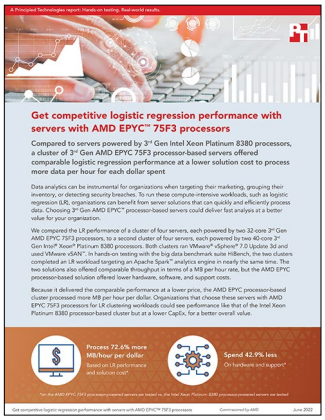

The science behind the report:

Get competitive logistic regression clustering performance at a better value with AMD EPYC™ 75F3 processor-powered servers

This document describes what we tested, how we tested, and what we found. To learn how these facts translate into real-world benefits, read the report Get competitive logistic regression performance with servers with AMD EPYC™ 75F3 processors.

We concluded our hands-on testing on May 3, 2022. During testing, we determined the appropriate hardware and software configurations and applied updates as they became available. The results in this report reflect configurations that we finalized on April 28, 2022 or earlier. Unavoidably, these configurations may not represent the latest versions available when this report appears.

Our results

To learn more about how we have calculated the wins in this report, go to https://facts.pt/calculating-and-highlighting-wins. Unless we state otherwise, we have followed the rules and principles we outline in that document.

Table 1: HiBench - Logistic Regression results.

	Cluster of AMD EPYC 75F3 processor-based servers	Cluster of Intel Xeon Platinum 8380 processor-based servers	Percentage difference for AMD solution
Runtime (sec)	1,676	1,650	-1.551%
Throughput (MB/s)	31.8	32.3	-1.548%
Throughput (MB/hr)	114,480	116,280	-1.548%
Solution cost (USD)	$199,181.68	$349,269.76	42.97%
Performance per dollar (MB/s)/$	0.00015965	0.00009248	72.64%
Performance per dollar (MB/hr)/$	0.5747	0.3329	72.64%

For the VMware® HCI Kit 6 Enterprise and its three-year production support costs, we contacted a third-party reseller who sent us the prices on April 7, 2022. According to these calculations, our estimated three-year cost per license is:

- VMware HCI Kit 6 Enterprise + Production support for three years (Licensed for the max of 32 cores/CPU socket): $14,244.76

Based on these prices, we calculated the licensing cost for each server and cluster, which we show in Table 2.

The table also includes hardware and support cost comparisons. On February 2, 2022, we received a quote from Supermicro for the servers we tested. We added the cost of the 3.84TB drive—which we received in a quote from Supermicro on September 1, 2021—as well as the cost of the 1.92TB drive, which we retrieved from https://www.serversupply.com/SSD/NVMe/1.92TB/SAMSUNG/MZWLJ1T9HBJR_332140.htm?gclid=CjwKCAjw6dmSBhBkEiwA_W-EoAsVdrKju-vf9SUusZ6F6prAdH6Jyfar2n2_RVwHwXUb3TBmI8pJsRoCOpsQAvD_BwE on April 13, 2022.

Table 2: A comparison of licensing, hardware, and support costs in USD for the AMD EPYC and Intel Xeon processor-based solutions.

	Supermicro AS-1124US-TNRP cluster powered by AMD EPYC 75F3 processors	Supermicro SYS-620U-TNR cluster powered by Intel Xeon Platinum 8380 processors	Less cost for Supermicro AS-1124US-TNRP cluster powered by AMD EPYC 75F3 processors
VMware HCI Kit 6 Enterprise + Production support for three years (Licensed for the max of 32 cores/CPU socket)	$28,489.52	$56,979.04	$28,489.52
Per server total	**$28,489.52**	**$56,979.04**	**$28,489.52**
Per cluster total	**$113,958.08**	**$227,916.16**	**$113,958.08**
Hardware and support: Single server	$21,305.90	$30,338.40	$9,032.50
Hardware and support: Cluster	$85,223.60	$121,353.60	$36,130.00
Total cost for three years	**$199,181.68**	**$349,269.76**	**$150,088.08**

CPU utilization statistics

Figures 1 and 2 show the CPU utilization during testing for the two clusters.

