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

Using HPE ProLiant DL325 Gen11 Server and HPE ProLiant DL345 Gen11 Server

CONTENTS

EXECUTIVE SUMMARY

Big Data Analytics have moved far beyond just utilizing legacy Hadoop clusters for running batch analytics. There is now a plethora of workloads highly dependent upon the Data Pipeline and Data Analytics, including interactive analytics, streaming analytics, AI/ML, etc. It's no surprise to any enterprise that data continues to grow exponentially, and that data growth is accelerating as more data comes in off the edge from remote locations. The challenge is the majority of enterprises are struggling to effectively manage all this data and to ultimately exploit that data to deliver better business outcomes, improved customer experiences, and to deliver new business insights. Enterprises have to do more than to just store this data. They need to effectively manage that data across a data pipeline to enable advanced analytics, train models, and to serve data intensive workloads. Hewlett Packard Enterprise (HPE) and Cloudera allow you to derive new business insights from all your data by providing a platform to store, manage, and process dense data at scale. This Reference Architecture provides several performance-optimized configurations for deploying Cloudera Data Platform (CDP) on clusters using 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 highly modular, disaggregated architecture to address the need for scalable, dense storage, and to act as an 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 Analytics workloads. It supports a wide range of workloads ranging from Extraction, Transform and 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 (Symmetric) or disaggregated (Asymmetric) 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 system (Asymmetric Architecture)** – HPE's recommended direction is to deploy Asymmetric or **Workload and Density Optimized** (WDO) architectures. These harness 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. In this testing the HPE WDO system is based on the HPE ProLiant DL325 Gen11 and HPE ProLiant DL345 Gen11.
- **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 frequently utilizes the HPE ProLiant DL325 Gen11 and DL345 Gen11 server platform, with density optimized variants.

This Reference Architecture has been created to assist in the rapid design and deployment of CDP Private Cloud Base v7.1.9 on HPE Elastic Platform for Big Data Analytics (EPA) with Asymmetric architecture for various sizes of HDFS clusters. This paper highlights recognizable benefits and provides guidance on building CDP for HDFS clusters that meets business needs.

Document purpose: This Reference Architecture provides comprehensive architectural guidelines and implementation of Big Data Analytics solutions on HPE EPA asymmetric architecture using CDP Private Cloud Base 7.1.9. 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 batch and real-time processing capabilities in their existing or new deployments.

This document describes the high-level design, performance results, and best practices for deploying CDP Private Cloud Base on bare metal HPE infrastructure.

The purpose of this document is to describe a Reference Architecture, highlighting recognizable benefits to technical audiences.

This Reference Architecture describes solution testing performed in December 2023.

SOLUTION OVERVIEW

Modern big data analytics environments must operate at large scale with high performance and are usually implemented as scalable clusters with compute. As these environments grow, operational and infrastructure complexity can limit the agility and flexibility that are required to support changing workloads and storage demands.

One technique used to simplify these environments is to separate the compute and storage functions. This approach allows independent scaling and management of compute and storage but must be carefully designed to avoid performance bottlenecks.

This architecture describes an implementation of CDP Private Cloud Base 7.1.9 on HPE infrastructure with Compute.

The Reference Architecture configurations are based on the CDP Private Cloud Base version 7.1.9 and HPE EPA WDO systems which include the HPE ProLiant DL325 Gen11 and HPE ProLiant DL345 Gen11 servers. Our solution, HPE EPA combining with CDP addresses the challenges and helps the organizations implement Big Data Analytics solution on-prem with Hadoop Distributed File System (HDFS).

This document describes the high-level design, performance results, and best practices for deploying CDP Private Cloud Base on bare-metal infrastructure with HPE ProLiant D325 Gen11as compute and HPE ProLiant DL345 as Storage.

HPE Elastic Platform for Big Data Analytics

Historically, Hadoop and HDFS have been deployed with storage and compute co-located and the main workload was a batch analytics process. The challenge with this design model was 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 cost unnecessarily. 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 requires a disaggregated, scalable, and flexible design running on a high-performing platform.

Figure 1 illustrates an example of the data flow from the edge to the 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 AI modeling.

- Edge Acquisition and Edge Analytics are where data are 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, it 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 are deployed at the core data center or in 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.

