independent scaling of compute and storage through infrastructure building blocks that are optimized for density and workloads. Hewlett Packard Enterprise supports two different deployment models under this platform:

- **HPE Workload and Density Optimized (WDO) system** – Harnesses the power of faster Ethernet networks that enable a building block approach to independently scale compute and storage and lets you consolidate your data and workloads growing at different rates. The HPE WDO system is based on the HPE Apollo 4200 storage block and the HPE Apollo 2000 compute block.

- **HPE Balanced and Density Optimized (BDO) system** – Supports Hadoop deployments that scale compute and storage together, with some flexibility in choice of memory, processor, and storage capacity. This is primarily based on the HPE ProLiant DL380 server platform, with density optimized variants using HPE Apollo 4200 servers. In this paper we have tested with BDO systems – standard solution, the HPE ProLiant DL380 Gen10 server platform.

This Reference Configuration (RC) describes deployment options for the Hortonworks HDP 2.6 using the HPE Elastic Platform for Big Data Analytics - modular building blocks of compute and storage optimized for modern workloads white paper - [HPE Reference Configuration for Elastic Platform for Big Data Analytics](http://h17007.www1.hpe.com/us/en/enterprise/reference-architecture/info-library/index.aspx?workload=big_data). This RC also provides suggested configurations that highlight the benefits of a building block approach to address the diverse processing and storage requirements typical of modern Big Data platforms.


**Target audience:** This document is intended for decision makers, system and solution architects, system administrators and experienced users who are interested in reducing the time to design and purchase an HPE and Hortonworks HDP solution. An intermediate knowledge of Apache Hadoop and scale out infrastructure is recommended. Those already possessing expert knowledge about these topics may proceed directly to the Solution components section.

**Document purpose:** The purpose of this document is to describe a Reference Configuration, highlighting recognizable benefits to technical audiences and providing guidance for end users on selecting the right configuration for building their Hadoop cluster needs.

This white paper describes testing performed in January 2018.
HPE Pointnext Services

HPE recommends that customers purchase the option of services from HPE Pointnext, as detailed in Appendix B: HPE Pointnext value-added services and support, to install and configure the operating system, verify if all firmware and versions are installed correctly, and run a suite of tests that verify that the configuration is performing optimally. Once this has been done, the customer can perform a standard Hortonworks HDP installation using the recommended guidelines in this document.

Hortonworks Data Platform Overview

Hortonworks is a major contributor to Apache Hadoop, the world’s most popular Big Data platform. Hortonworks focuses on further accelerating the development and adoption of Apache Hadoop by making the software more robust and easier to consume for enterprises and more open and extensible for solution providers. The Hortonworks Data Platform (HDP), powered by Apache Hadoop, is a massively scalable and 100% open source platform for storing, processing and analyzing large volumes of data. It is designed to deal with data from many sources and formats in a very quick, easy and cost-effective manner.

HDP is a platform for multi-workload data processing across an array of processing methods – from batch through interactive and real-time – all supported with solutions for governance, integration, security and operations. As the only completely open Hadoop data platform available, HDP integrates with and augments your existing best-of-breed applications and systems so you can gain value from your enterprise Big Data, with minimal changes to your data architectures. Finally, HDP allows you to deploy Hadoop wherever you want it – from cloud or on-premises as an appliance, and Linux®. Figure 1 shows the Hortonworks Data Platform.

HDP enables enterprises to deploy, integrate and work with unprecedented volumes of structured and unstructured data. It is a platform that is based on a centralized architecture supported by YARN that allocates resources among various applications. HDP is interoperable with a broad ecosystem of data center and cloud providers. As shown in Figure 1, the platform can utilize a range of processing methods – from batch to interactive and real-time – all supported by solutions for governance, integration, security, and operations. HDP integrates with and augments solutions like HPE WDO system, allowing you to maximize the value of big data. HDP enables Open Enterprise Hadoop, a full suite of essential Hadoop capabilities in the following functional areas: data management, data access, data governance and integration, security, and operations, refer to https://hortonworks.com/products/data-platforms/hdp/. Spark is part of the HDP and is certified as YARN Ready. Memory and CPU-intensive Spark-based applications can coexist with other workloads deployed in a YARN-enabled cluster as shown in Figure 1. This approach avoids the need to create and manage dedicated Spark clusters and allows for more efficient resource use within a single cluster.

Hortonworks approached Spark the same way as other data access engines like Storm, Hive, and HBase: outline a strategy, rally the community, and contribute key features within the Apache Software Foundation’s process.

![Figure 1. Hortonworks Data Platform](image-url)
This release incorporates the most recent innovations that have happened in Hadoop and its supporting ecosystem of projects. HDP 2.6 packages more than a hundred new features across all Apache Hadoop open source existing projects. Every component is updated and Hortonworks has added some key technologies and capabilities to HDP 2.6. Figure 2 shows Hortonworks Data Platform 2.6.