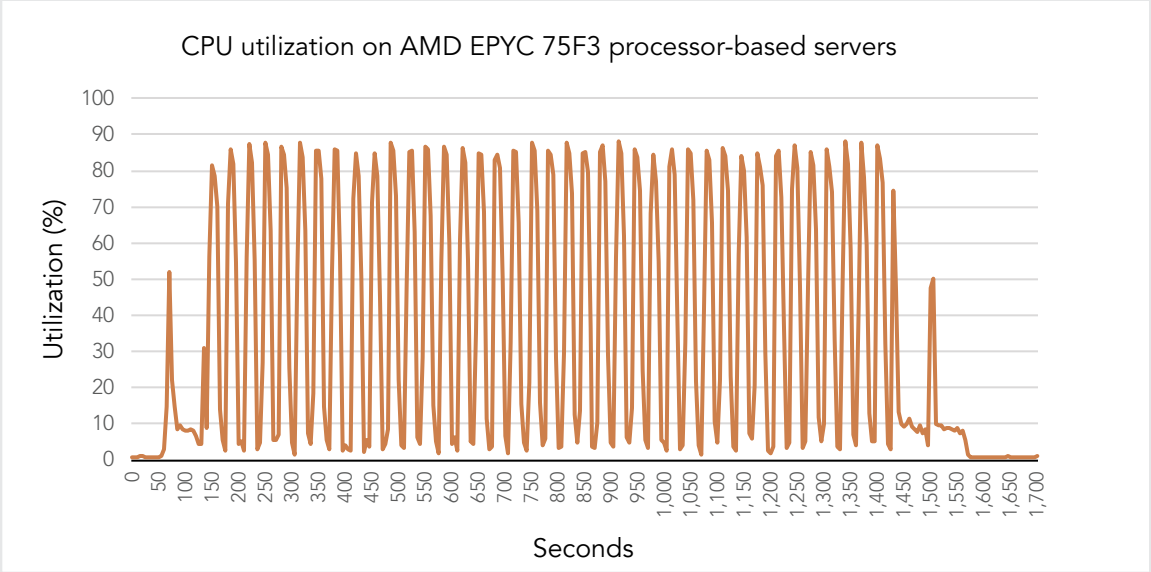

Figure 1: CPU utilization during testing for the AMD EPYC 75F3 processor-based cluster.

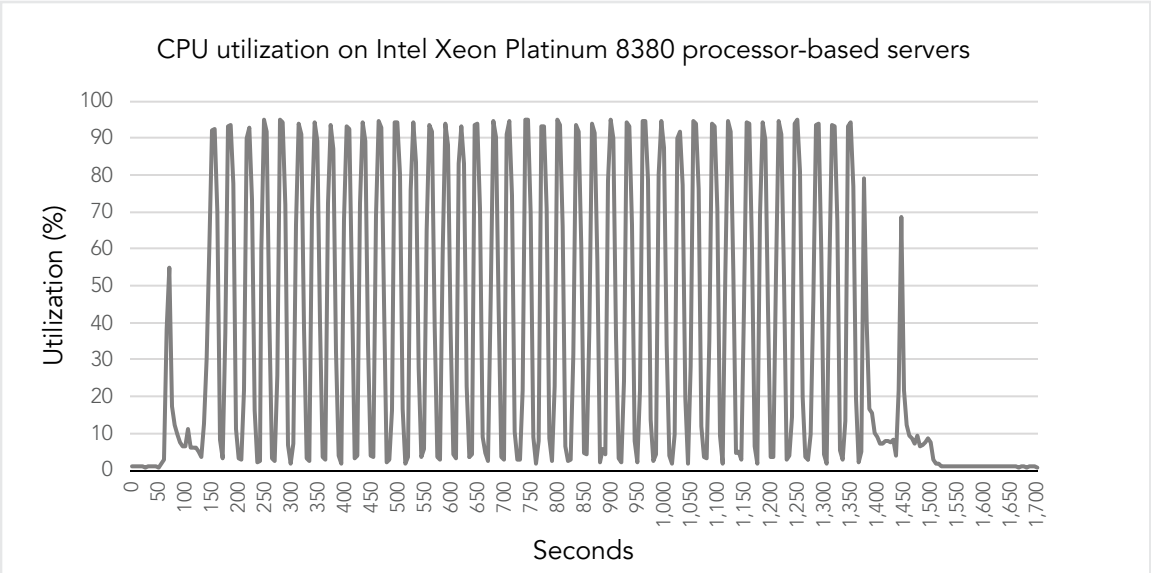

Figure 2: CPU utilization during testing for the Intel Xeon Platinum 8380 processor-based cluster.

