Creating the Hadoop VMs on Azure

Course 20773A, Analyzing Big Data with Microsoft R, requires that each student has access to a VM running Hadoop and Microsoft R server. This VM runs using Microsoft's Azure cloud platform. To save resources and setup effort, each student shares access to this VM. You should create this VM before the course starts, but the instructor should start it up and shut it down only when required by the labs and demos, to help minimize costs.

This guide describes the steps for creating the Hadoop VM. The steps are correct as of the time of publishing.

**As Azure is regularly updated and improved, there is a possibility that this guide may be out of date.**

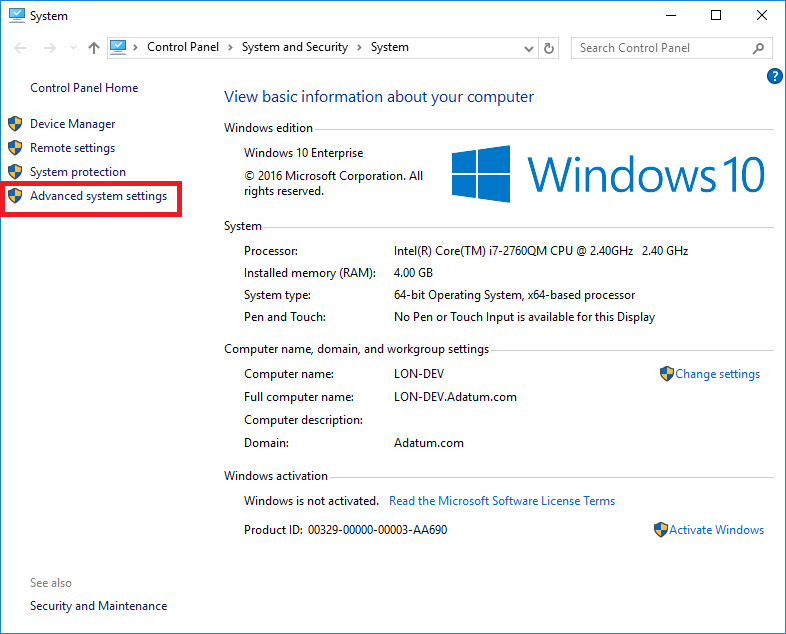
**Before following the steps below, please follow the details of how to acquire a Microsoft Azure pass for you and your class here:** [**http://go.microsoft.com/fwlink/?LinkId=512034**](http://go.microsoft.com/fwlink/?LinkId=512034)

**You also require Microsoft R Server 9.1.0 for Hadoop. This is available as a subscriber download from MSDN. You should download the gzip file containing this software to the desktop machine before starting this process.**

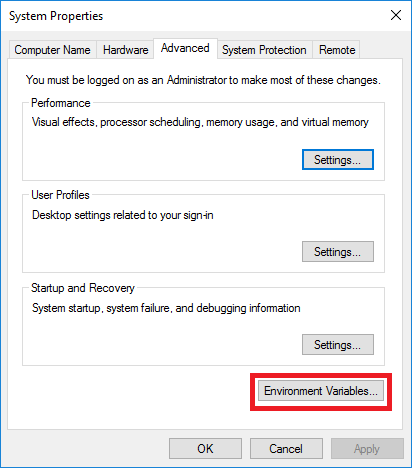
# Install PuTTY

The Hadoop VM runs CentOS Linux. You cannot connect to a Linux VM by using Remote Desktop without installing and configuring additional software, which can be a time-consuming process. Therefore this document uses SSH connections from an SSH client running on the Windows desktop. The simplest SSH client to install and use is PuTTY, a freely available open-source package. Follow these instructions to download and install PuTTY on the desktop machine.

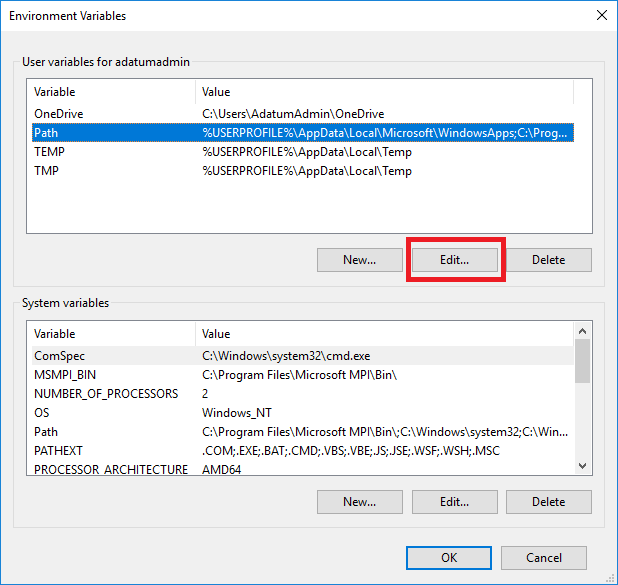
1. In Internet Explorer, browse to <https://the.earth.li/~sgtatham/putty/0.68/w64/putty-64bit-0.68-installer.msi>.
2. In the Internet Explorer message box, click **Run**.
3. In the **PuTTY Setup** wizard, on the **Welcome** page, click **Next**.
4. On the **Destination Folder** page, click **Next**.
5. On the **Product Features** page, click **Install**.
6. In the **User Account Control** dialog box, click **Yes**.
7. When the wizard has completed, clear **View README file**, and then click **Finish**.
8. Right-click the Windows **Start** button, and then click **System**.
9. In the **System** dialog box, click **Advanced System Settings**.



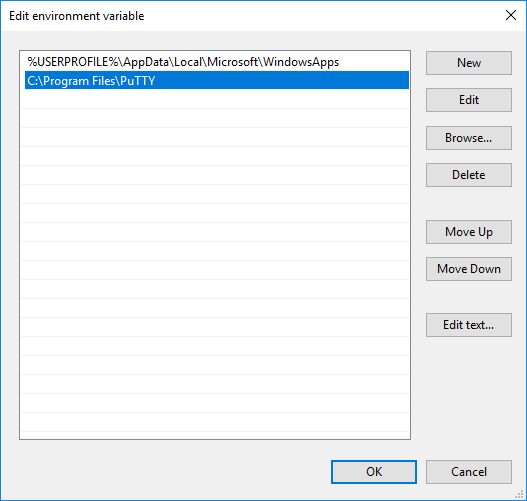
1. In the **System Properties** dialog box, click **Environment Variables**.



1. In the **Environment Variables** dialog box, click **Path**, and then click **Edit**.



1. In the **Edit User Variable** dialog box, append the path **C:\Program Files\PuTTY** to the **Variable value**, and then click **OK**.



1. In the **Environment Variables** dialog box, click **OK**.
2. In the **System Properties** dialog box, click **OK**.
3. Close the **System** dialog box.

# Log into Azure

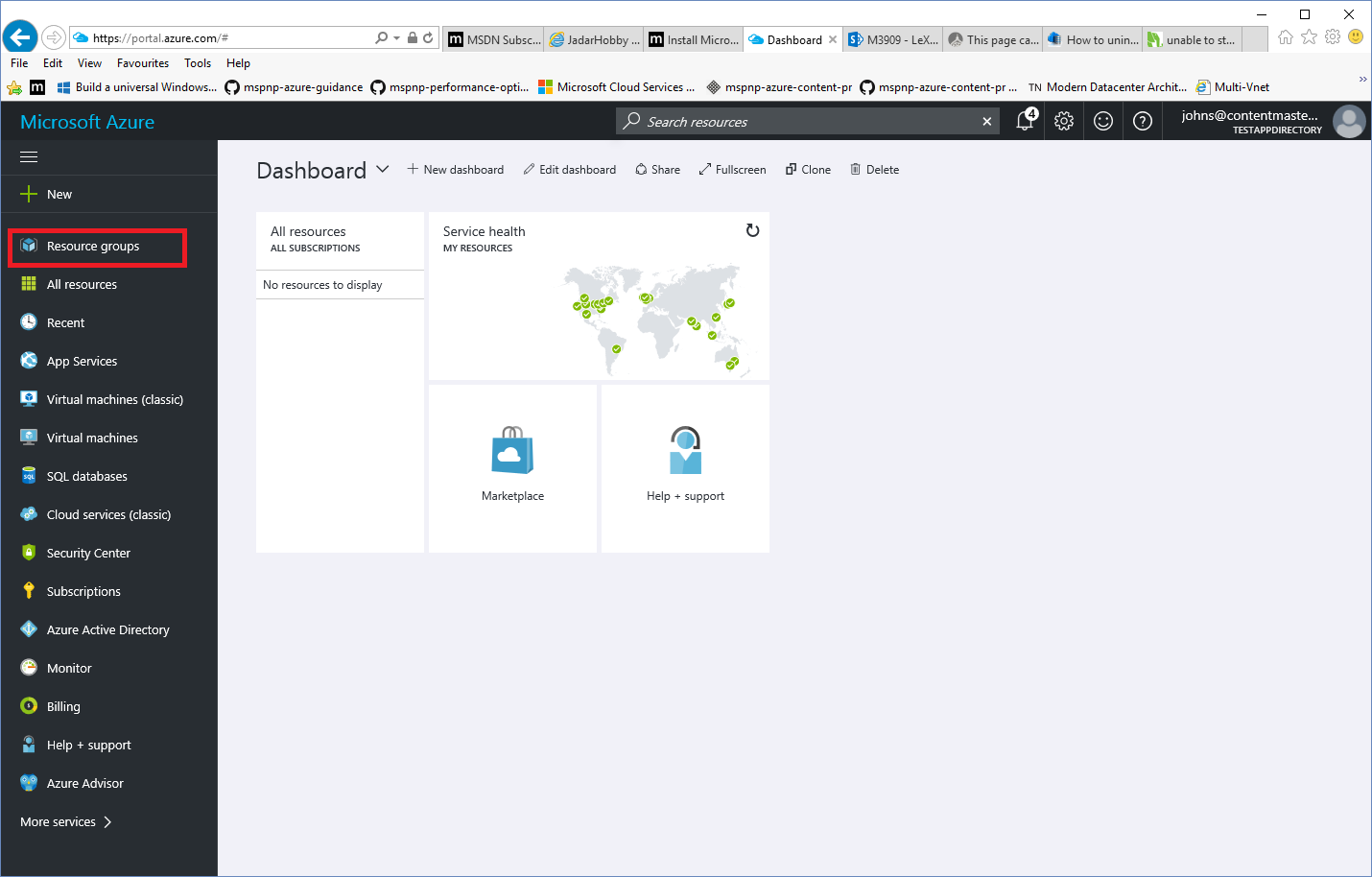
1. You will require a Microsoft account to login to the Azure Portal. The following steps assume you have already created these credentials.
2. On the Start menu, type **Internet Explorer**, and then click **Internet Explorer**.
3. In the address bar, type **portal.azure.com**, and then press Enter.
4. Enter your Microsoft account credentials to log in.