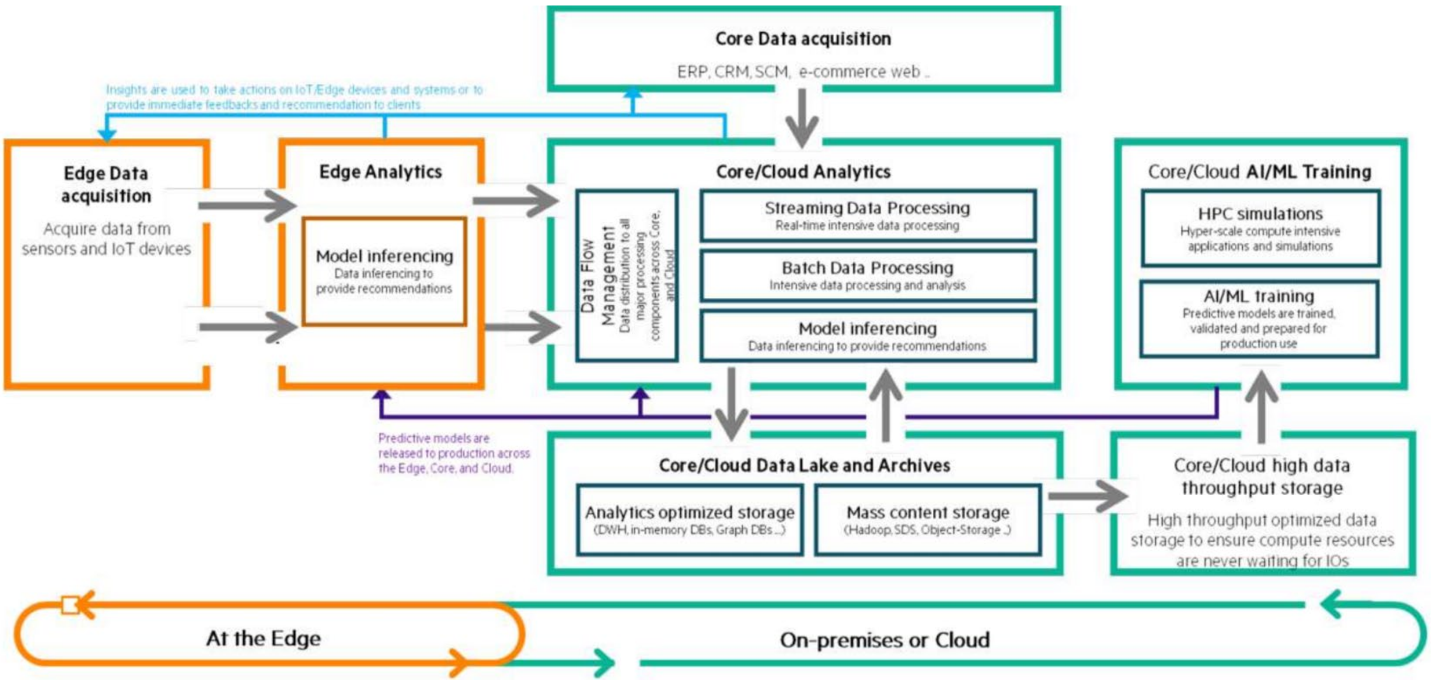

FIGURE 1. HPE Elastic Platform for Big Data Analytics (EPA) end-to-end data pipeline

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

CDP Private Cloud Base

CDP Private Cloud Base is an on-premises version of Cloudera Data Platform.

This new offering combines the best of Cloudera Enterprise Data Hub and Hortonworks Data Platform Enterprise, as well as new features and enhancements throughout the stack. This unified distribution is a scalable and adaptable platform through which we can securely run a wide range of workloads.

CDP Private Cloud Base supports a variety of hybrid solutions, including workloads created with CDP Private Cloud Data Services, in which compute tasks are separated from data storage and data can be accessed from remote clusters. By managing storage, table schema, authentication, authorization, and governance, this hybrid approach serves as a foundation for containerized applications.

The CDP Private Cloud Base shown in Figure 2 is made up of many different components, including Apache HDFS, Apache Hive 3, Apache HBase, and Apache Impala, as well as many others for specialized workloads. You can use any combination of these services to build clusters that meet your specific business needs and workloads. For common workloads, several pre-configured service packages are also available.

FIGURE 2. Architecture of CDP Private Cloud Base

These include the following:

Regular (Base) clusters

Develop and serve predictive models using the data engineering process. HDFS, Yet Another Resource Negotiator (YARN), YARN Queue Manager, Ranger, Atlas, Hive, Hive on Tez, Spark, Oozie, Hue and Iceberg were among the services provided. Interactively browse, query, and explore your data.

Data engineering

Process develop and serve predictive models.

Services included: HDFS, YARN, YARN Queue Manager, Ranger, Atlas, Hive, Hive on Tez, Spark, Oozie and Hue.

Data mart

Browse, query, and explore your data in an interactive way.

Services included: HDFS, Ranger, Atlas, Hive, and Hue

Operational database

Real-time insights for modern data-driven business.

Services included: HDFS, Ranger, Atlas, and HBase

Custom services

When you install a CDP Private Cloud Base cluster, you only need to install Cloudera Runtime, which contains all of the components. See Cloudera Runtime Component Versions for a complete list of the included components.

CDP Private Cloud Base includes powerful tools to help manage, govern, and secure your cluster in addition to the Cloudera Runtime components.

CDP Private Cloud Base tools

Cloudera Manager

Cloudera Manger is used by CDP Private Cloud Base to manage one or more clusters and their configurations, as well as to monitor cluster performance. Cloudera Manager is also used for managing installations, upgrades, maintenance workflows, encryption, access controls, and data replication. Cloudera Manager can also be used to manage Cloudera Enterprise CDH clusters. Cloudera Manager can be used to set up a virtual private cluster, which allows you to separate compute resources from data storage and share data storage among compute resources.

Apache Atlas

Apache Atlas, which is used to provide data governance, is also included in the CDP Private Cloud Base. Apache Atlas acts as a common metadata store, allowing metadata to be exchanged both inside and outside of the CDP stack. Because Atlas and Apache Ranger are tightly integrated, you can define, administer, and manage security and compliance policies consistently across all components of the CDP stack. Atlas replaces the Cloudera Navigator Metadata Server for Cloudera Enterprise customers. It offers the following capabilities:

- Flexible metadata models
- Entity search using model attributes, classifications (tags), and free text.
- Lineage across entities is based on processes applied to the entities.

Apache Ranger

Apache Ranger manages your CDP Private Cloud Base cluster auditing, authentication, and authorization.

Apache Ranger offers a centralized framework for collecting access audit history and reporting data, including parameter filtering. Through this centralized reporting capability, Ranger enhances audit information obtained from CDP components and provides insights.

Apache Ranger also manages access control via the user interface, ensuring that policies are applied consistently across CDP Private Cloud Base components. Security administrators can define database, table, column, and file security policies, as well as manage permissions for specific LDAP-based groups or individual users. Dynamic conditions such as time or geolocation can also be used to supplement an existing policy rule. Using a service-based definition, the Ranger authorization model can be easily extended to any data source.