System configuration information

Table 3: Detailed information on the systems we tested.

System configuration information	4x Supermicro AS-1124US-TNRP	4x Supermicro SYS-620U-TNR
BIOS name and version	American Megatrends Inc. 2.1	1.1A
Operating system name and version/build number	VMware ESXi-7.0U2e-19290878-standard	VMware ESXi-7.0U2e-19290878-standard
Date of last hypervisor updates/patches applied	2/22/2022	2/22/2022
Power management policy	Determinism Slider: Power	Maximum Performance
Processor		
Number of processors	2	2
Vendor and model	AMD EPYC 75F3	Intel Xeon Platinum 8380
Core count (per processor)	32	40
Core frequency (GHz)	2.95	2.30
Stepping	N/A	N/A
Memory module(s)		
Total memory in system (GB)	1,024	1,024
Number of memory modules	16	16
Vendor and model	Micron MTA36ASF8G72PZ-3G2E1	Micron MTA36ASF8G72PZ-3G2E1
Size (GB)	64	64
Type	PC4-3200	PC4-3200
Speed (MHz)	3,200	3,200
Speed running in the server (MHz)	3,200	3,200
Local storage (hypervisor)		
Number of drives	1	1
Drive vendor and model	Micron MTFDHBA256TCK	Intel SSD SSDSC2KB240G8
Drive size (GB)	240	240
Drive information (speed, interface, type)	PCIe® 3.0, M2 NVMe™	6Gb, SATA, SSD
Local storage (VMware vSAN™ capacity)		
Number of drives	2	2
Drive vendor and model	KIOXIA KCD6XLUL3T84	KIOXIA KCD6XLUL3T84
Drive size (TB)	3.84	3.84
Drive information (speed, interface, type)	PCIe® 4.0, NVMe™ 1.4, 64GT/s	PCIe 4.0, NVMe 1.4, 64GT/s

System configuration information	4x Supermicro AS-1124US-TNRP	4x Supermicro SYS-620U-TNR
Local storage (vSAN cache)		
Number of drives	2	2
Drive vendor and model	Samsung MZ-WLJ1T90	Samsung MZ-WLJ1T90
Drive size (TB)	1.92	1.92
Drive information (speed, interface, type)	PCIe 4.0, NVMe SSD, 2.4	PCIe 4.0, NVMe SSD, 2.4
Network adapter		
Vendor and model	Supermicro AOC-URG4N4-i4XTS	Supermicro AOC-2UR68G4-I2XT
Number and type of ports	4x 10GbE	2x 10GbE
Driver version	8.10	8.10
Network adapter		
Vendor and model	Supermicro AOC-S25G-M2S-O	Supermicro AOC-S25G-M2S-O
Number and type of ports	2x 25GbE SFP28	2x 25GbE SFP28
Driver version	14.28.2006	14.28.2006
Cooling fans		
Vendor and model	Sunon PF40561BX-Q40U-S9H	Sunon VF80381B1
Number of cooling fans	8	4
Power supplies		
Vendor and model	Supermicro PWS-1K22A-1R	Supermicro PWS-2K08A-1R
Number of power supplies	2	2
Wattage of each (W)	1,200	2,000

How we tested

Testing overview

We installed and configured the latest available version of VMware vSphere 7.0 Update 3 on a cluster of four Supermicro AS-1124US-TNRP servers with two AMD EPYC 75F3 processors each and a cluster of four Supermicro SYS-620U-TNR servers with two Intel Xeon Platium 8380 processors each. We installed the hypervisor on internal SSD drives. On each cluster, we configured and created a 28TB VMware vSAN™ datastore using four NVMe drives on each server. Each datastore used two disk groups: One 1.92 TB drive for cache and one 3.84TB drive for capacity. The vSAN datastore served as a shared datastore for Hadoop/Hadoop Distributed File System (HDFS) deployment. We deployed a Hadoop cluster with a manager node and four worker nodes (one worker node per host) on each host cluster. On the AMD processor-based cluster hosts, each worker node had 128 vCPUs and 512 GB of memory. On the Intel processor-based cluster hosts, each worker node had 160 vCPUs and 512 GB of memory. We ran the Logistic Regression workload from the HiBench suite and measured run-time and throughput for each cluster.

