cloudera®

Cloudera Enterprise Reference Architecture for Azure Deployments

Important Notice

Table of Contents

Executive Summary

This document is a high-level design and best-practices guide for deploying the Cloudera Enterprise distribution on Microsoft Azure cloud infrastructure. It describes Cloudera Enterprise and Microsoft Azure capabilities and deployment architecture recommendations.

Cloudera Reference Architecture documents illustrate example cluster configurations and certified partner products. The Cloud RAs are not replacements for official statements of supportability, rather they're guides to assist with deployment and sizing options. Statements regarding supported configurations in the RA are informational and should be cross-referenced with the latest documentation.

Audience and Scope

This guide is for IT and Cloud Architects who are responsible for the design and deployment of Apache Hadoop solutions in Microsoft Azure, as well as for Apache Hadoop administrators and architects who are data center architects/engineers or collaborate with specialists in that space.

This document describes Cloudera recommendations on the following topics:

- Instance type selection
- Storage and network considerations
- High availability
- One-click deployment for POCs, prototypes, and production clusters
- Deployment strategies for the Cloudera software stack on Azure

Overview

This section describes software and cloud infrastructure that enables a Cloudera cluster running on Azure. Specific deployment details are discussed later.

Cloudera Enterprise

Cloudera, an enterprise data management company, introduced the concept of the enterprise data hub (EDH): a central system to store and work with all data. The EDH has the flexibility to run a variety of enterprise workloads (for example, batch processing, interactive SQL, enterprise search, and advanced analytics) while meeting enterprise requirements such as integrations to existing systems, robust security, governance, data protection, and management. The EDH is the emerging center of enterprise data management. EDH builds on Cloudera Enterprise, which consists of the open source Cloudera Distribution including Apache Hadoop (CDH), a suite of management software and enterprise-class support.

In addition to needing an enterprise data hub, enterprises are looking to move or add this powerful data management infrastructure to the cloud for operation efficiency, cost reduction, compute and capacity flexibility, and speed and agility.

As organizations embrace Hadoop-powered big data deployments in cloud environments, they also want enterprise-grade security, management tools, and technical support--all of which are part of Cloudera Enterprise.

Customers of Cloudera and Amazon Web Services (AWS) can now run the EDH in the AWS public cloud, leveraging the power of the Cloudera Enterprise platform and the flexibility of the AWS cloud. Cloudera Director enables users to manage and deploy Cloudera Manager and EDH clusters in AWS. Users can create and save templates for desired instance types, spin up and spin down Cloudera Manager and EDH

as well as clone clusters. Users can also deploy multiple clusters and can scale up or down to adjust to demand.

Cloudera Reference Architecture documents illustrate example cluster configurations and certified partner products. The Cloud RAs are not replacements for official statements of supportability, rather they're guides to assist with deployment and sizing options. Statements regarding supported configurations in the RA are informational and should be cross-referenced with the latest documentation.

Cloudera Enterprise can be deployed in the Microsoft Azure infrastructure using the reference architecture described in this document.

Microsoft Azure

Microsoft Azure is an industry-leading cloud service for both infrastructure-as-a-service (IaaS) and platform-as-a-service (PaaS), with data centers spanning the globe. Microsoft Azure supports a diverse set of Linux as well as Windows based applications and has the necessary infrastructure to serve big-data workloads.

The offering consists of several services, including virtual machines, virtual networks, and storage services, as well as higher-level services such as web applications and databases. For Cloudera Enterprise deployments, the following service offerings are relevant:

Azure Virtual Machines

Azure Virtual Machines enable end users to rent virtual machines of different configurations on demand and pay for the amount of time they use them. Azure offers several types of virtual machines with different pricing options. For Cloudera Enterprise deployments, each virtual machine instance conceptually maps to an individual server. This document recommends specific virtual machine instances for Azure deployment. As service offerings change, this document will be updated to indicate instances best suited for various workloads.

Azure Storage

Azure storage provides the persistence layer for data in Microsoft Azure. Azure supports several different options for storage, including Blob storage, Table storage, Queue storage, and File storage. Storage options in Azure are tied to a storage account, which provides a unique namespace to manage up to 500 TB of storage. Up to 250 (default 200) unique storage accounts can be created per subscription. For more information on subscription level and per-account limits on services, see the Azure links in the References section.

Blob Storage

Blob storage stores file data. A blob can be any type of text or binary data, such as a document, media file, or application installer. Blobs are available in two forms: block blobs and page blobs (disks). Block blobs are optimized for streaming and storing cloud objects, and are a good choice for storing documents, media files, and backups. Page blobs are optimized for representing IaaS disks and supporting random writes, and can be up to 1 TB in size. An Azure virtual machine network-attached IaaS disk is a virtual hard disk (VHD) stored as a page blob.

Azure Data Lake Store

Azure Data Lake Store (ADLS) is a storage service that allows for storing large sized files in the range of petabytes and trillions of objects using a simple API while being scalable and consistent.

Virtual Network (VNet)

Azure Virtual Network (VNet) is a logical network overlay that can include services and VMs. Resources in a VNet can access each other. You can configure a VNet to connect to your on-premises network through a VPN or VNet Gateway to make the VNet function as an extension to your data center.

