



HPE Reference Architecture for Cloudera Data Platform Private Cloud Base on HPE EPA Symmetric Architecture for Big Data Analytics

Using HPE ProLiant DL380 Gen10 server for Apache Hadoop
Distributed File System (HDFS) and HPE Apollo 4510 for Hadoop Ozone

CONTENTS

Executive summary.....	3
Solution overview.....	4
HPE Elastic Platform for Big Data Analytics.....	4
Cloudera Data Platform Private Cloud Base.....	5
Solution Architecture for HPE Elastic platform for big data analytics with hdfs.....	7
Solution components.....	7
Rack configuration.....	7
Hardware.....	8
Software.....	8
Solution Architecture for HPE Elastic platform for big data analytics with Cloudera data Platform with Ozone.....	10
Solution components.....	10
Rack configuration.....	10
Hardware.....	11
Software.....	11
Best practices and configuration guidance for the solution.....	13
General hardware and software guidelines.....	13
Server best practices.....	15
HPE Sizer for the Elastic Platform for Big Data Analytics.....	16
Summary.....	16
Resources and additional links.....	18



EXECUTIVE SUMMARY

Today we are in a digital world that has created opportunities for Big Data Analytics beyond just being a legacy Hadoop cluster for running batch analytics. Many workloads are closely associated with Big Data Analytics like interactive analytics, streaming analytics, AI/ML, etc. Also, data is growing every day, and organizations are looking to store this data to run analytics on top of it. Hewlett Packard Enterprise and Cloudera allow you to derive new business insights from all your data by providing a platform to store dense data, manage, and process data at scale. This Reference Architecture provides several performance-optimized configurations for deploying the Cloudera clusters on Hewlett Packard Enterprise infrastructure that provides a significant reduction in complexity and a recognized increase in value and performance.

The HPE Elastic Platform for Analytics (EPA) is designed as a modular infrastructure architecture to address the need for scalable, dense storage, and elastic multi-tenant Big Data Platform for on-premises deployment. It has a wide selection of building blocks based on density, capacity, and performance to reduce the total cost of ownership and data center footprint while optimizing performance for Big Data workloads. It supports a wide range of workloads ranging from extract, transform, load (ETL) processing offloaded from traditional data warehouses (DW), SQL-based interactive analytics, and near real-time event processing of data streams to machine and deep learning applications with GPUs. Customers can select modular building blocks of compute, storage, and networking from Hewlett Packard Enterprise's diverse product portfolio, and integrate these blocks with software that enables an on-demand and elastic infrastructure foundation for Big Data. These building blocks can be deployed with co-located compute and storage on the same node or disaggregated over a high-bandwidth network as more blocks of compute and storage are added. Hewlett Packard Enterprise supports two different deployment models under this platform:

- **HPE Workload and Density Optimized (WDO) system (Asymmetric Architecture)** – Harnesses the power of a faster Ethernet network that enables a building block approach to independently scale, compute, and storage, and lets you consolidate your data and workloads growing at different rates. The HPE WDO system is based on the HPE Apollo 4200 storage block and the HPE Synergy 480 or HPE Apollo 2000 compute block.
- **HPE Balanced and Density Optimized (BDO) system (Symmetric Architecture)** – Supports Hadoop deployments that scale, compute, and storage together, with some flexibility in the choice of memory, processor, and storage capacity. This is primarily based on the HPE ProLiant DL380 server platform, with density optimized variants using HPE Apollo 4200 and HPE Apollo 4510 servers. In this paper we have tested with BDO systems – standard solution, the HPE ProLiant DL380 Gen10 server platform for HDFS and HPE Apollo 4510 for Ozone.

This Reference Architecture has been created to assist in the rapid design and deployment of Cloudera Data Platform Private Cloud Base v7.1 on HPE Elastic Platform for Big Data Analytics, with symmetric architecture for clusters of various sizes (HDFS and Ozone). This paper highlights recognizable benefits and provides guidance on the building Cloudera Data Platform for both HDFS and Ozone clusters that meets business needs. Customers who are looking to deploy Hadoop clusters of up to 100TB of capacity and node can opt for HDFS. Apache Ozone is designed to overcome the capacity and scaling limitations of HDFS. Hence customers who are looking to deploy denser nodes beyond 100TB, can opt for Ozone. We have tested Ozone for the highest density, with each storage node equipped with a maximum capacity of 700TB/node. With this Ozone is now able to take full advantage of HPE Apollo 4510 capability of hosting 48 LFF drives. We have also provided recommended configuration and best practices to deploy Ozone at this scale.

Document purpose: This Reference Architecture provides comprehensive architectural guidelines and implementation of Big Data Analytics solutions on HPE EPA symmetric architecture using Cloudera Data Platform Private Cloud Base 7.1. In addition to outlining the key solution components, this document also provides guidelines for optimizing infrastructure configuration as well as performance analysis for distributed training and streaming analytics.

Target audience: This document is intended for subject matter experts, domain experts, data engineers, IT managers, pre-sales engineers, services consultants, partner engineers, and customers who are interested in implementing Big Data Analytics workloads for real-time processing capabilities in their existing or new deployments.

This solution testing performed in July 2021.



