Technical white paper

HP Reference Architecture for Hortonworks Data Platform 2.1 on ProLiant DL Servers – SUSE Linux Enterprise Server

HP Converged Infrastructure with Hortonworks Data Platform 2.1 for Apache Hadoop

Table of contents

Executive summary

HP and Apache Hadoop allow you to derive new business insights from Big Data by providing a platform to store, manage and process data at scale. However, Apache Hadoop is complex to deploy, configure, manage and monitor. This white paper provides several performance optimized configurations for deploying Hortonworks Data Platform (HDP) clusters of varying sizes on HP infrastructure that provide a significant reduction in complexity and increase in value and performance.

The configurations are based on Hortonworks distribution of Hadoop (HDP), specifically HDP 2.1 and the HP ProLiant DL Gen8 server platform. The configurations reflected in this document have been jointly designed and developed by HP, SUSE and Hortonworks to provide optimum computational performance for Hadoop and are also compatible with other HDP 2.x releases.

HP Big Data solutions provide best-in-class performance and availability, with integrated software, services, infrastructure, and management – all delivered as one proven configuration as described at hp.com/go/hadoop. In addition to the benefits described above, the Reference Architecture in this white paper also includes the following features that are unique to HP:

- Servers: HP ProLiant DL360p Gen8 and DL380p Gen8 include:
  - The HP Smart Array P420i controller which provides increased[1] I/O throughput performance resulting in a significant performance increase for I/O bound Hadoop workloads (a common use case) and the flexibility for the customer to choose the desired amount of resilience in the Hadoop Cluster with either JBOD or various RAID configurations.
  - Two sockets with 10 core processors, using Intel® Xeon® E5-2600 v2 product family, provide the high performance required for faster completion of CPU bound Hadoop workloads. For Management and Head nodes, the same CPU family with 8 core processors are used.
  - The HP iLO Management Engine on the servers contains HP Integrated Lights-Out 4 (iLO 4) and features a complete set of embedded management features for HP Power/Cooling, Agentless Management, Active Health System, and Intelligent Provisioning which reduces node and cluster level administration costs for Hadoop.
- Cluster Management: HP Insight Cluster Management Utility (CMU) provides push-button scale out and provisioning with industry leading provisioning performance, reducing deployments from days to hours. HP CMU provides real-time and historical infrastructure and Hadoop monitoring with 3D visualizations allowing customers to easily characterize Hadoop workloads and cluster performance. This allows customers to further reduce complexity and improve system optimization leading to improved performance and reduced cost. In addition, HP Insight Management and HP Service Pack for ProLiant, allow for easy management of firmware and servers.
- Networking: The HP 5900AF-48XGT-4QSFP+ 10GbE Top of Rack switch has 48 RJ-45 1/10GbE ports and 4 QSFP+ 40GbE ports. It provides IRF bonding and sFlow for simplified management, monitoring and resiliency of Hadoop network. The 512MB flash, 2GB SDRAM and packet buffer size of 9MB provide excellent performance of 952 million pps throughput and switching capacity of 1280 Gb/s with very low 10Gb/s latency of less than 1.5 µs (64-byte packets).

  HP FlexFabric 5930-32QSFP+ 40GbE Aggregation switch provides IRF Bonding and sFlow which simplifies the management, monitoring and resiliency of the customer's Hadoop network. The 512MB flash, 4GB SDRAM memory and packet buffer size of 12.2MB provide excellent performance of 1905 million pps throughput and routing/switching capacity of 2560Gb/s with very low 10Gb/s latency of less than 1 µs (64-byte packets). The switch seamlessly handles burst scenarios such as shuffle, sort and block replication which are common in Hadoop clusters.
- Analytics database, the HP Vertica connectors for Hadoop allow seamless integration of both structured and unstructured data providing end-to-end analytics thereby simplifying bi-directional data movement for Hadoop and reducing customer integration costs. Vertica is a leading real-time, scalable, analytical platform for structured data.

All of these features reflect HP's balanced building blocks of servers, storage and networking, along with integrated management software.

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 or purchase an HP and Hortonworks 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 Solution components.

Document purpose: The purpose of this document is to describe a reference architecture, 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 May-Aug 2014.

[1] Compared to the previous generation of Smart Array controllers

Introduction

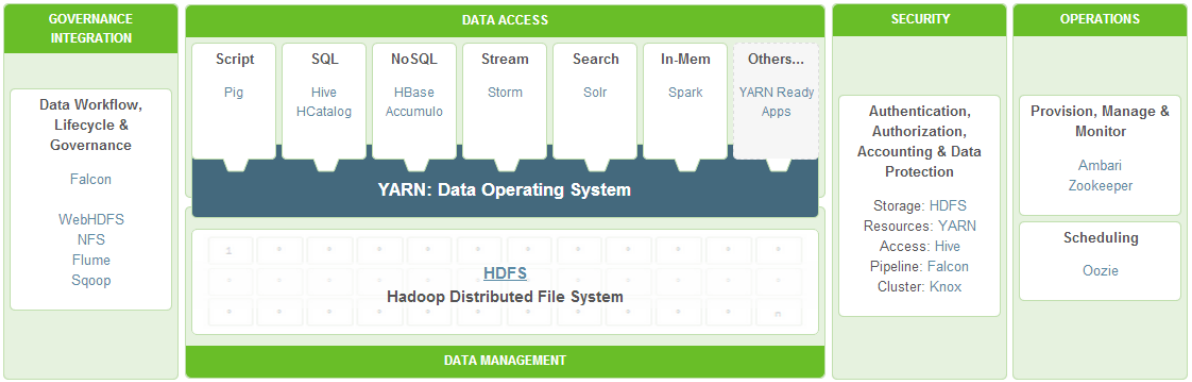
This white paper has been created to assist in the rapid design and deployment of Hortonworks Data Platform software on HP infrastructure for clusters of various sizes. It is also intended to identify the software and hardware components required in a solution to simplify the procurement process. The recommended HP Software, HP ProLiant servers, and HP Networking switches and their respective configurations have been carefully tested with a variety of I/O, CPU, network, and memory bound workloads. The configurations included provide the best value for optimum MapReduce, YARN, Hive, HBase and Solr computational performance, resulting in a significant performance increase at an optimum cost.

Hortonworks Data Platform delivers enterprise Hadoop

Apache Hadoop is an open-source project administered by the Apache Software Foundation. Hadoop's contributors work for some of the world's biggest technology companies. That diverse, motivated community has produced a genuinely innovative platform for consolidating, combining and understanding large-scale data in order to better comprehend the data deluge. Enterprises today collect and generate more data than ever before. Relational and data warehouse products excel at OLAP and OLTP workloads over structured data. Hadoop, however, was designed to solve a different problem: the fast, reliable analysis of both structured data and complex data. As a result, many enterprises deploy Hadoop alongside their existing IT systems, which allows them to combine old data and new data sets in powerful new ways.

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 across both Linux® and Microsoft® Windows®.

Figure 1. Hortonworks Data Platform: A full Enterprise Hadoop Data Platform



Source: hortonworks.com/hdp

Hortonworks Data Platform enables Enterprise Hadoop: the full suite of essential Hadoop capabilities that are required by the enterprise and that serve as the functional definition of any data platform technology. This comprehensive set of capabilities is aligned to the following functional areas: Data Management, Data Access, Data Governance and Integration, Security, and Operations.

Key highlights for Hortonworks Data Platform 2.1

- Interactive query with Apache Hive and Apache Tez
- Stream processing with Apache Storm
- Search with Apache Solr
- Operations with Apache Ambari
- Data Governance with Apache Falcon
- Perimeter Security with Apache Knox

For detailed information on Hortonworks Data Platform, please see hortonworks.com/hdp

Hortonworks platform

The platform functions within Hortonworks Data Platform are provided by two key groups of services, namely the Management and Worker Services. Management Services manage the cluster and coordinate the jobs whereas Worker Services are responsible for the actual execution of work on the individual scale out nodes. Tables 1 and 2 below specify which services are management services and which services are worker services. Each table contains two columns. The first column is the description of the service and the second column specifies the number of nodes the service can be distributed to. The Reference Architectures (RAs) we provide in this document will map the Management and Worker services onto HP infrastructure for clusters of varying sizes. The RAs factor in the scalability requirements for each service.

Management services

Table 1. HDP Base Management Services

Service	Maximum Distribution across Nodes
Ambari	1
HueServer	1
ResourceManager	2
JobHistoryServer	1
HBaseMaster	Varies
NameNode	2
Oozie	1
ZooKeeper	Varies

Worker services

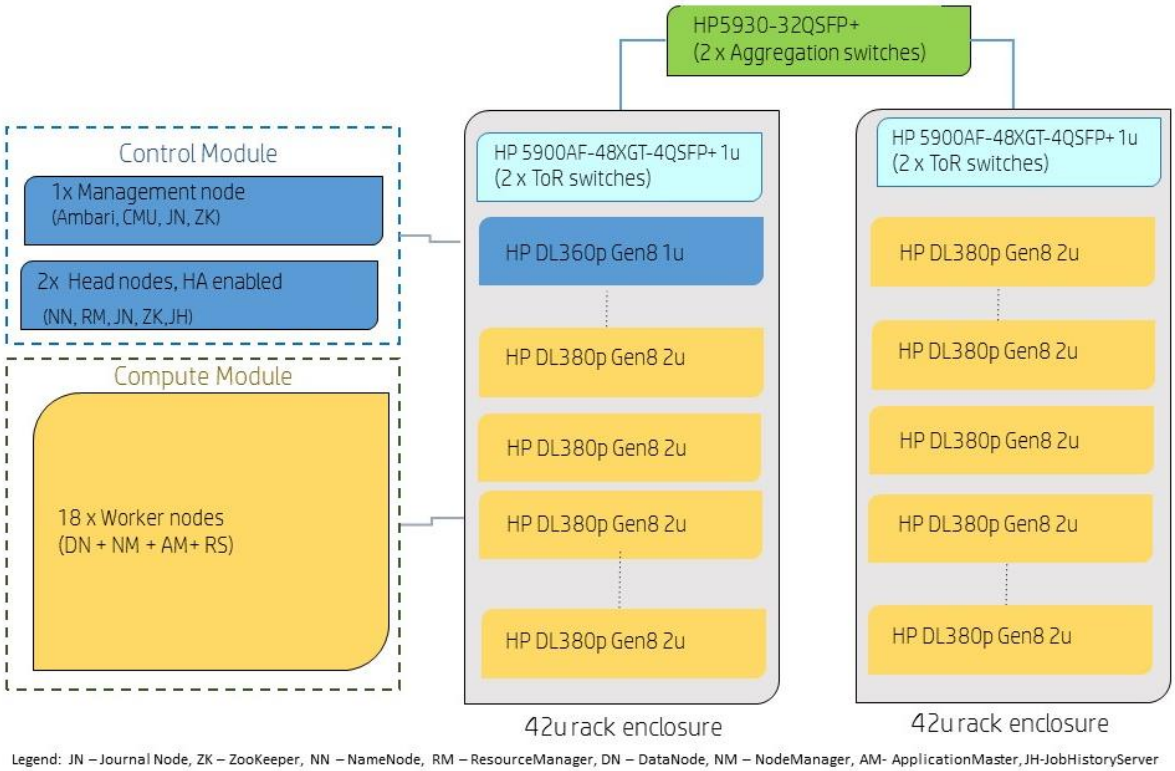
Table 2. HDP Base Enterprise Worker Services

Service	Maximum Distribution across Nodes
DataNode	Most or all nodes
NodeManager	Most or all nodes
ApplicationMaster	One for each job
HBaseRegionServer	Varies

A full Hadoop cluster will have component-specific services on each worker node and management node.

Solution components

Figure 2. Basic conceptual diagram



Each of the components is discussed at length below. For full BOM listing on products selected, please refer to Appendix C.

High-availability considerations

The following are some of the high-availability features considered in this reference architecture configuration:

- Hadoop NameNode HA – The configurations in this white paper utilize quorum-based journaling high-availability features in HDP 2.1. For this feature, servers should have similar I/O subsystems and server profiles so that each NameNode server could 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.
- OS availability and reliability – For the reliability of the server, the OS disk is configured in a RAID 1+0 configuration thus preventing failure of the system from OS hard disk failures.
- Network reliability – The reference architecture configuration uses two HP 5900AF-48XGT switches for redundancy, resiliency and scalability through using Intelligent Resilient Framework (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).