ExpressRoute

You can use ExpressRoute to establish a direct, dedicated link between your on-premises infrastructure and Azure. This provides more secure, faster, and more reliable connectivity than a VPN. ExpressRoute is particularly useful for moving large volumes of data between your on-premises infrastructure and Azure.

Glossary of Terms

Term	Description
ARM	Azure Resource Manager.
Availability Set/AS	A logical grouping of machines in Azure used to automatically separate machines into distinct failure domains.
Cloudera Manager Agent/CMA	The Cloudera Manager Agent is a Cloudera Manager component that works with the Cloudera Manager Server to manage the processes that map to role instances.
DataNode/DN	Worker nodes of the cluster to which the HDFS data is written.
Failover Controller/FC	HDFS Failover Controller (AKA: ZKFailoverController/ZKFC) is a ZooKeeper client that monitors and manages the state of the NameNode. Each of the hosts that run a NameNode in an HA configuration also run a ZKFC.
HDD	Hard disk drive.
HDFS	Hadoop Distributed File System
High-Availability/HA	Configuration that addresses availability issues in a cluster. In a standard configuration, the NameNode is a single point of failure (SPOF). Each cluster has a single NameNode, and if that machine or process became unavailable, the cluster as a whole is unavailable until the NameNode is either restarted or brought up on a new host. The secondary NameNode does not provide failover capability. High availability enables running two NameNodes in the same cluster: the active NameNode and the standby NameNode. The standby NameNode allows a fast failover to a new NameNode in case of machine crash or planned maintenance.
History Server	Process that archives job metrics and metadata for YARN or Spark.
IaaS	Infrastructure as a Service

NameNode/NN	The metadata master processes of HDFS essential for the integrity and proper functioning of the distributed file system.
Network Security Group (NSG)	A list of access control rules that can allow or deny traffic to a Subnet in an Azure Virtual Network or to a Virtual Machine.
NIC	Network interface card
NodeManager	The process that starts application processes and manages resources on the DataNodes.
NUMA	Non-uniform memory access. Addresses memory access latency in multi-socket servers, where memory that is remote to a core (that is, local to another socket) needs to be accessed. This is typical of SMP (symmetric multiprocessing) systems, and there are several strategies to optimize applications and operating systems.
PaaS	Platform as a service.
QJM **QJN**	Quorum Journal Manager. Provides a fencing mechanism for high availability in a Hadoop cluster. This service is used to distribute HDFS edit logs to multiple hosts (at least three are required) from the active NameNode. The standby NameNode reads the edits from the JournalNodes and constantly applies them to its own namespace. In case of a failover, the standby NameNode applies all of the edits from the JournalNodes before promoting itself to the active state. Quorum JournalNodes. Nodes on which the journal services are installed.
Resource Manager/RM	The YARN resource management component, which initiates application startup and controls scheduling on the worker nodes of the cluster (one per cluster).
VHD	Virtual hard disk. Azure storage for page blobs.
VM	Virtual machine.
VNet	Azure Virtual Network. Isolated logical network in Azure used for cluster deployment.
ZooKeeper/ZK	ZooKeeper is a distributed, open-source coordination service that exposes a simple set of primitives that distributed applications can use for synchronization, configuration maintenance, and groups and naming.

Deployment Architecture

System Architecture Best Practices

This section describes Cloudera's recommendations and best practices applicable to Hadoop cluster system architecture.

Java

Cloudera Manager and CDH are certified to run on Oracle JDK. At this time OpenJDK is not supported. Cloudera distributes a compatible version of the Oracle JDK through the Cloudera Manager repository. Customers are also free to install a compatible version of the Oracle JDK distributed by Oracle.

Refer to CDH and Cloudera Manager Supported JDK Versions for a list of supported JDK versions.

Right-size Server Configurations

Cloudera recommends deploying three or four machine types into production:
- **Master Node.** Runs the Hadoop master daemons: NameNode, Standby NameNode, YARN Resource Manager and History Server, the HBase Master daemon, Sentry server, and the Impala StateStore Server and Catalog Server. Master nodes are also the location where Zookeeper and JournalNodes are installed. The daemons can often share single pool of servers. Depending on the cluster size, the roles can instead each be run on a dedicated server. Kudu Master Servers should also be deployed on master nodes.
- **Worker Node.** Runs the HDFS DataNode, YARN NodeManager, HBase RegionServer, Impala impalad, Search worker daemons and Kudu Tablet Servers.
- **Utility Node.** Runs Cloudera Manager and the Cloudera Management Services. It can also host a MySQL (or another supported) database instance, which is used by Cloudera Manager, Hive, Sentry and other Hadoop-related projects.
- **Edge Node.** Contains all client-facing configurations and services, including gateway configurations for HDFS, YARN, Impala, Hive, and HBase. The edge node is also a good place for Hue, Oozie, HiveServer2, and Impala HAProxy. HiveServer2 and Impala HAProxy serve as a gateway to external applications such as Business Intelligence (BI) tools.

For more information refer to Cluster Hosts and Role Assignments.

> **Note:**
>
> The edge and utility nodes can be combined in smaller clusters, however in cloud environments it's often more practical to provision dedicated instances for each.