SOLUTION OVERVIEW

The Reference Architecture configurations are based on the Cloudera Data Platform Private Cloud Base edition, specifically, CDP 7.1 version and HPE EPA BDO systems which include the HPE ProLiant DL380 Gen10, HPE ProLiant 360 Gen10, and HPE Apollo 4510 Gen10 servers. Our solution HPE EPA combining with Cloudera Data Platform addresses the challenges and helps the organizations to implement Big Data Analytics solution on-prem with HDFS and Ozone.

HPE Elastic Platform for Big Data Analytics

In the traditional Hadoop approach, storage and compute were co-located and the main workload was a batch analytics process. The challenge with this design model is that as the cluster scaled out, compute and storage grew together, and typically one of the resources was over-provisioned to satisfy the needs of the other, which increased unnecessary cost. With the rapid advancement of the next generation analytics applications and workload patterns (for example, streaming analytics, interactive analytics, end-to-end data pipelines, ML, AI, etc.), the architecture design requires a disaggregated, scalable, flexible, and high-performing platform.

Figure 1 illustrates an example of the data flow from edge to core to cloud, routed through a data pipeline that provides an infrastructure for data to not only flow bi-directionally, but also allow for the implementation of analytic processes in real-time, near real-time, at rest, and also AI modeling.

- “Edge Acquisition and Edge Analytics” is where data is acquired, processed, queued at the edge, and replicated in real-time to the core.
- “Core/Cloud Analytics” is deployed at the core data center or the cloud and integrated with an edge, and is where data is streamed, enriched, and analyzed in real-time, as well as being retained indefinitely for future analytics operations.
- “AI/ML Training” is deployed at the core data center or in the cloud and is where the processed data is available for data scientists and data technicians, to understand the data, build machine learning models, train, and validate the models.
- “Data Lake and Archives” is deployed at the core data center or at the cloud to provide optimized storage; it is where all the ETL is performed and data is prepared for machine learning and deep learning in particular.

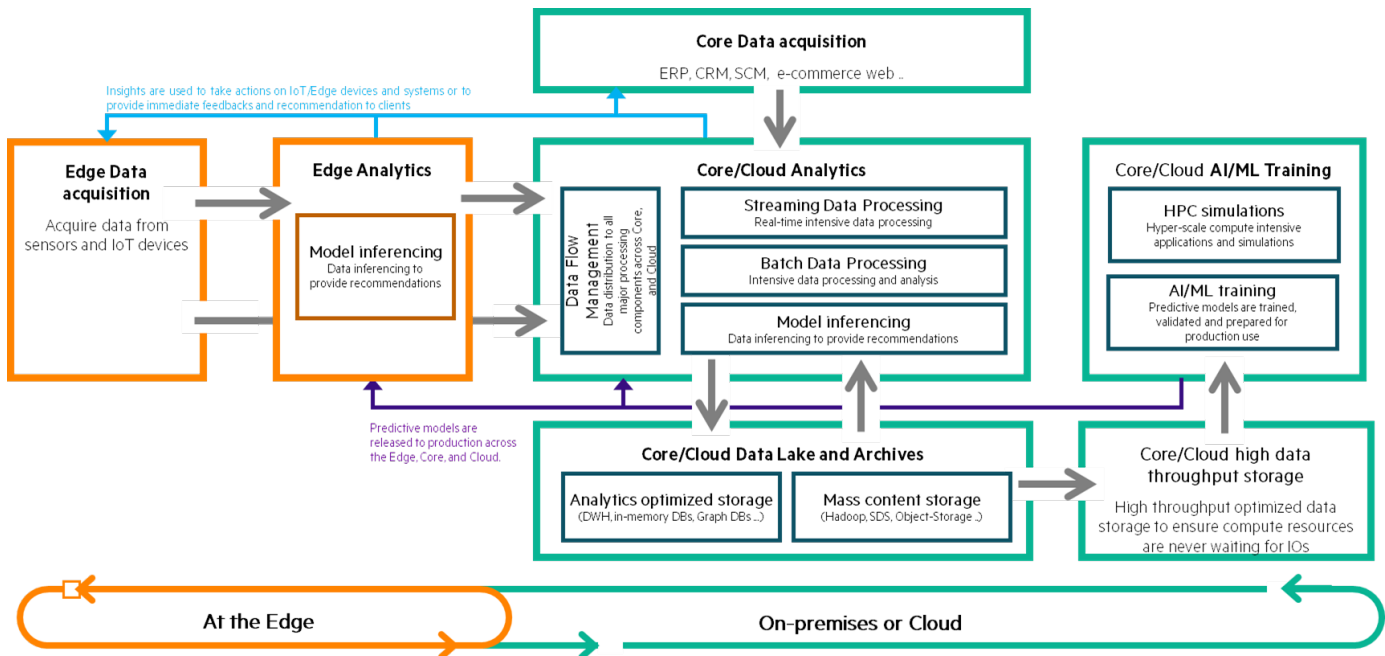


FIGURE 1. HPE Elastic Platform for Big Data Analytics (EPA) end-to-end data pipeline

For more information, see the HPE Reference Configuration for Elastic Platform for Analytics (EPA) document located at https://www.hpe.com/psnow/doc/4AA6-8931ENW?jumpid=in_lit-psnow-red



Cloudera Data Platform Private Cloud Base

Cloudera Data Platform (CDP) Private Cloud Base is the on-premises version of the Cloudera Data Platform. This new product combines the best of Cloudera Enterprise Data Hub and Hortonworks Data Platform Enterprise along with new features and enhancements across the stack. This unified distribution is a scalable and customizable platform where you can securely run many types of workloads.