We used a Dell Networking 1Gb X1052 switch for the VMware management network. We used a Dell 5048-ON 25Gb switch for vMotion®, vSAN, and the VM network. On the 25Gb switch, we configured the following VLANs for the AMD EPYC processor-based cluster: 20 for the vMotion network, and 30 for the vSAN network. We configured the following VLANs for the Intel processor-based cluster: 40 for the vMotion network and 60 for the vSAN. We also configured VLAN 2 for the VM network, which the Intel Xeon processor-based and AMD EPYC processor-based clusters shared for VM traffic. We configured a port channel for each host using the two 25Gb ports from each host. We set the 25Gb switch MTU setting to 9216. We configured a two port LAG on a Distributed vSwitch using two 25Gb ports from each host connected to the 25Gb switch. On the Distributed vSwitch, we created three port groups: vSAN, vMotion, and workload.

The following sections describes the steps we took to configure the test environment and run the test.

Installing vSphere 7.0 Update 3d on a Supermicro server

1. Download ESXi 7.0 Update 3d from the following link: https://my.vmware.com/group/vmware/evalcenter?p=vsphere-eval-7#tab_download.
2. Open a new browser tab, and connect to the IP address of the Supermicro server BMC.
3. Log in with the BMC credentials.
4. In the main screen, click Launch Virtual Console.
5. In the console menu bar, select Virtual Media, and select Virtual Storage.
6. From the drop-down menu in the Virtual Storage pop-up window, select ISO, and click Open Image to browse your local computer and select the image you downloaded in step 1.
7. To mount the ISO image, click Plug In, and click OK.
8. On the console menu bar, click the Power Control, and select Power Reset.
9. The system will boot to the mounted image, and the Loading ESXi installer screen will appear. When prompted, press Enter to continue.
10. To Accept the EULA and Continue, press F11.
11. Select the storage device to target for installation. We selected the internal SD card. To continue, press Enter.
12. To confirm the storage target, press Enter.
13. Select the keyboard layout, and press Enter.
14. Provide a root password, and confirm it. To continue, press Enter.
15. To install vSphere, press F11.
16. To reboot the server upon completion, press Enter.
17. Complete the previous steps for each server under test.

Installing VMware vCenter Server Appliance 7.0 Update 3d

1. Download vCenter 7.0 Update 3d from the VMware support portal: https://my.vmware.com.
2. Mount the image on your local system, and browse to the vcsa-ui-installer folder. Expand the folder for your OS, and launch the installer if it doesn't automatically begin.
3. When the vCenter Server Installer wizard opens, click Install.
4. To begin installation of the new vCenter server appliance, click Next.
5. Check the box to accept the license agreement, and click Next.
6. Enter the IP address of the infrastructure server with ESXi 7.0 Update 2. Provide the root password, and click Next.
7. To accept the SHA1 thumbprint of the server's certificate, click Yes.
8. Accept the VM name, and provide and confirm the root password for the VCSA. Click Next.
9. Set the size for the environment you're planning to deploy. We selected Medium. Click Next.
10. Select the datastore on which to install vCenter. Accept the datastore defaults, and click Next.
11. Enter the FQDN, IP address information, and DNS servers you want to use for the vCenter server appliance. Click Next.

12. To begin deployment, click Finish.
13. Once Stage 1 completes, click Close. To confirm, click Yes.
14. Open a browser window, and connect to https://[vcenter.FQDN:5480/.
15. On the Getting Started - vCenter Server page, click Set up.
16. Enter the root password, and click Log in.
17. Click Next.
18. Enable SSH access, and click Next.
19. To confirm the changes, click OK.
20. Type `vsphere.local` for the Single Sign-On domain name. Enter a password for the administrator account, confirm it, and click Next.
21. Click Next.
22. Click Finish.