For customers familiar with Cloudera Enterprise, Apache Ranger replaces Sentry and Navigator Audit Server and also provides the following capabilities:

- Better fine-grained access controls:
  - Dynamic Row Filtering
  - Dynamic Column Masking
  - Attribute-based Access Control
  - Spark SQL fine-grained access control
- Rich policy features
  - Allow/Deny constructs, Custom policy conditions/context enrichers, time bound policies, Atlas integration (for tag-based policies)
- Extensive Access Auditing with rich event metadata

HPE ProLiant DL325 Gen11 Server and HPE ProLiant DL345 Gen11 Server

The HPE ProLiant DL325 Gen11 server is a low-cost 1U 1P solution that delivers exceptional value balancing compute, memory, and network bandwidth at 1P economics. Powered by 4th Generation AMD EPYC™ Processors with up to 128 cores, increased memory bandwidth (up to 3 TB), high-speed PCIe Gen5 I/O and EDSFF storage, and supporting up to 2 GPUs at the front, this server is a superb low-cost, 1U 1P, performance solution for your virtualized workloads. The silicon root of trust anchors the server firmware, creating a fingerprint for the AMD Secure Processor that must be matched exactly before the server boot.

FIGURE 3. HPE ProLiant DL325 Gen11 server

The HPE ProLiant DL345 Gen11 server is a scalable 2U 1P solution that delivers exceptional compute performance and large-capacity storage options at 1P economics. The server is powered by 4th Generation AMD EPYC™ Processors with up to 128 cores, increased memory bandwidth (up to 3 TB), high-speed PCIe Gen5 I/O, and EDSFF storage up to 20 LFF/ 34 SFF/ 36 EDSFF drives, and up to four GPUs at the front. Provides impressive storage performance and options for data-intensive workloads such as software-defined storage, video transcoding, and virtual apps

FIGURE 4. HPE ProLiant DL345 Gen11 server

Aruba CX 8325 Switch Series

The Aruba CX 8325 Switch Series offers a flexible and innovative approach to addressing the application, security, and scalability demands of the mobile, cloud and IoT era. These switches serve the needs of the next generation core and aggregation layer, as well as emerging data center requirements at the Top of Rack (ToR) and End of Row (EoR). They provide over 6.4Tbps of capacity, with line-rate Gigabit Ethernet interfaces including 1Gbps, 10Gbps, 25Gbps, 40Gbps, and 100Gbps. The Aruba CX 8325 switch series includes industry-leading line rate ports 1/10/25GbE (SFP/SFP+/SFP28) and 40/100GbE (QSFP+/QSFP28) with connectivity in a compact 1U form factor. These switches offer a fantastic investment for customers wanting to migrate from older 1GbE/10GbE to faster 25GbE, or 10GbE/40GbE to 100GbE ports.

FIGURE 5. Aruba CX 8325 Switch

Aruba 6300F Switch series

The Aruba CX 6300 Switch series is a modern, flexible, and intelligent family of AOS-CX stackable switches ideal for access, aggregation, and data center top-of-rack (TOR) deployments. With a cloud-centric design that combines a fully programmable OS with the Aruba Network Analytics Engine, the Aruba CX 6300 extends industry-leading monitoring and troubleshooting capabilities to the access layer. Support of Aruba NetEdit and the Aruba CX Mobile App verify that configurations are flawless and easy to deploy.

A powerful Aruba Gen7 ASIC architecture delivers fast, non-blocking performance, meaning your network is ready for tomorrow's unpredictable demands. Aruba Virtual Stacking Framework (VSF) allows for stacking of up to 10 switches, providing scale and simplified management. This flexible series has built-in highspeed uplinks and supports high density IEEE 802.3bt high power PoE with HPE Smart Rate multi-gigabit Ethernet for high-speed APs and IoT devices.

FIGURE 6. Aruba 6300F Switch

Solution architecture

The HPE EPA building blocks are the foundation of this Reference Architecture. The components can be combined in various ways to solve and address unique requirements. This section explains the detailed compute, and control blocks based on the HPE ProLiant DL325 Gen11 and HPE ProLiant DL345 Gen11 servers required to build and implement a CDP Big Data Analytics solution on the HPE EPA WDO 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 EPA WDO solution infrastructure blueprints are composed of five blocks: control blocks, storage, compute blocks, network blocks, and rack blocks.

Following is the infrastructure of building blocks required to implement the solution with HDFS:

In this solution architecture, 7 (Seven) HPE ProLiant DL325 Gen 11 Servers and 4 (four) HPE ProLiant DL345 Gen11 servers were used to build Cloudera Data Platform (CDP) Private Cloud Base Cluster. Deployment of the cluster should be planned and this reference link CDP Private Cloud Base Installation Guide | CDP Private Cloud (cloudera.com) provides guidance to build the cluster.

Figure 7 depicts the single-rack configuration for HPE EPA for WDO solution with CDP Private Cloud Base.