General Cluster Architecture

This section contains general guidelines for cluster layout, assuming the usage of HDFS, YARN, Hive, Impala, Hue, Oozie, and ZooKeeper.

In this reference architecture, we consider different kinds of workloads that are run on top of an Enterprise Data Hub. The initial requirements focus on host types that are suitable for a diverse set of workloads. As service offerings change, these requirements may change to specify host types that are

unique to specific workloads. You choose host types based on the workload you run on the cluster. You should also do a cost-performance analysis.

Small Cluster (up to 20 worker nodes)
- 2 Master Nodes
  - HDFS NameNode (2, HA)
  - HDFS JournalNode (3)
  - HDFS FailoverController (2, collocated with NN services)
  - YARN Resource Manager (2, HA)
  - YARN History Server
  - HBase Master (2, HA)
  - Sentry Service
  - ZooKeeper (2)
  - Kudu Master Server (2)
- 1 Utility/Edge Node
  - Cloudera Manager
  - Cloudera Management Services
  - HDFS JournalNode
  - Hive Metastore Server
  - HiveServer2
  - HBase Thrift or REST Server
  - Impala Catalog Server
  - Impala StateStore Server
  - Hue
  - Oozie
  - Flume Agent
  - Database server
  - Gateway (client) configurations
  - ZooKeeper Node[1]
  - Kudu Master Server (1)
- 20 Worker Nodes
  - HDFS DataNode
  - YARN NodeManager
  - HBase Region Server
  - Impalad
  - Kudu Tablet Server

Medium Cluster (up to 200 worker nodes)
- 3 Master Nodes
  - HDFS NameNode (2, HA)
  - HDFS JournalNode (3)
  - HDFS FailoverController (2, collocated with NN services)
  - YARN Resource Manager (2, HA)
  - YARN History Server

---

[1] When using ephemeral instances, ZooKeeper Node and JournalNode should have their own dedicated spindle in this configuration.

- ○ HBase Master (3, HA)
- ○ Sentry Service
- ○ ZooKeeper (3)
- ○ Kudu Master Servers (3)
- 2 Utility Nodes
  - ○ Cloudera Manager (node 1)
  - ○ Cloudera Management Services (node 2)
  - ○ Hive Metastore Server (node 1)
  - ○ Impala Catalog Server (node 1)
  - ○ Impala StateStore Server (node 1)
  - ○ Oozie (node 1)
  - ○ RDBMS servers (both)
- 1+ Edge Nodes
  - ○ Hue
  - ○ HiveServer2
  - ○ HBase Thrift or REST Server
  - ○ Flume Agent
  - ○ Gateway (Client) configurations
- 200 Worker Nodes
  - ○ HDFS DataNode
  - ○ YARN NodeManager
  - ○ HBase Region Server
  - ○ Impalad
  - ○ Kudu Tablet Server

Large Cluster (up to 500 worker nodes)
- 5 Master Nodes
  - ○ HDFS NameNode (2, HA)
  - ○ HDFS JournalNode (5)
  - ○ HDFS FailoverController (collocated with NN services)
  - ○ YARN Resource Manager (2, HA)
  - ○ HBase Master (5, HA)
  - ○ Sentry Service
  - ○ ZooKeeper Node (5)
  - ○ Kudu Master Server (5)
- 2 Utility Nodes
  - ○ Cloudera Manager (node 1)
  - ○ Cloudera Management Services (node 2)
  - ○ Hive Metastore Server (node 1)
  - ○ Impala Catalog Server (node 1)
  - ○ Impala StateStore Server (node 1)
  - ○ Oozie (node 1)
  - ○ Database server (both)
- 1+ Edge Nodes
  - ○ Hue
  - ○ HiveServer2
  - ○ HBase Thrift or REST Server
  - ○ Flume Agent

- ○ Gateway (Client) configurations
- ● 500 Worker Nodes
  - ○ HDFS DataNode
  - ○ YARN NodeManager
  - ○ HBase Region Server
  - ○ Impalad
  - ○ Kudu Tablet Server

Very Large Cluster (500+ worker nodes)

Very Large clusters have a layout similar to Large clusters but require detailed environment and workload analysis and custom configuration. Contact Cloudera for additional information.

Encryption Infrastructure (for all cluster sizes)
- ● 4 Utility nodes
  - ○ Key Management Server (2)
  - ○ Key Trustee Server (2)

Further detail can be found in the [Data at Rest Encryption](#) documentation.

> **Note:**
>
> Edge nodes in the medium and large cluster architectures are mostly driven by use cases. Complex ingest pipelines and/or lots of client access will require more edge nodes to handle the load.

> **Note:**
>
> Without three master servers Kudu does not provide HA. Five masters can be utilized; in this case the loss of 2 Kudu Master Servers will be tolerated.

Deployment Options

Cloudera has expanded support of the Azure platform in several areas:

- ● Support for deployment via Cloudera Director
- ● Support for additional instance sizes and disk configurations
- ● Additional guidance on availability for CDH clusters in Azure

Cloudera currently allows deployment into Azure via several methods:

Requirements:

- ● All deployments require a cluster to be placed into a single Virtual Network so the instances can communicate with one another with low latency.