**Figure 2.** Hortonworks Data Platform HDP 2.6

### Solution overview

The Reference Configurations are based on the Hortonworks HDP edition, specifically version 2.6, and BDO systems which includes the HPE ProLiant DL380 Gen10 server platform.

**HPE Balanced and Density Optimized (BDO) solution (Standard)**

The HPE BDO solution infrastructure blueprints are composed of four blocks: storage/compute blocks, control blocks, network blocks, and rack blocks. Listed below are the blocks and model in a BDO solution:

<table>
<thead>
<tr>
<th>Blocks</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Block</td>
<td>HPE ProLiant DL360 Gen10</td>
</tr>
<tr>
<td>Compute/Storage Block</td>
<td>HPE ProLiant DL380 Gen10</td>
</tr>
<tr>
<td>Network Block</td>
<td>HPE FlexFabric S940 48XGT 6QSFP28 switch</td>
</tr>
<tr>
<td>Rack Block</td>
<td>1200mm or 1075mm</td>
</tr>
</tbody>
</table>

**NOTE**

Use one 150GB SATA RI M.2 DS SSD for the OS disk in SATA AHCI mode, and not software RAID. If software RAID is used, the two 150GB M.2 SSD disks are managed by an HPE Smart Array S100i SR Gen10 SW RAID controller using in-distro open-source software to create a two-disk RAID1 boot volume. Software RAID can require a significant amount of the server’s resources and harm performance. For more information you can access the Implementing Linux Software RAID1 on HPE ProLiant Servers with RHEL 7.3 and SLES 12 SP2 document.

If RAID1 for OS drives are required, HPE recommends configuring the HPE ProLiant DL380 Gen10 2SFF rear drives for OS.

For detailed information, refer to HPE Reference Configuration for Elastic Platform for Big Data Analytics.
Solution components and configuration guide

Single-rack Reference Configuration

The single-rack Hadoop Reference Configuration (RC) is designed to perform well as a single-rack cluster design but also form the basis for a much larger multi-rack design. When moving from the single-rack to multi-rack design, one can simply add racks to the cluster without having to change any components within the single-rack. The RC reflects the following.

Single-rack network block

The HPE FlexFabric 5940 48XGT 6QSFP28 switch is a high density ToR switch available as a 1RU 48-port 10GbE. This switch can be used for high-density 10GbE ToR with 100GbE/40GbE/25GbE/10GbE spine/ToR connectivity. 100GbE ports may be split into four 25GbE ports and can also support 40GbE which can be split into four by 10GbE for a total of 80 25/10GbE ports. The HPE FlexFabric 5940 48XGT 6QSFP28 switch includes six 100GbE uplinks which can be used to connect the switches in the rack into the desired network or to the 100GbE HPE FlexFabric 5950 32QSFP28 aggregation switch. Keep in mind that if IRF bonding is used, it requires 2x 100GbE ports per switch, which would leave 4x 100GbE ports on each HPE FlexFabric 5940 48XGT 6QSFP28 switch for uplinks.

Power and cooling

In planning for large clusters, it is important to properly manage power redundancy and distribution. To ensure the servers and racks have adequate power redundancy we recommend that each server have a backup power supply, and each rack have at least two Power Distribution Units (PDUs). There is an additional cost associated with procuring redundant power supplies.

Reference Configuration for standard BDO with HPE ProLiant DL380 Gen10 balanced block

Refer to Figure 3 for a rack-level view of the single-rack Reference Configuration for this solution with HPE ProLiant DL380 balanced block.

For more information on configuration for the Balanced Optimization solution refer to the HPE Reference Configuration for Elastic Platform for Big Data Analytics.

Control Block - 1 Edge Node
HPE DL360 Gen10 8SFF
Dual 12-Core Intel Xeon-S 4116 2.10GHz
192GB RAM (12x 16GB 2Rx8 PC4-2666V-R)
7.2TB Disks (8x 900GB 12G SAS 10K HDD)
1x Smart Array P408i-a
1x Ethernet 10Gb 2-port 535FLR-T Adapter

Control Block - 2 Ethernet Switches
HPE FlexFabric 5940 48XGT 6QSFP28 switch

Control Block - 1 Management Node
HPE DL360 Gen10 8SFF
Dual 12-Core Intel Xeon-S 4116 2.10 GHz
192GB RAM (12x 16GB 2Rx8 PC4-2666V-R)
7.2TB Disks (8x 900GB 12G SAS 10K HDD)
1x Smart Array P408i-a
1x Ethernet 10Gb 2-port 535FLR-T Adapter

Network Block - 2 Head nodes - NameNode/Resource manager
HPE DL360 Gen10 8SFF
Dual 12-Core Intel Xeon-S 4116 2.10 GHz
192GB RAM (12x 16GB 2Rx8 PC4-2666V-R)
7.2TB Disks (8x 900GB 12G SAS 10K HDD)
1x Smart Array P408i-a
1x Ethernet 10Gb 2-port 535FLR-T Adapter

Software
Operating System: Red Hat Enterprise Linux 7.3
Hortonworks HDP 2.6
HPE Insight Cluster Management Utility v8.2
Optional:
Intelligent PDU