Pre-deployment considerations / system selection

There are a number of key factors you should consider prior to designing and deploying a Hadoop Cluster. The following subsections articulate the design decisions in creating the baseline configurations for the reference architectures. The rationale provided includes the necessary information for you to take the configurations and modify them to suit a particular custom scenario.

Table 3. Pre-deployment considerations

Functional Component	Value
Operating system	Improves Availability and Reliability
Computation	Ability to balance Price with Performance
Memory	Ability to balance Price with Capacity and Performance
Storage	Ability to balance Price with Capacity and Performance
Network	Ability to balance Price with Performance

Operating system

Hortonworks HDP 2.1 supports the following 64-bit operating systems:

- For SUSE systems, Hortonworks provides 64-bit packages for SUSE Linux Enterprise Server 11 Service Pack 3, the current release as of this reference architecture publication, or later is required.
- For Red Hat® systems, Hortonworks provides 64-bit packages for Red Hat Enterprise Linux (RHEL) 5 and Red Hat Enterprise Linux 6. Compatible versions of RHEL 6 are RHEL 6.2 and RHEL 6.4. Hortonworks recommends using update 7 or later for Red Hat Enterprise Linux 5.
- For Ubuntu systems, Hortonworks provides 64-bit packages for Precise (12.04) LTS.

Full details on supported OS, databases and JDK versions are available at the Hortonworks site.

Recommendation

HP recommends using a 64-bit operating system to avoid constraining the amount of memory that can be used on worker nodes. The 64-bit version of SUSE Linux Enterprise Server 11 SP3 is recommended due to its superior filesystem, performance and scalability characteristics, plus the comprehensive, certified support of the Hortonworks and HP supplied software used in Big Data clusters. The Reference Architectures listed in this document were tested with 64-bit SUSE Linux Enterprise Server (SLES 11 SP3). For Ambari and its supporting services database, HP best practice is to use a single database such as PostgreSQL.

Computation

Unlike MR1, where the processing or computational capacity of a Hadoop cluster is determined by the aggregate number of Resource Containers available across all the worker nodes, under YARN/MR2, the notion of slots has been discarded and resources are now configured in terms of amounts of memory and CPU (virtual cores).There is no distinction between resources available for map, and resources available for reduces – all MR2 resources are available for both in the form of containers. Employing Hyper-Threading increases your effective core count, potentially allowing ResourceManager to assign more cores as needed.

Resource request tasks that will use multiple threads can request more than one core with the mapreduce.map.cpu.vcores and mapreduce.reduce.cpu.vcores properties. Hortonworks supports the use of the Fair and FIFO schedulers in MR2. The minMaps, maxMaps, minReduces, and maxReduces queue properties have been replaced with a minResources property and a maxResources property. Instead of taking a number of slots, these properties take a value like “1024MB, 3 vcores”.

Recommendation

When computation performance is of primary concern, HP recommends higher CPU powered DL380p servers for worker nodes with 256 GB RAM. HDP 2.1 components, such as HBase, HDFS Caching, Storm and Solr, benefit from large amounts of memory.

Memory
Use of Error Correcting Memory (ECC) is a practical requirement for Apache Hadoop and is standard on all HP ProLiant servers. Memory requirements differ between the management nodes and the worker nodes. The management nodes typically run one or more memory intensive management processes and therefore have higher memory requirements. Worker nodes need sufficient memory to manage the NodeManager and Container processes. If you have a memory bound YARN job we recommend that you increase the amount of memory on all the worker nodes. In addition, a high memory cluster can also be used for Spark, HBase or interactive Hive, which could be memory intensive.

Storage
Fundamentally, Hadoop is designed to achieve performance and scalability by moving the compute activity to the data. It does this by distributing the Hadoop job to worker nodes close to their data, ideally running the tasks against data on local disks.

Best practice
HP recommends choosing LFF drives over SFF drives (same angular speed) as the LFF drives have better performance than the SFF drives due to faster tangential speed that leads to higher disk I/O. Given the architecture of Hadoop, the data storage requirements for the worker nodes are best met by direct attached storage (DAS) in a Just a Bunch of Disks (JBOD) configuration and not as DAS with RAID or Network Attached Storage (NAS). As a consequence they fit into a Hadoop environment nicely.

There are several factors to consider and balance when determining the number of disks a Hadoop worker node requires.

- Storage capacity – The number of disks and their corresponding storage capacity determines the total amount of the HDFS 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 16 TB of block data fails, 16 TB 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 – The management nodes are configured differently from the worker nodes because the management processes are generally not redundant and as scalable as the worker processes. For management nodes, storage reliability is therefore 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. If performance and reliability are important, we recommend SAS MDL disks; otherwise, we recommend SATA MDL disks. Specific details around disk and RAID configurations will be provided in the Server selection section.

Network
Configuring a single Top of Rack (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 HP 5900AF-48XGT switch supports this through HP's Intelligent Resilient Framework (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. Multi-Rack Hadoop clusters, that are not using IRF bonding for inter-rack traffic, will benefit from having ToR switches connected by 40 GbE uplinks to core aggregation switches. Large Hadoop clusters introduce multiple issues that are not typically present in small to medium sized clusters. To understand the reasons for this, it is helpful to review the network activity associated with running Hadoop jobs and with exception events such as server failure.

A more detailed white paper for Hadoop Networking best practices is available. See the Resource tab at hp.com/go/hadoop.

Best practice
HP recommends ToR switches with packet buffering and connected by 40 GbE uplinks to core aggregation switches for large clusters. For MapReduce jobs, during the shuffle phase, the intermediate data has to be pulled by the reduce tasks from mapper output files across the cluster. While network load can be reduced if partitioners and combiners are used, it is possible that the shuffle phase will place the core and ToR switches under a large traffic load.

Each reduce task can concurrently request data from a default of five mapper output files. Thus, there is the possibility that servers will deliver more data than their network connections can handle which will result in dropped packets and can lead to a collapse in traffic throughput. ToR switches with packet buffering protect against this event.

Server selection

This section will provide topologies for the deployment of management and worker nodes for single and multi-rack clusters. Depending on the size of the cluster, a Hadoop deployment consists of one or more nodes running management services and a quantity of worker nodes. We have designed this reference architecture so that regardless of the size of the cluster, the server used for the management nodes and the server used for the worker nodes remains consistent. This section specifies which server to use and the rationale behind it.

Management nodes

Management services are not distributed redundantly across as many nodes as the services that run on the worker nodes and therefore benefit from a server that contains redundant fans and power supplies. In addition, Management nodes require storage only for local management services and the OS, unlike worker nodes that store data in HDFS, and so do not require large amounts of internal storage. However, as these local management services are not replicated across multiple servers, an array controller supporting a variety of RAID schemes and SAS direct attached storage is required. In addition, the management services are memory and CPU intensive; therefore, a server capable of supporting a large amount of memory is also required.

Best practice
HP recommends that all Management nodes in a cluster be deployed with either identical or highly similar configurations. The configurations reflected in this white paper are also cognizant of the high availability feature in Hortonworks HDP 2.1. For this feature, servers should have similar I/O subsystems and server profiles so that each management server could potentially take the role of another. Similar configurations will also 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.

This section contains 4 subsections:

- Server platform
- Management node
- ResourceManager server
- NameNode server

Server platform: HP ProLiant DL360p Gen8
The HP ProLiant DL360p Gen8 (1U) is an excellent choice as the server platform for the management nodes and head nodes.

Figure 3. HP ProLiant DL360p Gen8 Server

Processor configuration
The configuration features two sockets with 8 core processors and the Intel E5-2600 v2 product family, which provide 16 physical cores and 32 Hyper-Threaded cores per server. We recommend that Hyper-Threading be turned on.

The reference architecture was tested using the Intel Xeon E5-2650 v2 processors for the management servers with the ResourceManager, NameNode and Ambari services. The configurations for these servers are designed to be able to handle an increasing load as your Hadoop cluster grows. Choosing a powerful processor such as this to begin with sets the stage for managing growth seamlessly. An alternative CPU option is a 10-core E5-2660 v2 for head nodes to support large numbers of services/processes.

Drive configuration
Apache Hadoop does not provide software redundancy for the management servers of a Hadoop cluster the same way it does for the workers and thus RAID is appropriate. The Smart Array P420i Controller is specified to drive four 900GB 2.5" SAS disks on the Management node, ResourceManager and NameNode servers. Hot pluggable drives are specified so that drives can be replaced without restarting the server. Due to this design, one should configure the P420i controller to apply the following RAID schemes:

- Management node: 4 disks with RAID 1+0 for OS and PostgreSQL database, and management stack software.
- ResourceManager and NameNode Servers: 4 disks with RAID 1+0 for OS and Hadoop software.

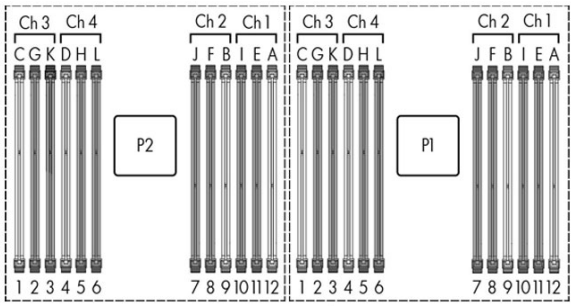
Best practice
For a performance oriented solution HP recommends SAS drives as they offer a significant read and write performance enhancement over SATA disks. The Smart Array P420i controller provides two port connectors per controller with each containing 4 SAS links. The drive cage for the DL360p contains 8 disks slots and thus each disk slot has a dedicated SAS link which ensures the server provides the maximum throughput that each drive can give you.

Memory configuration
Servers running management services such as the HBaseMaster, ResourceManager, NameNode and Ambari should have sufficient memory as they can be memory intensive. When configuring memory, one should always attempt to populate all the memory channels available to ensure optimum performance. The dual Intel Xeon E5-2600 v2 series processors in the HP ProLiant DL360p Gen8 have 4 memory channels per processor which equates to 8 channels per server. The configurations for the management and head node servers were tested with 128GB of RAM, which equated to eight 16GB DIMMs.

Best practice
Configure all the memory channels according to recommended guidelines to ensure optimal use of memory bandwidth.

For example, on a two socket processor with eight memory channels available per server one would typically populate channels with 16GB DIMMs resulting in a configuration of 128GB in sequential alphabetical order balanced between the two processors: P1-A, P2-A, P1-B, P2-B, P1-C, P2-C, P1-D, P2-D, as shown in Figure 4.

Figure 4. Memory configuration



Network configuration
The HP ProLiant DL360p Gen8 is designed for network connectivity to be provided via a FlexibleLOM. The FlexibleLOM can be ordered as a 2 x 10GbE NIC configuration. This Reference Architecture was tested using the 2 x 10GbE NIC configuration (as specified in the server configuration below).

Best practice
For each management server HP recommends bonding and cabling two 10GbE NICs to create a single bonded pair which will provide 20GbE of throughput as well as a measure of NIC redundancy. In the reference architecture configurations later in the document you will notice that we use two IRF Bonded switches. In order to ensure the best level of redundancy we recommend cabling NIC 1 to Switch 1 and NIC 2 to Switch 2.

Management node components

The Management node hosts the applications that submit jobs to the Hadoop Cluster. We recommend that you install with the following software components:

Table 4. Management node Basic Software components

Software	Description
SUSE Linux Enterprise Server 11 SP3	Recommended Operating System
HP Insight CMU 7.1	Infrastructure Deployment, Management, and Monitoring
Oracle JDK 1.7.0_45	Java Development Kit
PostgreSQL 8.4	Database Server for Ambari
Ambari 1.6.x	Hortonworks Hadoop Cluster Management Software
Hue Server	Web Interface for Applications
NameNode HA	NameNode HA (Journal Node)
Apache Pig and Apache Hive	Analytical interfaces to the Hadoop Cluster
HiveServer2	Hue application to run queries on Hive with authentication
ZooKeeper	Cluster coordination service

Please see the following link for the Ambari Installation guide http://docs.hortonworks.com/HDPDocuments/Ambari-1.6.1.0/bk_using_Ambari_book/content/ambari-chap2.1.3.html.