# Create the Resource Group for the Virtual Machine

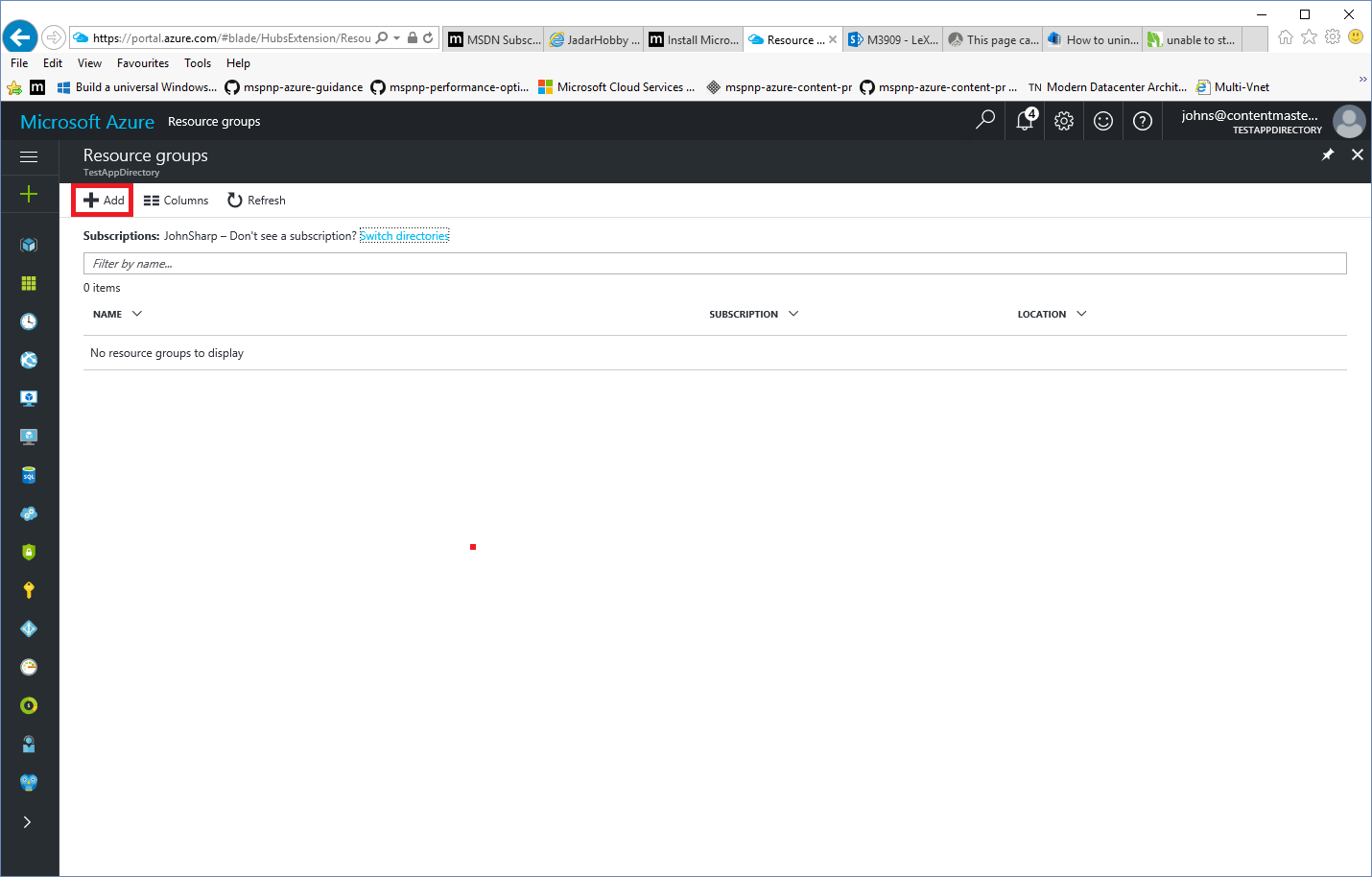
You create the VM and its resources in the same resource group. This helps to make management easier.

The following steps create the resource group.

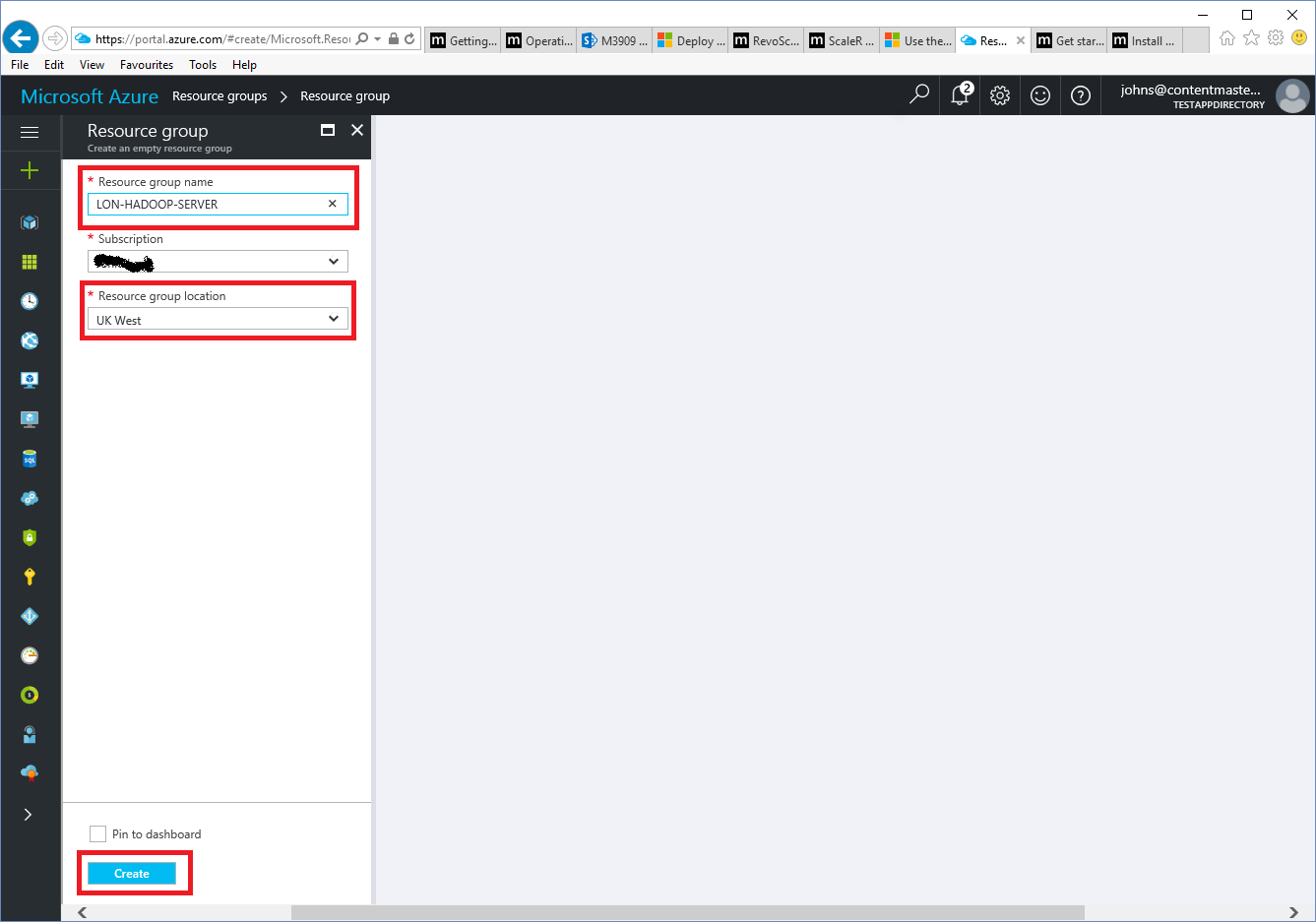
1. In the navigation blade on the left side of the portal, click **Resource groups**.



1. In the **Resource groups** blade, click **Add**.



1. In the **Resource group** blade, in the **Resource group name** box, type **LON-HADOOP-SERVER**, select your nearest location, and then click **Create**.