CDP Private Cloud Base shown in Figure 2 supports a variety of hybrid solutions where compute tasks are separated from data storage and where data can be accessed from remote clusters. This hybrid approach provides a foundation for containerized applications by managing storage, table schema, authentication, authorization, and governance.



FIGURE 2. Cloudera Data Platform Private Cloud Base

CDP Private Cloud Base comprises a variety of components such as Apache HDFS, Apache Hive 3, Apache HBase, and Apache Impala, along with many other components for specialized workloads. You can select any combination of these services to create clusters that address your business requirements and workloads. Several pre-configured packages of services are also available for common workloads. These include:

- Data Engineering
 - Ingest, transform, and analyze data
 - Services: HDFS, YARN, YARN Queue Manager, Ranger, Atlas, Hive metastore, Hive on Tez, Spark
 - Oozie, Hue, and Data Analytics Studio
- Data Mart
 - Browse, query, and explore your data in an interactive way
 - Services: HDFS, YARN, YARN Queue Manager, Ranger, Atlas, Hive metastore, Impala, and Hue
- Operational Database
 - Low latency writes, reads, and persistent access to data for Online Transactional Processing (OLTP) use cases
 - Services: HDFS, Ranger, Atlas, and Hbase



Apache Ozone

Apache Ozone is also a part of Cloudera Data Platform Private Cloud Base. It is a new distributed object store designed to make hybrid cloud computing at scale a reality as shown in Figure 3. Customers can choose to deploy on-premises and easily enable a cloud-native architecture for their Big Data workloads that are tested, integrated with their entire data lifecycle. Ozone provides customers the capability to use S3 API protocols for their on-premises Big Data workloads and scale up to 10x the capacity and size of HDFS with billions of objects and 100's of Petabytes with the traditional performance of Big Data and the scale of a modern object-store fully supporting multiple API protocols. Ozone offers 4x times denser nodes and 30x times object-store capability that provides 50% of the cost benefits in infrastructure, operations, and management.

Ozone is a distributed key-value store that can efficiently manage both small and large files alike. Ozone was designed and built to overcome HDFS NameNode's limitations of handling more than 300M files. Ozone is designed to work well with the existing Apache Hadoop ecosystem and also fill the need for a new open-source object store that is designed for ease of operational use and scale to thousands of nodes and billions of objects in a single cluster. Ozone is a strongly consistent object store which is achieved by using protocols like RAFT.

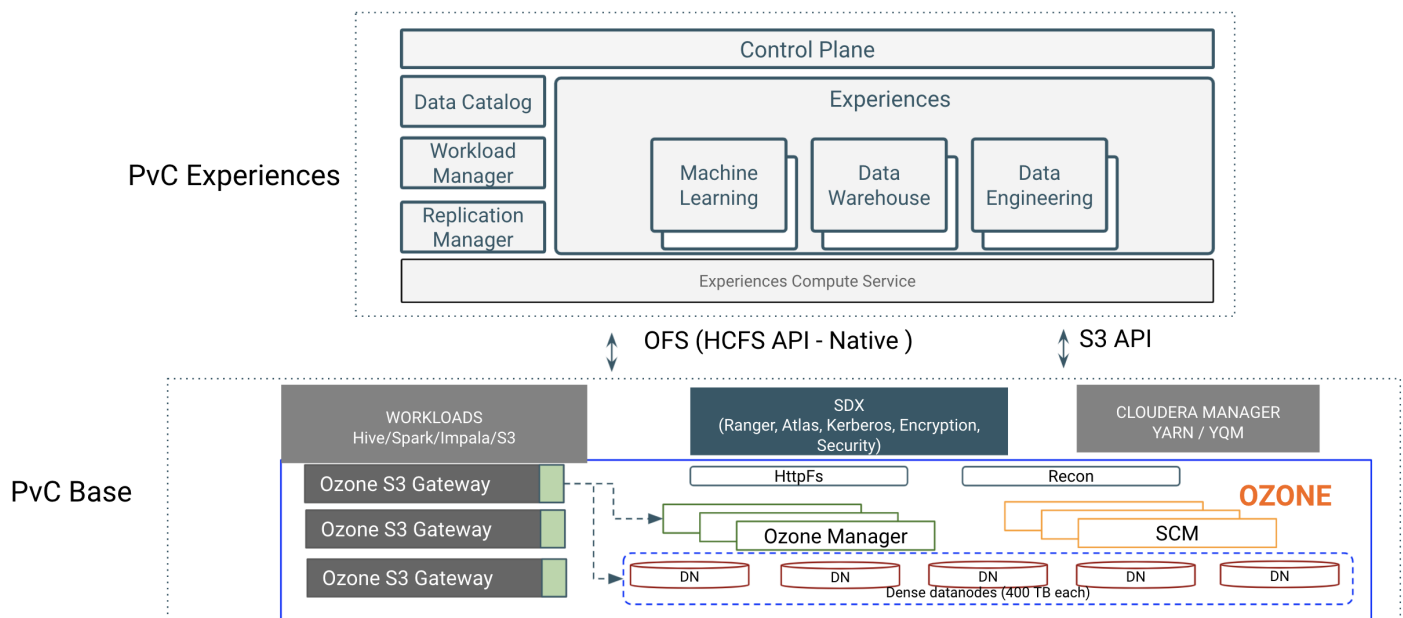


FIGURE 3. CDP Private Cloud Base with Ozone and experience

Ozone stores data in volumes, buckets, and keys:

- Volumes are user accounts. Only administrators can create or delete volumes.
- Buckets are directories. A bucket can contain any number of keys, but buckets cannot contain other buckets.
- Keys are like files.