The Management node and head nodes, as tested in the Reference Architecture, contain the following base configuration:

- 2 x Eight-Core Intel Xeon E5-2650 v2 Processors
- Smart Array P420i Controller with 512MB FBWC
- 3.6TB – 4 x 900GB SFF SAS 10K RPM disks
- 128 GB DDR3 Memory – 8 x 16GB 2Rx4 PC3-14900R-13
- 10GbE 2P NIC 561FLR-T card

A BOM for the Management node is available in Appendix C – Table 12.

Head node 1 – ResourceManager server

The ResourceManager server contains the following software components. Please see the following link for more information on installing and configuring the ResourceManager and NameNode HA: http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.5/bk_system-admin-guide/content/ch_hadoop-ha-rm.html.

Table 5. ResourceManager Server Base Software components

Software	Description
SUSE Linux Enterprise Server 11 SP3	Recommended Operating System
Oracle JDK 1.7.0_45	Java Development Kit
ResourceManager	YARN ResourceManager
NameNode HA	NameNode HA (Failover Controller, Journal Node, NameNode Standby)
Oozie	Oozie Workflow scheduler service
HBaseMaster	The HBase Master for the Hadoop Cluster (Only if running HBase)
ZooKeeper	Cluster coordination service
Flume	Flume

Head node 2 – NameNode server

The NameNode server contains the following software components. Please see the following link for more information on installing and configuring the NameNode and ResourceManager HA.
http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.5/bk_system-admin-guide/content/ch_hadoop-ha-1.html

Table 6. NameNode Server Base Software Components

Software	Description
SUSE Linux Enterprise Server 11 SP3	Recommended Operating System
Oracle JDK 1.7.0_45	Java Development Kit
NameNode	The NameNode for the Hadoop Cluster (NameNode Active, Journal node, Failover Controller)
JobHistoryServer	Job History for ResourceManager
ResourceManager HA	Failover, Passive mode
Flume	Flume agent (if required)
HBMaster	HBase Master (Only if running HBase)
ZooKeeper	Cluster coordination service

Worker nodes

The worker nodes run the NodeManager and YARN container processes and thus storage capacity and performance are important factors.

Server platform: HP ProLiant DL380p Gen8

The HP ProLiant DL380p Gen8 (2U) is an excellent choice as the server platform for the worker nodes. For ease of management we recommend a homogenous server infrastructure for your worker nodes.

Figure 5. HP ProLiant DL380p Gen8 Server

Processor selection

The configuration features two processors from the Intel Xeon E5-2600 v2 family. The Base configuration provides 16 physical or 32 Hyper-Threaded cores per server. For increased performance applications 10 core processors can be selected to provide 20 physical or 40 Hyper-Threaded cores per server. Hadoop manages the amount of work each server is able to undertake via the ResourceManager configured for that server. The more cores available to the server, the better for ResourceManager Utilization (see the Computation section for more detail). We recommend that Hyper-Threading be turned on. For this RA, we had picked 2 x E5-2670 v2 (10 cores/2.5GHz) CPUs.

Memory selection

Servers running the worker node processes should have sufficient memory for either HBase or for the amount of MapReduce Slots configured on the server. The dual Intel Xeon E5-2600 v2 series processors in the HP ProLiant DL380p Gen8 have 4 memory channels per processor which equates to 8 channels per server. When configuring memory, one should always attempt to populate all the memory channels available to ensure optimum performance.

With the advent of YARN in HDP 2.1, the memory requirement has gone up significantly to support a new generation of Hadoop applications. A base configuration of 128 GB is recommended, and for certain high memory capacity applications 256 GB is recommended. For this RA, we had picked 128 GB memory (8 x 16GB 2Rx4 PC3-14900R-13).

Best practice

To ensure optimal memory performance and bandwidth, HP recommends using 16GB DIMMs to populate each of the 4 memory channels on the processor which will provide an aggregate of 128GB of RAM. We used 16GB DIMMs (which gave us an aggregate of 128GB of RAM per server) in our testing. For 256GB capacity, we recommend adding an additional 8 x 16GB DIMMs, one in each memory channel. For any applications requiring more than 256GB capacity, we recommend going with 32GB DIMMs, not to populate the third slot on the memory channels to maintain full memory channel speed.

Drive configuration

Redundancy is built into the Apache Hadoop architecture and thus there is no need for RAID schemes to improve redundancy on the worker nodes as it is all coordinated and managed by Hadoop. Drives should use a Just a Bunch of Disks (JBOD) configuration, which can be achieved with the HP Smart Array P420i controller by configuring each individual disk as a separate RAID 0 volume. Additionally array acceleration features on the P420i should be turned off for the RAID 0 data volumes. The first two positions on the drive cage allow the OS drive to be placed in RAID1.

The HP Smart Array P420i controller provides two port connectors per controller with each containing 4 SAS links. The drive cage picked for the DL380p supports 12 Large Form Factor (LFF) disks. A total of 8 SAS links ensures the server provides the maximum throughput that each drive can give. For a performance oriented solution, we recommend SAS drives as they offer a significant read and write throughput performance enhancement over SATA disks. For this RA, we had picked 2TB LFF SATA drives.

Best practice

For OS drives, HP recommends using two 2TB SATA MDL LFF disks with the HP Smart Array P420i controller configured as one RAID1 mirrored 500GB logical drive for the OS, and RAID0 for a 2TB logical drive for HDFS on the remaining capacity. The other 10 disks are each configured for HDFS as 2TB in RAID0, for a total of 11 logical drives for Hadoop data. Protecting the OS provides an additional measure of redundancy on the worker nodes.

DataNode settings
By default, the failure of a single dfs.data.dir or dfs.datanode.data.dir will cause the HDFS DataNode process to shut down, which results in the NameNode scheduling additional replicas for each block that is present on the DataNode. This causes needless replications of blocks that reside on disks that have not failed. To prevent this, you can configure DataNodes to tolerate the failure of dfs.data.dir or dfs.datanode.data.dir directories; use the dfs.datanode.failed.volumes.tolerated parameter in hdfs-site.xml. For example, if the value for this parameter is 3, the DataNode will only shut down after four or more data directories have failed. This value is respected on DataNode startup; in this example the DataNode will start up as long as no more than three directories have failed.

Note
For configuring YARN, update the default values of the following attributes with ones that reflect the cores and memory available on a worker node.

```
yarn.nodemanager.resource.cpu-vcores
yarn.nodemanager.resource.memory-mb
```

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.

```
mapreduce.map.cpu.vcores
mapreduce.map.memory.mb
mapreduce.reduce.cpu.vcores
mapreduce.reduce.memory.mb
```

Similarly, specify the appropriate size for map and reduce task heap sizes using the following attributes:

```
mapreduce.map.java.opts.max.heap
mapreduce.reduce.java.opts.max.heap
```

Network configuration
For 10GbE networks, we recommend that the two 10GbE NICs be bonded to improve throughput performance to 20GbE/s and thereby improve performance. In addition, in the reference architecture configurations later on in the document you will notice that we use two IRF Bonded switches. In order to ensure the best level of redundancy we recommend cabling NIC 1 to Switch 1 and NIC 2 to Switch 2.

Worker node components

The worker node contains the following software. Please see the following link for more information on installing and configuring the NodeManager (or HBaseRegionServer) and DataNode manually: For adding worker nodes: http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.5/bk_system-admin-guide/content/admin_add-nodes-2.html. For adding HBase RegionServer: http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.5/bk_system-admin-guide/content/admin_add-nodes-3.html

Table 7. Worker Node Base Software components

Software	Description
SUSE Linux Enterprise Server 11 Service Pack 3	Recommended Operating System
Oracle JDK 1.7.0_45	Java Development Kit
NodeManager	The NodeManager process for MR2/YARN
DataNode	The DataNode process for HDFS
HBaseRegionServer	*The HBaseRegionServer for HBase (Only if running HBase)*

The ProLiant DL380p Gen8 (2U) as configured for the Reference Architecture as a worker node has the following configuration:

- Dual 10-Core Intel Xeon E5-2670 v2 Processors with Hyper-Threading
- Twelve 2TB 3.5" 7.2K LFF SATA MDL (22 TB for Data)
- 128 GB DDR3 Memory (8 x HP 16GB), 4 channels per socket
- 1 x 10GbE 2 Port NIC FlexibleLOM (Bonded)
- 1 x Smart Array P420i Controller with 512MB FBWC

Note
Customers also have the option of purchasing a second power supply for additional power redundancy. This is especially appropriate for single rack clusters where the loss of a node represents a noticeable percentage of the cluster.

The BOM for the Worker node is provided in Appendix C Table 13.

Note
HP iLO Management Engine – Easy to use Integrated Manage and Support

HP iLO Management Engine is a complete set of embedded features, standard on all ProLiant Gen8 servers. Includes HP iLO, HP Agentless Management, HP Active Health System, HP Intelligent Provisioning, and HP Embedded Remote Support.

HP Insight Online with HP Insight Remote Support provides 24x7 remote monitoring and anywhere, anytime personalized access to your IT and support status.

HP Smart Update, including HP Smart Update Manager (HP SUM), HP Service Pack for ProLiant (SPP) and other products, reduces deployment time and update complexity by systematically and securely updating server infrastructure in the data center; in most cases the downtime is limited to a single reboot.

Switch selection

Hadoop clusters contain two types of switches, namely Top of Rack (ToR) switches and Aggregation switches. Top of Rack switches route the traffic between the nodes in each rack and Aggregation switches route the traffic between the racks.

Top of Rack (ToR) switches
The HP 5900AF-48XGT-4QSFP+, 10GbE, is an ideal ToR switch with forty eight 10GbE ports and four 40GbE uplinks providing resiliency, high availability and scalability support. In addition this model comes with support for CAT6 cables (copper wires) and Software Defined Networking (SDN). A dedicated management switch for iLO traffic is not required as the ProLiant DL360p Gen8 and DL380p Gen8 are able to share iLO traffic over NIC 1. The volume of iLO traffic is minimal and does not degrade performance over that port. For more information on the 5900AF-48XGT-4QSFP+, 10GbE switch, please see hp.com/networking/5900

The BOM for the HP 5900AF-48XGT switch is provided in Appendix C Table 14.

Customers who would like to separate iLO and PXE traffic from the data/Hadoop network traffic can add a 1GbE network switch for example HP 5900AF-48G-4XG-2QSFP+ Switch (JG510A) and order HP Ethernet 10Gb 2-port 561T Adapter NIC and HP Ethernet 1Gb 4-port 331FLR FIO Adapter FlexLOM NIC instead of HP Ethernet 10Gb 2P 561FLR-T FIO Adptr FlexLOM NIC so that the iLO and PXE network traffic is directed to a 1GbE network switch. The BOMs for the 1GbE switch and network cards are provided in Appendix C Table 16 and 17.

Figure 6. HP 5900AF-48XGT-4QSFP+ Aggregation switch

Aggregation switches

The HP FlexFabric 5930-32QSFP+, 40GbE, switch is an ideal aggregation switch as it is well suited to handle very large volumes of inter-rack traffic such as can occur during shuffle and sort operations, or large scale block replication to recreate a failed node. The switch has better connectivity with 32 40GbE ports, supporting up to 104 ports of 10GbE via breakout cables and six 40GbE uplink ports, aggregation switch redundancy and high availability (HA) support with IRF bonding ports. SDN ready with OpenFlow 1.3 and overlay networks with VXLAN and NVGRE support. For more information on the HP 5930-32QSFP+ please see hp.com/networking/5930

The BOM for the HP 5930-32QSFP+ switch is provided in Appendix C Table 15.

Figure 7. HP 5930-32QSFP+ 40GbE Aggregation switch

HP Insight Cluster Management Utility

HP Insight Cluster Management Utility (CMU) is an efficient and robust hyper-scale cluster lifecycle management framework and suite of tools for large Linux clusters such as those found in High Performance Computing (HPC) and Big Data environments. 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. HP Insight CMU makes the management of a cluster more user friendly, efficient, and error free than if it were being managed by scripts, or on a node-by-node basis. HP Insight CMU offers full support for iLO 2, iLO 3, iLO 4 and LO100i adapters on all ProLiant servers in the cluster.

Best practice

HP recommends using HP Insight CMU for all Hadoop clusters. HP 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. 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, 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. Insight CMU is able to deploy 800 nodes in 30 minutes.

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

For more information, please see hp.com/go/cmu.

For the CMU BOM please see Appendix C Table 19.