Compute/Storage Block-18 Compute/Worker Nodes
18 x HPE ProLiant DL380 Gen10 19LFF
Dual 14-Core Intel Xeon-G 5120 2.2GHz
384GB RAM (12x 32GB 2Rx4 PC4-2666V-R)
76TB Disks (19x 4TB 6G SATA 7.2k LFF HDD)
2x 1TB 6G SATA 7.2k SFF HDD
Smart Array P408i-a + SAS Expander
1x Ethernet 10Gb 2-port 535FLR-T Adapter

Three-phase PDU (4 PDUs per rack):
HPE 4.9k VA/L6-30i/NA/J PDU

For multi-rack architecture, refer to HPE Reference Configuration for Elastic Platform for Big Data Analytics.
**Best practice**
For each server, Hewlett Packard Enterprise recommends that each power supply is connected to a different PDU than the other power supply on the same server. Furthermore, the PDUs in the rack can each be connected to a separate data center power line to protect the infrastructure from a data center power line failure.

Additionally, distributing the server power supply connections evenly to the in-rack PDUs, as well as distributing the PDU connections evenly to the data center power lines, ensures an even power distribution in the data center and avoids overloading any single data center power line. When designing a cluster, check the maximum power and cooling that the data center can supply to each rack and ensure that the rack does not require more power and cooling than is available.

**Pre-deployment considerations**
The operating system and the network are key factors you need to consider prior to designing and deploying a HDP cluster. The following subsections articulate the design decisions in creating the baseline configurations for the Reference Configurations.

**Operating system**
Hortonworks 2.6.4 supports 64-bit operating systems, visit [https://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.6.4/bk_support-matrices/content/ch01.html](https://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.6.4/bk_support-matrices/content/ch01.html) for the minimum requirements. In this RC, we have tested with Red Hat Enterprise Linux® (RHEL) 7.3.

**Key point**

**Computations**
Employing Hyper-Threading increases effective core count, potentially allowing the YARN ResourceManager to assign more cores as needed.

**Storage capacity**
The number of disks and their corresponding storage capacity determines the total amount of the storage capacity for your cluster.

**Redundancy**
Hadoop ensures that a certain number of block copies are consistently available. This number is configurable in the block replication factor setting, which is typically set to three. If a Hadoop worker node goes down, Hadoop will replicate the blocks that had been on that server onto other servers in the cluster to maintain the consistency of the number of block copies. For example, if the NIC (Network Interface Card) on a server with 16TB of block data fails, 16TB of block data will be replicated between other servers in the cluster to ensure the appropriate number of replicas exist. Furthermore, the failure of a non-redundant ToR (Top of Rack) switch will generate even more replication traffic. Hadoop provides data throttling capability in the event of a node/disk failure so as to not overload the network.

**I/O performance**
The more disks you have, the less likely it is that you will have multiple tasks accessing a given disk at the same time. This avoids queued I/O requests and incurring the resulting I/O performance degradation.

**Disk configuration**
For management nodes, storage reliability is important and SAS drives are recommended. For worker nodes, one has the choice of SAS or SATA and as with any component there is a cost/performance tradeoff. Specific details around disk and RAID configurations will be provided in the server tuning section in Appendix A.

**Network**
Configuring a single ToR switch per rack introduces a single point of failure for each rack. In a multi-rack system such a failure will result in a very long replication recovery time as Hadoop rebalances storage; and, in a single-rack system such a failure could bring down the whole cluster. Consequently, configuring two ToR switches per rack is recommended for all production configurations as it provides an additional measure of redundancy. This can be further improved by configuring link aggregation between the switches. The most desirable way to configure link aggregation is by bonding the two physical NICs on each server. Port1 wired to the first ToR switch and Port2 wired to the second ToR switch, with the two switches IRF bonded. When done properly, this allows the bandwidth of both links to be used. If either of the switches fail, the
servers will still have full network functionality, but with the performance of only a single link. Not all switches have the ability to do link aggregation from individual servers to multiple switches; however, the HPE FlexFabric 5940 48XGT 6QSFP28+ switch supports this through HPE Intelligent Resilient Fabric (IRF) technology. In addition, switch failures can be further mitigated by incorporating dual power supplies for the switches.

Hadoop is rack-aware and tries to limit the amount of network traffic between racks. The bandwidth and latency provided by two bonded 10 Gigabit Ethernet (GbE) connections from the worker nodes to the ToR switch is more than adequate for most Hadoop configurations.


High Availability considerations

The following are some of the High Availability (HA) features considered in this Reference Configuration:

- **NameNode HA** – The configurations in this white paper utilize quorum-based journaling high-availability feature. For this feature, servers should have similar I/O subsystems and server profiles so that each NameNode server can potentially take the role of another. Another reason to have similar configurations is to ensure that ZooKeeper's quorum algorithm is not affected by a machine in the quorum that cannot make a decision as fast as its quorum peers.