- ● Resources must all be located in a single Azure region.

- Resources for a cluster will be organized into one or more Resource Groups and configured with Network Security Groups (NSG) to control network access.

All deployments require a cluster to be placed into a single Virtual Network so the instances can communicate with one another with low latency. Resources must all be located in a single Azure region. Resources for a cluster will be organized into one or more Resource Groups and configured with Network Security Groups (NSG) to control network access.

Supported:

- One-click deployment using Azure Marketplace
- Using deployment scripts hosted on the Azure GitHub Repository
- Installation with Cloudera Manager on provisioned VMs, following guidance in this document regarding role placement and storage configuration

Recommended:

- Using Cloudera Director 2.1 or greater

All services within Cloudera Enterprise are fully supported except those that may be mentioned in the Special Considerations section.

Cloudera Director

Cloudera Director allows you to easily deploy, monitor and modify clusters in Azure and other cloud providers. Detailed information regarding environment setup and prerequisites are provided in Director's Getting Started on Microsoft Azure documentation. Cloudera Director provides additional capabilities to grow or shrink your cluster to match changing needs in your workload.

Cloudera Director deployments in Azure **require** the use of specific settings and bootstrap actions to set networking and Cloudera Enterprise configurations properly. The latest versions of necessary script files and sample configurations will be published to the Cloudera Director-Scripts GitHub Repository.

For production deployments you must use the Director Command Line Interface and configuration files based on the published examples. The log and data locations must not change, except as needed to reflect the number of data disks.

Azure Marketplace

The Cloudera Enterprise Data Hub offering in the Azure Marketplace allows you to quickly deploy a properly configured cluster in Azure.The automation logic assigns the correct number of master and worker nodes. The same configuration can also be created using the high availability (HA) option from the Azure portal. This offering provides simple, one-click provisioning of a cluster for proof-of-concept, prototype, development, and production environments. The provisioned cluster runs on the latest distribution of CDH with services such as HDFS, YARN, Impala, Oozie, Hive, Hue, and ZooKeeper with Cloudera Manager.

Deployment Scripts

You can use the Cloudera on CentOS **Azure Quickstart Template** as a starting point for a simple or customized deployment to Azure. The scripts demonstrate proper setup of Nodes, Role Assignment, and Configuration.

Cloudera Manager

Installation with Cloudera Manager is covered in the [Cloudera Installation](#) documentation. Special consideration should be taken to ensure that role placement aligns with Table 1, and that no data directories are placed on the Azure VM Temporary Disk (usually mounted at /mnt/resource).

Figure 1 shows a deployment scenario using Cloudera Director, Azure Marketplace offer or Deployment Scripts to launch a Cloudera cluster.

Figure 1: Deployment Scenario
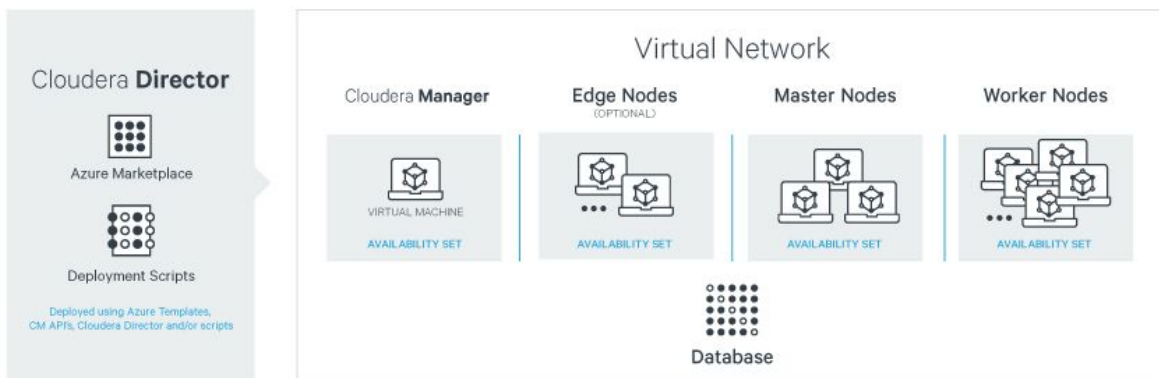
Edge Security

The accessibility of the Cloudera Enterprise cluster is defined by the Azure Network Security Group (NSG) configurations assigned to VMs and/or Virtual Network subnets and depends on security requirements and workload. Typically, edge nodes (also known as client nodes or gateway nodes) have direct access to the cluster's internal network. Through client applications on these edge nodes, users interact with the cluster and its data. These edge nodes can run web applications for real-time serving workloads, BI tools, or the Hadoop command-line client used to interact with HDFS.

By default the Network Security Groups in the [Cloudera Enterprise Data Hub](#) and the [Cloudera Director](#) offerings on Azure Marketplace allow public IP access only to port 22 (SSH). To access these configurations you can log in via an ssh terminal or [set up a tunnel](#) to connect to web interfaces.

Azure Resource Quotas

The default quota for cores within an Azure subscription is 20 cores per region; for more than 20 cores, customers will need engage Microsoft Support within the Azure portal to request an increased quota. The available quota for the region you are deploying to has to be greater than the number of cores used by all VMs you are launching.