Figure 8. HP Insight CMU Interface – real-time view

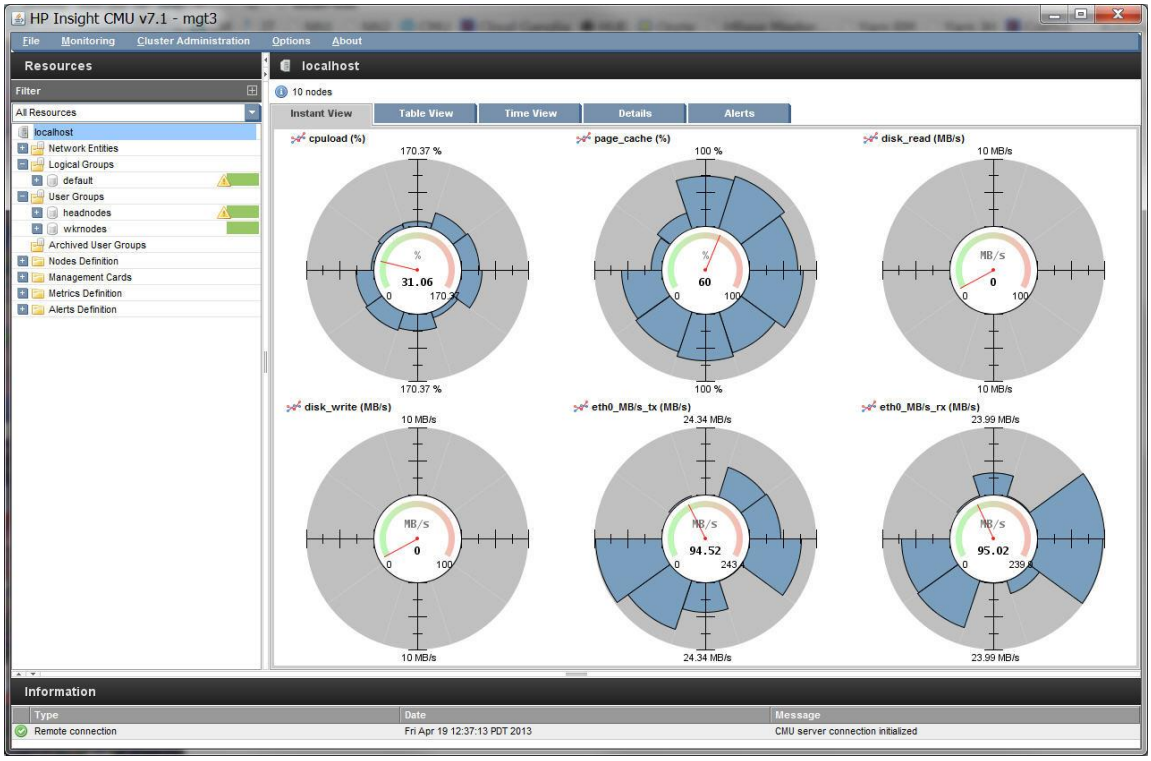


Figure 9. HP Insight CMU Interface – real-time view – BIOS settings

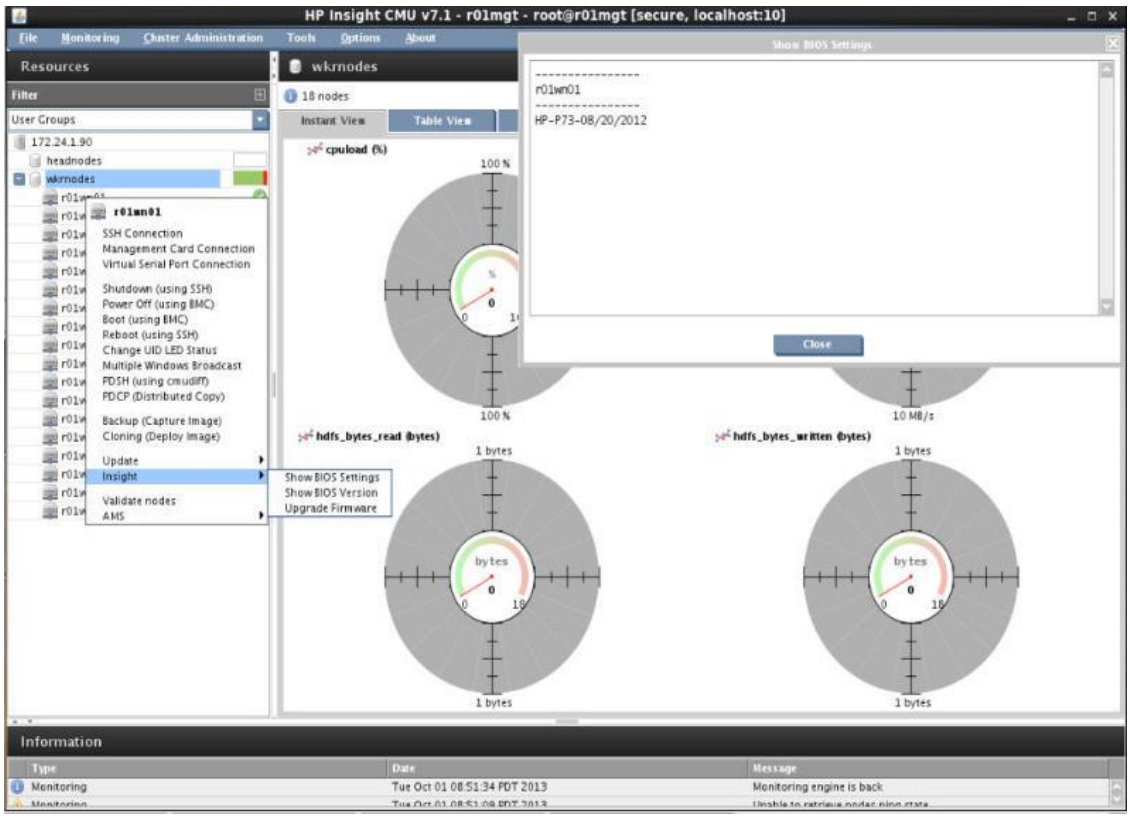
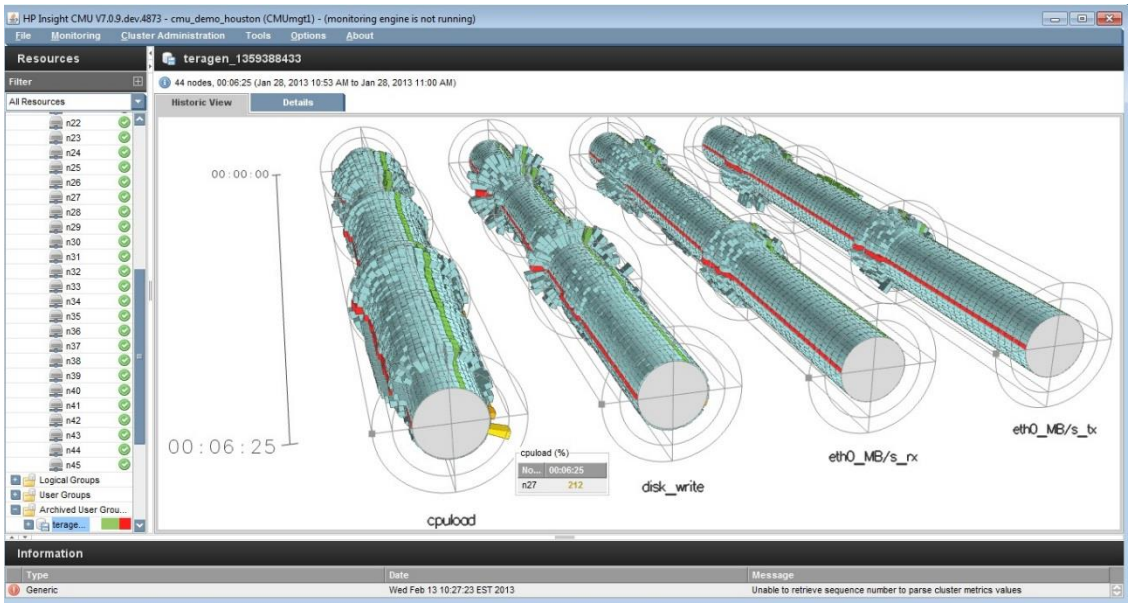


Figure 10. HP Insight CMU Interface – Time View



CMU can be configured to support High Availability with an Active-Passive cluster.

SUSE Linux Enterprise Server

For this Reference Architecture SUSE Linux Enterprise Server version 11, SP3 was used.

SUSE Linux Enterprise Server is designed to harness the volume and velocity of Big Data. As a result, enterprise Linux products from SUSE have been chosen to run today's most innovative Big Data technologies, including in-memory database, data warehouse and Apache Hadoop-based solutions. SUSE Linux Enterprise Server, optimized for massively data intensive workloads, delivers the highest performing, open source Linux operating system in the market today and provides a rock-solid foundation for deploying Hadoop clusters. The flexibility of SUSE Linux Enterprise Server enables you to optimize system performance with fine-grained controls for CPU, memory, storage, file systems, and networking. The scale out capability of SUSE Linux Enterprise Server means you can grow your Hadoop cluster while maintaining the performance you need for real time data analysis. Built-in security features, backed up by security certifications, ensure that your critical business data stays safe. SUSE Linux Enterprise Server supports more hardware architectures and software applications than any other Linux distribution, and includes advanced features for interoperability with Windows and UNIX® environments.

SUSE also maintains documented "Best Practices for Deploying Hadoop on SUSE Linux Enterprise Server", https://www.suse.com/communities/conversations/best-practices-deploying-hadoop-suse-linux-enterprise-server/

- Scale out – SUSE Linux Enterprise Server supports the latest open source high-speed interconnects, protocols, adapters and drivers for improved communication between your scale-out solution components which are crucial performance components in a Big Data environment. Also, updated LVM support with improved thin provisioning to optimize shared storage usage, and btrfs enhancements with quota support for subvolumes and faster snapshotting, improving administrator efficiency, speed, and flexibility of deployment.
- Scale up – SUSE Linux Enterprise Server supports the latest industry-standard, scale-up hardware and supports up to 4096 logical CPUs, 32TiB RAM and high-speed interconnect making it the ideal scale-up choice. Multiple file system options exist in SUSE Linux Enterprise Server such as XFS which excels for large-scale, heavy load and parallel read/write operations that exist in Big Data implementations.
- High performance computing – SUSE is a leader in high performance computing, with more than half of the world's largest super computer clusters and 80% of all Linux mainframes running on SUSE Linux Enterprise Server. As the foundation for market leading analytics solutions such as SAP HANA and Teradata Aster, SUSE Linux Enterprise Server delivers the reliability, performance and scalability necessary for high performance and Big Data computing. The engineering required for this leadership carries over into Big Data solutions, where the data and computing are brought together on a very granular, localized level across the elements of your cluster.
- Security – SUSE Linux Enterprise Server includes Linux security frameworks apparmor and SELinux, and has achieved EAL4+ Common Criteria Certification and FIPS 140-2 validation. With the SUSE Linux Enterprise Server security

updates available as part of the product subscription and the best practices outlined in this reference architecture, core data and Big Data computing infrastructure that business decisions rely upon remains secure.

- Certified and supported – SUSE Linux Enterprise Server is a certified and supported platform for leading Hadoop vendors Cloudera, Hortonworks and WANDisco, undergoing additional levels of testing and quality assurance to improve integration with partner software and providing customers with an assurance of interoperability. The SUSE YES Certified Program assures your HP system platform, including those used in this reference architecture, is compatible, certified, and supported with SUSE Linux Enterprise Server. By using SUSE Linux Enterprise Server in a Big Data solution, there is a completely supported stack of hardware, operating system, and application stack

With over 20 years of Linux experience, leadership in high-performance, high availability and cloud computing solutions and leading partnerships, SUSE provides the right foundation to allow you to capitalize on the valuable data that exists in your organization.

Reference Architectures

The following illustrates a reference progression of Hadoop clusters from a single rack to a multi-rack configuration. Best practices for each of the components within the configurations specified have been articulated earlier in this document.

Single Rack Reference Architecture

The Single Rack Hortonworks Enterprise Reference Architecture (RA) 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 Reference Architecture reflects the following.