- **ResourceManager HA** – To make a YARN cluster highly available (similar to JobTracker HA in MR1), the underlying architecture of an Active/Standby pair is configured, hence the completed tasks of in-flight MapReduce jobs are not re-run on recovery after the ResourceManager is restarted or failed over. One ResourceManager is Active and one or more ResourceManagers are in standby mode waiting to take over should anything happen to the Active ResourceManager. Ambari provides a simple wizard to enable HA for YARN ResourceManager.

- **OS availability and reliability** – For the reliability of the server, the OS disk is configured in a RAID1+0 configuration thus preventing failure of the system from OS hard disk failures.

- **Network reliability** – The Reference Configuration uses the standard HPE BDO network block with two HPE FlexFabric 5940 48XGT 6QSFP28 switches for redundancy, resiliency and scalability through using Intelligent Resilient Fabric (IRF) bonding. We recommend using redundant power supplies.

- **Power supply** – To ensure the servers and racks have adequate power redundancy we recommend that each server have a backup power supply, and each rack have at least two Power Distribution Units (PDUs).

Software components for control blocks

The control block is made up of three HPE ProLiant DL360 Gen10 servers, with an optional fourth server acting as an edge or gateway node depending on the customer enterprise network requirements.

Management node

The management node hosts the applications that submit jobs to the Hadoop cluster. We recommend that you install with the software components shown in Table 2.

<table>
<thead>
<tr>
<th>Software</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red Hat Enterprise Linux 7.3</td>
<td>Recommended Operating System</td>
</tr>
<tr>
<td>HPE Insight CMU 8.2</td>
<td>Infrastructure Deployment, Management, and Monitoring</td>
</tr>
<tr>
<td>Oracle JDK 1.8.0.151</td>
<td>Java Development Kit</td>
</tr>
<tr>
<td>PostgreSQL 9.2</td>
<td>Database Server for Ambari</td>
</tr>
<tr>
<td>Ambari 2.6</td>
<td>Ambari Management Software</td>
</tr>
<tr>
<td>Hue Server</td>
<td>Web Interface for Hadoop Applications</td>
</tr>
<tr>
<td>ZooKeeper</td>
<td>Cluster coordination service</td>
</tr>
<tr>
<td>NameNode HA</td>
<td>NameNode HA (Journal Node)</td>
</tr>
</tbody>
</table>
Head nodes
The head node servers contain the following software components with HA feature enabled. See the following link for more information on installing and configuring the service, https://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.6.4/index.html.

Table 3 shows the head node servers base software components.

### Table 3. Head node servers base software components

<table>
<thead>
<tr>
<th>Software</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red Hat Enterprise Linux 7.3</td>
<td>Recommended Operating System</td>
</tr>
<tr>
<td>Oracle JDK 1.8.0.151</td>
<td>Java Development Kit</td>
</tr>
<tr>
<td>ResourceManager</td>
<td>YARN ResourceManager (Active or Failover)</td>
</tr>
<tr>
<td>NameNode HA</td>
<td>NameNode HA (Journal Node, NameNode Active or Failover)</td>
</tr>
<tr>
<td>Oozie</td>
<td>Oozie Workflow scheduler service</td>
</tr>
<tr>
<td>HBaseMaster</td>
<td>The HBase Master for the Hadoop cluster (Only if running HBase)</td>
</tr>
<tr>
<td>ZooKeeper</td>
<td>Cluster coordination service</td>
</tr>
<tr>
<td>Flume</td>
<td>Flume</td>
</tr>
</tbody>
</table>

Edge node
The edge node hosts the client configurations that submit jobs to the Hadoop cluster, but this optional control block depending on the customer enterprise network requirements. We recommend that you install the following software components shown in Table 4.

### Table 4. Edge node basic software components

<table>
<thead>
<tr>
<th>Software</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red Hat Enterprise Linux 7.3</td>
<td>Recommended Operating System</td>
</tr>
<tr>
<td>Oracle JDK 1.8</td>
<td>Java Development Kit</td>
</tr>
<tr>
<td>Client Services</td>
<td>Hadoop Gateway Services (HDFS, YARN, MapReduce, HBase, and others)</td>
</tr>
</tbody>
</table>

Software components for compute/storage blocks
The worker nodes run the DataNode, NodeManager and YARN container processes and thus storage capacity and compute performance are important factors.

Balanced block software components
Table 5 lists the worker node software components. See the following link for more information on installing and configuring the NodeManager and DataNode, https://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.6.4/index.html.

### Table 5. Compute/storage node base software components

<table>
<thead>
<tr>
<th>Software</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red Hat Enterprise Linux 7.3</td>
<td>Recommended Operating System</td>
</tr>
<tr>
<td>Oracle JDK 1.8.0.151</td>
<td>Java Development Kit</td>
</tr>
<tr>
<td>NodeManager</td>
<td>The NodeManager process for MR2/YARN</td>
</tr>
<tr>
<td>DataNode</td>
<td>DataNode process for HDFS</td>
</tr>
</tbody>
</table>

For Hardware configuration guidelines, refer to BDO Standard block, HPE Reference Configuration for Elastic Platform for Big Data Analytics.
Capacity and sizing

Hadoop cluster storage sizing requires careful planning and identifying the current and future storage and compute needs. Use the following as general guidelines for data inventory:

- Sources of data
- Frequency of data
- Raw storage
- Processed HDFS storage
- Replication factor
- Default compression turned on
- Space for intermediate files

Hadoop Best practices and tuning guidelines

HDFS configuration optimizations

Make the following changes to the HDFS configuration.