Creating a vSphere cluster in vCenter

1. Open a browser, and enter the address of the vCenter server you deployed. For example: `https://[vcenter.FQDN]/ui`.
2. In the left panel, select the vCenter server, right-click, and select New Datacenter.
3. Provide a name for the new data center, and click OK.
4. Select the data center you just created, right-click, and select New Cluster.
5. Give a name to the cluster, and enable vSphere DRS. Click OK.
6. In the cluster configuration panel, under Add hosts, click Add.
7. Check the box for Use the same credentials for all hosts. Enter the IP address and root credentials for the first host, and enter the IP addresses of all remaining hosts. Click Next.
8. To select all hosts, check the box beside Hostname/IP Address. Click OK.
9. Click Next.
10. Click Finish.

Creating a distributed vSwitch and port group

1. From vSphere client, click Home→Networking.
2. Select your Datastore.
3. In the Actions drop-down menu in the right panel, select Distributed vSwitch→New Distributed vSwitch.
4. Give your vSwitch a name, or accept the default. Click Next.
5. Select 7.0.2 - ESXi 7.0.2 and later as the version, and click Next.
6. Select the number of uplinks per ESXi host you'll give to the vSwitch. We selected 2. Click Next.
7. Click Finish.
8. Select new DSwitch, and click Configure.
9. Select Properties, and click Edit.
10. Under advanced, set MTU (Bytes) to 9,000.
11. Click OK.
12. Select LACP, click + NEW, and set mode to Active.
13. Click OK.
14. Enter an appropriate name, set ports to two, and click OK.
15. Right-click the new DSwitch, and select Add and Manage Hosts.
16. Leave Add hosts selected, and click Next.
17. To add new hosts, click +.
18. To select all the hosts in your target cluster, check the box beside Host. Click OK. Click Next.
19. Select the NIC you want to use for this DSwitch, and click Assign Uplink.
20. Select the first LAG port.
21. Check the box for Apply this uplink assignment to the rest of the hosts. Click OK.
22. Complete steps 16 through 18 for the second NIC, and assign it to the second LAG port.
23. Click Next.
24. Do not assign VMkernel adapters at this time. Click Next.
25. Do not migrate any VM networking at this time. Click Next.
26. Click Finish.
27. Right-click the DvSwitch, and select Distributed Port Group→New Distributed Port Group.
28. Name it `Workload Network`, and click Next.
29. Change the VLAN type to VLAN, and set the VLAN ID to VLAN 2. Click Next.

30. Click Finish.
31. On the AMD EPYC processor-based cluster, complete steps 17 through 20 two times using the following settings:
    - Name: vMotion; VLAN ID 20
    - Name: vSAN; VLAN ID 30
32. For the Intel Xeon processor-based cluster, use the following port group settings:
    - Name: vMotion; VLAN ID 40
    - Name vSAN; VLAN ID 60

Creating a VMkernel Adapter

Note: We completed these steps on each server under test for the private network, vSAN network, and vMotion network using the appropriate IPv4 settings for each network and enabling the corresponding service for each VMkernel adapter.

1. From the vCenter UI in the Hosts and Clusters view, right-click a host, and click Add Networking.
2. Select VMkernel Network Adapter, and click Next.
3. Click Browse, and select existing network.
4. Select the desired Distributed Port Group that you created in the previous section.
5. If necessary, assign a different network label.
6. Select the desired service or services for the VMkernel, and click Next.
7. Select Use static IPv4 settings, and enter the desired IPv4 address, subnet mask, and default gateway.
8. Click Next.
9. Click Finish.