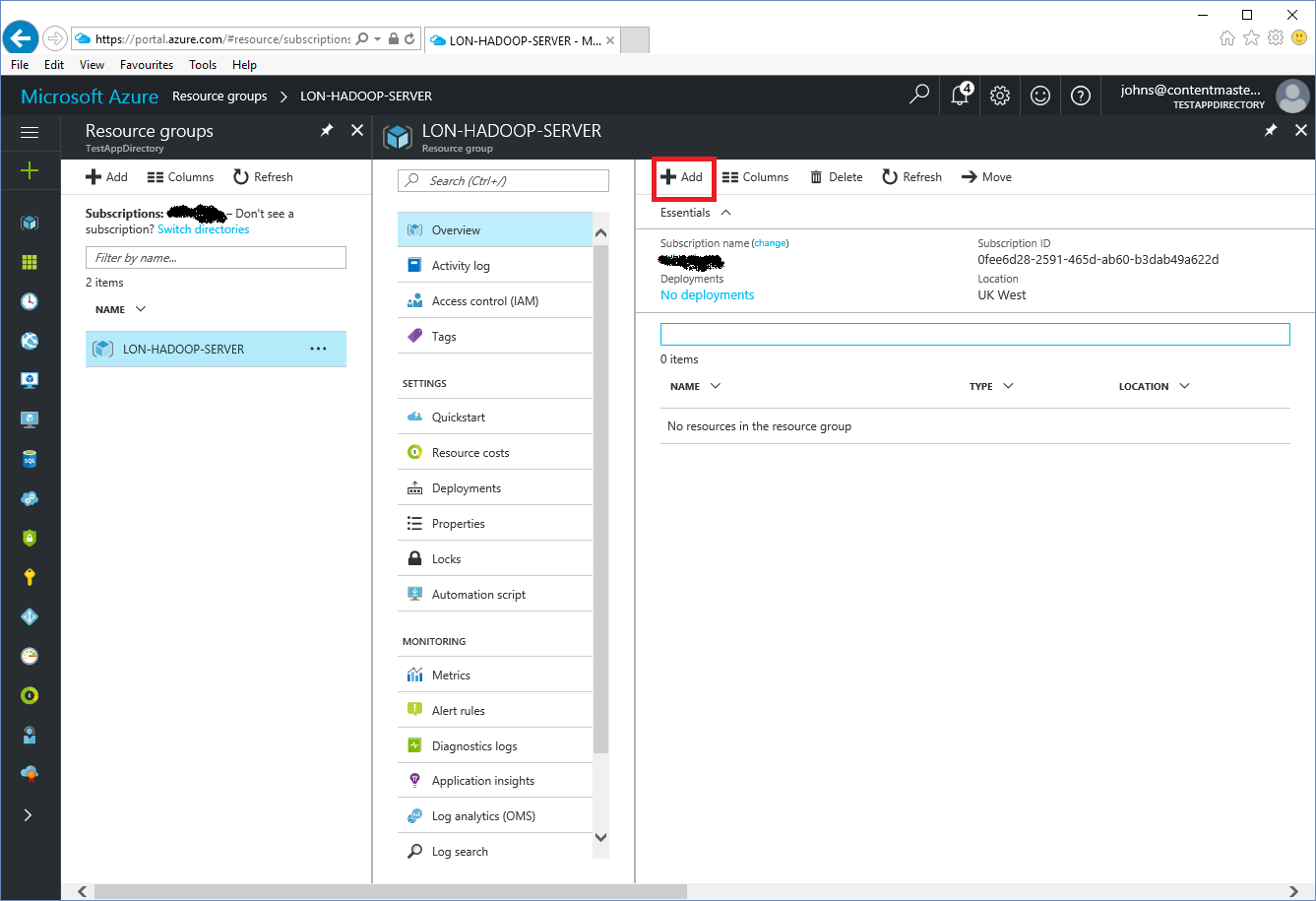


# Create the LON-HADOOPVirtual Machine

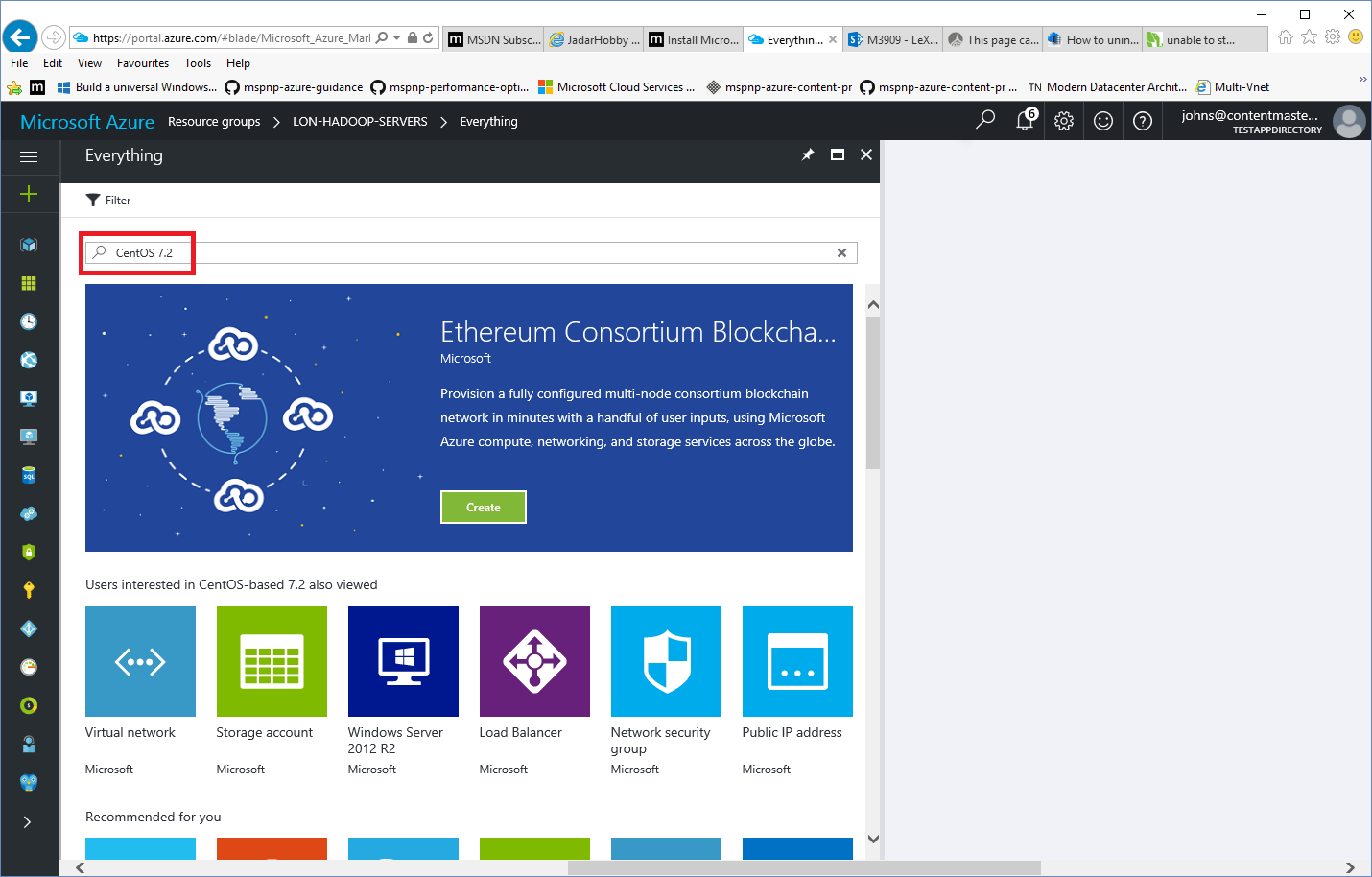
1. In the navigation blade on the left side of the portal, click **Resource groups**.
2. Click the **LON-HADOOP-SERVER** resource group.



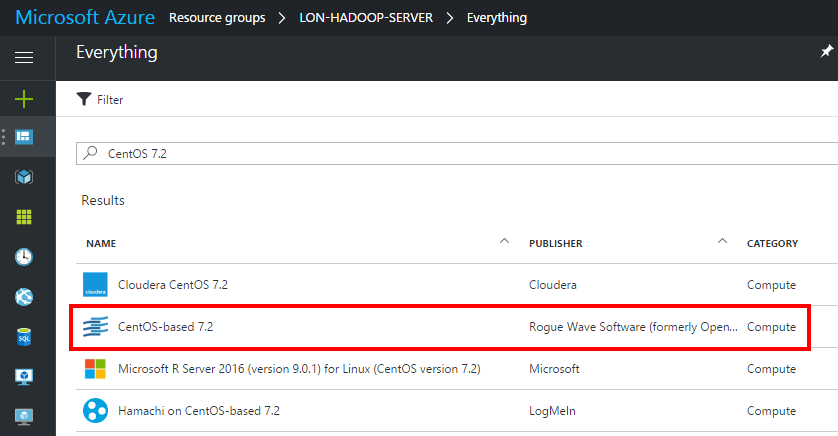
1. In the **LON-HADOOP-SERVER** blade, click **Add**.



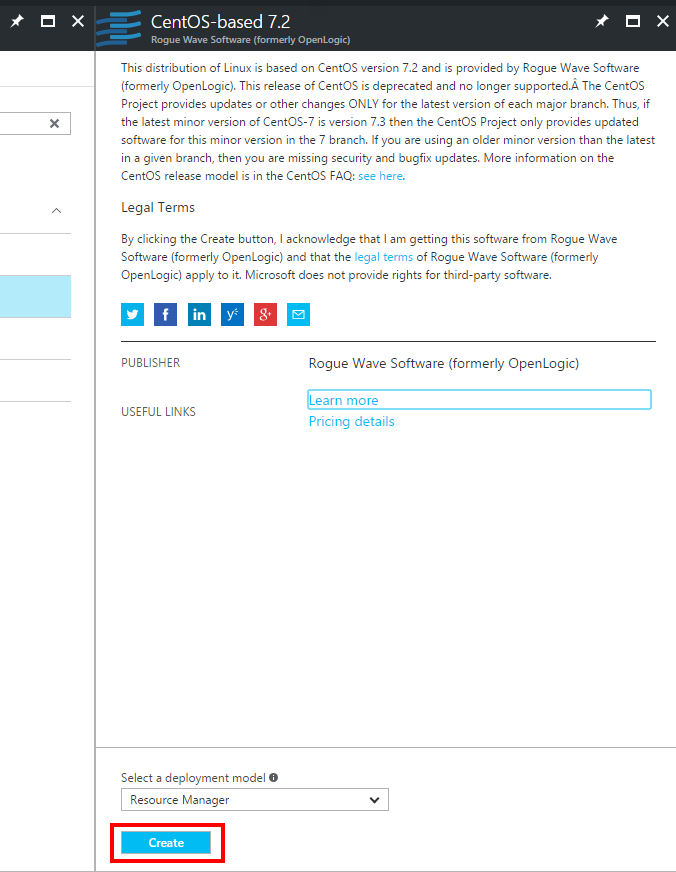
1. In the search box, type **CentOS 7.2**, and then press Enter.