- Increase the dfs.blocksize value to allow more data to be processed by each map task, thus reducing the total number of mappers and NameNode memory consumption:
  
dfs.blocksize 268435456 or 536870912 depending on workloads

- Increase the dfs.namenode.handler.count value to better manage multiple HDFS operations from multiple clients:
  
dfs.namenode.handler.count 120

YARN/MapReduce2 configurations

While configuring YARN for MapReduce jobs, make sure that the following attributes have been specified with sufficient vcores and memory. They represent resource allocation attributes for map and reduce containers.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>mapreduce.map.cpu.vcores</td>
<td>72</td>
</tr>
<tr>
<td>mapreduce.map.memory.mb</td>
<td>5120</td>
</tr>
<tr>
<td>mapreduce.reduce.cpu.vcores</td>
<td>72</td>
</tr>
<tr>
<td>mapreduce.reduce.memory.mb</td>
<td>5120</td>
</tr>
<tr>
<td>mapreduce.task.io.sort.factor</td>
<td>100</td>
</tr>
<tr>
<td>yarn.nodemanager.resource.memory-mb</td>
<td>204800</td>
</tr>
<tr>
<td>yarn.nodemanager.resource.cpu-vcores</td>
<td>72</td>
</tr>
<tr>
<td>mapreduce.job.reduce.slowstart.completedmaps</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Note

Optimum values for these attributes depend on the nature of workload/use case.

mapreduce.map.java.opts.max.heap -Xmx4096m
mapreduce.reduce.java.opts.max.heap -Xmx4096m
**Best practice**

These servers have 6 memory channels per proc and in order to get optimal performance all channels should be used, which means 12 DIMMs (or 24 DIMMs). For details on the HPE Server Memory Options Population Rules, visit http://www.hpe.com/docs/memory-population-rules.

---

**HPE Sizer for the Elastic Platform for Big Data Analytics**

HPE has developed the [HPE Sizer for the Elastic Platform for Big Data Analytics](http://www.hpe.com/info/sizers) to assist customers with proper sizing of these environments. Based on design requirements, the sizer will provide a suggested bill of materials (BOM) and metrics data for an HPE EPA WDO cluster which can be modified further to meet customer requirements.

To download the [HPE Sizer for the Elastic Platform for Big Data Analytics](http://www.hpe.com/info/sizers), visit hpe.com/info/sizers.

---

**HPE Insight Cluster Management Utility**

The HPE Insight Cluster Management Utility (Insight CMU) is a collection of tools used to manage and monitor a large group of nodes, specifically High performance Computing and large Linux Clusters such as Big data environments. Insight CMU helps manage, install, and monitor the nodes of your cluster from a single interface. A simple graphical interface enables an “at-a-glance” real-time or 3D historical view of the entire cluster for both infrastructure and application (including Hadoop) metrics, provides frictionless scalable remote management and analysis, and allows rapid provisioning of software to all nodes of the system.

**Best practice**

HPE recommends using HPE Insight CMU for all Hadoop clusters. HPE Insight CMU allows one to easily correlate Hadoop metrics with cluster infrastructure metrics, such as CPU Utilization, Network Transmit/Receive, Memory Utilization and I/O Read/Write. This allows characterization of Hadoop workloads and optimization of the system thereby improving the performance of the Hadoop cluster. HPE Insight CMU Time View Metric Visualizations will help you understand, based on your workloads, whether your cluster needs more memory, a faster network or processors with faster clock speeds. In addition, HPE Insight CMU also greatly simplifies the deployment of Hadoop, with its ability to create a golden Image from a node and then deploy that image to up to 4000 nodes. HPE Insight CMU is able to deploy 800 nodes in 30 minutes.

HPE Insight CMU is highly flexible and customizable, offers both GUI and CLI interfaces supports for Ansible, and can be used to deploy a range of software environments, from simple compute farms to highly customized, application-specific configurations. HPE Insight CMU is available for HPE ProLiant and HPE BladeSystem servers, and is supported on a variety of Linux operating systems, including Red Hat Enterprise Linux, SUSE Linux Enterprise Server, CentOS, and Ubuntu. HPE Insight CMU also includes options for monitoring graphical processing units (GPUs) and for installing GPU drivers and software. Figures 5 and 6 show views of the HPE Insight CMU.