Single Rack network

As previously described in the Network section, two IRF Bonded HP 5900AF-48XGT ToR switches are specified for performance and redundancy. The HP 5900AF-48XGT includes four 40GbE uplinks which can be used to connect the switches in the rack into the desired network or to the 40GbE HP 5930-32QSFP+ aggregation switch. Keep in mind that if IRF bonding is used, it requires 2x 40GbE ports per switch, which would leave 2x 40GbE ports on each switch for uplinks.

Cluster isolation and access configuration

It is important to isolate the Hadoop Cluster on the network so that external network traffic does not affect the performance of the cluster. In addition, this also allows the Hadoop cluster to be managed independently from that of its users, which ensures that the cluster administrator is the only one capable of making changes to the cluster configurations. To achieve this, we recommend isolating the ResourceManager, NameNode and Worker nodes on their own private Hadoop Cluster subnet.

Key point

Once a Hadoop cluster is isolated, the users of the cluster will still need a way to access the cluster and submit jobs to it. To achieve this we recommend multi-homing the Management node so that it participates in both the Hadoop Cluster subnet and a subnet belonging to the users of the cluster. Ambari is a web application that runs on the Management node and allows users to manage and configure the Hadoop cluster (including seeing the status of jobs) without being on the same subnet, provided the Management node is multi-homed. Furthermore, this allows users to shell into the Management node and run the Apache Pig or Apache Hive command line interfaces and submit jobs to the cluster that way.

Staging data

In addition, once the Hadoop Cluster is on its own private network one needs to think about how to be able to reach the HDFS in order to ingest data. The HDFS client needs the ability to reach every Hadoop DataNode in the cluster in order to stream blocks of data onto the HDFS. The Reference Architecture provides two options to do this.

The first option is to use the already multi-homed Management node, which can be configured with 4 additional disks to provide twice the amount of disk capacity (an additional 3.6TB) compared to the other management servers in order to provide a staging area for ingesting data into the Hadoop Cluster from another subnet.

The other option is to make use of the open ports that have been left available in the switch. This Reference Architecture has been designed such that if both NICs are used on each worker node and 2 NICs are used on each management node it leaves 26 ports still available across both the switches in the rack. These 26 10GbE ports or the remaining 40GbE ports on the switches can be used by other multi-homed systems outside of the Hadoop cluster to move data into the Hadoop Cluster.

Note
The benefit of using dual-homed edge node(s) to isolate the in-cluster Hadoop traffic from the ETL traffic flowing to the cluster is often debated. One benefit of doing so is better security. However, the downside of a dual-homed network architecture is ETL performance/connectivity issues, since a relatively few number of nodes in the cluster are capable of ingesting data. For example, the customer may want to kick off Sqoop tasks on the worker nodes to ingest data from external RDBMs, which will maximize the ingest rate. However this requires that the worker nodes be exposed to the external network to parallelize data ingestion, which is less secure. The customer has to weigh their options before committing to an optimal network design for their environments.

Lastly, one can leverage WebHDFS which provides an HTTP proxy to securely read and write data to and from the Hadoop Distributed File System. For more information on WebHDFS, please see http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.5/bk_system-admin-guide/content/sysadminguides_webhdfs.html

Rack enclosure
The rack contains eighteen HP ProLiant DL380p servers, three HP ProLiant DL360p servers and two HP 5900AF-48XGT switches within a 42U rack. This leaves 1U open for a 1U KVM switch.

Network
As previously described in the Switch selection section, two HP 5900AF-48XGT switches are specified for performance and redundancy. The HP 5900AF-48XGT includes up to four 40GbE uplinks which can be used to connect the switches in the rack.

Management nodes
Three ProLiant DL360p Gen8 management nodes are specified:

- The Management Node
- The ResourceManager / NameNode HA
- The NameNode / ResourceManager HA

Detailed information on the hardware and software configurations is available in the Server selection section of this document.

Worker nodes
As specified in this design, eighteen ProLiant DL380p Gen8 worker nodes will fully populate a rack.

Best practice
Although it is possible to deploy with as few nodes as a single worker node, HP recommends starting with a minimum of three worker nodes in order to provide the redundancy that comes with the default replication factor of 3 for availability. Performance improves with additional worker nodes as ResourceManager can leverage idle nodes to land jobs on servers that have the appropriate blocks, leveraging data locally rather than pulling data across the network. These servers are homogenous and run the DataNode and the NodeManager (or HBaseRegionServer) processes.

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. This is less important for larger clusters as the inherent redundancy within the Hortonworks Distribution of Hadoop will ensure there is less impact.

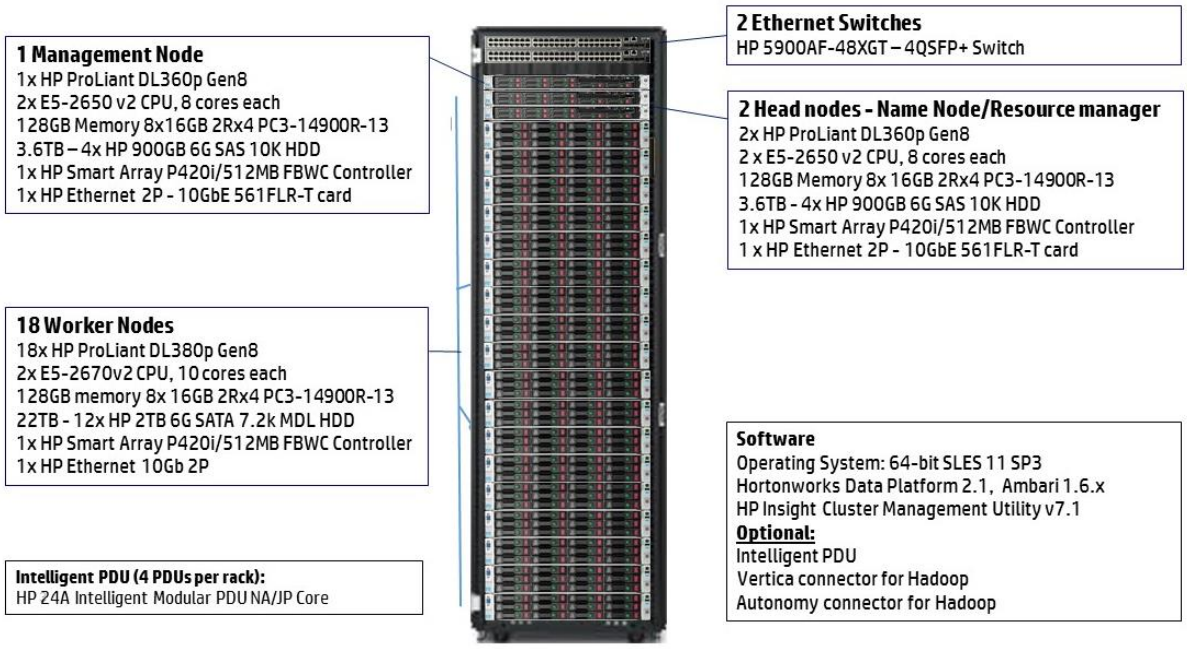
Best practice
For each server, HP recommends that each power supply is connected to a different PDU than the other power supply on the same server. Furthermore, the PDUs in the rack can each be connected to a separate data center power line to protect the infrastructure from a data center power line 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.

Open rack space
The design leaves 1U open in the rack allowing for a KVM switch when using a standard 42U rack.

Figure 11. Single Rack Reference Architecture – Rack Level View



Multi-Rack Reference Architecture

The Multi-Rack design assumes the Single Rack RA Cluster design is already in place and extends its scalability. The Single Rack configuration ensures the required amount of management services are in place for large scale out. For Multi-Rack clusters, one simply adds more racks of the configuration provided below to the Single Rack configuration. This section reflects the design of those racks.

Rack enclosure
The rack contains eighteen HP ProLiant DL380p Gen8 servers and two HP 5900AF-48XGT switches within a 42U rack. 4U remains open and can accommodate an additional 2x DL380p servers (2U each) or a KVM switch (1U); or, install two HP 5930-32QSFP+ (1U) aggregation switches in the first expansion rack.

Multi-Rack Network
As previously described in the Switch selection section, two HP 5900AF-48XGT ToR switches are specified per each expansion rack for performance and redundancy. The HP 5900AF-48XGT includes up to four 40GbE uplinks which can be used to connect the switches in the rack into the desired network, via a pair of HP 5930-32QSFP+ aggregation switches.

Software
The ProLiant DL380p servers in the rack all are configured as Worker nodes in the cluster, as all required management processes are already configured in the Single Rack RA. Aside from the OS, each worker node typically runs DataNode, NodeManager (and HBaseRegionServer if you are using HBase).

Note
While much of the architecture for Multi-Rack Hadoop cluster was borrowed from the Single Rack design, the architecture suggested here for multi-rack is based on previous iterations of testing on the DL380e platform. It is provided here as a general guideline for designing multi-rack Hadoop clusters.

Figure 12. Multi Rack Reference Architecture – Rack Level View

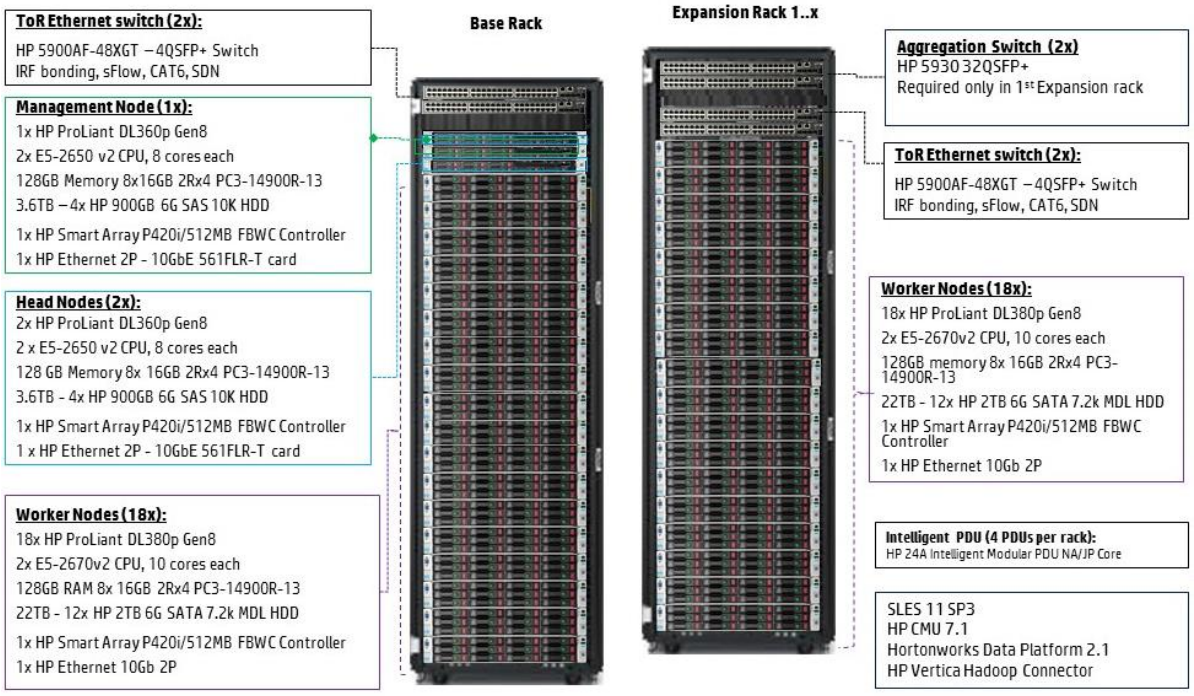
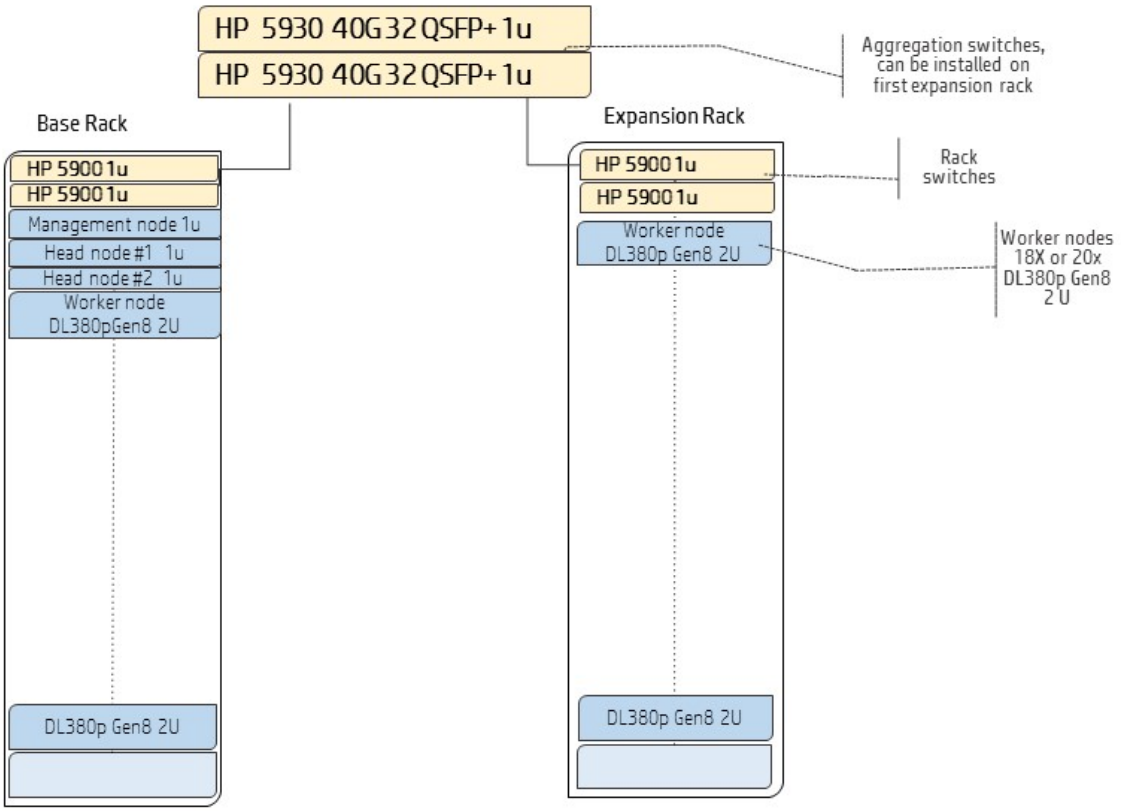