Workloads & Roles

In this reference architecture, we consider different kinds of workloads that are run on top of an Enterprise Data Hub. The initial requirements focus on instance types that are suitable for a diverse set of workloads. As service offerings change, these requirements may change to specify instance types that are unique to specific workloads. Table 1 identifies service roles for node types.

Table 1: Node Types and Service Roles

	Master Node 1/2	Master Node 3	Worker Nodes	Edge/Utility Node(s)
HDFS	NN, QJN,FC	QJN, HttpFS	DataNode	
YARN	RM	JobHistory Server	NodeManager	
Hive		MetaStore WebHCat HiveServer2		
Hue		Hue Server		
Spark		History Server	(Runs on YARN)	
Impala		Catalog Server State Store	Impala Daemon	
Cloudera Search			Solr Server	
HBase	HBase Master	Thrift Server	HBase Region Server	
ZooKeeper	ZK	ZK		
Flume				Flume Agent
Gateway Roles				HDFS, YARN, Hue, Hive, Sqoop, etc. As needed
Management	CMA	CMA,Oozie, Management Roles, Cloudera Manager[2]	CMA	
Utility				KTS, KMS, CDSW, Other

Master Node Services

Master nodes for a Cloudera Enterprise deployment run the master daemons, management, and coordination services, which may include:

- Cloudera Manager
- ResourceManager
- NameNode
- Standby NameNode
- JournalNodes
- ZooKeeper
- Oozie
- Impala Catalog Server
- Impala State Store
- Job History Server

Worker Node Services

Worker nodes for a Cloudera Enterprise deployment run worker services. These may include:

---

[2] CM will be on dedicated node for Director Deployments

- DataNode
- NodeManager
- RegionServer
- Impala Daemon

Edge/Utility Node Services

An edge node contains client-facing configurations and services, including gateway configurations for HDFS, MapReduce, Hive, and HBase. Hadoop client services run on edge nodes. Utility nodes provide additional services that may have resource requirements that are different from master or worker nodes. These services may include:

- Third-party tools
- Hadoop command-line client
- Beeline client
- Flume agents
- Hue Server
- KTS, KMS, CDSW

Note: Edge/Utility node services are typically deployed to the same type of hardware as those responsible for master node services, however any instance type can be used for an edge or utility node so long as it has sufficient resources for your use. Depending on the size of the cluster, there may be numerous systems designated as edge or utility nodes.

Kafka Broker Services

For dedicated Kafka Broker nodes we require an instance type and disk configuration from Table 2 (similar to Worker nodes). The instance type and count of broker nodes should take into consideration the documented VM Storage throughput per-node. Logging and data directories must be configured to use persistent Premium Storage or Standard Storage disks.

Instance Types

Cloudera recommends the following instance types for CDH workloads on Microsoft Azure.

Table 2: Recommended Worker Instance Types with Premium Storage Disks

Instance type For Director config prefix with STANDARD_	Memory/Cores	Storage Limits		
		Documented VM Storage Throughput	Min count P30 1 TB Data Disks	Max count P30 1 TB Data Disks
DS12_v2 * Director 2.1.1 or later	28 GB / 4 Cores	192 MB/s	1	4
DS13, DS13_V2	56 GB / 8 Cores	V1: 256 MB/s V2: 384 MB/s	2	8
DS14, DS14_V2	112 GB / 16 Cores	V1: 512 MB/s V2: 768 MB/s	4	14
DS15_v2	140 GB / 20 Cores	960 MB/s	5	20
GS4	224 GB / 16 Cores	1000 MB/s	5	20
GS5	448 GB / 32 Cores	2000 MB/s	10	31
	* HDFS replicated write throughput may be network-limited.			
Each node **requires** a dedicated P20 or P30 log disk that is not included in the Data Disk count above.				

Table 3: Recommended Worker Instance Types with Standard Storage Disks (Director 2.2 and later)

Instance type For Director config prefix with STANDARD_	Memory/Cores	Storage Limits	
		Min count Data Disks	Max count Data Disks
D12_v2	28 GB / 4 Cores	3	7
D13, D13_v2	56 GB / 8 Cores	5	15
D14, D14_v2	112 GB / 16 Cores	7	24
D15_v2	140 GB / 20 Cores	9	24

Each node **requires** a dedicated 128 GB or greater log disk that is not included in the Data Disk count above.
Standard Disks are billed for used space + transactions. (Azure Storage Standard Disk Pricing)
TRIM/UNMAP should be implemented to free space from deleted files. (TRIM/UNMAP Support in Azure)
Similar DS Instances are supported but D are preferred for Standard Storage disks.

Instance type For Director config prefix with STANDARD_	Memory/Cores	Storage Limits		
		Documented VM Storage Throughput	Dedicated P20 512 GB or P30 1 TB Disks for Master Services (JN, NN, ZK)	
			Min count	Max count
DS12_v2 * Director 2.1.1 or later	28 GB / 4 Cores	192 MB/s	One for each Master Service requiring dedicated storage (JN, NN, ZK, etc). Typically 2 - 3.	4
DS13, DS13_V2	56 GB / 8 Cores	V1: 256 MB/s V2: 384 MB/s		8
DS14, DS14_V2	112 GB / 16 Cores	V1: 512 MB/s V2: 768 MB/s		14
DS15_v2	140 GB / 20 Cores	960 MB/s		20
GS4	224 GB / 16 Cores	1000 MB/s		20
GS5	448 GB / 32 Cores	2000 MB/s		31
Each node **requires** a dedicated P20 or P30 log disk that is not included in the Master Disk count above.				