HPE Insight CMU is free for managing up to 32 nodes.
HPE Insight CMU can be configured to support High Availability with an active-passive cluster. For more information, see hpe.com/info/cmu.
Summary

HPE and Hortonworks HDP allow one to derive new business insights from Big Data by providing a platform to store, manage and process data at scale. However, designing and ordering Hadoop clusters can be both complex and time consuming. This white paper provides several Reference Configurations for deploying clusters of varying sizes with the Hortonworks HDP 2.6 on HPE infrastructure and management software. These configurations leverage HPE balanced building blocks of servers, storage and networking, along with integrated management software and bundled support. In addition, this white paper has been created to assist in the rapid design and deployment of Hortonworks HDP software on HPE infrastructure for clusters of various sizes.

Appendix A: Hadoop cluster tuning/optimization

Gen10 Server BIOS Configuration

HPE recommends changing the default BIOS setting to the following using Workload Profile HighPerformanceCompute (HPC) on all ProLiant servers hosting Hadoop to ensure highest performance.

![HPE ProLiant DL380 Gen10 BIOS workload profile](image)

Figure 7: HPE ProLiant DL380 Gen10 BIOS workload profile
The following are the BIOS setting recommended for HPE ProLiant DL380 Gen10.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boot_Mode</td>
<td>Recommended Operating System</td>
<td>UEFI_Mode</td>
</tr>
<tr>
<td>UEFI_Optimized_Boot</td>
<td>Java Development Kit</td>
<td>Enabled</td>
</tr>
<tr>
<td>Workload_Profile</td>
<td>Sets power and performance settings for application workloads.</td>
<td>High_Performance_Compute</td>
</tr>
<tr>
<td>Processor_x2APIC_Support</td>
<td>Enables or disables x2APIC support.</td>
<td>Disabled</td>
</tr>
<tr>
<td>Intel_VT</td>
<td>Controls whether a Virtual Machine Manager (VMM) supporting Virtualization Technology can use hardware capabilities provided by UEFI Intel processors</td>
<td>Disabled</td>
</tr>
<tr>
<td>Intel_VT-d</td>
<td>Enables or disables Intel Virtualization Technology for Directed I/O (VT-d) on a Virtual Machine Manager (VMM).</td>
<td>Disabled</td>
</tr>
<tr>
<td>SR-IOV</td>
<td>Enables or disables the BIOS to allocate more PCI resources to PCIe devices.</td>
<td>Disabled</td>
</tr>
<tr>
<td>Power_Regulator</td>
<td>Sets the power regulator mode.</td>
<td>Static_High_Performance_Mode</td>
</tr>
<tr>
<td>Minimum_Processor_Idle_Power_Core_C-State</td>
<td>Sets the lowest processor idle power state (C-State).</td>
<td>No_C-States</td>
</tr>
<tr>
<td>Minimum_Processor_Idle_Power_Package_C-State</td>
<td>Sets the lowest processor idle power state (C-State).</td>
<td>No_Package_State</td>
</tr>
<tr>
<td>Energy/Performance_Bias</td>
<td>To optimize the processor's performance and power usage.</td>
<td>Maximum_Performance</td>
</tr>
<tr>
<td>Collaborative_Power_Control</td>
<td>Enables or disables collaborative power control for operating systems that support the Processor Clocking Control (PCC) interface.</td>
<td>Disabled</td>
</tr>
<tr>
<td>Intel_Hyper-Threading</td>
<td>Enables or disables the logical processor cores on processors supporting Intel Hyperthreading technology.</td>
<td>Enabled</td>
</tr>
<tr>
<td>Intel_Turbo_Boost_Technology</td>
<td>Enables or disables Intel Turbo Boost Technology to control whether the processor transitions to a higher frequency than the processor's rated speed if the processor has available power and is within temperature.</td>
<td>Enabled</td>
</tr>
<tr>
<td>Energy_Efficient_Turbo</td>
<td>Controls whether the processor uses an energy efficient based policy.</td>
<td>Disabled</td>
</tr>
<tr>
<td>Maximum_Memory_Bus_Frequency</td>
<td>Configures the system to run memory at a lower maximum speed than that supported by the installed processor and DIMM configuration.</td>
<td>Auto</td>
</tr>
<tr>
<td>Channel_Interleaving</td>
<td>Enables or disables a higher level of memory interleaving.</td>
<td>Enabled</td>
</tr>
<tr>
<td>Intel_UPI_Link_Power_Management</td>
<td>To place the Ultra Path Interconnect (UPI) links into a low power state when the links are not being used.</td>
<td>Disabled</td>
</tr>
<tr>
<td>Intel_TXT_support</td>
<td>Enables or disable Intel TXT (Trusted Execution Technology) support for servers with Intel processors.</td>
<td>Disabled</td>
</tr>
<tr>
<td>Embedded_SATA_Configuration</td>
<td>Sets the mode for the embedded SATA controller.</td>
<td>Enable_AHCI_Support</td>
</tr>
<tr>
<td>NUMA_Group_Size_Optimization</td>
<td>The number of logical processors in a NUMA (Non-Uniform Memory Access) node.</td>
<td>Clustered</td>
</tr>
<tr>
<td>Uncore_Frequency_Scaling</td>
<td>Controls the frequency scaling of the processor's internal buses (the uncore)</td>
<td>Maximum</td>
</tr>
</tbody>
</table>
Server tuning

Below are some general guidelines for tuning the server OS and the storage controller for a typical Hadoop proof-of-concept (POC). Please note that these parameters are recommended for YARN workloads which are most prevalent in Hadoop environments. Please note that there is no silver bullet performance tuning. Modifications will be needed for other types of workloads.