1. Click **CentOS-based 7.2**, published by **Rogue Wave Software (formerly OpenLogic)**.

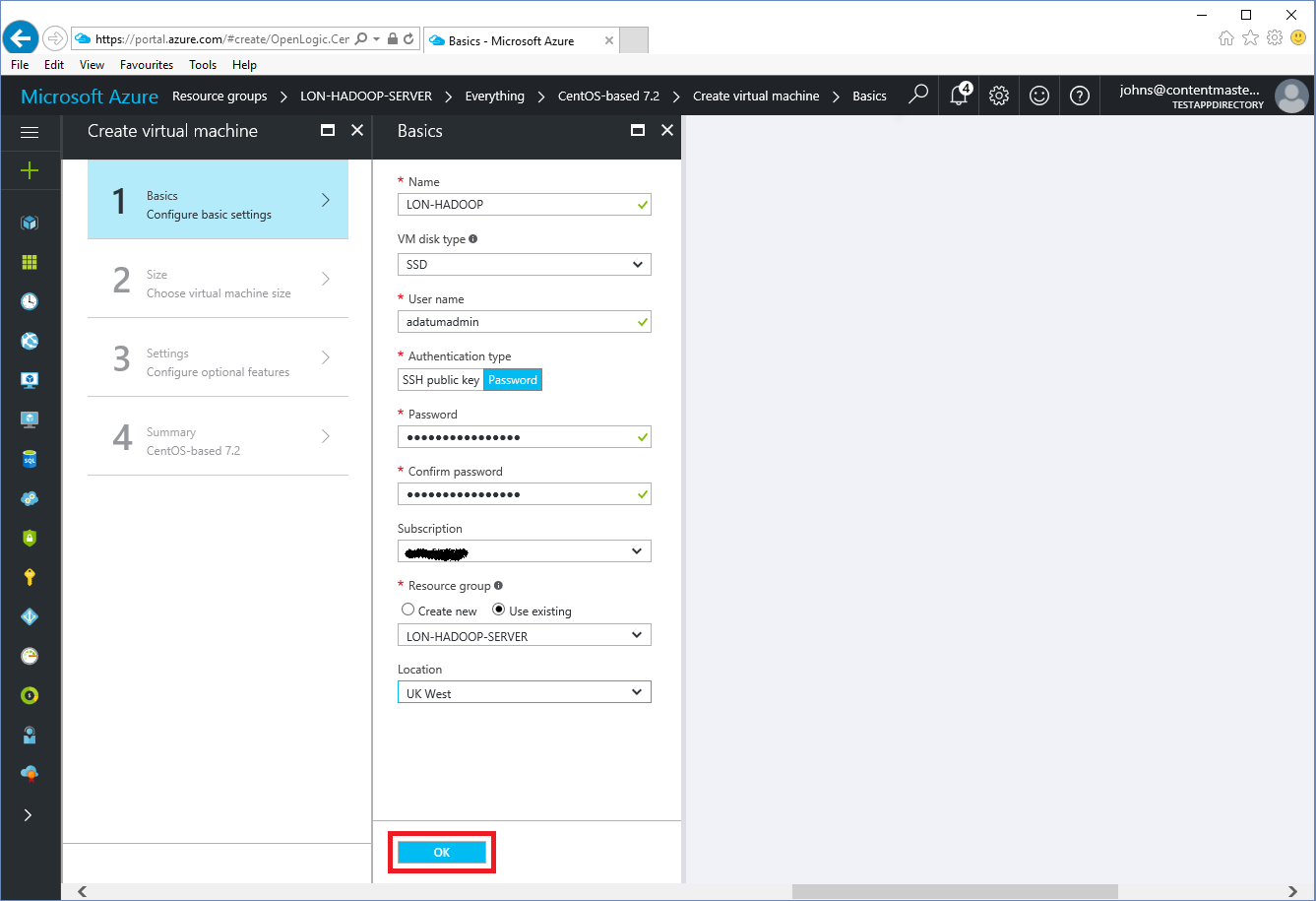


1. In the **CentOS-base 7.2** blade, Click **Create**.

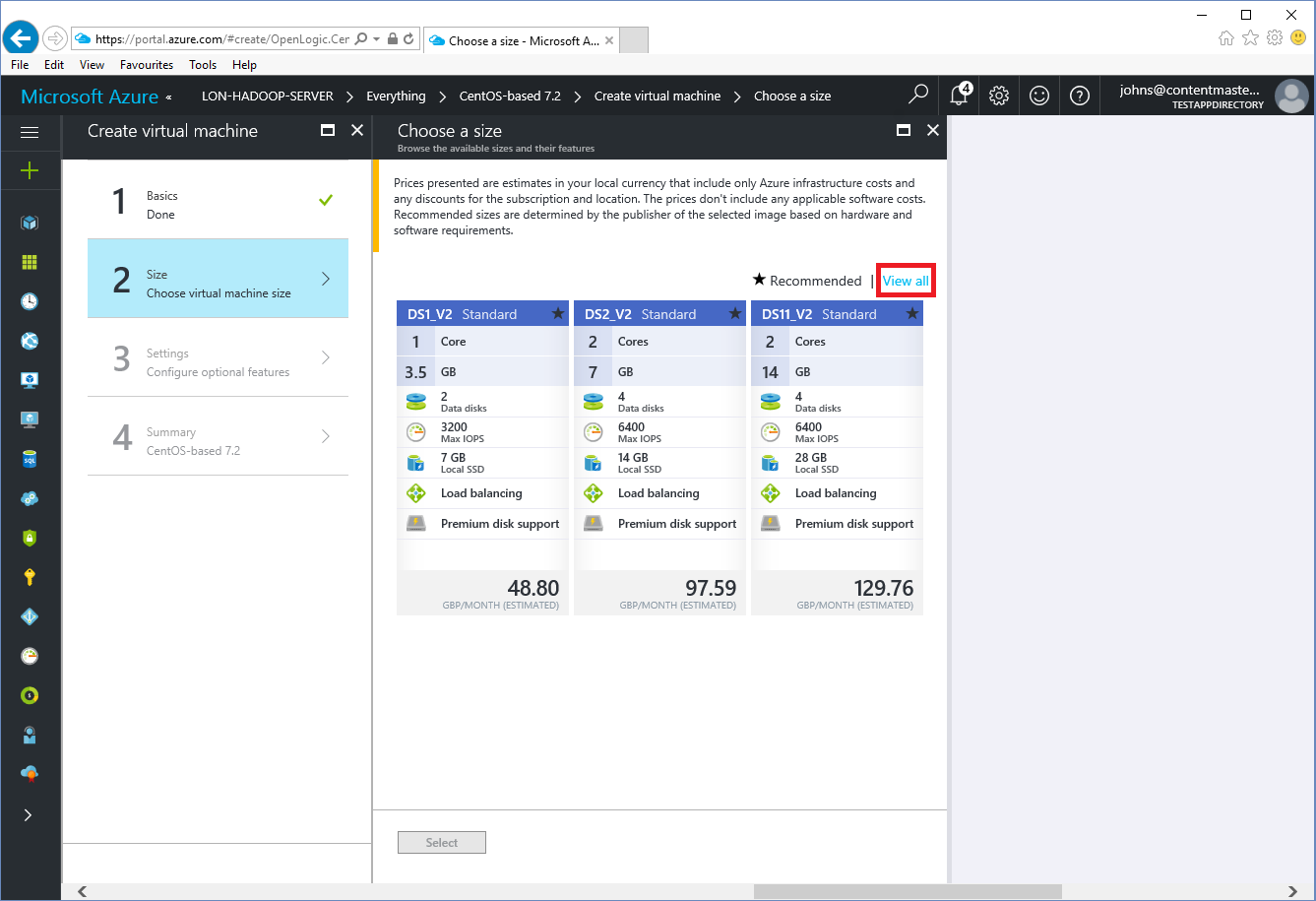


1. In the **Basics** blade, enter the values shown in the following table and then click **OK**.

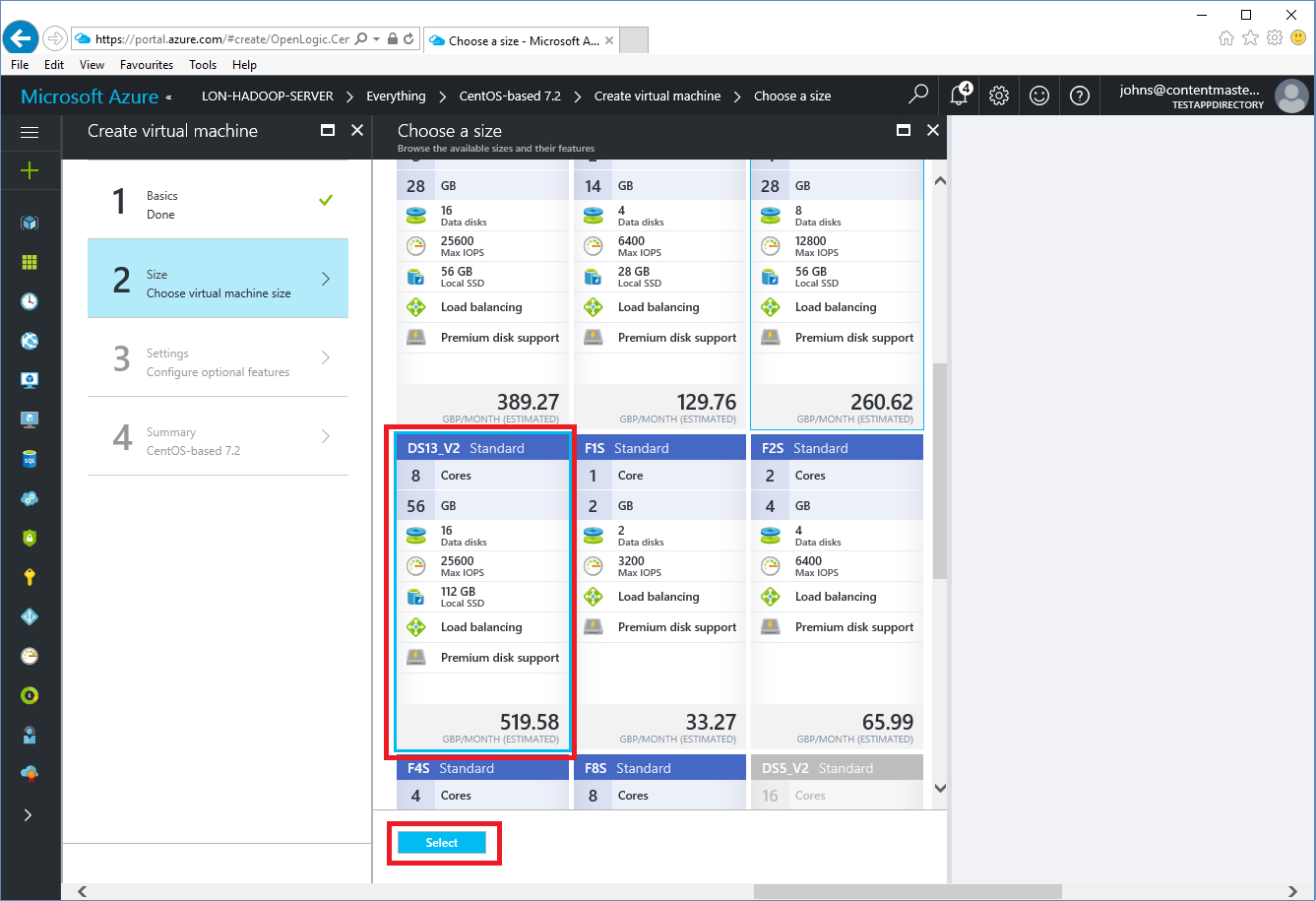
|  |  |
| --- | --- |
| Property | Value |
| Name | LON-HADOOP |
| VM disk type | SSD |
| User name | adatumadmin |
| Authentication type | Password |