Ozone can be accessed through Java client libraries and supports different protocols like S3 and Hadoop File System APIs. Applications using frameworks like Apache Spark, YARN, and Hive work natively without any modifications.



SOLUTION ARCHITECTURE FOR HPE ELASTIC PLATFORM FOR BIG DATA ANALYTICS WITH HDFS

The HPE Elastic Platform for Analytics (EPA) with Cloudera Data Platform (CDP) is designed to be a modular infrastructure foundation that can be scaled in a building block fashion to support different customer requirements. This platform supports various Big Data eco-system components such as Hive, Impala, Spark, etc.

Solution components

In this Reference Architecture, HPE EPA building blocks are the foundation of the solution that can be built together in various ways to solve the unique requirements. This section explains the detailed storage/compute, and control blocks based on the HPE ProLiant DL380 Gen10 and HPE ProLiant DL360 Gen10 servers required to build and implement a Big Data Analytics solution on the HPE EPA BDO Platform. The blocks defined in this section may be modified (for example, processor model, memory, etc.) to address or meet the customer requirements accordingly.

The HPE BDO solution infrastructure blueprints are composed of four blocks: storage/compute blocks, control blocks, network blocks, and rack blocks. Below is the infrastructure of building blocks required to implement the solution with HDFS:

Table 1 shows the solution components.

TABLE 1. Solution components

Block	Model
Control Block	HPE ProLiant DL360 Gen10
Storage/Worker Block	HPE ProLiant DL380 Gen10
Network Block	HPE Aruba 8325-48Y8C (25G x 40P SFP+ 40G/100G x 8P QSFP)
Rack Block	1200mm or 1075mm

NOTE

If the solution requires more than 100TB capacity per node, or if the solution is expected to handle more than 300M files, then it is recommended to go with [Hadoop Ozone](#) instead of HDFS.

Rack configuration

The single-rack Hadoop is designed to perform well as a single-rack cluster also forms 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 solution is deployed with the design priorities like performance, ease of use, and fault tolerance, etc.



Figure 4 depicts the single-rack configuration for HPE EPA for BDO solution with CDP Private Base.

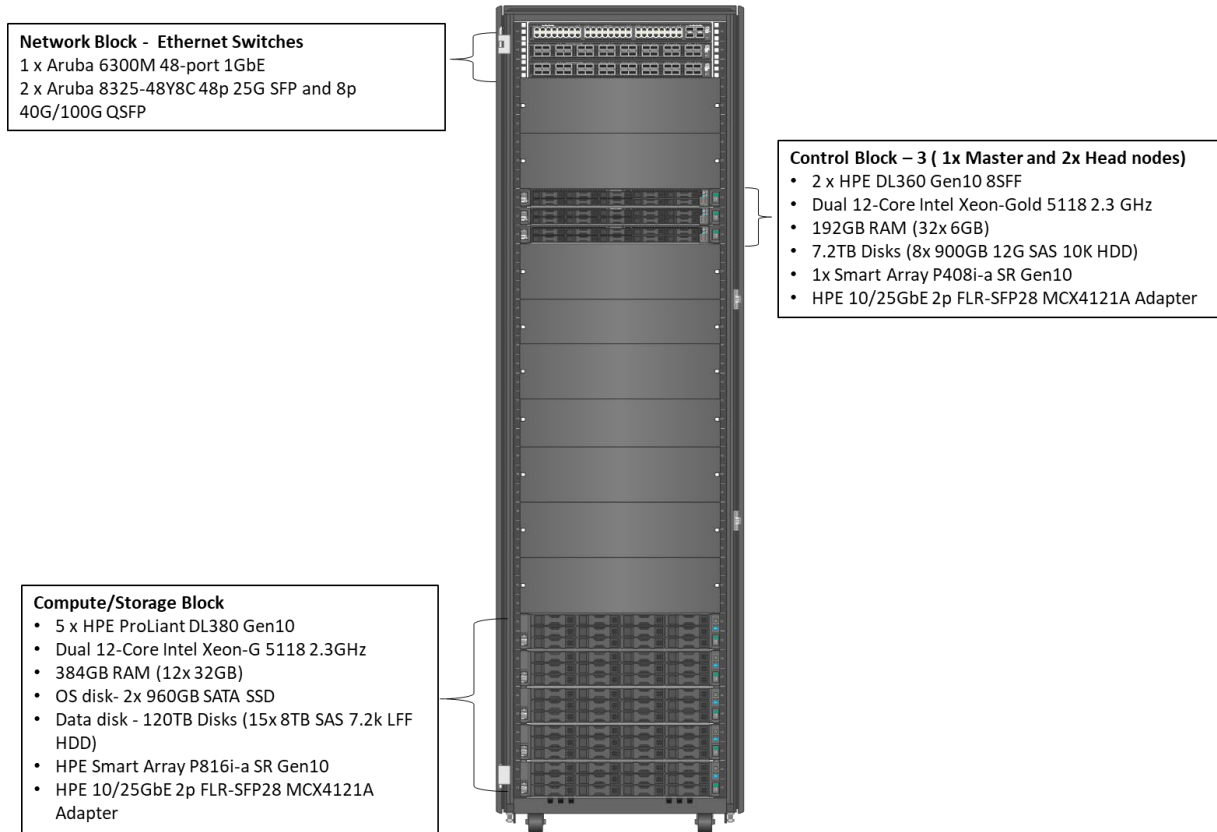


FIGURE 4. Basic conceptual diagram of single-rack for HPE EPA with Cloudera Data Platform Private Cloud Base with HDFS