Configuring the vSAN datastore

1. In the left panel of the vCenter, right-click the test cluster, select vSAN, and select configuration.
2. Select single site cluster, and click Next.
3. In the Services screen, click Next.
4. In the Claim disks screen, click Group by, and select Host.
5. Select disks for each host. We selected one drive for cache and two drives for capacity. Click Next.
6. In the Review screen, click Finish.
7. Click the test cluster from vCenter, and in the right panel, click the Configure tab .
8. In the vSAN section, select Services.
9. To create a distributed switch, in the Configure Cluster section, click Configure.
10. In the Physical adapters section, choose two uplink adapters for the distributed switch, and click Next.
11. In the Storage traffic screen, set the VLAN ID assigned for vSAN, choose Static IPs, and fill in the static IP address, Subnet mask, and default gateway. Click Next.
12. In the Advanced options screen, click Next.
13. In the Claim disks screen, click Next.
14. In the Proxy settings screen, click Next.
15. Click Finish.

Installing and configuring a base CentOS VM

1. Log into vCenter, and from Menu, click Storage.
2. Select datastore1, and click Files.
3. Click Upload Files, and upload the CentOS stream 8 ISO image.
4. Right-click the cluster, and click New Virtual Machine.
5. Click Next.
6. Enter a name for the VM, and click Next.
7. Click Next.
8. Select the vSAN Datastore, and click Next.
9. Click Next.
10. From the Guest OS Family drop-down menu, select Linux.
11. From the Guest OS version drop-down menu, select CentOS 8 (64 bit), and click Next.

12. Assign the VM two vCPUs, 8 GB of memory, and a 512GB hard disk.
13. From the New CD/DVD Drive drop-down, select Datastore ISO File, and select the CentOS ISO you uploaded to the datastore previously. Ensure Connect At Power On is checked, and click Next.
14. Click Finish.
15. Power on the VM, and click Launch Remote Console.
16. Select Install CentOS Linux 8, and press Enter.
17. Select English as language, and click Continue.
18. On the Installation summary screen, click Network & Host Name.
19. Turn on the switch for the ethernet port, and click Done.
20. Click Root Password, and set a password for the root user.
21. Click Installation Source, select http:// from the drop-down menu, and type `mirror.centos.org/centos/8-stream/BaseOS/x86_64/os/`.
22. Click Done.
23. Click Software Selection, select Server with GUI, and click Done.
24. Click Installation Destination, and select the disk for OS.
25. Click Begin Installation.
26. When installation completes, click Reboot System.

Configuring Centos 8 and installing Apache Hadoop and Spark

1. Log into the base CentOS VM via SSH as user root.
2. Modify SSH with the following configurations to allow a pre-shared key login:

```
mkdir -p /root/.ssh
chmod 700 /root/.ssh
cd /root/.ssh
ssh-keygen -t rsa -q
cp id_rsa.pub authorized_keys
echo "StrictHostKeyChecking=no" > config
```

3. Type the following command to set the hostname:

```
hostnamectl set-hostname [HOSTNAME]
```

4. Modify your hosts file to add your hostname to your IP address.
5. Type the following command to turn off and disable your firewall:

```
systemctl stop firewalld
systemctl disable firewalld
```

6. Type the following command to edit your selinux so it disables its enforcing:

```
setenforce 0
vi /etc/selinux/config (modify "enforcing" to "disabled" in the file)
```

7. Type the following to update your OS:

```
yum upgrade -y
```

8. Type the following commands to install the prerequisites via yum:

```
yum install https://dl.fedoraproject.org/pub/epel/epel-release-latest-8.noarch.rpm
yum install -y nmon python3 vim tar wget java-1.8.0-openjdk maven git blas64 lapack64
python2 bc curl
```

9. Type the following commands to download Hadoop and Spark:

```
wget http://www.gtlib.gatech.edu/pub/apache/spark/spark-3.1.2/spark-3.1.2-bin-hadoop3.2.tgz
wget http://www.gtlib.gatech.edu/pub/apache/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz
```

10. Modify your bash profile with `~/.bashrc`, and add the following lines:

```
JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.322.b06-11.el8.x86_64/jre
PATH=$PATH:$HOME/bin:/opt/yarn/hadoop-3.3.0/bin
HADOOP_HOME=/opt/yarn/hadoop-3.3.0
```

11. Type the following commands to add the Hadoop users:

```
groupadd hadoop
useradd -g hadoop yarn
useradd -g hadoop hdfs
useradd -g hadoop mapred
```