| Password | Pa55w.rdPa55w.rd *(Note: The repetition is intentional)* |
| Confirm password | Pa55w.rdPa55w.rd |
| Subscription | *Specify your subscription* |
| Resource group | Use existing, LON-HADOOP-SERVER |
| Location | *Specify the same location that* you used when you created the resource group |



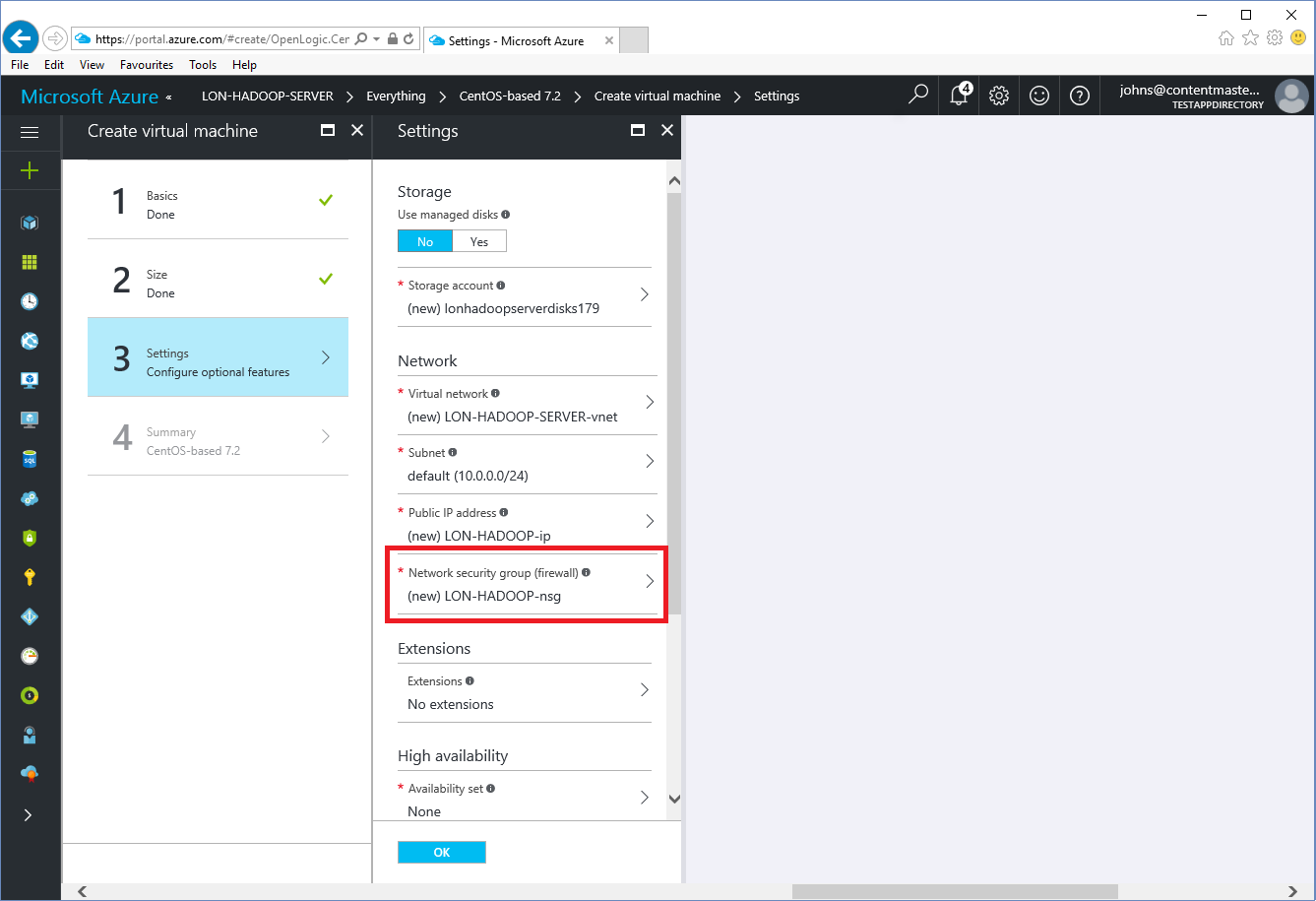
1. In the **Choose a size** blade, click **View all**.



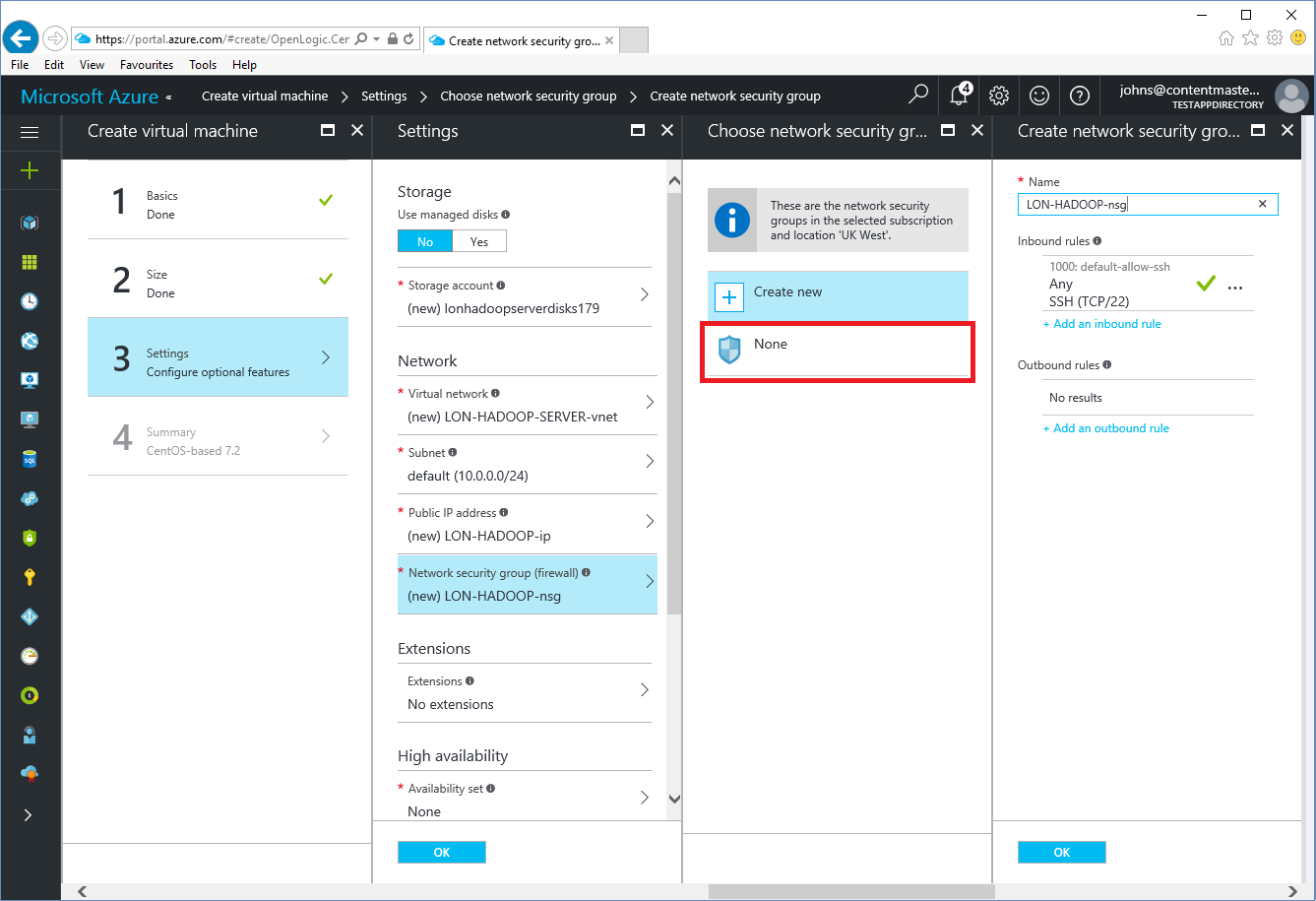
1. Scroll down, click **DS13\_V2 Standard**, and then click **Select**