Hardware

- **HPE ProLiant DL380 (Storage/Compute Block):** HPE ProLiant DL380 Gen10 servers are cost-effective industry-standard servers and can serve as a storage and compute node in the HPE EPA BDO solution. The HPE ProLiant DL380 Gen10 server performs as elastic storage nodes for HDFS and allows organizations to build multi-temperature data lake environments. The HPE ProLiant DL380 Gen10 server can be leveraged as a storage block for low latency workloads, and for environments where failure domain considerations prevent higher density storage nodes.
- **HPE ProLiant DL360 (Control Block):** HPE ProLiant DL360 Gen10 server delivers security, agility, and flexibility without compromise. The HPE ProLiant DL360 Gen10 server is a highly flexible, performance-driven, and two-socket server that features Intel's new Scalable family of processors in a densely-packed 1U form factor.

HPE ProLiant 380 Gen10 server, <https://www.hpe.com/psnow/doc/a00008180enw>

HPE ProLiant 360 Gen10 server, <https://www.hpe.com/psnow/doc/a00008159enw>

Software

The following software packages were installed for our configuration.

Software components deployment

The tables below provide the details of software components deployed on storage or compute and control nodes, which are distributed across the cluster as we tested on a single-rack solution.

Management node

The management node of the Hadoop runs the Hadoop master processes such as Cloudera Manager, and applications that submit jobs to the cluster.



Table 2 shows the management node software requirements.

TABLE 2. Management node basic software components

Software	Description
Red Hat® Enterprise Linux 7.9	Recommended Operating System
Open JDK 1.8.0	Java Development Kit
CM 7.3	Cloudera Manager
Postgress SQL	Database to store the Cloudera Manager configuration
ZooKeeper	Cluster coordination service

Head node

The head node of the Hadoop cluster runs the Hadoop master processes such as NameNode, YARN ResourceManager, and other master processes including gateway services of each component.

Table 3 shows the head node software requirements.

TABLE 3. Head node basic software components

Software	Description
Red Hat Enterprise Linux 7.9	Recommended Operating System
Open JDK 1.8.0	Java Development Kit
ResourceManager	YARN ResourceManager
NameNode	HDFS NameNode
Secondary Name Node	Secondary HDFS Name Node service (on the second head node)
History Server	Job History for ResourceManager
Hive server2	Hive Service to run SQL-like ad hoc queries
Hive Metastore	Central Repository of Apache Hive (tables and schema) metadata
Impala Catalog server	Metadata and metastore for Impala
Impala state store	Checks on the health of Impala daemons
ZooKeeper™	Cluster coordination service
Oozie	Oozie Workflow scheduler service
HBase Master	The HBase Master for the Hadoop cluster (Only if running HBase)
Spark Thrift Server	The tool on Hive Server2 for multiple remote client access
Spark History Server	Web UI for Spark application which enables events collection

HDFS Data Node

The storage node hosts are used for storing huge amounts of data in a distributed manner along with fault tolerance. It includes core services of Hadoop and cluster management services.

Table 4 shows the storage or compute node software requirements.

TABLE 4. Storage or compute node software components

Software	Description
Red Hat® Enterprise Linux® 8.1	Recommended Operating System
Open JDK 1.8.0	Java Development Kit
Node Manager	Yarn node manager service
DataNode	DataNode process for HDFS



Software	Description
Hive client	Hive client services for communication
Impala Daemon	Service to accept queries and distribute work across the cluster
Spark SQL	Processing Engine for structured data processing
Spark Streaming	Stream Processing engine for data streams

SOLUTION ARCHITECTURE FOR HPE ELASTIC PLATFORM FOR BIG DATA ANALYTICS WITH CLUDERA DATA PLATFORM WITH OZONE

The HPE Elastic Platform for Analytics (EPA) with Cloudera Data Platform and Apache Ozone is designed to be a modular infrastructure foundation that can be scaled in a building block fashion to support different customer requirements. This platform supports various Big Data eco-system components such as Hive, Impala, Spark, etc. Ozone on HPE EPA has been tested and validated with 768TB/node storage capacity with 10 nodes.

Below are the pointers to consider Apache Ozone as an object store instead of HDFS for Hadoop clusters:

- Support for billions of files (to overcome the limitation of HDFS up to 300 million files).
- Looking for more than 100TB/node storage capacity.
- Need to handle a large number of small files.

Solution components

In this Reference Architecture, HPE EPA building blocks are the foundation of the solution that can be built together in various ways to solve the unique requirements. This section explains the detailed storage/compute, and control blocks based on the HPE Apollo 4510 and HPE ProLiant DL360 Gen10 server required to build and implement a Big Data Analytics solution with Apache Ozone on the HPE EPA BDO Platform. The blocks defined in this section may be modified (for example, processor model, memory, etc.) to address or meet the customer requirements accordingly.

The HPE BDO solution infrastructure blueprints are composed of four blocks: storage/compute blocks, control blocks, network blocks, and rack blocks. Below are the infrastructure building blocks required to implement the Hadoop solution with Apache Ozone.