Note:

Changing the Instance Type is Not Supported in Azure
- Changing the instance type of an already-deployed VM is not supported in Azure.
- Changing the instance type of a VM through a tool external to Cloudera Director is not supported in Azure. You cannot, for example, use the Azure Portal to change the instance type. Cloudera Director is not updated when instance type changes are made with external tools, and your cluster will show errors in Cloudera Director.

Regions

Azure Regions are geographical locations where Microsoft has deployed infrastructure for customers. Each region is self contained and has a deployment of the various Azure services. Cloudera Enterprise clusters are deployed within a single region. Cluster configurations that span multiple regions are not supported.

Azure Government and Sovereign Clouds

Microsoft has several independent sovereign clouds which are managed and run in different geographies to address data sovereignty and other specific regional regulations.

Cloudera Enterprise deployments are now supported on Azure Government and Azure Germany. Supported deployment methods may vary for the different sovereign clouds. You should confirm that the specific instance sizes and services are available in the Azure Region where you intend to deploy. (Notes: Premium Storage is required for master nodes and is not available in all regions. Azure Germany does not support Azure Marketplace.)

Supported Virtual Machine Images

Virtual Machines Images are the base hard disk images that are used to initialize a VM running in the Microsoft Azure cloud. They consist of the operating system and any other software that the virtual machine image creator bundles with them.

Cloudera currently supports the following images for running Cloudera Enterprise in Azure:

Deployed via Marketplace or Azure Quickstart Deployment Scripts:

- Cloudera Published CentOS 6.6 and 6.7

Deployed via Director:

- Cloudera Published CentOS 6.6, 6.7, 6.8, 7.2, and 7.4
- Red Hat Published Red Hat 6.7, 6.8, 7.2, and 7.4

The Cloudera Published CentOS Images are pre-configured to meet the OS-level prerequisites for Cloudera Enterprise. The Red Hat images **require** the use of additional bootstrap actions to set OS-level prerequisites. The latest versions of these scripts will be published to the Cloudera Director-Scripts GitHub Repository.

Note: The Red Hat published images include an additional per-hour surcharge billed by Microsoft for each VM.

Storage Options and Configuration

Azure offers different storage options that vary in performance, durability, and cost.

Microsoft Azure VHDs/Page Blobs and Premium Storage

Virtual hard disks (VHDs) are network attached disks that can be attached to an instance to expand the available storage on a virtual machine. The recommended number of attached Premium Storage VHDs used for HDFS/worker data directories varies by instance size; refer to Table 2 for specifics. Master services for HDFS and ZooKeeper require dedicated storage devices. The minimum count will vary depending on the master role assignment. Maximum disk count for master nodes by instance size can be found in Table 4. Premium Storage is the only supported primary storage on Azure for master nodes and is recommended on worker nodes for workloads requiring low latency or high throughput.

Premium Storage Disks can deliver a target throughput that is metered based on the disk size. 1 TB P30 disks are provisioned for 200MB/s throughput. This limit, as well as IOPs and Throughput targets for other disk sizes can be found in the Microsoft documentation. The storage throughput may also be limited per-node based on the **Max uncached disk throughput: IOPS / MBps** column of the VM size tables. Numbers for all recommended instance sizes are included as **Documented VM Storage Throughput** in Table 2. Both managed and Unmanaged Premium Disks are supported.

> **Note:**
>
> Managed disks can only be used for master nodes in regions with at least 3 fault domains available for managed disks (Number of Fault Domains per region).

Microsoft Azure VHDs/Page Blobs and Standard Storage for Worker Nodes

Virtual hard disks (VHDs) on Microsoft Azure Standard Storage do not provide the same throughput and performance guarantees as Premium Storage disks, but are a lower cost option for workloads that are

price sensitive. The recommended number of attached Standard Storage VHDs used for HDFS/worker data directories varies by instance size; refer to Table 3 for specifics.   Both managed and Unmanaged Standard Disks are supported.

Standard Storage Disks can deliver a target throughput of up to 60MB/s regardless of size since they are implemented as a single Standard Storage Page Blob per disk. Actual storage throughput from a given node may be constrained by limits of the storage account as documented in Azure Storage Scalability and Performance Targets, or by other load in the datacenter since standard storage bandwidth is not provisioned. Effective throughput with Standard Storage disks is generally lower than what can be achieved on a similarly sized VM using Premium Storage Disks and is more variable.

Temporary (or Local) Disk Storage

Microsoft Azure virtual machine instances have storage attached at the instance level, similar to disks on a physical server. The storage is virtualized and referred to as a *temporary or ephemeral* disk because the lifetime of the storage is the same as that of the virtual machine instance. If you stop or terminate the Azure instance, or if the virtual machine is relocated because of maintenance or failure in Azure, all data contained on the local storage is lost. For Cloudera Enterprise clusters, the HDFS data directories, log directories, metadata directories, and any other data that must persist must use the attached Premium Storage or Standard Storage disks. The temporary disk must not be used by Cloudera Enterprise components.