****

1. In the **Settings** blade, click **Network security group**.



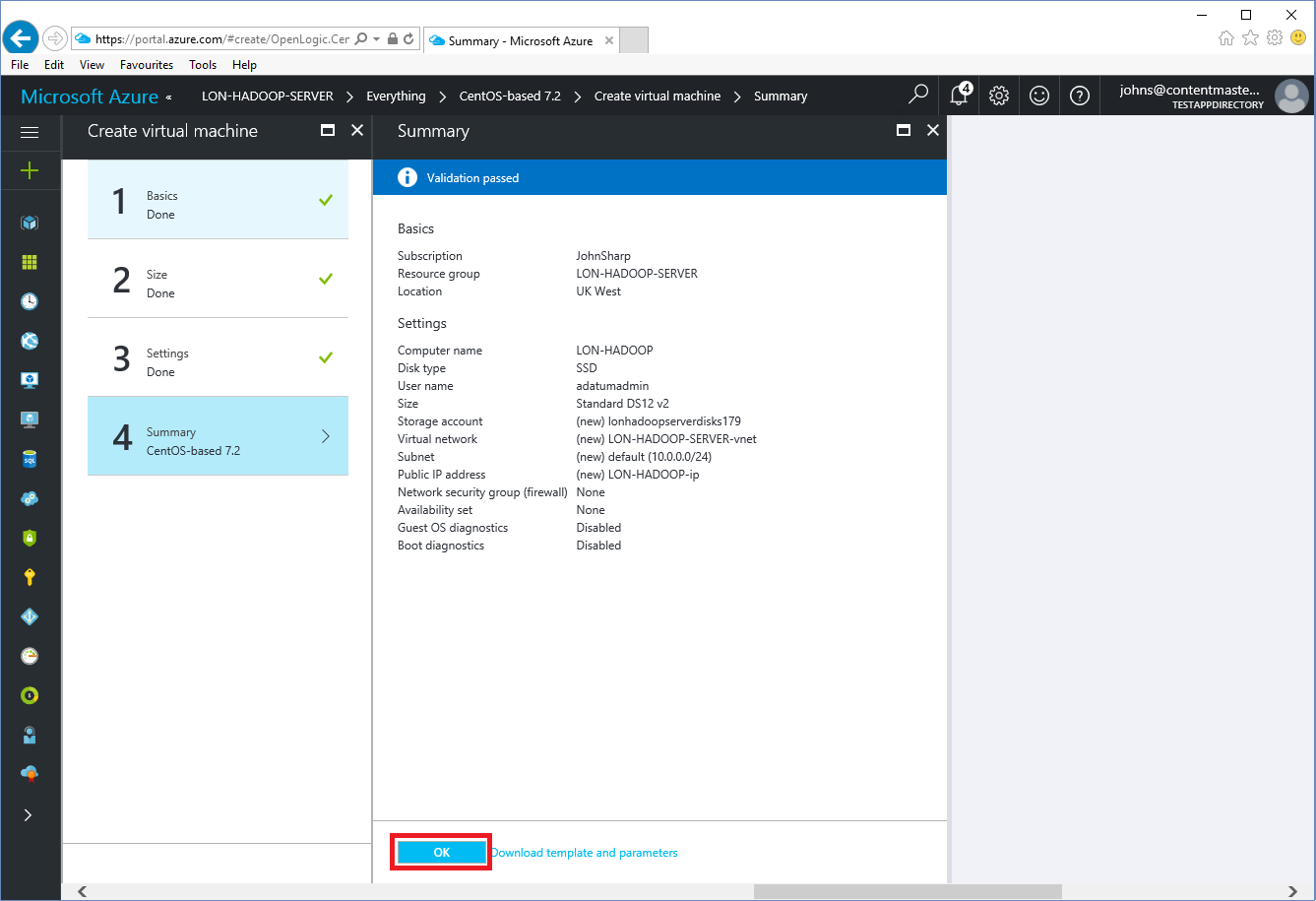
1. In the **Choose network security group** blade, click **None**.



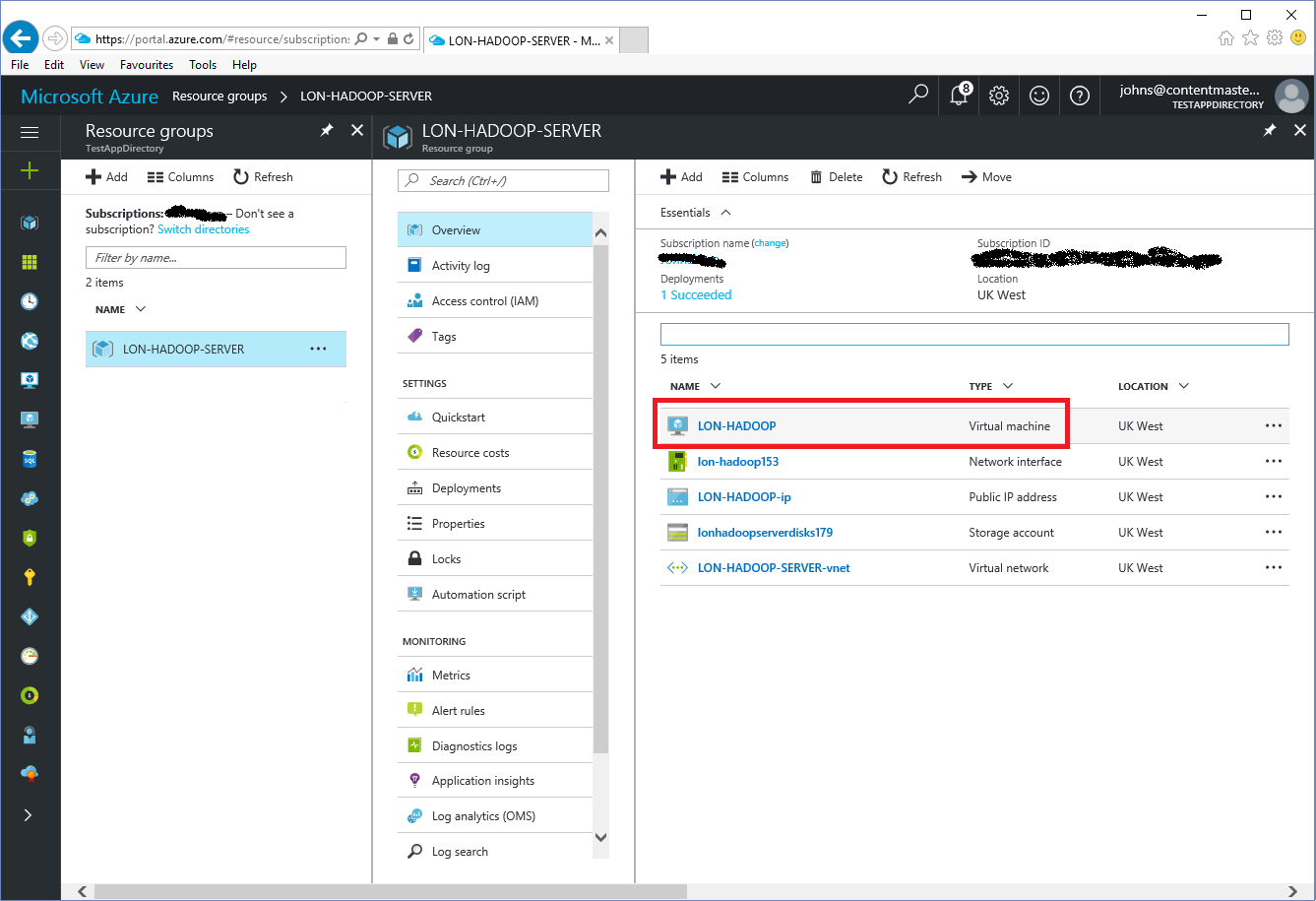
1. In the **Settings** blade, under **Monitoring Boot diagnostics**, click **Disabled**, and then click **OK**.