Table 5 shows the solution components.

TABLE 5. Solution components

Block	Model
Control Block	HPE ProLiant DL360 Gen10
Storage/Worker Block	HPE Apollo 4510
Network Block	HPE Aruba 8325-48Y8C (25G x 40P SFP+ 40G/100G x 8P QSFP)
Rack Block	1200mm or 1075mm

Rack configuration

The single-rack Hadoop is designed to perform well as a single-rack cluster design also forms 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 solution is deployed with the design priorities like performance, ease of use, and fault tolerance, etc.



Figure 5 depicts the single-rack configuration for HPE EPA for BDO solution with CDP Private Base.

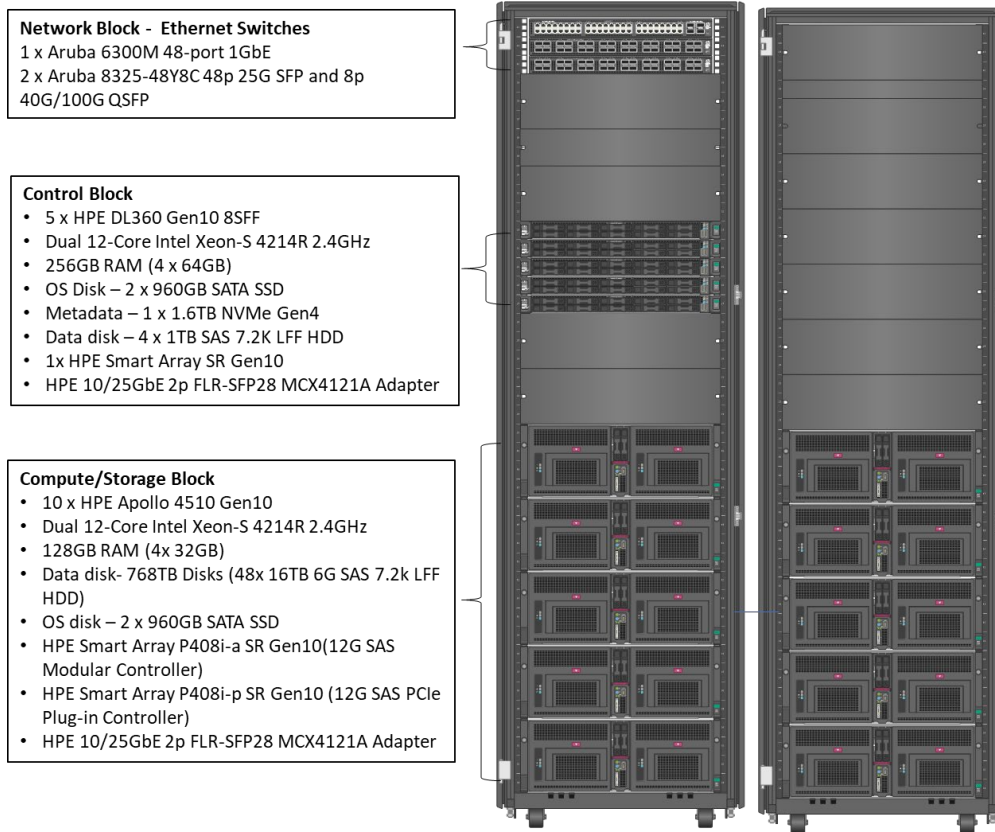


FIGURE 5. Basic conceptual diagram of single-rack for HPE EPA with Cloudera Data Platform Private Cloud Base with Ozone

Hardware

- **HPE Apollo 4510 (Storage/Compute Block):** The HPE Apollo 4510 Gen10 System offers revolutionary storage density and accessibility in a 4U rackmount form factor. Fitting in HPE Standard 1075 mm racks and featuring a front-drawer design for its bulk data drives, it has one of the highest capacities of any 4U storage server with standard rack depth. When you are running Big Data solutions, such as object storage, data analytics, deep archive, or other data-intensive workloads, the HPE Apollo 4510 Gen10 System allows you to save valuable data center space.
- **HPE ProLiant DL360 (Control Block):** HPE ProLiant DL360 Gen10 server delivers security, agility, and flexibility without compromise. The HPE ProLiant DL360 Gen10 is a highly flexible, performance-driven, and two-socket server that features Intel's new Scalable family of processors in a densely-packed 1U form factor.

Software

The following software packages were installed for our configuration.

Software components deployment

The tables below provide the details of software components deployed on storage or compute and control compute nodes, which are distributed across the cluster as we tested on a single-rack solution.

Management node

The management node of the Hadoop runs the Hadoop master processes such as Cloudera Manager, and client applications that submit jobs to the cluster.



Table 6 shows the management node software requirements.

TABLE 6. Management node basic software components

Software	Description
Red Hat Enterprise Linux 7.9	Recommended Operating System
Open JDK 1.8.0	Java Development Kit
CM	Cloudera Manager
Postgress SQL	Database to store the Cloudera Manager configuration
ZooKeeper	Cluster coordination service

Master nodes

The Ozone master nodes of the cluster run Ozone master services like Ozone Manager, Storage Container Manager, Ozone Recon, etc. along with other Hadoop master processes such as YARN ResourceManager and other master processes including gateway services of each component.

Table 7 shows the master node software requirements.