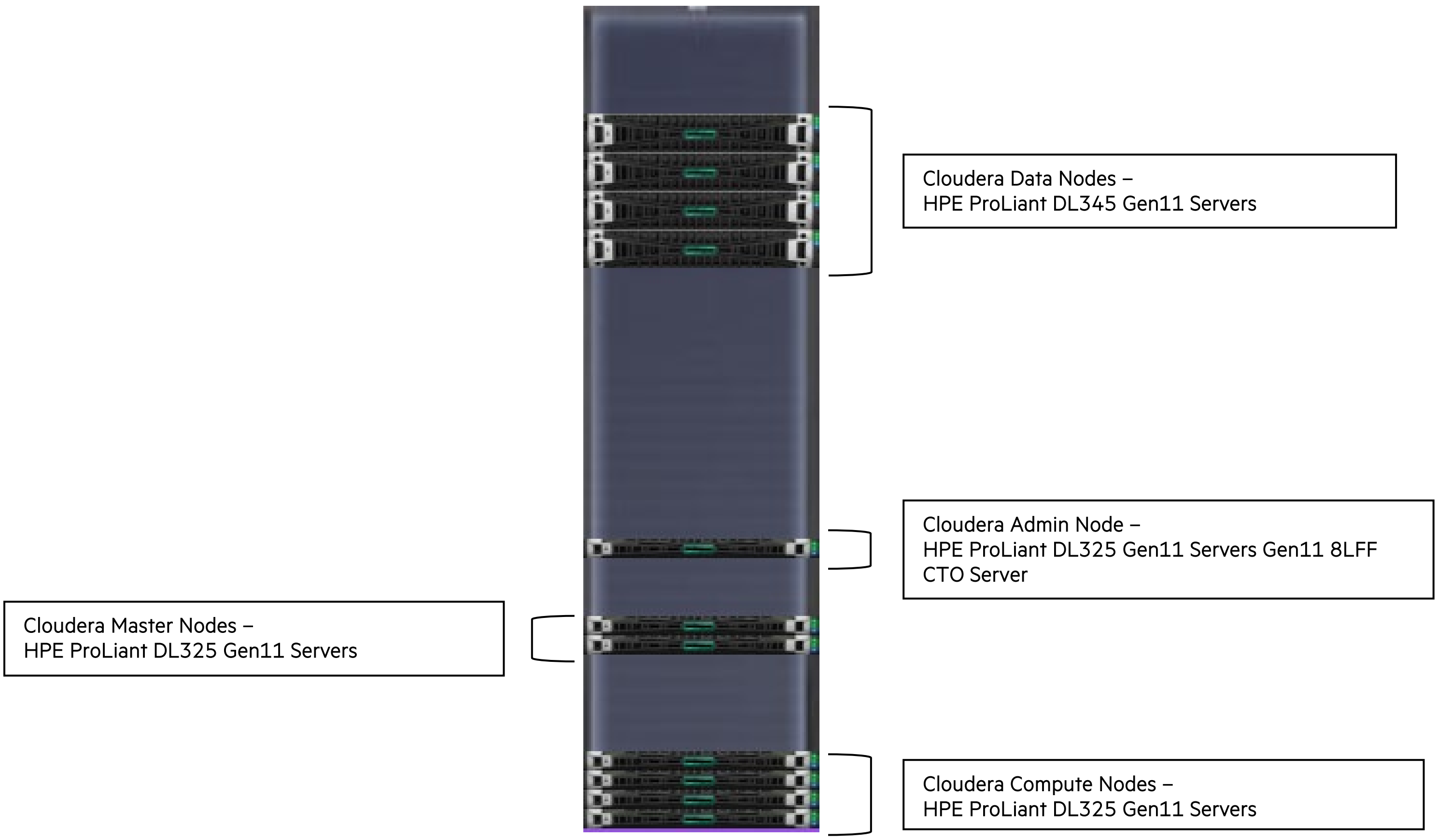

FIGURE 7. Basic conceptual diagram of single rack for HPE EPA with CDP Private Cloud Base with HDFS

Initially all the servers were deployed with Red Hat Enterprise Linux 9.3 Operating system and prerequisites were configured to meet the requirements. It is important to prepare each node before deploying the CDP Private Cloud Base cluster.

TABLE 1. Nodes for each Block

Block	Model
Control Block	HPE ProLiant DL325 Gen11 Servers
Storage Block	HPE ProLiant DL345 Gen11 Servers
Compute Block	HPE ProLiant DL325 Gen11 Servers
Network Block	Aruba 8325-48Y8C 48p 25G SFP/+/28 8p 25G QSFP Aruba 6300F 48-port 1GbE and 4-port SFP56
Rack Block	1200mm or 1075mm

In this solution, the Cloudera CDP Cluster was designed with two master nodes, four compute nodes and four data nodes

Table 2, Table 3, and Table 4 show the services configured on the servers and roles.

TABLE 2. HPE Cloudera Data Platform 7.1.9 Services and roles

HPE ProLiant DL325 Gen11 Servers	Role	Services
Server No.1	Management Node	Cloudera Manager
Server No.2	Utility Node	Hive Metastore Server, Hue Load Balancer, Hue Server, Hue Kerberos Ticket Renewer, Impala Catalog Server, Cloudera Management Service Alert Publisher, Cloudera Management Service Event Server, Cloudera Management Service Host Monitor, Cloudera Management Service Service Monitor, ZooKeeper Server
Server No.3	Master node1	Atlas Server, Hbase RegionServer, HDFS NameNode, Hive Gateway, Impala Daemon, Solr Server, Tez Gateway, YARN JobHistory Server, YARN ResourceManager, ZooKeeper Server
Server No.4	Master Node2	HDFS SecondaryNameNode, Hive on Tez HiveServer2, YARN ResourceManager, ZooKeeper Server

TABLE 3. HPE ProLiant DL325 Gen11 Servers

HPE ProLiant DL325 Gen11 Servers	Role	Services
Server No.5	Compute Node / Worker Node	YARN NodeManager
Server No.6	Compute Node / Worker Node	YARN NodeManager
Server No.7	Compute Node / Worker Node	YARN NodeManager
Server No.8	Compute Node / Worker Node	HBase Master, Hive Gateway, Tez Gateway, YARN NodeManager

TABLE 4. HPE Proliant DL3325 Gen11

HPE Proliant DL325 Gen11 Servers	Role	Services
Server No.9	Storage Node / Worker Node	HDFS DataNode
Server No.10	Storage Node / Worker Node	HDFS DataNode, NiFi Node
Server No.11	Storage Node / Worker Node	HDFS DataNode, NiFi Registry
Server No.12	Storage Node / Worker Node	HDFS DataNode, HBase Thrift Server

SOLUTION COMPONENTS

The solution consists of seven HPE ProLiant DL325 Gen11 servers, 4 HPE ProLiant DL345 Gen11 servers