Blob Storage

Cloudera supports using Windows Azure Storage Blob (WASB) to keep a copy of the data you have in HDFS for disaster recovery. WASB is backed by Microsoft Azure Storage which provides durability and availability guarantees that make it ideal for a cold backup of your data.

WASB is only supported for backups using Hadoop's DistCP. Other services are not supported running directly against WASB.

Azure Data Lake Store

Azure Data Lake Store (ADLS) is supported by Cloudera to be used as a storage location for data. ADLS provides storage of large files up to petabytes in size to scale for parallel analytics. It provides high-availability and reliability. Data is stored durably by making redundant copies to guard against any unexpected failures. Check here to see how to configure ADLS for CDH and for a list of the components that support using ADLS as a storage layer. Data in ADLS can be stored and accessed using the AzureDataLakeFilesystem implementation using the adl:// URI prefix.

ADLS is only available in certain Azure regions. Check here for the latest list of regions that support Azure Data Lake Store.

Encryption at Rest

Several encryption at rest options are available in Azure. Azure Storage Service Encryption can be enabled at the storage account level for all VM-associated storage accounts and will encrypt any further writes to the disks with Microsoft-managed keys.

HDFS Transparent Encryption can be used if the primary concern is to encrypt data that is stored in HDFS.

You can also use Cloudera Navigator Encrypt which leverages Linux-native disk encryption technologies. Navigator Encrypt will incur some performance overhead, and may require selection of specific VM Images to provide the necessary dependencies and disk partitioning layout. Refer to the Navigator Encrypt documentation for specific requirements.

Cluster Availability through Azure Availability Sets

Availability Sets

Azure manages the availability of VMs using Azure Availability Sets. Each VM can only belong to a single Availability Set. Microsoft recommends that machines with identical function be placed into an Availability Set, especially if they are configured in a high availability grouping.

The Microsoft documentation is good background reading on this topic:

- Manage the availability of virtual machines (Linux)
- Planned maintenance for Linux virtual machines in Azure

Inside an Availability Set

Availability Sets currently implement two uncorrelated failure domains. The first is the Fault Domain and is associated with the underlying hardware, networking, and power that are required for the VM to be operational. A single hardware failure can make all machines in a given Fault Domain unavailable. The second failure domain is the Update Domain and it is used for planned maintenance that requires reboot of the hypervisor host. Hypervisor host updates requiring guest reboot are infrequent, but will be processed one update domain at a time. Hypervisor host updates utilizing Memory-preserving updates will pause/resume VMs one update domain at a time (typically a 9 to 30 second clock jump is seen on VM).

Figure 2: Layout of Worker Nodes in an Availability Set with 3 Fault Domains and 4 Update Domains
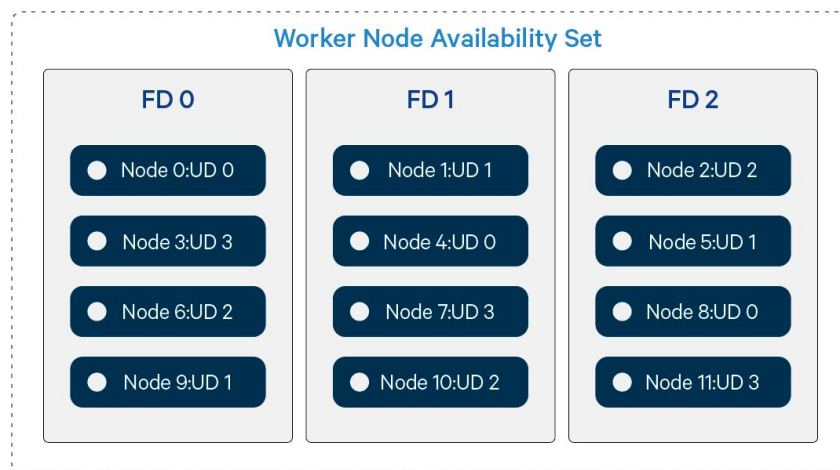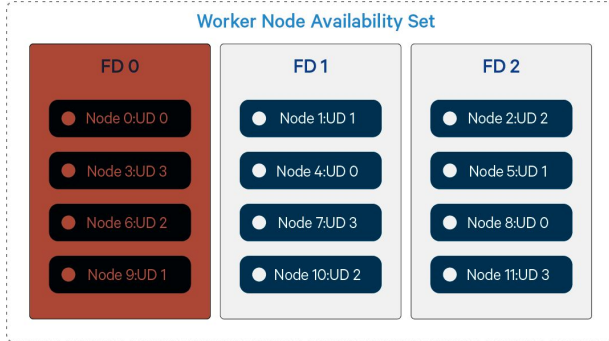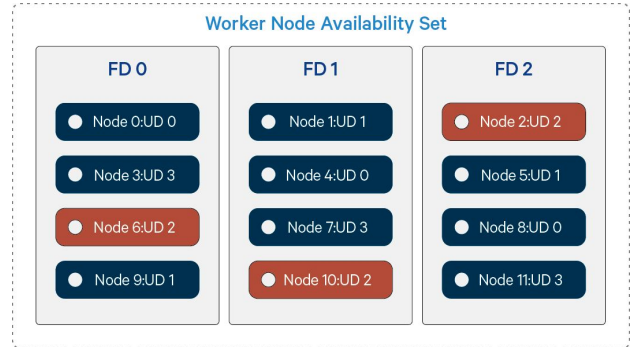
Figure 3: Nodes Offline from Failure in FD 0