TABLE 7. Master node basic software components

Software	Description
Red Hat Enterprise Linux 7.9	Recommended Operating System
Open JDK 1.8.0	Java Development Kit
Ozone Gateway	Daemon service
Storage Container Manager	Block Manager
Ozone Manager	Responsible for all volume, bucket, and key operations
Ozone Recon	Web UI for Ozone analytics
Ozone Prometheus	Metric collector for Ozone
ResourceManager	YARN ResourceManager
History Server	Job History for ResourceManager
Hive server2	Hive Service to run SQL-like ad hoc queries
Hive Metastore	Central Repository of Apache Hive (tables and schema) metadata
Impala Catalog server	Metadata and metastore for Impala
Impala state store	Checks on the health of Impala daemons
ZooKeeper	Cluster coordination service
Oozie	Oozie Workflow scheduler service
HBase Master	The HBase Master for the Hadoop cluster (Only if running HBase)
Spark Thrift Server	The tool on Hive Server2 for multiple remote client access
Spark History Server	Web UI for Spark application which enables events collection

Apache Ozone DataNode

The storage node hosts are used for storing huge amounts of data in a distributed manner along with fault tolerance. It includes core services of Apache Ozone to support scaling billions of objects of varying sizes. The below services deployed on Ozone DataNodes assume – the customer is running only Ozone cluster services and not Hadoop.

If the customer is planning to deploy co-located Ozone and HDFS services, see the [best practices](#) section for the data node service layout on the cluster.



Table 8 shows the Storage or Compute node software requirements.

TABLE 8. Storage or compute node software components

Software	Description
Red Hat® Enterprise Linux® 7.9	Recommended Operating System
Oracle® JDK 11	Java Development Kit
Node Manager	Yarn node manager service
Ozone DataNode	DataNode process for Ozone
Hive client	Hive client services for communication
Impala Daemon	Service to accept queries and distribute work across the cluster

BEST PRACTICES AND CONFIGURATION GUIDANCE FOR THE SOLUTION

General hardware and software guidelines

Base cluster set-up requires system-level optimizations, detailed below while setting up the initial cluster to obtain a production-grade environment to scale to the high-throughput requirement for streaming analytics. These settings may require additional, specific optimizations as per the type of workloads.

OS tuning

As a general recommendation, update to the latest patch level available to improve stability and optimize performance.

1. Enable defaults, nodiratime, noatime (/etc/fstab).
2. Do not use logical volume manager (LVM).
3. Tune the OS block read ahead to 8K (/etc/rc.d/rc.local):

```

        blockdev --setra 8192 <storage device>
for f in /sys/block/sd*; do
    val=$(cat $f/queue/rotational)
    if [[ "$val" -eq "0" ]]; then
        echo Disk $f configured as SSD
        echo noop > $f/queue/scheduler
        echo 4096 > $f/queue/read_ahead_kb
        echo 32 > $f/queue/nr_requests
        echo 2 > $f/queue/nomerges
        echo 2 > $f/queue/rq_affinity
        echo 1024 > $f/device/queue_depth
    else
        echo Disk $f configured as HDD
        echo deadline > $f/queue/scheduler
        echo 4096 > $f/queue/read_ahead_kb
        echo 16 > $f/queue/nr_requests
        echo 0 > $f/queue/nomerges
        echo 2 > $f/queue/rq_affinity
        echo 1024 > $f/device/queue_depth
    fi
done

```

4. Decrease kernel swappiness to minimum 1:

```
Set sysctl vm.swappiness=1 in /etc/sysctl.conf
```

5. Tune ulimits for number of open files to a high number:

```

Example: in /etc/security/limits.conf:
### Hadoop specific optimisations
soft nproc 65536

```



```

hard nproc 65536
soft nofile 262144
hard nofile 262144
# End of file
soft memlock unlimited
hard memlock unlimited
nofile 65536
nproc 64000
memlock unlimited
core unlimited
nice -10

```

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 the boot, append the following to the kernel boot line in /etc/grub.conf:

```
elevator=deadline
```

- Install and configure Chrony to ensure clocks on each node are in sync with the management node.

- Set tuned profile network-latency for the server. The profile for low latency network tuning also disables Transparent Huge Pages (THP), NUMA balancing, and tunes several other network related sysctl parameters:

```
tuned-adm profile network-latency
```

- For good performance improvements, disable transparent huge page compaction:

```
echo never > /sys/kernel/mm/transparent_hugepage/enabled
```

- Disable SELinux on RHEL 7 by editing /etc/selinux/config and setting SELINUX=disabled and can be enabled post-Cloudera installation.

HDFS Tuning

Below are some of the changes to the cluster configuration.

- Yarn Specific configurations:

```

mapreduce.map.cpu.vcores
yarn.scheduler.maximum-allocation-vcores 90
yarn.scheduler.maximum-allocation-mb 124732 [121.81 GiB]
yarn.nodemanager.resource.cpu-vcores 64
yarn.nodemanager.resource.memory-mb 200
mapreduce.reduce.memory.mb 5GiB
mapreduce.map.memory.mb 2GiB
mapreduce.task.io.sort.factor 65
mapreduce.job.reduce.slowstart.completedmaps 0.7

```

- Hive Specific configurations:

```

hive.exec.reducers.bytes.per.reducer 512MiB
hive.metastore.server.max.message.size 819 MiB
hive.exec.reducers.bytes.per.reducer 256 MiB