- **OS tuning**

  As a general recommendation, update to the latest patch level available to improve stability and optimize performance. The recommended Linux file system is ext4, 64 bit OS:
  - Enable defaults, nodiratime, noatime [/etc/fstab]
  - Do not use logical volume management (LVM)
  - Tune OS block readahead to 8K (/etc/rc/local):
    ```bash
    blockdev --setra 8192 <storage device>
    ```
  - Decrease kernel swappiness to minimum 1:
    ```bash
    Set sysctl vm.swappiness=1 in /etc/sysctl.conf
    ```
  - Tune ulimits for number of open files to a high number:
    ```bash
    Example: in /etc/security/limits.conf:
    soft nofile 65536
    hard nofile 65536
    Set nproc = 65536
    Add it to end of (/etc/security/limits.conf)
    ```
  - Set IO scheduler policy to deadline on all the data drives:
    ```bash
    echo deadline > /sys/block/<device>/queue/scheduler
    ```
  - For persistency across boot, append the following to kernel boot line in /etc/grub.conf:
    ```bash
    elevator=deadline
    ```
  - Configure network bonding on two 10GbE server ports, for 20GbE throughput.
  - Ensure forward and reverse DNS is working properly.
  - Install and configure ntp to ensure clocks on each node are in sync to the management node.
  - Setting tuned profile network-latency for the server. Profile for low latency network tuning it additionally disables transparent hugepages, NUMA balancing and tunes several other network related sysctl parameters:
    ```bash
    tuned-adm profile network-latency
    ```
  - For good performance improvements, disable transparent huge page compaction:
    ```bash
    echo never > /sys/kernel/mm/transparent_hugepage/enabled
    ```
  - Disable SELinux on RHEL 7 by editing /etc/selinux/config and setting SELINUX=disabled

- **HPE Smart Array E208i-a/ P408i-a/ P816i-a**

  - Configure each Hadoop data drive as a separate RAID0 array with stripe size of 1024KB
    ```bash
    ssaci ctrl slot=<slot number> ld <ld number> modify ss=1024
    ```
  - Set power mode be set to maxperformance:
    ```bash
    ssaci ctrl slot=0 modify powermode=maxperformance
    ```
- For data drivers we recommend enabling drive write cache:
  
  ```
  ssacli ctrl slot=0 modify dwc=enable
  ```

- Turn Off “Array Acceleration” / “Caching” for all data drives

  Example:
  
  ```
  ctrl slot=<slot number> ld all modify arrayaccelerator=disable
  ```

  disable arrayaccelerator on all logical drives on 1st ctrlr

  ```
  ctrl slot=<slot number> ld 1 modify arrayaccelerator=enable
  ```

  enable arrayaccelerator on the OS logical drive on 1st ctrlr

- Oracle Java

  ```
  java.net.preferIPv4Stack set to true
  ```

- Patch common security vulnerabilities

  Check Red Hat Enterprise Linux and SUSE security bulletins for more information.

**Appendix B: HPE Pointnext value-added services and support**

In order to help customers jump-start their Big Data solution development, HPE Pointnext offers flexible, value-added services, including Factory Express and Big Data Consulting services which can accommodate and end-to-end customer experience.

**HPE Pointnext Factory Express Services**

Factory-integration services are available for customers seeking a streamlined deployment experience. With the purchase of Factory Express services, your cluster will arrive racked and cabled, with software installed and configured per an agreed upon custom statement of work, for the easiest deployment possible. HPE Factory Express Level 4 Service (HA4S4A1) is the recommended Factory Integration service for Big Data covering hardware and software integration, as well as end-to-end delivery project management. Please engage HPE Pointnext Factory Express for details and quoting assistance. For more information and assistance on Factory Integration services, you can go to:


Or contact:

- AMS: easy.solutions.americas@hpe.com
- APJ: ap.fe-engagement@hpe.com
- EMEA: sol_eng_support@hpe.com

**HPE Pointnext Big Data Consulting – Reference Configurations Implementation Service for Hadoop**

With the HPE Reference Configurations Implementation Service for Hadoop, experienced HPE Big Data consultants install, configure, deploy, and test your Hadoop environment based on the HPE Reference Configurations for Hadoop. HPE will implement a Hadoop design: naming, hardware, networking, software, administration, backup and operating procedures and work with you to configure the environment according to your goals and needs. HPE will also conduct an acceptance test to validate and prove that the system is operating as defined in the Reference Configurations.