Figure 13. Multi-Rack Reference Architecture (extension of the single rack reference architecture)



Capacity and sizing

Hadoop cluster storage sizing requires careful planning and identifying the current and future storage and compute needs.

Here is a general guideline on data inventory:

- Sources of data
- Frequency of data
- Raw storage
- Processed HDFS storage
- Replication factor
- Default compression turned on
- Space for intermediate files
- How to calculate storage needs – guidelines

To calculate the storage needs find the number of TB of data per day, week, month and year. Then add the ingestion rate of all data sources. It would make sense to identify storage requirements for short term, medium term and long term.

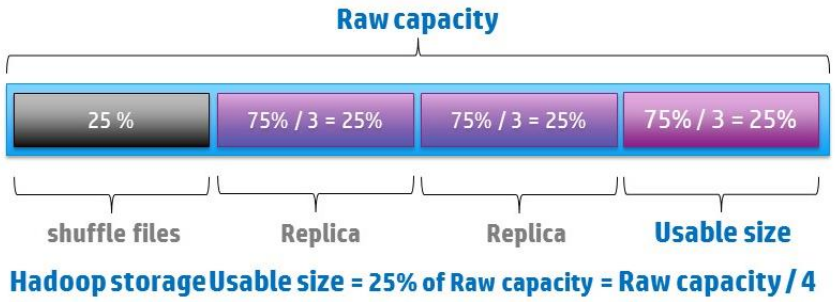
Another important consideration is data retention both size and duration. What data is required to be kept and for how long. Consider the maximum fill-rate and file system format space requirements on hard drives while estimating size of storage.

Usable HDFS capacity sizing

Some customers may have a specific requirement for usable HDFS capacity. Generally speaking there is a 10% deduction (decimal to binary conversion) when converting from raw to usable.

For example, a full rack cluster (18 x DL380p Gen8) that has 11x 2TB drives per node, the raw capacity would be 2TB *11 * 18 = 396TB, subtract 10% from the raw space to calculate usable HDFS space of 356TB. 25% of usable space should be set aside for MapReduce leaving 267TB. If the replication factor is **3**, then you would get 267TB/3 or approximately 89TB of usable space. Following this rule of thumb would help decide how many data nodes to order. Compression would provide additional usable space, depending on the type of compression, such as snappy or gzip.

Figure 14. Usable storage for replication factor of 3.



Key point

HP recommends using compression as it reduces file size on disks and speeds up data transfer to disks and network. The parameter to set to configure job output files to be compressed:

```
mapreduce.output.fileoutputformat.compress=true
```

System configuration guidance

Workload matters

Hadoop distributes data across a cluster of balanced machines and uses replication to ensure data reliability and fault tolerance. Because data is distributed on machines with compute power, processing can be sent directly to the machines storing the data. Since each machine in a Hadoop cluster stores as well as processes data, those machines need to be configured to satisfy both data storage and processing requirements.

Table 8. Examples of CPU and I/O bound workloads

I/O bound jobs	CPU bound job
Sorting	Classification
Grouping	Clustering
Data import and export	Complex text mining
Data Movement and transformation	Natural language processing
	Feature extraction

Based on feedback from the field, most users looking to build a Hadoop cluster are not aware of the eventual profile of their workload. Often the first jobs that an organization runs with Hadoop differ very much from the jobs that Hadoop is ultimately used for as proficiency increases.

Building a cluster appropriate for the workload is key to optimizing the Hadoop cluster.

Processor options

For workloads that are CPU intensive it is recommended to choose higher capacity processors with more cores. Typically workloads such as interactive Hive, Spark and Solr Search will benefit from higher capacity CPUs. On the selected DL380p the following table shows alternative CPUs.

Table 9. CPU recommendations

CPU	Description
2 x E5-2650 v2	Base configuration (8 cores/2.6GHz)
2 x E5-2670 v2	Enhanced (10 cores/2.5GHz)
2 x E5-2680 v2	High Performance (10 cores/ 2.8GHz)

See Appendix D Table 22 for BOM details on various CPU choices.

Memory options

When calculating memory requirements, remember that Java uses up to 10 percent of memory to manage the virtual machine. HP recommends to configure Hadoop to use strict heap size restrictions to avoid memory swapping to disk.

It is important to optimize RAM for the memory channel width. For example, when using dual-channel memory, each machine should be configured with pairs of DIMMs. With triple-channel memory each machine should have triplets of DIMMs. Similarly, quad-channel DIMMs should be in groups of four.

Key point: Memory intensive workloads

Interactive Hive, Spark, Storm workloads are generally more memory intensive, hence it is suggested to use 256 GB RAM on each DL380p server.

Table 10. Memory recommendations

Memory	Description
128 GB 8 x 16GB 2Rx4 PC3-14900R-13	Base configuration
256 GB 16 x 16GB 2Rx4 PC3-14900R-13	High capacity configuration

See Appendix D Table 23 for BOM details on alternative memory configurations.

Storage options

For workloads such as ETL and similar long running queries where the amount of storage is likely to grow, it is recommended to pick higher capacity and faster drives. For a performance oriented solution, SAS drives are recommend as they offer a significant read and write throughput performance enhancement over SATA disks. On the selected DL380p the following table shows alternative storage options.

Table 11. HDD recommendations

HDD	Description
2/3/4 TB LFF SATA	Base configuration
2/3/4 TB LFF SAS	Performance configuration

See Appendix D Table 24 for BOM details of alternative hard disks.

Network planning

Hadoop is sensitive to network speed in addition to CPU, memory and disk I/O, hence 10GbE is ideal, considering current and future needs for additional workloads. Switches with deep buffer caching are very useful. Network redundancy is a must for Hadoop. Hadoop shuffle phase does generate a lot of network traffic.

Generally accepted oversubscription ratios are around 2-4:1. Lower/higher oversubscription ratios can be considered if higher performance is required. A separate white paper on networking best practices for Hadoop is available at http://h20195.www2.hp.com/V2/GetDocument.aspx?docname=4AA5-3279ENW

Recommendation: BIOS/Firmware

HP recommends all ProLiant systems be upgraded to the latest BIOS and firmware versions before installing the OS. HP Service Pack for ProLiant (SPP) is a comprehensive systems software and firmware update solution, which is delivered as a single ISO image. The minimum SPP version recommended is 2014.02.0 (B). This updated version includes a fix for the OpenSSL HeartBleed Vulnerability. The latest version of SPP can be obtained from: http://h18004.www1.hp.com/products/servers/service_packs/en/index.html

Vertica and Hadoop

Relational database management systems such as HP Vertica excel at analytic processing for big volumes of structured data including call detail records, financial tick streams and parsed weblog data. HP Vertica is designed for high speed load and query when the database schema and relationships are well defined. Hortonworks Distribution for Hadoop, built on the popular open source Apache Software Foundation project, addresses the need for large-scale batch processing of unstructured or semi-structured data. When the schema or relationships are not well defined, Hadoop can be used to employ massive MapReduce style processing to derive structure out of data. The Hortonworks Distribution simplifies installation, configuration, deployment and management of the powerful Hadoop framework for enterprise users.

Each can be used standalone – HP Vertica for high-speed loads and ad-hoc queries over relational data, Hortonworks Distribution for general-purpose batch processing, for example from log files. Combining Hadoop and Vertica creates a nearly infinitely scalable platform for tackling the challenges of big data.

Note

Vertica was the first analytic database company to deliver a bi-directional Hadoop Connector enabling seamless integration and job scheduling between the two distributed environments. With Vertica's Hadoop and Pig Connectors, users have unprecedented flexibility and speed in loading data from Hadoop to Vertica and querying data from Vertica in Hadoop as part of MapReduce jobs for example. The Vertica Hadoop and Pig Connectors are supported by Vertica, and available for download.

For more information, please see vertica.com/industries/hadoop-acceleration

HP IDOL and Hadoop

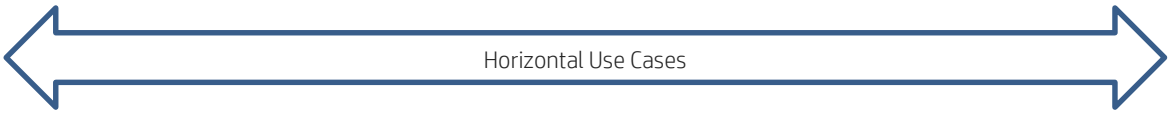
HP IDOL for Hadoop from HP Autonomy is designed to make Hadoop data come alive by helping manage, control, and enable greater insight so businesses can profit from information. Based on the technology of HP IDOL 10 (Intelligent Data Operating Layer), a market-leading information analytics platform currently being used to solve the toughest information challenges across a wide range of industries, HP IDOL for Hadoop tightly integrates with Hadoop to enrich its unstructured content capabilities. By using HP IDOL for Hadoop, components in an existing Hadoop environment and a programming framework (MapReduce), you can reduce coding and implementation costs while performing complex analysis that improve business outcomes. Data analysis techniques include inquire (search), investigate (explore), interact (engage), and improve (filter) data at scale to extract timely and actionable insights from big data in Hadoop.

HP IDOL for Hadoop leverages the market-leading file filtering technology of HP IDOL KeyView to extract text, metadata, and security information from all forms of data. The product can then apply its entity extraction function to identify sensitive fields such as social security and credit card numbers. For instance, if you are a healthcare provider with hundreds of thousands of patient forms stored in your Hadoop farms, you can ensure that no sensitive information is being stored. HP IDOL for Hadoop helps you remain compliant with your company policies, as well as HIPAA, PCI DSS, Sarbanes-Oxley, and more.

For more information, please see autonomy.com/idolhadoop

Use cases

Finance	Government	Telecom	Manufacturing	Energy	Healthcare
Fraud detection	Fraud detection	Broadcast monitoring	Supply chain optimization	Weather forecasting	Drug development
Anti-money laundering	Anti-money laundering	Churn prevention	Defect tracking	Natural resource exploration	Scientific research
Risk management	Risk management	Advertising optimization	RFID Correlation		Evidence based medicine
			Warranty management		Healthcare outcomes analysis



Sentiment analysis	Marketing campaign optimization	Logistics optimization
Social CRM / network analysis	Brand management	Clickstream analysis
Churn mitigation	Social media analytics	Influencer analysis
Brand monitoring	Pricing optimization	IT infrastructure analysis
Cross and Up sell	Internal risk assessment	Legal discovery
Loyalty & promotion analysis	Customer behavior analysis	Equipment monitoring
Web application optimization	Revenue assurance	Enterprise search

HP value added services and support

In order to help customers jump-start their Hadoop solution development, HP offers several Big Data services, including Factory Express and Technical Services (TS) Consulting. With the purchase of Factory Express services, your Hadoop cluster will arrive racked and cabled, with software installed and configured per an agreed upon custom Statement of Work. TS Consulting offers specialized Hadoop design, implementation, and installation and setup services. HP offers a variety of support levels to meet your needs.