```

Ozone Tuning

- Below are some of the configurations changes recommended on Ozone:

```

dfs.container.ratis.leader.pending.bytes.limit 2GiB
ozone.scm.datanode.pipeline.limit 8
ozone.scm.pipeline.owner.container.count 10
hdds.container.report.interval 1h
CM config 100000 [ulimit]
om_max_heap_size 31GB
scm_max_heap_size 31GB
ozone_datanode_heap_size 31GB

```



```
ozone_recon_max_heap_size 16GB
ozone_S3_gateway_max_heap_size 16GB
```

- Below services need to be configured to point to the SSDs:

```
dfs.container.ratis.datanode.storage.dir: Should be setup on SSD on each worker node
ozone.scm.ratis.storage.dir: Should be setup on SSD on 1 master node
ozone.scm.db.dirs: Should be setup on SSD on 3 master nodes
ozone.om.ratis.storage.dir: Should be setup on SSD on 3 master nodes
ozone.om.db.dirs: Should be setup on SSD on 3 master nodes
```

Table 9 shows the detailed Ozone services deployed on master nodes of the cluster.

TABLE 9. Ozone Master node services

Software	Ozone Master1	Ozone Master2	Ozone Master3	Ozone Master4	Ozone Master5
Ozone Gateway	Y	Y	Y	Y	Y
SCM	Y				
Ozone Prometheus					Y
Recon					Y
Ozone Ratis		Y	Y	Y	
Ozone S3 Gateway					Y
Ozone Manager		Y	Y	Y	

Additional scenarios and considerations for running Ozone and HDFS

Apache Ozone can be co-located with HDFS with unified security and governance policies, and offer seamless application portability. Depending on the workload requirement, Ozone can either be run on dedicated nodes or can be co-located with HDFS.

When we have dedicated Ozone DataNodes, we can leverage YARN queues. YARN queues can be set up to reflect the resource isolation requirements or access restrictions needed by various organizations. To enforce resource isolation, we can create multiple queues, and we can attach compute resources to each queue. Hence workloads placed on a specific queue will be executed on resources associated with that specific queue. In the case of dedicated Ozone nodes, the compute workload can be made to run on Ozone Storage nodes for better performance and isolation. To segregate the workloads, YARN queue managers could be configured. To start with we can have a default queue for HDFS and an Ozone Queue. The Ozone workloads can be placed on Ozone Queues.

The following services need to be included if HDFS and Ozone are being co-located.

- On the master node, we need to add HDFS services such as NameNode and Secondary NameNode along with the other service mentioned in Table 7.
- On compute or storage nodes add HDFS gateway client service along with other components shown in Table 8.

NOTE

The Apache Ozone has **High Availability (HA)** tested for Ozone data nodes when the cluster capacity was 70%. The 500TB data of a single data node, was successfully rebuilt during the node failure test.

Server best practices

HPE ProLiant DL380 Gen10 and HPE Apollo 4510 Gen10 Server BIOS Configuration

Hewlett Packard Enterprise recommends changing the default BIOS settings to the following using Workload Profile High-Performance Compute (HPC) on all servers hosting Hadoop to ensure the highest performance.



TABLE 10. HPE Apollo 4510 and HPE ProLiant DL380, DL360 Gen10 BIOS settings

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

HPE SIZER FOR THE ELASTIC PLATFORM FOR BIG DATA ANALYTICS

Hewlett Packard Enterprise has developed the [HPE Sizer for the Elastic Platform for Big Data Analytics](#) to assist customers with the proper sizing of these environments. Based on design requirements, the sizer will provide a suggested bill of materials (BOM) and metrics data for a Big Data cluster which can be modified further to meet customer requirements.

SUMMARY

Hewlett Packard Enterprise and Cloudera 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 Reference Architecture provides several configurations for deploying clusters of varying sizes with the Cloudera Data Platform Private Cloud Base 7.1 on HPE infrastructure and management software. These configurations leverage Hewlett Packard Enterprise to balance building blocks of servers, storage, and networking, along with integrated management software and bundled support. In addition, this Reference Architecture has been



created to assist in the rapid design and deployment of Cloudera Data Platform Private Cloud Base software on HPE infrastructure for clusters of various sizes.

NOTE

Part numbers are at the time of publication/testing 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 Hewlett Packard Enterprise Reseller or Hewlett Packard Enterprise Sales Representative for more details. hpe.com/us/en/services/consulting.html.



RESOURCES AND ADDITIONAL LINKS

HPE Servers for Big Data Analytics and Hadoop, hpe.com/info/Hadoop

HPE products link, <https://buy.hpe.com/us/en/>

HPE ProLiant 380 Gen10, <https://www.hpe.com/psnow/doc/a00008180enw>

HPE ProLiant 360 Gen10, <https://www.hpe.com/psnow/doc/a00008159enw>

Cloudera Data Platform: <https://www.cloudera.com/products/cloudera-data-platform.html>

HPE Networking, hpe.com/networking

HPE Pointnext, hpe.com/services

HPE Education Services, <https://education.hpe.com/ww/en/training/portfolio/bigdata.html>

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

© Copyright 2021, 2023-2024 Hewlett Packard Enterprise Development LP. The information contained herein is subject to change without notice. The only warranties for Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty.

Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

Intel is a trademark of Intel Corporation in the U.S. and other countries. Red Hat Enterprise Linux is a registered trademark of Red Hat, Inc. in the United States and other countries. All third-party marks are property of their respective owners.

a50004899enw, August 2024