**HPE GreenLake Big Data---your complete end-to-end solution**

HPE Pointnext offers a scalable solution that radically simplifies your experience with Hadoop. It takes much of the complexity and cost off your back, so that you can focus purely on deriving intelligence from your Hadoop cluster(s). Offering support for both symmetrical and asymmetrical environments, HPE GreenLake Big Data offers complete end-to-end solution that includes hardware, software, and HPE Pointnext services. HPE Pointnext experts will get you set up and operational, and help you manage and maintain your cluster(s). They will also simplify billing, aligning it with business KPIs. With HPE's unique pricing and billing method, it's much easier to understand your existing Hadoop costs and better predict future costs associated with your solution. HPE GreenLake Big Data covers the whole Hadoop lifecycle. It is composed of the required hardware, software, and HPE Pointnext services to provide a comprehensive, end-to-end solution — including data migration, if needed.
HPE Pointnext Advisory, Transform and Manage - Big Data Consulting Services

HPE Pointnext Big Data Consulting Services cover the spectrum of services to advise, transform, and manage your Hadoop environment, helping you to reshape your IT infrastructure to corral increasing volumes of bytes – from e-mails, social media, and website downloads – and convert them into beneficial information. Our Big Data solutions encompass strategy, design, implementation, protection and compliance. We deliver these solutions in three steps.

1. **Big Data Architecture Strategy and Roadmap:** We’ll define the functionalities and capabilities needed to align your IT with your Big Data initiatives. Through transformation workshops and roadmap services, you’ll learn to capture, consolidate, manage and protect business-aligned information, including structured, semi-structured and unstructured data.

2. **Big Data System Infrastructure:** HPE experts will design and implement a high-performance, integrated platform to support a strategic architecture for Big Data. Choose from design and implementation services, Reference Configurations implementations and integration services. Your flexible, scalable infrastructure will support Big Data variety, consolidation, analysis, share and search on HPE platforms.

3. **Big Data Protection:** Ensure availability, security and compliance of Big Data systems. Our consultants can help you safeguard your data, achieve regulatory compliance and lifecycle protection across your Big Data landscape, as well as improve your backup and continuity measures.


**Hewlett Packard Enterprise Support options**

HPE offers a variety of support levels to meet your needs:

- **HPE Datacenter Care** - HPE Datacenter Care provides a more personalized, customized approach for large, complex environments, with one solution for reactive, proactive, and multi-vendor support needs.

- **HPE Support Plus 24** - For a higher return on your server and storage technology, our combined reactive support service delivers integrated onsite hardware/software support services available 24x7x365, including access to HPE technical resources, 4-hour response onsite hardware support and software updates.

- **HPE Proactive Care** - HPE Proactive Care begins with providing all of the benefits of proactive monitoring and reporting along with rapid reactive care. You also receive enhanced reactive support, through access to HPE's expert reactive support specialists. You can customize your reactive support level by selecting either 6 hour call-to-repair or 24x7 with 4 hour onsite response. You may also choose DMR (Defective Media Retention) option.

- **HPE Proactive Care with the HPE Personalized Support Option** - Adding the Personalized Support Option for HPE Proactive Care is highly recommended. The Personalized Support option builds on the benefits of HPE Proactive Care Service, providing you an assigned Account Support Manager who knows your environment and delivers support planning, regular reviews, and technical and operational advice specific to your environment. These proactive services will be coordinated with Microsoft's proactive services that come with Microsoft® Premier Mission Critical, if applicable.

- **HPE Proactive Select** - And to address your ongoing/changing needs, HPE recommends adding Proactive Select credits to provide tailored support options from a wide menu of services, designed to help you optimize capacity, performance, and management of your environment. These credits may also be used for assistance in implementing updates for the solution. As your needs change over time you flexibly choose the specific services best suited to address your current IT challenges.

- **Other offerings** - In addition, Hewlett Packard Enterprise highly recommends HPE Education Services (for customer training and education) and additional Pointnext, as well as in-depth installation or implementation services as may be needed.
Resources and additional links

Hortonworks HDP, https://www.hortonworks.com/

Hortonworks HDP 2.6, https://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.6.4/index.html

HPE Solutions for Apache Hadoop, hpe.com/info/hadoop

HPE Insight Cluster Management Utility (CMU), hpe.com/info/cmu

HPE FlexFabric 5900 switch series, hpe.com/networking/5900


HPE ProLiant servers, hpe.com/info/proliant

HPE Networking, hpe.com/networking

HPE Services, hpe.com/services

Red Hat, redhat.com


To help us improve our documents, please provide feedback at hpe.com/contact/feedback.

About Hortonworks

Hortonworks is a leading innovator in the industry, creating, distributing and supporting enterprise-ready open data platforms and modern data applications. Our mission is to manage the world's data. We have a single-minded focus on driving innovation in open source communities such as Apache Hadoop, NiFi, and Spark. We along with our 1600+ partners provide the expertise, training and services that allow our customers to unlock transformational value for their organizations across any line of business. Our connected data platforms powers modern data applications that deliver actionable intelligence from all data: data-in-motion and data-at-rest.