Factory Express Services
Factory Integration services are available for customers seeking a streamlined deployment experience. With the purchase of Factory Express services, your Hadoop cluster will arrive racked and cabled, with software installed and configured per an agreed upon custom Statement of Work, for the easiest deployment possible. Please engage TS Consulting for details and quoting assistance.

Technical Services Consulting – Reference Architecture Implementation Service for Hadoop (Hortonworks)
With HP Reference Architecture Implementation Service for Hadoop, experienced HP Big Data consultants install, configure, deploy, and test your Hadoop environment based on the HP Reference Architecture. We'll implement all the details of the original Hadoop design: naming, hardware, networking, software, administration, backup, disaster recovery, and operating procedures. Where options exist, or the best choice is not clear, we'll work with you to configure the environment according to your goals and needs. We'll also conduct an acceptance test to validate and prove that the system is operating to your satisfaction.

Technical Services Consulting – Big Data services
HP Big Data Services can help you reshape your IT infrastructure to corral these 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: 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: HP 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 architecture implementations and integration services. Your flexible, scalable infrastructure will support Big Data variety, consolidation, analysis, share and search on HP 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.

For additional information, please visit:
hp.com/services/bigdata

HP Support options

HP offers a variety of support levels to meet your needs. HP Proactive Care helps prevent problems, resolve problems faster, and improve productivity. It helps customers identify and address IT problems before they cause performance issues or outages through analysis, reports, and update recommendations. Customers experiencing any performance issues are rapidly connected to experts for faster resolution. HP recommends adding HP Personalized Support option and HP Proactive Select with Proactive Care. Personalized support option is delivered by a local Account Support Manager who helps customers plan support options specific to their environment, deliver services, and review results. Proactive Select credits can be purchased upfront and used for optimization and improvement services related to health-checks, availability, and firmware updates throughout the year.

HP Datacenter Care provides a more personalized, customized approach for large, complex environments, with one solution for reactive, proactive, and multi-vendor support needs. You may also choose DMR (Defective Media Retention) option.

HP 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 HP technical resources, 4-hour response onsite hardware support and software updates.

HP Proactive Care

HP Proactive Care – HP 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 HP'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.

HP Proactive Care with the HP Personalized Support Option – Adding the Personalized Support Option for HP Proactive Care is highly recommended. The Personalized Support option builds on the benefits of HP 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.

HP Proactive Select
And to address your ongoing/changing needs, HP 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.

In addition, HP highly recommends HP Education Services (for customer training and education) and additional Technical Services, as well as in-depth installation or implementation services as may be needed.

For additional information, please visit:

HP Education Services: http://h10076.www1.hp.com/education/bigdata.htm

HP Technology Consulting Services: hp.com/services/bigdata

HP Deployment Services: hp.com/services/deployment

Summary

HP and Hortonworks 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 provided several reference architecture configurations for deploying clusters of varying sizes with Hortonworks Data Platform 2.1 on HP infrastructure and management software. These configurations leverage HP's 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 Data Platform software on HP infrastructure for clusters of various sizes.

Appendix A: Cluster design – heat map for server platforms

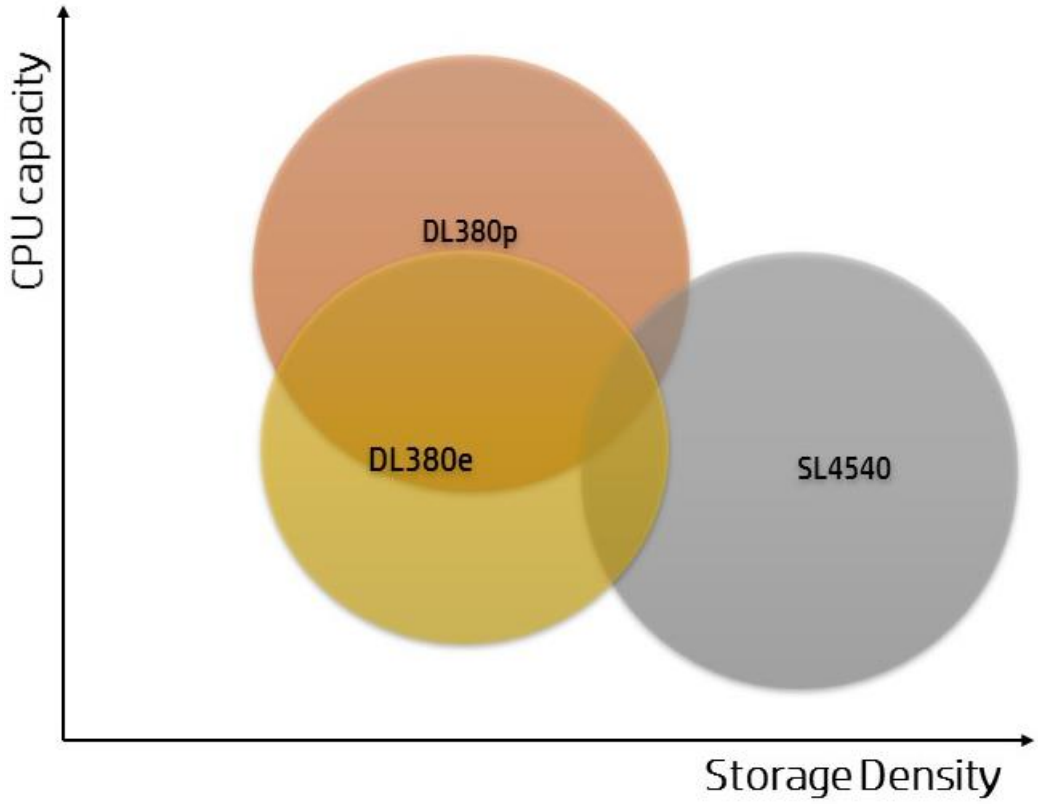
Tiered compute/storage deployment
Generally speaking, a data center has tiered compute/storage deployments. For simplicity reasons, dissect the data center into three major tiers according to their computation characteristics: high performance, high capacity and balanced zones. Please refer to Figure 15 below.

HP ProLiant SL4540 Gen8 Servers: A highly efficient converged design that delivers the right combination of capacity and performance, in the least amount of space and at lower cost, with the reliability and manageability you expect from HP ProLiant Gen8, the world's most intelligent servers.

HP ProLiant DL380e Gen8 Server: A cost-effective, enterprise-class server that delivers essential performance for data centers to meet their compute and storage needs today with scalability to grow as business requirements change.

HP ProLiant DL380p Gen8 Server: A high-performance, enterprise-class server that delivers top-notch processing capacity with fast memory and storage options ideally suited for big data applications with scalability to grow as business requirements change.

Figure 15. Compute/Storage Tiers in the Datacenter



High performance – interactive/SQL
Compute power optimized systems such as the HP ProLiant DL380p are ideal candidates when price/performance with high I/O bandwidth workloads are of primary importance. The data crunching rate in this tier is extremely high, so the demand for CPU power in this tier is high, as well as the memory size. Typical applications in this tier could be ETL apps, real time streaming (Spark Streaming), NoSQL applications such as MongoDB, Cassandra, Hbase, and in-memory processing such as Spark. Additionally, machine learning apps like K-means clustering and search indexing apps such as Solr could make best use in this tier. Both require significant computing power. The DL380p Gen8 server is a perfect selection for this tier as it has high CPU clock rate and large memory configuration. The data footprint in this tier ranges from hundreds of gigabytes to multi terabytes. If the data volume grows over time, it may not be economical to store data in this tier, we recommend migrating infrequently accessed data (cold or warm data) to high capacity or balanced tiers.

High capacity – storage optimized
Storage optimized systems, such as HP ProLiant SL4540 servers, are an ideal consideration when low cost per terabyte is the prime factor. This is a perfect place for data warehousing analytics as the bigger the data volume is, the higher accuracy of the analytics outcome. Typical workloads are web log analytics, sentiment and click stream analysis. Applications such as Hive, Pig, and map-reduce often find their home in this tier. The SL4540 (3x15) server platform is an ideal candidate for this tier as it is slanted toward higher storage density to tackle the large data volume. The other variants of the SL4540, 2x25 and 1x60, are also very good for mirroring cold data for backup/DR purpose.

Balanced performance and capacity – batch focused
A mixture of workloads typically found in high performance and high capacity tiers are also seen in this tier. The compute and storage density requirement for the server is equally important. The data footprint in this tier is from multi terabytes to a couple of petabytes. The HP ProLiant DL380e Gen8 server is a good selection in this tier as its compute and storage architecture is more balanced. Typical workloads are similar to cold tier, such as web log analytics, sentiment and click stream analysis, but compute and storage needs are more balanced.

Overlapped tiers
As one can see in Figure 15, there are no distinct boundaries between these tiers. When there is a variety or mixture of workloads a hybrid approach may be useful, a combination of server platforms can help fulfillment of the performance and data storage requirements. It is recommended to have early discussions with IT infrastructure and other related teams to fully discuss the implications of workload requirements vs. platform selection, in order to arrive at the right Hadoop system.

Appendix B: Hadoop cluster tuning/optimization

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 map-reduce 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.

a) 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 XFS, 64 bit OS:
  - Enable allocsize=128k,noatime,nobarrier,inode64,nodev,noatime (/etc/fstab)
  - Do not use logical volume management (LVM)
- Tune OS block readahead to 8K (/etc/rc/local):
  `blockdev --setra 8192 <storage device>`
- Turn off disk swappiness or to min=5:
  `Set sysctl vm.swappiness=0 in /etc/sysctl.conf`
- Tune ulimits for number of open files to a high number:
  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
  `echo deadline > /sys/block/<device>/queue/scheduler`

  For persistency across boot, append the following to kernel boot line in `/etc/grub.conf`:
  `elevator=deadline`
- Configure network bonding on a minimum two 10 GbE server ports, for up to a max 20 GbE throughput.
- Ensure forward and reverse DNS is working properly.
- Install and configure ntp to ensure clocks on each node are in sync to management node.

- For good performance improvements, disable transparent huge page compaction:
  `echo never > /sys/kernel/mm/transparent_hugepage/enabled`

b) Storage controller tuning

- Tune array controller stripe size to 1024MB:
  `hpssacli ctrl slot=<slot number> ld <ld number> modify ss=1024`

- Disable array accelerator(caching) (aa=disable):
  `hpssacli ctrl slot=<slot number> ld <ld number> modify aa=disable`

c) Power settings

Please note for a performance driven POC, we recommend using settings that help boost performance but could have negative impact on power consumption measurement:

- HP Power Profile → Maximum Performance
- HP Power regulator → Static High Performance mode
- Intel_QPI_Link_Mgt → Disabled
- Min_Proc_Idle_power_Core_State → No C-states
- Mem_Power_Saving → Max Perf
- Thermal Configuration → increased cooling
- Min_Proc_Idle Power Package state → No Package state
- Energy/Performance Bios → Disabled
- Collaborative Power Control → Disabled
- Dynamic Power Capping Functionality → Disabled
- DIMM Voltage Preference → Optimized for Performance

d) CPU tuning

The default BIOS settings for CPU should be adequate for most Hadoop workloads. Make sure that Hyper-Threading is turned on as it will help with additional performance gain.

e) HP ProLiant BIOS

- SPP version >= 2014.02.0 (B)
- Update System BIOS version to be >= P71
- Update Integrated Lights-Out (iLO) version to be >= 1.1
- Intel Virtualization Technology → Disabled
- Intel VT-d → Disabled

f) HP Smart Array P420i

- Update controller firmware to be >= v5.22
- Configure each Hadoop data drive as a separate RAID 0 array with stripe size of 1024KB
- Turn Off “Array Acceleration” / “Caching” for all data drives

Example: (with two controllers)

- ctrl slot=0 ld all modify caching=disable ← disable caching on all logical drives on 1st ctrlr
- ctrl slot=0 ld1 modify caching=enable ← enable caching on the OS logical drive on 1st ctrlr
- ctrl slot=2 ld all modify caching=disable ← disable caching on all logical drives on 2nd ctrlr

g) Network cards

- DL380p Ethernet driver ixgbe, version >=3.19.0.46 and firmware version >= 0x800004ec, 1.464.0
- DL360p Ethernet driver bnx2x, version >=1.78.80 and firmware version >= bc 7.4.22 phy 1.34

h) Oracle Java

- java.net.preferIPv4Stack set to true

i) Patch common security vulnerabilities

- Bash 9740 ‘Shellshock’ fix for bash shell vulnerability as per instructions on
  https://download.suse.com/Download?buildid=sjAu9hv-X1A~
  http://support.novell.com/security/cve/CVE-2014-0475.html
- For Heartbleed vulnerability, all versions of OpenSSL 1.0.1 prior to 1.0.1g need to be updated to 1.0.1g.
  http://support.novell.com/security/cve/CVE-2014-0160.html

Appendix C: Bill of materials

The BOMs outlined below are based on the tested configuration for a Single-Rack Reference Architecture with 1 management node, 2 Head nodes, 18 worker nodes and 2 ToR switches. Quantities specified on each table are based on per server configuration.