TABLE 5. Hardware

Component	Qty	description
HPE ProLiant DL325 Gen11 8SFF CTO Server 1 x AMD EPYC 9224P (2.5GHz/24-core) processor 12x 16GB DDR5-4800MHz (192GB total) 1 x HPE MR416-p Gen11 x16 Lanes 8GB Cache SPDM PCI Storage Controller 1 x HPE NS204i-u Gen11 NVMe Hot Plug Boot Optimized storage device 2 x 1.92TB 12G SAS RI SFF SSD (Hive Metastore DB) RAID-1 2 x 1.92TB 12G SAS RI SFF SSD (NameNode metadata) RAID-1 2 x 1.92TB 12G SAS RI SFF SSD (ZooKeeper) 2 x 1.92TB 12G SAS RI SFF SSD (Quorum JournalNode) 1 x Mellanox MCX631102AS-ADAT Ethernet 10/25Gb 2-port SFP28 PCIe Adapter 1 x Broadcom BCM5719 Ethernet 1Gb 4-port Base-T OCP3 Adapter 2 x HPE 800W Flex Slot Platinum Hot Plug Low Halogen Power Supply Kit	1	These servers were used for HPE Cloudera CDP Admin
HPE ProLiant DL325 Gen11 8SFF CTO Server 1 x AMD EPYC 9224P (2.5GHz/24-core) processor 12x 16GB DDR5-4800MHz (192GB total) 1 x HPE MR416-p Gen11 x16 Lanes 8GB Cache SPDM PCI Storage Controller 1 x HPE NS204i-u Gen11 NVMe Hot Plug Boot Optimized storage device 2 x 1.92TB 12G SAS RI SFF SSD (Hive Metastore DB) RAID-1 2 x 1.92TB 12G SAS RI SFF SSD (NameNode metadata) RAID-1 2 x 1.92TB 12G SAS RI SFF SSD (ZooKeeper) 2 x 1.92TB 12G SAS RI SFF SSD (Quorum JournalNode) 1 x Mellanox MCX631102AS-ADAT Ethernet 10/25Gb 2-port SFP28 PCIe Adapter 1 x Broadcom BCM5719 Ethernet 1Gb 4-port Base-T OCP3 Adapter 2 x HPE 800W Flex Slot Platinum Hot Plug Low Halogen Power Supply Kit	2	These servers were used for HPE Cloudera CDP Master
HPE ProLiant DL325 Gen11 Servers 1 x AMD EPYC 9654P (2.4GHz/96-core) processor 12x 64GB DDR5-4800MHz (768GB total) 1 x HPE MR216-p Gen11 x16 Lanes No Cache SPDM PCI Storage Controller 1 x HPE NS204i-u Gen11 NVMe Hot Plug Boot Optimized storage device 2 x 1.92TB 12G SAS RI SFF SSD (Spill-to-disk cache) JBOD 1 x Mellanox MCX631102AS-ADAT Ethernet 10/25Gb 2-port SFP28 PCIe Adapter 1 x Broadcom BCM5719 Ethernet 1Gb 4-port Base-T OCP3 Adapter 2 x HPE 800W Flex Slot Platinum Hot Plug Low Halogen Power Supply Kit	4	These servers were used for HPE Cloudera CDP compute
HPE ProLiant DL345 Gen11 Servers 1 x AMD EPYC 9354P (3.25GHz/32-core) processor 8x 32GB DDR5-4800MHz (256GB total) 1 x HPE MR216-p Gen11 x16 Lanes No Cache SPDM PCI Storage Controller 1 x HPE NS204i-u Gen11 NVMe Hot Plug Boot Optimized storage device 1 x HPE ProLiant DL345 Gen11 4LFF x1 SAS/SATA 12G Front Backplane Kit 12 x 8TB SAS 12G Business Critical 7.2k LFF HDD (HDFS Space) JBOD 1 x Mellanox MCX631102AS-ADAT Ethernet 10/25Gb 2-port SFP28 PCIe Adapter 1 x Broadcom BCM5719 Ethernet 1Gb 4-port Base-T OCP3 Adapter 2 x HPE 800W Flex Slot Platinum Hot Plug Low Halogen Power Supply Kit	4	These servers were used for HPE Cloudera CDP data

TABLE 6. Software

Software

Component	Version	Description
Red Hat Enterprise Linux	8.6	Operating System
Open JDK	1.8.0_262	Open Java Development Kit
CDP Private Cloud	7.1.9	Cloudera Data Platform Private Cloud

BEST PRACTICES AND CONFIGURATION GUIDANCE FOR SOLUTION

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 workload.

This section provides best practices and a recommended configuration for HPE CDP Private Cloud Base cluster running on HPE ProLiant DL325 Gen11 Servers and HPE ProLiant DL345 Gen11 servers. The settings were performed on all the nodes in the cluster.

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

1. Power on the server. During the POST, Press F9 for the server to boot to BIOS (System Utilities).
2. Select System Configuration menu.
3. Select BIOS/Platform Configuration (RBSU) menu.
4. Select Workload Profile. Select the Workload Profile as High_Performance_Compute option from the drop down menu.
5. Select Power and Performance Options. Select the Power Regulator as Static_High_Performance_Mode.

Red Hat Enterprise Linux settings

We have performed the following CPU, RAM, Security, and Network settings for optimal performance on all the nodes in the CDP Private Cloud Base Cluster.

CPU

1. Enable Multi-Threading.
2. Set the BIOS settings for CPU and memory to Maximum Performance mode.

Memory

1. Minimize anonymous page faults. Minimize anonymous page faults by setting `vm. swappiness = 1,` which frees them from the page cache before swapping application pages (this reduces the OOM-killer invocation).
2. Edit the /etc/sysctl.conf file in your editor of choice and add the following line.

```
vm. swappiness=1
```

Then, run the following:

```
# sysctl -p # sysctl -a|grep "vm. swappiness"
```

3. Disable transparent huge-page compaction.

```
echo "never" > /sys/kernel/mm/redhat_transparent_hugepage/enabled
```

4. Disable transparent huge-page defragmentation.

```
echo "never" > /sys/kernel/mm/redhat_transparent_hugepage/defrag
```

5. Add these commands to /etc/rc.local to ensure that transparent huge page compaction and defragmentation remain disabled across reboots.