12. Type the following commands to create default Hadoop directories and set their permissions:

```
mkdir -p /var/data/hadoop/hdfs/nn
mkdir -p /var/data/hadoop/hdfs/snn
mkdir -p /var/data/hadoop/hdfs/dn
chown hdfs:hadoop /var/data/hadoop/hdfs/ -R
mkdir -p /var/data/hadoop/hdfs/tmp
mkdir -p /var/log/hadoop/yarn
chown yarn:hadoop /var/log/hadoop/yarn/ -R
mkdir -p /opt/yarn
```

13. Type the following commands to extract the Hadoop and Spark compressed files:

```
cd /opt/yarn
tar xvzf /root/hadoop-3.3.0.tar.gz
tar -xvzf ~/spark-3.1.2-bin-hadoop3.2.tgz
```

14. Type the following commands to move into the Hadoop directory and make a yarn directory:

```
cd hadoop-3.3.0/
mkdir logs
chmod g+w logs
chown yarn:hadoop . -R
```

15. Append the following line to `/opt/yarn/spark-3.1.2-bin-hadoop3.2/conf/spark-env.sh` on each VM:

```
export SPARK_LOCAL_DIRS=/var/data/hadoop/hdfs/tmp
```

16. Type the following command to navigate into the Hadoop configuration directory:

```
cd etc/hadoop/
```

17. Modify the Hadoop configuration files with the following settings:

```
core-site.xml
<configuration>
       <property>
              <name>fs.default.name</name>
              <value>hdfs://[MANAGER IP ADDRESS]:9000</value>
       </property>
```

```
       <property>
               <name>hadoop.http.staticuser.user</name>
               <value>hdfs</value>
       </property>
</configuration>

hdfs-site.xml
<configuration>
 <property>
   <name>dfs.replication</name>
   <value>3</value>
 </property>
 <property>
   <name>dfs.namenode.name.dir</name>
   <value>file:/var/data/hadoop/hdfs/nn</value>
 </property>
 <property>
   <name>fs.checkpoint.dir</name>
   <value>file:/var/data/hadoop/hdfs/snn</value>
 </property>
 <property>
   <name>fs.checkpoint.edits.dir</name>
   <value>file:/var/data/hadoop/hdfs/snn</value>
 </property>
 <property>
   <name>dfs.datanode.data.dir</name>
   <value>file:/var/data/hadoop/hdfs/dn</value>
 </property>
</configuration>

mapred-site.xml
<configuration>
 <property>
   <name>mapreduce.framework.name</name>
   <value>yarn</value>
 </property>
    <property>
            <name>yarn.app.mapreduce.am.env</name>
            <value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
    </property>
    <property>
            <name>mapreduce.map.env</name>
            <value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
    </property>
    <property>
            <name>mapreduce.reduce.env</name>
            <value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
    </property>
</configuration>

yarn-site.xml
<configuration>
        <property>
                <name>yarn.resourcemanager.hostname</name>
                <value>[MANAGER HOSTNAME HERE]</value>
        </property>
 <property>
   <name>yarn.nodemanager.aux-services</name>
   <value>mapreduce_shuffle</value>
 </property>
 <property>
   <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
   <value>org.apache.hadoop.mapred.ShuffleHandler</value>
 </property>
</configuration>

  hadoop-env.sh
  Uncomment the JAVA_HOME line and add the following information:
JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.322.b06-11.el8.x86_64/jre
Type the following command to power off the instance:
poweroff
```

Creating manager and worker nodes for Hadoop and Spark

1. Log into vCenter, and right-click the base CentOS VM.
2. Select Clone, and select Clone to Virtual Machine.
3. Enter a name for the virtual machine, and click Next.
4. Select a host for the virtual machine, and click Next.
5. Select the vSAN Datastore, and click Next.
6. Select Customize this virtual machine's hardware, and click Next.
7. For the Manager VM, select two vCPUs and 64 GB of memory, and click Next.
8. For Worker VMs, select 128 vCPUs on the AMD EPYC processor-based servers or 160 vCPUs on the Intel Xeon processor-based servers, and 512 GB of memory. Click Next.
9. Click Finish.
10. Once cloning completes, power on the virtual machine.