Management Node and Head Node BOM

Table 12. The HP ProLiant DL360p Gen8 Server Configuration

Qty	Part Number	Description
1	654081-B21	HP DL360p Gen8 8-SFF CTO Chassis
1	712726-L21	HP DL360p Gen8 E5-2650v2SDHS FIO Kit
1	712726-B21	HP DL360p Gen8 E5-2650v2SDHS Kit
8	708641-B21	HP 16GB 2Rx4 PC3-14900R-13 Kit
4	652589-B21	HP 900GB 6G SAS 10K 2.5in SC ENT HDD
1	700700-B21	HP Ethernet 10Gb 2P 561FLR-T FIO Adptr
1	661069-B21	HP 512MB FBWC for P-Series Smart Array
2	656362-B21	HP 460W CS Plat PL Ht Plg Pwr Supply Kit
1	SG506A	HP C13 - C14 WW 250V 10Amp IPD 0.76m 1pc Jumper Cord
1	SG508A	HP C13 - C14 WW 250V 10Amp IPD 1.37m 1pc Jumper Cord
1	663201-B21	HP 1U SFF BB Gen8 Rail Kit
1	C6N36AAE	HP Insight Control ML/DL/BL Bundle E-LTU
	C6N36A	HP Insight Control ML/DL/BL FIO Bndl Lic (optional if E-LTU is not available)
1	BD783AAE	SUSE Linux Enterprise Server, 1-2 Sockets, Physical, 1 Year Subscription, 24x7 Support E-LTU

Worker Node BOM

Table 13. The HP ProLiant DL380p Gen8 Server Configuration

Qty	Part Number	Description
1	665552-B21	HP DL380p Gen8 12-LFF CTO Server
1	715216-L21	HP DL380p Gen8 E5-2670v2 FIO Kit
1	715216-B21	HP DL380p Gen8 E5-2670v2 Kit
8	708641-B21	HP 16GB 2Rx4 PC3-14900R-13 Kit
12	658079-B21	HP 2TB 6G SATA 7.2k 3.5in SC MDL HDD
1	700700-B21	HP Ethernet 10Gb 2P 561FLR-T FIO Adptr
1	661069-B21	HP 512MB FBWC for P-Series Smart Array
1	720864-B21	HP 2U LFF BB Gen8 Rail Kit
2	656363-B21	HP 750W CS Plat PL Ht Plg Pwr Supply Kit
1	SG506A	HP C13 - C14 WW 250V 10Amp IPD 0.76m 1pc Jumper Cord
1	SG508A	HP C13 - C14 WW 250V 10Amp IPD 1.37m 1pc Jumper Cord
1	C6N36AAE	HP Insight Control ML/DL/BL Bundle E-LTU
	C6N36A	HP Insight Control ML/DL/BL FIO Bndl Lic (optional if E-LTU is not available)
1	BD783AAE	SUSE Linux Enterprise Server, 1-2 Sockets, Physical, 1 Year Subscription 24x7 Support E-LTU

Network BOMs

Table 14. Network – Top of Rack Switch

Qty	Part Number	Description
2	JG336A	HP 5900AF-48XGT-4QSFP+ Switch
2	JG326A	HP X240 40G QSFP+ QSFP+ 1m DAC Cable
4	JC680A	HP A58x0AF 650W AC Power Supply
4	JG553A	HP X712 Bck(pwr)-Frt(prt) HV Fan Tray

Table 15. Network – Aggregation/Spine Switch (Only required for first Expansion Rack. Not required for Single Rack Architecture.)

Qty	Part Number	Description
2	JG726A	HP FF 5930-32QSFP+
4	JC680A	HP 58x0AF 650W AC Power Supply
4	JG553A	HP X712 Bck(pwr)-Frt(prt) HV Fan Tray
2	JG326A	HP X240 40G QSFP+ QSFP+ 1m DAC Cabl
8	JG328A	HP X240 40G QSFP+ QSFP+ 5m DAC Cable

For separate iLO and PXE network, use the following BOM to replace the 10GbE FlexLOM NIC

Table 16. Modified BOM for HP ProLiant DL380p Gen8 Server Configuration

Qty	Part Number	Description
1	716591-B21	HP Ethernet 10Gb 2-port 561T Adapter
1	684208-B21	HP Ethernet 1Gb 4-port 331FLR FIO Adapter

Table 17. Network – Top of Rack Switch for separate iLO and PXE network

Qty	Part Number	Description
1	JG510A	HP 5900AF-48G-4XG-2QSFP+ Switch
1	JC680A	HP A58x0AF 650W AC Power Supply
2	JC682A	HP A58x0AF Back (power side) to Front (port side) Airflow Fan Tray

Other hardware and software BOMs

Table 18. Hardware – Rack and PDU

Note: The quantity specified below is for a full rack with 2 switches, 3x DL360p and 18x DL380p.

Qty	Part Number	Description
4	AF520A	HP Intelligent Mod PDU 24a Na/Jpn Core
6	AF547A	HP 5xC13 Intelligent PDU Extension Bar G2 Kit
1	BW946A	HP 42U Location Discovery Kit
1	BW904A	HP 642 1075mm Shock Intelligent Series Rack
1	BW932A	HP 600mm Rack Stabilizer Kit
1	BW930A	HP Air Flow Optimization Kit
1	BW906A	HP 42U 1075mm Side Panel Kit
1	BW891A	HP Rack Grounding Kit

Table 19. Software – HP Insight Cluster Management Utility (CMU) options

Note: The quantity specified below is for a single node.

Qty	Part Number	Description
1	QL803B	HP Insight CMU 1yr 24x7 Flex Lic
1	QL803BAE	HP Insight CMU 1yr 24x7 Flex E-LTU
1	BD476A	HP Insight CMU 3yr 24x7 Flex Lic
1	BD476AAE	HP Insight CMU 3yr 24x7 Flex E-LTU
1	BD477A	HP Insight CMU Media

Table 20. Software – SUSE Linux Enterprise Server

Qty	Part Number	Description
21	BD783AAE	SUSE Linux Enterprise Server, 1-2 Sockets, Physical, 1 Year Subscription 24x7 Support E-LTU

Note
While HP is a certified reseller of Hortonworks software subscription, all application support (L1-L3) for Hortonworks software is provided by Hortonworks. The HP ProLiant DL380p platform is Hortonworks Certified.

Table 21. Software – Hortonworks Subscription options

Qty	Part Number	Description
5	F5Z52A	Hortonworks Data Platform Enterprise 4 Nodes or 50TB Raw Storage 1 year 24x7 Support LTU.

Note
Part numbers are at time of publication and subject to change. The bill of materials does not include complete support options or other rack and power requirements. If you have questions regarding ordering, please consult with your HP Reseller or HP Sales Representative for more details. hp.com/large/contact/enterprise/index.html

Appendix D: Alternate parts

Table 22. Alternate Processors – DL380p

Qty/Node	Part Number	Description
1	715218-L21	HP DL380p Gen8 E5 2650v2 FIO Kit 8 cores at 2.6GHz
1	715218-B21	HP DL380p Gen8 E5 2650v2 Kit
1	715216-L21	HP DL380p Gen8 E5 2670v2 FIO Kit 10 cores at 2.5GHz
1	715216-B21	HP DL380p Gen8 E5 2670v2 Kit
1	715215-L21	HP DL380p Gen8 E5 2680v2 FIO Kit 10 cores at 2.8GHz
1	715215-B21	HP DL380p Gen8 E5 2680v2 Kit

Table 23. Alternate Memory – DL380p

Qty/Node	Part Number	Description
8	708641-B21	HP 16GB 2Rx4 PC3 14900R 13 Kit for 128GB of Memory
16	708641-B21	HP 16GB 2Rx4 PC3 14900R 13 Kit for 256GB of Memory

Table 24. Alternate Disk Drives – DL380p

Qty/Node	Part Number	Description
12	652757-B21	HP 2TB 6G SAS 7.2K 3.5in SC MDL HDD
12	652766-B21	HP 3TB 6G SAS 7.2K 3.5in SC MDL HDD
12	695510-B21	HP 4TB 6G SAS 7.2K 3.5in SC MDL HDD
12	658079-B21	HP 2TB 6G SATA 7.2k 3.5in SC MDL HDD
12	628061-B21	HP 3TB 6G SATA 7.2k 3.5in SC MDL HDD
12	693687-B21	HP 4TB 6G SATA 7.2k 3.5in SC MDL HDD

Table 25. Alternate Network Cards – DL380p

Qty/Node	Part Number	Description
1	684210-B21	HP Ethernet 10GbE 530FLR SFP+ FIO Adapter for 10Gb networking only
1	652503-B21	HP Ethernet 10Gb 2P 530SFP+ Adapter

Note
The SFP+ network cards are used with DAC cabling and will not work with CAT6 cabling. If SFP+ network cards are used the 5900 SFP+ equivalent ToR network switches are required (HP 5900AF-48XG-4QSFP+ Part number JC772A).

Table 26. Alternate Controller Cards – DL380p and DL360p

Qty/Node	Part Number	Description
1	698529-B21	HP Smart Array P430/2GB FBWC 12Gb 1-port Int SAS Controller
1/Drive	D8S84A	HP Secure Encryption No Media Flexible License per Drive
1/Drive	D8S85AAE	HP Secure Encryption No Media E-LTU per Drive

For more information

Hortonworks, hortonworks.com

HP Solutions for Apache Hadoop, hp.com/go/hadoop

Hadoop and Vertica, vertica.com/industries/hadoop-acceleration

HP Insight Cluster Management Utility (CMU), hp.com/go/cmu

HP 5900 Switch Series, hp.com/networking/5900

HP FlexFabric 5930 Switch Series, hp.com/networking/5930

HP ProLiant servers, hp.com/go/proliant

HP Enterprise Software, hp.com/go/software

HP Networking, hp.com/go/networking

HP Integrated Lights-Out (iLO), hp.com/servers/ilo

HP Product Bulletin (QuickSpecs), hp.com/go/quickspecs

HP Services, hp.com/go/services

HP Support and Drivers, hp.com/go/support

HP Systems Insight Manager (HP SIM), hp.com/go/hpsim

SUSE, suse.com

To help us improve our documents, please provide feedback at hp.com/solutions/feedback.

Hortonworks develops, distributes and supports the only 100% open source Apache Hadoop data platform. Our team comprises the largest contingent of builders and architects within the Hadoop ecosystem who represent and lead the broader enterprise requirements within these communities. The Hortonworks Data Platform provides an open platform that deeply integrates with existing IT investments and upon which enterprises can build and deploy Hadoop-based applications. Hortonworks has deep relationships with the key strategic data center partners that enable our customers to unlock the broadest opportunities from Hadoop. For more information, visit hortonworks.com

SUSE is the original provider of the enterprise Linux distribution and the most interoperable platform for mission-critical computing. It's the only Linux recommended by VMware, Microsoft and SAP. And it's supported on more hardware and software than any other enterprise Linux distribution. With over 20 years of Linux experience, leadership in high-performance, high availability and cloud computing solutions and leading partnerships, SUSE provides the right foundation to allow you to capitalize on the valuable data that exists in your organization.

Sign up for updates
hp.com/go/getupdated

4AA5-4975ENW, December 2014, Rev. 2