Security

1. Disable SELinux - SELinux can be disabled on RHEL or CentOS by editing /etc/selinux/config and setting SELINUX=disabled. This change must be done as root (or with proper sudo access) and requires a reboot.
2. Disable Firewall - Disable the host-based firewalls on all the machines in the cluster.

```
systemctl stop firewalld. service
systemctl disable firewalld. Service
```

Network

1. Enable the NTP - Enable the NTP daemon by running the following commands.

```
systemctl start ntpd. service
systemctl enable ntpd.service
```

2. Compute nodes network tuning:
   a. Add the following parameters in /etc/sysctl.conf. Disable TCP timestamps to improve CPU utilization (this is an optional parameter and will depend on your NIC vendor).
   b. Disable TCP timestamps to improve CPU utilization (this is an optional parameter and depends on your NIC vendor): net.ipv4.tcp_timestamps=0.
3. Enable TCP sacks to improve throughput:

```
net.ipv4.tcp_sack=1
```

4. Increase the maximum length of processor input queues:

```
net. core. netdev_max_backlog=250000
```

5. Increase the TCP max and default buffer sizes using setsockopt ():

```
net. core. rmem_max=4194304
net. core. wmem_max=4194304
net. core. rmem_default=4194304
net. core. wmem_default=4194304
net. core. optmem_max=4194304
```

6. Increase memory thresholds to prevent packet dropping:

```
net.ipv4.tcp_rmem=4096 87380 4194304
net.ipv4.tcp_wmem=4096 65536 4194304
```

NOTE
If you want to run this from the command line, then quote the values being set. For example: sysctl -w net.ipv4.tcp_rmem=4096 87380 4194304.

7. Set the socket buffer to be divided evenly between TCP window size and application buffer.

```
net.ipv4.tcp_adv_win_scale=1
```

8. Verify NIC advanced features - Verify which features are available with your NIC using ethtool.

```
$ sudo ethtool -k <ethx>
```

9. There are various offload capabilities that modern NICs (and especially high-performance NICs) have, and it is always recommended to enable them. A few features such as tcp-segmentation-offload (TSO), scatter-gather (SG), and generic segmentation-offload (GSO) are good features to enable (if not enabled by default).
   a. NIC ring buffer configurations
   b. Check existing ring buffer sizes. ethtool -g <ethx>
10. After checking the preset maximum values and the current hardware settings, the following command can be used to resize the ring buffers:

```
# ethtool -G <interface> rx <newsize>
```

or

```
# ethtool -G <interface> tx <newsize>
```

CDP Private Cloud Base Cluster Configurations

CDP Private Cloud Base includes many default parameter settings. The Figure 8 depicts tuning parameters set to achieve the best performance from the testing cluster.

```
mapreduce.map.memory.mb=4096
mapreduce.map.java.opts=-Xmx3072m
mapreduce.map.cpu.vcores=1
mapreduce.map.output.compress=true
mapreduce.map.output.compress.codec=org.apache.hadoop.io.compress.Lz4Codec
mapreduce.map.speculative=true
mapreduce.reduce.memory.mb=4096
mapreduce.reduce.java.opts=-Xmx3072m
mapreduce.reduce.cpu.vcores=1
mapreduce.reduce.speculative=true
mapreduce.task.io.sort.factor=64
mapreduce.task.io.sort.mb=768
mapred.map.tasks=1248
mapred.reduce.tasks=624
mapreduce.job.reduce.slowstart.completedmaps=0.85
mapreduce.reduce.shuffle.parallelcopies=12
```

FIGURE 8. Tuning parameters

TESTING AND VALIDATION OF DEPLOYMENT

This section provides details of the validation done using HPE EPA to accomplish the following test objectives and use cases:

Testing the cluster:

1. Access cluster nodes via ssh – validate root level access to the cluster nodes.
2. Access Cloudera Manager – validate admin level access to Cloudera Manager.
3. In Cloudera Manager, ensure that required services are setup and running and that all health checks are Green.
4. Ensure that Cluster has been set up per proper Cloudera supported guidelines.
5. Run a PiEstimator job to manually verify that the CDP Private Cloud Base installation was successful.

NOTE

If you have a secure cluster, use the kinit command-line tool to authenticate Kerberos.

6. Log in to a host in the cluster.
7. Run the Hadoop PiEstimator example using the following command.

   ```
   yarn jar /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar
   pi 10 100
   ```
8. In Cloudera Manager, navigate to **Cluster > ClusterName > yarn Applications**.
9. Check the results of the job as shown in the following figure.

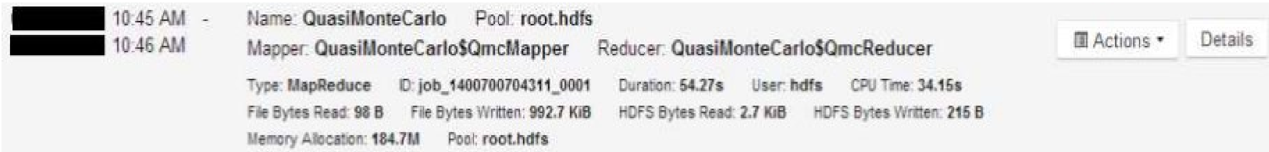

FIGURE 9. Example of the results

10. Validation of MapReduce operations using TeraSuite test:
    a. Run teragen with 1 mapper per Datanodes/Storage disk in the Cluster, with 1x replication.
    b. Run teragen with 1 mapper per datanode/Storage disk in the Cluster, with default (3x) replication.
    c. Run terasort with a varying number of reducers to find best throughput.

TeraSuite test

The solution was validated with the MapReduce benchmark tools Teragen and Terasort. The goal was to validate the functionality of different components and services rather than stress the cluster for maximum performance. This test was designed to exercise a default replication factor of 3. The solution was operationally validated with different workload sizes of 100GB, 500GB, and 1TB. The CDP Private Cloud Base Cluster performed optimally, indicating the resources such as I/O and network have been configured and tuned to optimally run and support different workload sizes.

The performance lab environment was configured with four (7) HPE ProLiant DL325 Gen11, and four(4) HPE ProLiant DL345 Gen11 servers.

The main configuration applied is the size of each mapper and reducer. Using some simple YARN calculations, we used 1 vCore for each container, and since the entire cluster has 240 vCores, this would yield 240 containers. Allocated memory was four (4) GB per container, resulting in 960 GB of the total 1.4 TB used for each run.