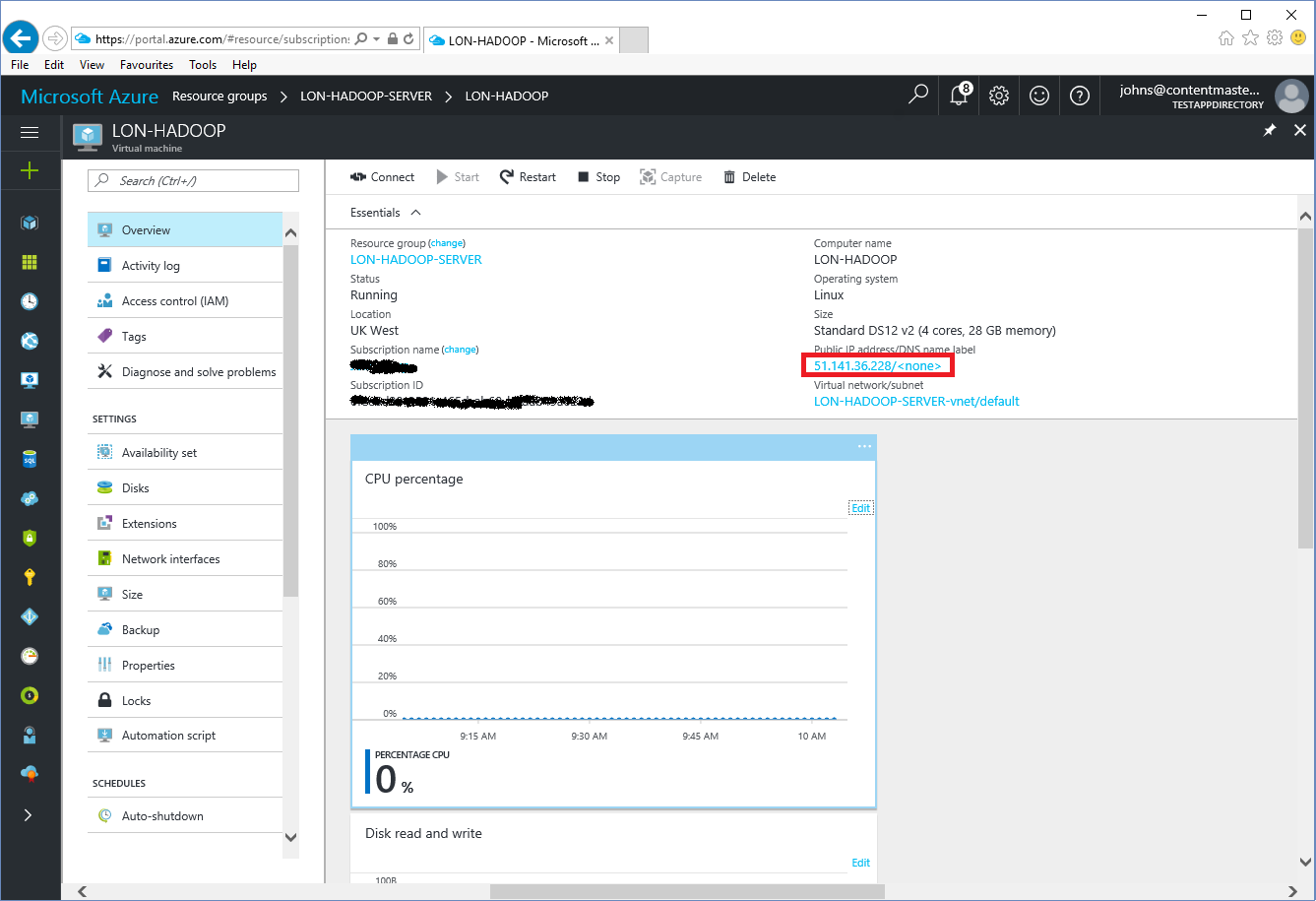
1. In the **Summary** blade, click **OK**.



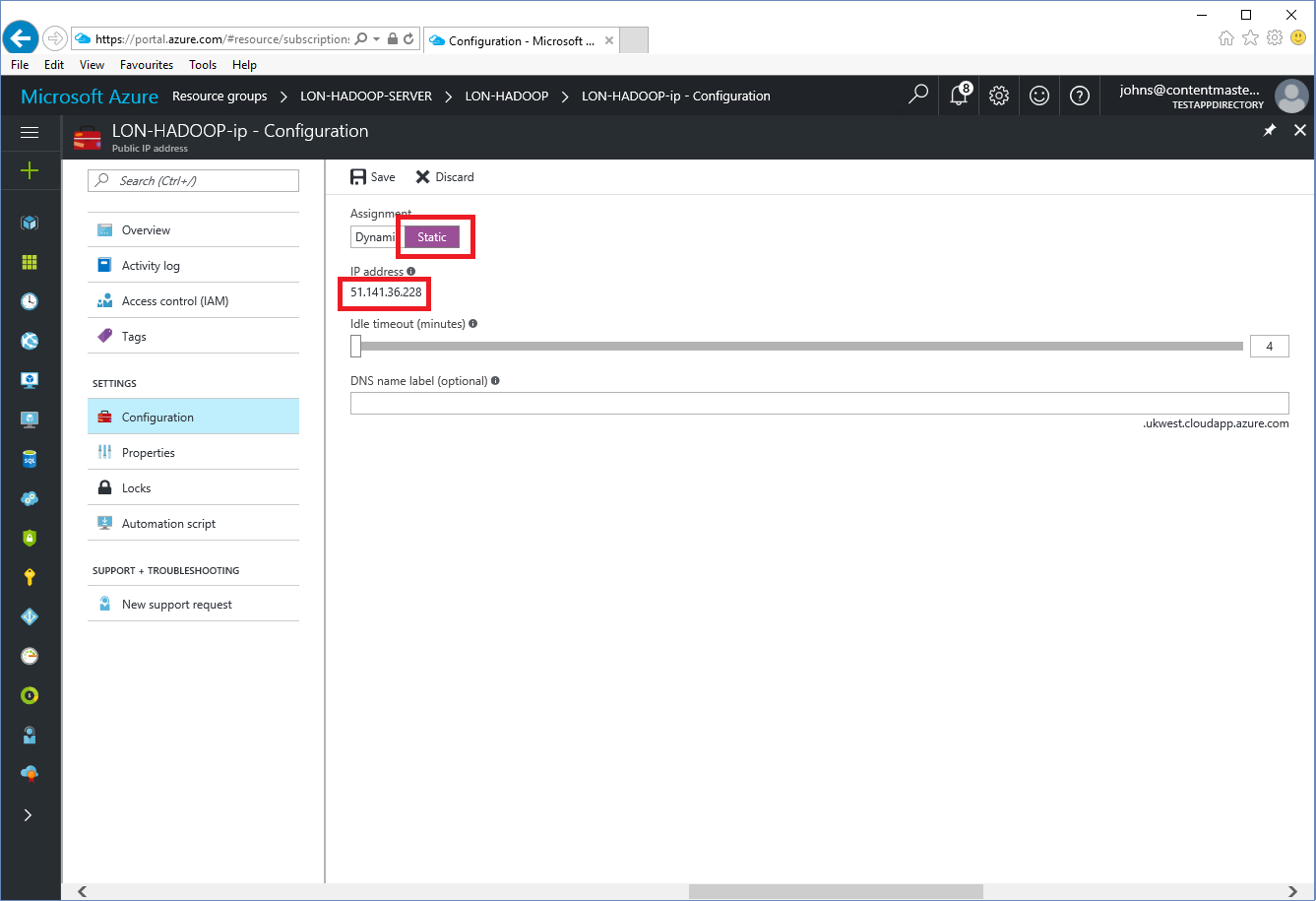
1. Wait while the VM is created and deployed.
2. In the navigation blade on the left side of the portal, click **Resource groups**, and then click the **LON-HADOOP-SERVER** resource group.
3. In the **LON-HADOOP-SERVER** blade, click the **LON-HADOOP** virtual machine.