Worker Node Availability Set

FD 0	FD 1	FD 2
● Node 0:UD 0	● Node 1:UD 1	● Node 2:UD 2
● Node 3:UD 3	● Node 4:UD 0	● Node 5:UD 1
● Node 6:UD 2	● Node 7:UD 3	● Node 8:UD 0
● Node 9:UD 1	● Node 10:UD 2	● Node 11:UD 3

Figure 4: Nodes Offline during Update in UD 2

Worker Node Availability Set

FD 0	FD 1	FD 2
● Node 0:UD 0	● Node 1:UD 1	● Node 2:UD 2
● Node 3:UD 3	● Node 4:UD 0	● Node 5:UD 1
● Node 6:UD 2	● Node 7:UD 3	● Node 8:UD 0
● Node 9:UD 1	● Node 10:UD 2	● Node 11:UD 3

Our Recommendation and Caveats

- We recommend using one Availability Set for Masters, another Availability Set for Workers, and additional Availability Sets for any other group of like-functioning nodes.
- The number of Fault Domains in each Availability Set should be changed to 3 when it is created.
- Use Classic/Unmanaged Availability Sets for deployments that leverage unmanaged disks and Aligned/managed Availability Sets for deployments that leverage managed disks.

If there are more than 3 VMs in a set you should associate the nodes with **rack locations** that match the Azure-assigned Fault Domain **or** Update Domain. You should be aware of the risk associated with the failure domain that is not mitigated by rack-awareness. Update Domain risk can be minimized by using a larger number of Update Domains when creating the Availability Set (up to 20).

Note: Azure Premium Storage and Standard Storage disks provide 3 in-region storage replicas limiting this to an availability problem.

Known Limitations

There is currently an Azure limitation of 100 VMs per Availability Set. There are also limitations on provisioning VMs from different series (D, DS, DS_V2) into a single Availability Set.

Relational Databases

Cloudera Enterprise deployments require relational databases for the following components:

- Cloudera Manager databases
- Hive metastore
- Hue database
- Oozie database

You can provision Microsoft Azure virtual machine instances and install and manage your own database instances. The list of supported database types and versions is available here. The Azure Marketplace deployment and Azure Quickstart deployment scripts setup a PostgreSQL server for use by the above components. Note: The database server must utilize Azure Premium Storage in order to meet throughput and latency requirements for Cloudera Enterprise.

Cloudera Enterprise Configuration Considerations

HDFS

Durability

For Cloudera Enterprise deployments in Azure, data must be stored on Premium Storage or Standard Storage data disks. Data availability and durability in HDFS is achieved by keeping HDFS replication at three; Cloudera does not recommend lowering the replication factor.

Availability

HDFS availability can be improved by deploying the NameNode with high availability with at least three JournalNodes. The nodes should be placed into an Availability Set to ensure they don't all go down due to hardware failure or host maintenance. Putting two or more nodes an availability sets guarantees that your nodes are spread across multiple racks in the Windows Azure Data Centers. This means redundant power supply, switches and servers.

ZooKeeper

Cloudera requires at least three ZooKeeper instances for availability and durability reasons. Cloudera requires using Premium Storage data disks for ZooKeeper data to ensure low and consistent latency.

Flume

For Flume agents, use memory channel or file channels. Flume's memory channel offers increased performance at the cost of no data durability guarantees. File channels offer a higher level of durability guarantee because data is persisted on disk in the form of files. Cloudera requires file channels be located on attached Premium Storage or Standard Storage data disks in Microsoft Azure.

Special Considerations

Cloudera Enterprise deployments are subject to Azure's subscription and service limits, quotas, and constraints. The default limits might impact your ability to create even a moderately sized cluster, so plan ahead. Some limits can be increased by submitting a support request to Microsoft, although these requests typically take a few days to process.

Special attention should be given to Disk Input/Output bandwidth limits, Network bandwidth limits, Ingress and Egress data rates, and upper limits for cores, Virtual Machines, Storage Accounts, or other elements within a subscription, region, or logical grouping.

References

Cloudera Enterprise

Cloudera homepage

Cloudera Enterprise Documentation

Cloudera Enterprise Support

Cloudera Enterprise on Azure Marketplace (Microsoft login required)

Cloudera Director Getting Started on Microsoft Azure

CDH 5 and Cloudera Manager 5 Requirements and Supported Versions

Microsoft Azure

[Azure homepage](#)

[Azure Virtual Machines](#)

[Azure Virtual Network](#)

[Azure Premium Storage](#)

[Azure Disks and Images](#)

[Azure ExpressRoute](#)

[Azure Subscription and Service Limits, Quotas, and Constraints](#)

Version: v5.15-20180619