Steps

1. The figure 10 shows the **teragen** command to generate 1TB of data.

```
time hadoop jar /opt/cloudera/parcels/CDH/jars/hadoop-mapreduce-examples-3.1.1.7.1.7.1000-141.jar teragen \
-Dmapreduce.map.log.level=INFO -Dmapreduce.reduce.log.level=INFO \
-Dyarn.app.mapreduce.am.log.level=INFO -Dio.file.buffer.size=131072 \
-Ddfs.blocksize=256M -Ddfs.replication=1 -Dmapreduce.map.max.attempts=1 \
-Dmapreduce.map.memory.mb=4096 -Dmapreduce.map.java.opts=-Xmx3072m \
-Dmapreduce.map.cpu.vcores=1 -Dmapreduce.map.output.compress=true \
-Dmapreduce.map.output.compress.codec=org.apache.hadoop.io.compress.Lz4Codec \
-Dmapreduce.map.speculative=true -Dmapreduce.task.io.sort.factor=64 \
-Dmapreduce.task.io.sort.mb=768 -Dmapreduce.job.maps=1248 10000000000 \
-Dyarn.app.mapreduce.am.resource.mb=1024  -Dyarn.app.mapreduce.am.command.opts=768 \
/user/data/sandbox/poc/teragen
```

FIGURE 10. Teragen command

2. The Figure 12 shows the **terasort** command to sort the generated 1TB data.

```
time hadoop jar /opt/cloudera/parcels/CDH/jars/hadoop-mapreduce-examples-3.1.1.7.1.7.1000-141.jar terasort \
-Dmapreduce.map.log.level=INFO -Dmapreduce.reduce.log.level=INFO \
-Dyarn.app.mapreduce.am.log.level=INFO -Dio.file.buffer.size=131072 \
-Ddfs.blocksize=256M -Ddfs.replication=1 -Dmapreduce.map.max.attempts=1 \
-Dmapreduce.map.memory.mb=4096 -Dmapreduce.map.java.opts=-Xmx3072m \
-Dmapreduce.map.cpu.vcores=1 -Dmapreduce.map.output.compress=true \
-Dmapreduce.map.output.compress.codec=org.apache.hadoop.io.compress.Lz4Codec \
-Dmapreduce.map.speculative=true -Dmapreduce.task.io.sort.factor=64 \
-Dmapreduce.task.io.sort.mb=768 -Dmapreduce.job.maps=1248 \
-Dmapreduce.reduce.memory.mb=4096 -Dmapreduce.reduce.java.opts=-Xmx3072m \
-Dmapreduce.reduce.cpu.vcores=1 -Dmapreduce.reduce.maxattempts=1 \
-Dmapreduce.reduce.speculative=true -Dyarn.app.mapreduce.am.resource.mb=1024 \
-Dmapreduce.job.reduces=624 -Dyarn.app.mapreduce.am.command.opts=-Xmx768m \
-Dmapreduce.job.reduce.slowstart.completedmaps=0.85 \
-Dmapreduce.reduce.shuffle.parallelcopies=12 -Dmapreduce.terasort.output.replication=1 \
/user/data/sandbox/poc/teragen  /user/data/sandbox/poc/terasort
```

FIGURE 11. Terasort command

3. The Figure 13 shows the **teravalidate** command for the generated data and has been sorted.

```
time hadoop jar /opt/cloudera/parcels/CDH/jars/hadoop-mapreduce-examples-3.1.1.7.1.7.1000-141.jar teravalidate \
-Dio.file.buffer.size=131072 -Ddfs.blocksize=256M \
-Dmapreduce.map.memory.mb=4096 -Dmapreduce.map.java.opts=-Xmx3072m \
-Dmapreduce.reduce.memory.mb=4096 -Dmapreduce.reduce.java.opts=-Xmx3072m \
-Dmapreduce.task.io.sort.factor=64 -Dmapreduce.task.io.sort.mb=768 \
-Dmapreduce.job.maps=1248 -Dmapreduce.job.reduces=624 \
-Dyarn.app.mapreduce.am.resource.mb=1024  -Dyarn.app.mapreduce.am.command.opts=768 \
/user/data/sandbox/poc/terasort  /user/data/sandbox/poc/teravalidate
```

FIGURE 12. teravalidate command

With these test results, HPE validated the CDP Private Base Cluster HDFS and MapReduce layers with effective use of compute and storage resources. The goal of these tests was not to measure performance metrics. They were to confirm that all services deployed in the CDP Private Cloud Base cluster were working correctly.

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 CDP Private Cloud Base 7.1.9 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 CDP Private Cloud Base software on HPE infrastructure for clusters of various sizes.

APPENDIX A: BILL OF MATERIALS

This section provides the details of the Bill of Materials used for the solution built for this Reference Architecture. Based on the user needs the solution is flexible to scale out by adding symmetric nodes of HPE ProLiant DL325 Gen11 and HPE ProLiant DL345 Gen11.