Configuring and starting the cluster

1. Type the following command to set the hostname on the primary and each of the worker nodes:

```
hostnamectl set-hostname <HOST_NAME>
```

2. Add the FQDN, hostname, and IP address of the manager VM and all worker VMs to the /etc/hosts file on the manager and all worker nodes.
3. Verify that you can do passwordless SSH into each VM.
4. On the Manager VM, format the hdfs filesystem with the following:

```
hdfs namenode -format
```

5. Start the Hadoop services and Spark on the manager node with the following commands:

```
/opt/yarn/hadoop-3.3.0/bin/hdfs --daemon start namenode
/opt/yarn/hadoop-3.3.0/bin/hdfs --daemon start secondarynamenode
/opt/yarn/hadoop-3.3.0/bin/yarn --daemon start resourcemanager
/opt/yarn/hadoop-3.3.0/bin/yarn --daemon start nodemanager
/opt/yarn/spark-3.1.2-bin-hadoop3.2/sbin/start-master.sh
```

6. Type the following commands to start the Hadoop services and Spark on each of the worker nodes:

```
/opt/yarn/hadoop-3.3.0/bin/hdfs --daemon start datanode
/opt/yarn/spark-3.1.2-bin-hadoop3.2/sbin/start-slave.sh spark://[MANAGER IP ADDRESS]:7077
```

7. Type the following commands to verify that HDFS is running with all four data nodes:

```
hdfs dfsadmin -report
```

8. Verify Spark is running by going to the remote console of the Manager VM. Open a web browser, and type `localhost:8080` for the address. Verify all four worker nodes have joined the Workers list.
9. To install and configure HiBench, perform the following steps on the manager node:
    a. Type the following commands to create the directories for HiBench:

```
hdfs dfs -mkdir -p /user/root
hdfs dfs -mkdir /HiBench
hdfs dfs -chown -R root:hadoop /HiBench
hdfs dfs -chown root /user/root
```

b. Type the following commands to navigate to your home directory and download HiBench:

```
cd ~
git clone https://github.com/intel-hadoop/HiBench.git
```

c. Install HiBench for Spark 3.0 with the following commands:

```
cd HiBench/
mvn -Dspark=3.0 -Dscala=2.12 clean package | tee hibench_build.log
cd conf/
```

d. Modify the HiBench configuration files with the following information:

```
hadoop.conf
# Hadoop home
hibench.hadoop.home     /opt/yarn/hadoop-3.3.0

# The path of hadoop executable
hibench.hadoop.executable     ${hibench.hadoop.home}/bin/hadoop

# Hadoop configraution directory
hibench.hadoop.configure.dir  ${hibench.hadoop.home}/etc/hadoop

# The root HDFS path to store HiBench data
hibench.hdfs.master       hdfs://[MANAGER IP ADDRESS]:9000


# Hadoop release provider. Supported value: apache, cdh5, hdp
hibench.hadoop.release    apache

spark.conf
# Spark home
hibench.spark.home      /opt/yarn/spark-3.1.2-bin-hadoop3.2/

# Spark master
#   standalone mode: spark://xxx:7077
#   YARN mode: yarn-client
hibench.spark.master spark://[MANAGER IP ADDRESS]:7077
```

Running the tests

In this section, we list the steps to run the logistic regression (LR) benchmark on the VMs under test. The manager node starts the benchmark using a script that automates the entire process at each instance size.

1. Log into the manager node via SSH.
2. Navigate to the HiBench prepare directory for LR, and generate the input dataset.

```
cd ~/HiBench/bin/workloads/ml/lr/prepare
./prepare.sh
```

3. Type the following to navigate to the Spark directory:

```
cd ~/HiBench/bin/workloads/ml/lr/spark
```

4. Type the following to clear the cache on the manager and worker nodes before running a test:

```
sync; echo 3 > /proc/sys/vm/drop_caches
```

5. Run the benchmark with the following:

```
./run_test.sh 2>&1 | tee run.log
```

6. Run the same test three times on each cluster. Report the median of these three runs.

Read the report at https://facts.pt/5D5qBFW ▶

This project was commissioned by AMD.

Principled Technologies®

Facts matter.®