NOTE

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

TABLE A1. Bill of materials

Part number	Quantity	Description
		Rack
P9K08A	1	HPE 42U 600mmx1075mm G2 Kitted Advanced Shock Rack with Side Panels and Baying
P9K08A 001	1	HPE Factory Express Base Racking Service
		Admin Nodes
P54199-B21	1	HPE DL325 Gen11 8SFF CTO Svr
P53704-B21	1	AMD EPYC 9354P 3.25GHz 32-core 280W Processor for HPE
P50311-B21	6	HPE 32GB (1x32GB) Dual Rank x8 DDR5-4800 CAS-40-39-39 EC8 Registered Smart Memory Kit
P47777-B21	1	HPE MR416-p Gen11 x16 Lanes 8GB Cache SPDM PCI Storage Controller
P48183-B21	1	HPE NS204i-u Gen11 NVMe Hot Plug Boot Optimized storage device
P16499-H21	2	HPE 3.2TB NVMe Gen4 High Performance Mixed Use SFF SC U.3 PM1735 SSD
P16499-H21	2	HPE 3.2TB NVMe Gen4 High Performance Mixed Use SFF SC U.3 PM1735 SSD
P16499-H21	2	HPE 3.2TB NVMe Gen4 High Performance Mixed Use SFF SC U.3 PM1735 SSD
P16499-H21	2	HPE 3.2TB NVMe Gen4 High Performance Mixed Use SFF SC U.3 PM1735 SSD
P42044-B21	1	Mellanox MCX631102AS-ADAT Ethernet 10/25Gb 2-port SFP28 PCIe Adapter
P51181-B21	1	Broadcom BCM5719 Ethernet 1Gb 4-port Base-T OCP3 Adapter
P38997-B21	2	HPE 1600W Flex Slot Platinum Hot Plug Low Halogen Power Supply Kit
		Master Nodes
P54199-B21	1	HPE DL325 Gen11 8SFF CTO Svr
P53704-B21	1	AMD EPYC 9354P 3.25GHz 32-core 280W Processor for HPE
P50311-B21	6	HPE 32GB (1x32GB) Dual Rank x8 DDR5-4800 CAS-40-39-39 EC8 Registered Smart Memory Kit
P47777-B21	1	HPE MR416-p Gen11 x16 Lanes 8GB Cache SPDM PCI Storage Controller
P48183-B21	1	HPE NS204i-u Gen11 NVMe Hot Plug Boot Optimized storage device
P16499-H21	2	HPE 3.2TB NVMe Gen4 High Performance Mixed Use SFF SC U.3 PM1735 SSD
P16499-H21	2	HPE 3.2TB NVMe Gen4 High Performance Mixed Use SFF SC U.3 PM1735 SSD
P16499-H21	2	HPE 3.2TB NVMe Gen4 High Performance Mixed Use SFF SC U.3 PM1735 SSD
P16499-H21	2	HPE 3.2TB NVMe Gen4 High Performance Mixed Use SFF SC U.3 PM1735 SSD
P42044-B21	1	Mellanox MCX631102AS-ADAT Ethernet 10/25Gb 2-port SFP28 PCIe Adapter
P51181-B21	1	Broadcom BCM5719 Ethernet 1Gb 4-port Base-T OCP3 Adapter

Part number	Quantity	Description
P38997-B21	2	HPE 1600W Flex Slot Platinum Hot Plug Low Halogen Power Supply Kit
		Compute Nodes
P54199-B21	1	HPE DL325 Gen11 8SFF CTO Svr
P53697-B21	1	AMD EPYC 9654P 2.4GHz 96-core 360W Processor for HPE
P50312-B21	12	HPE 64GB (1x64GB) Dual Rank x4 DDR5-4800 CAS-40-39-39 EC8 Registered Smart Memory Kit
P47789-B21	1	HPE MR216-p Gen11 x16 Lanes No Cache SPDM PCI Storage Controller
P48183-B21	1	HPE NS204i-u Gen11 NVMe Hot Plug Boot Optimized storage device
P40499-B21	2	1.92TB 12G SAS RI SFF SSD (Spill-to-disk cache) JBOD
P42044-B21	1	Mellanox MCX631102AS-ADAT Ethernet 10/25Gb 2-port SFP28 PCIe Adapter
P51181-B21	1	Broadcom BCM5719 Ethernet 1Gb 4-port Base-T OCP3 Adapter [USB - Ethernet adapter]
P38997-B21	2	HPE 1600W Flex Slot Platinum Hot Plug Low Halogen Power Supply Kit
		Data Nodes
P54205-B21	1	HPE ProLiant DL345 Gen11 8SFF Configure-to-order Server
P53704-B21	1	AMD EPYC 9354P 3.25GHz 32-core 280W Processor for HPE
P50311-B21	8	HPE 32GB (1x32GB) Dual Rank x8 DDR5-4800 CAS-40-39-39 EC8 Registered Smart Memory Kit
P47789-B21	1	HPE MR216-p Gen11 x16 Lanes No Cache SPDM PCI Storage Controller
P48183-B21	1	HPE NS204i-u Gen11 NVMe Hot Plug Boot Optimized storage device
P57114-B21	1	HPE ProLiant DL345 Gen11 4LFF x1 SAS/SATA 12G Front Backplane Kit
P47837-K21	8	8TB SAS 12G Business Critical 7.2k LFF HDD (HDFS Space) JBOD
P42044-B21	1	Mellanox MCX631102AS-ADAT Ethernet 10/25Gb 2-port SFP28 PCIe Adapter
P51181-B21	1	Broadcom BCM5719 Ethernet 1Gb 4-port Base-T OCP3 Adapter
P38995-B21	2	HPE 800W Flex Slot Platinum Hot Plug Low Halogen Power Supply Ki
		Network
JL624A	1	Aruba 8325-48Y8C 48p 25G SFP/+/28 8p 25G QSFP
R0Z26A	20	HPE Aruba 100G QSFP28 to QSFP28 5M DAC Cab
JL448A	1	HPE Aruba X2C2 RJ45 to DB9 Console Cable
JL667A	1	Aruba 6300F 48-port 1GbE and 4-port SFP56

RESOURCES AND ADDITIONAL LINKS

HPE Reference Architectures, hpe.com/info/ra

HPE Servers, hpe.com/servers

HPE Storage, hpe.com/storage

HPE Networking, hpe.com/networkingHPE Apollo 4000 Systems, https://www.hpe.com/us/en/storage/apollo-4000.html

HPE ProLiant 325 Gen11, https://www.hpe.com/psnow/doc/a50004297enw.pdf?jumpid=in_pdp-psnow-qs

HPE ProLiant 345 Gen11, https://www.hpe.com/psnow/doc/a50004298enw.pdf?jumpid=in_pdp-psnow-qs

AMD Database and Analytics Hub: https://www.amd.com/en/solutions/databases-and-analytics

AMD EPYC CPU product page: https://www.amd.com/en/processors/epyc-server-cpu-family

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

HPE Technology Consulting Services, hpe.com/us/en/services/consulting.html

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

a50010560enw

Hewlett Packard Enterprise