1. In the **LON-HADOOP** blade, click the public IP address of the VM.

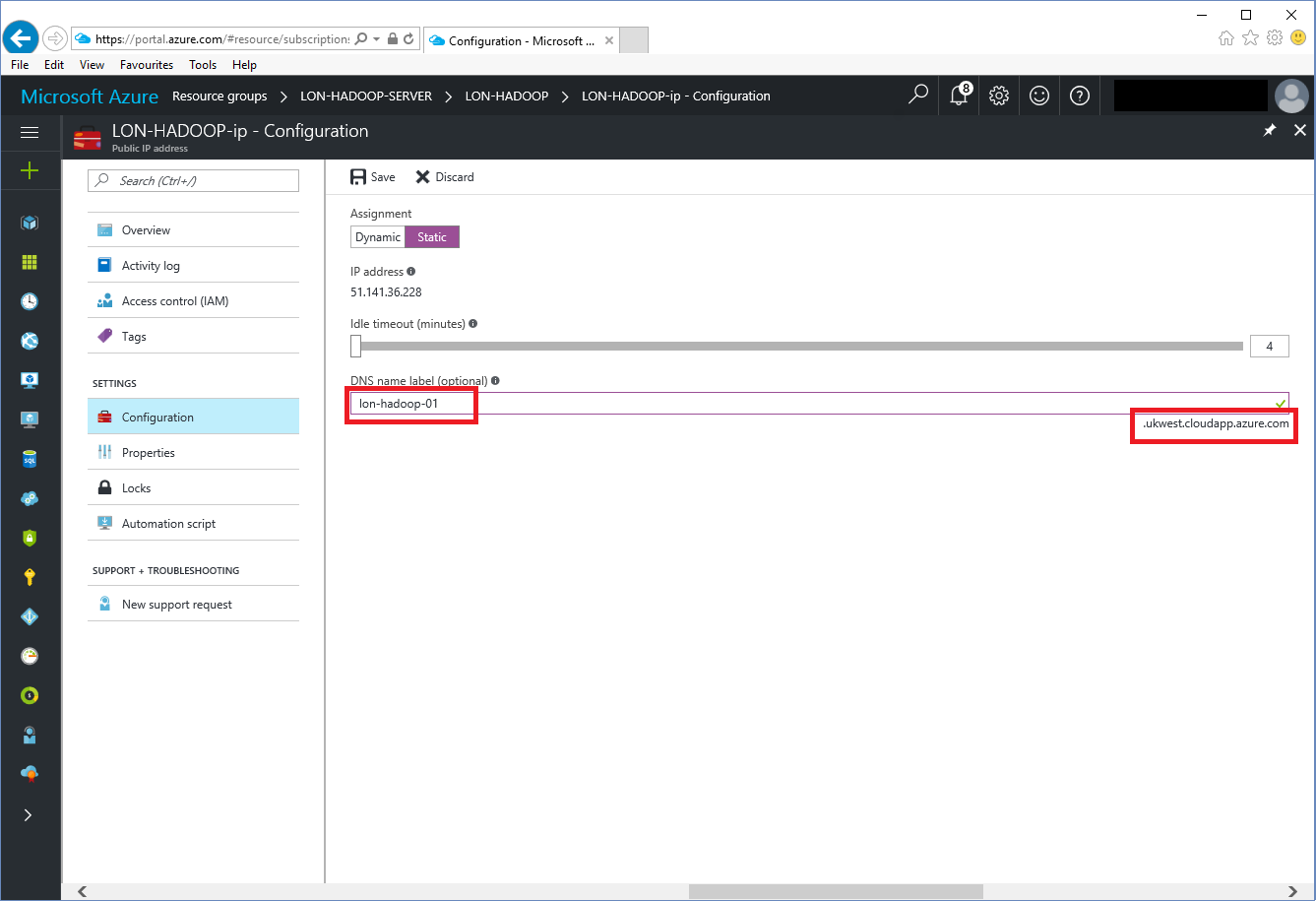


1. In the **LON-HADOOP-ip** - **Configuration** blade, under **Assignment**, click **Static**, and record the IP address allocated to the VM in the table at the end of this procedure.

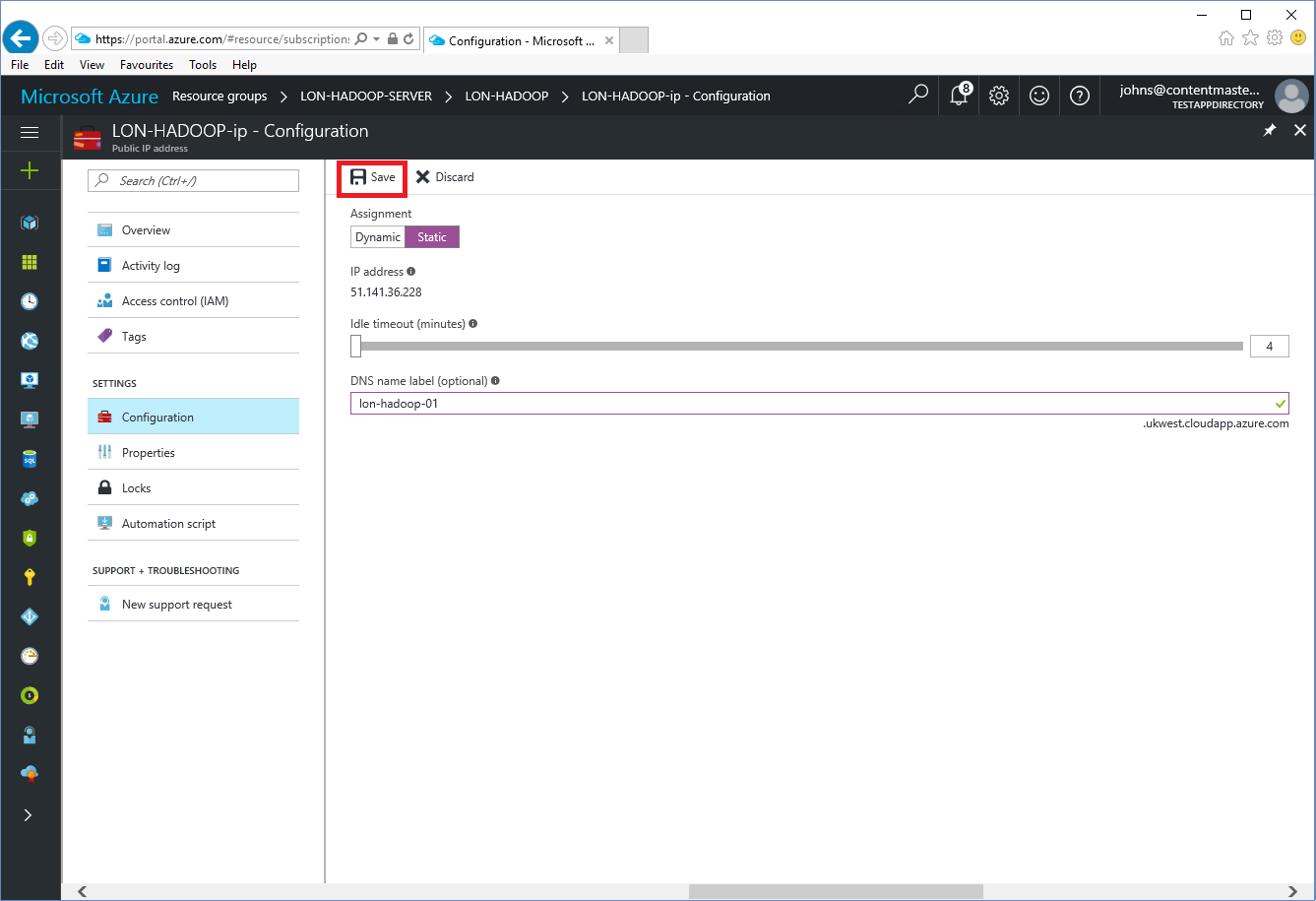


1. In the **DNS name label** box, type **lon-hadoop-*nn***, where ***nn***is a unique suffix for the VM. There might be other VMs in the same DNS domain with the name lon-hadoop, and this suffix is intended to avoid clashes. Pick your own suffix.

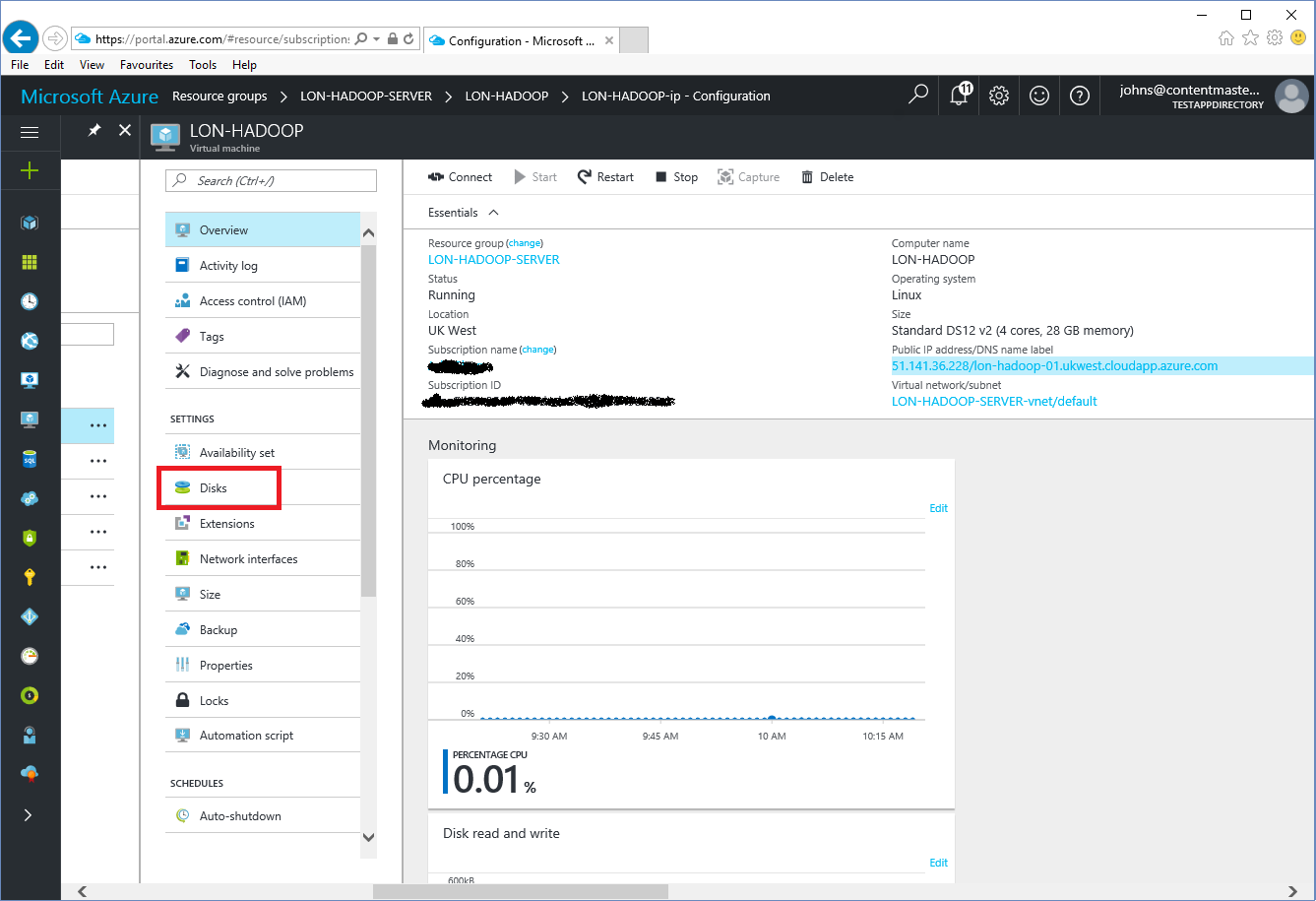
Record the full DNS name for the VM (for example, lon-hadoop-01.ukwest.cloudapp.azure.com) in the table at the end of this procedure.



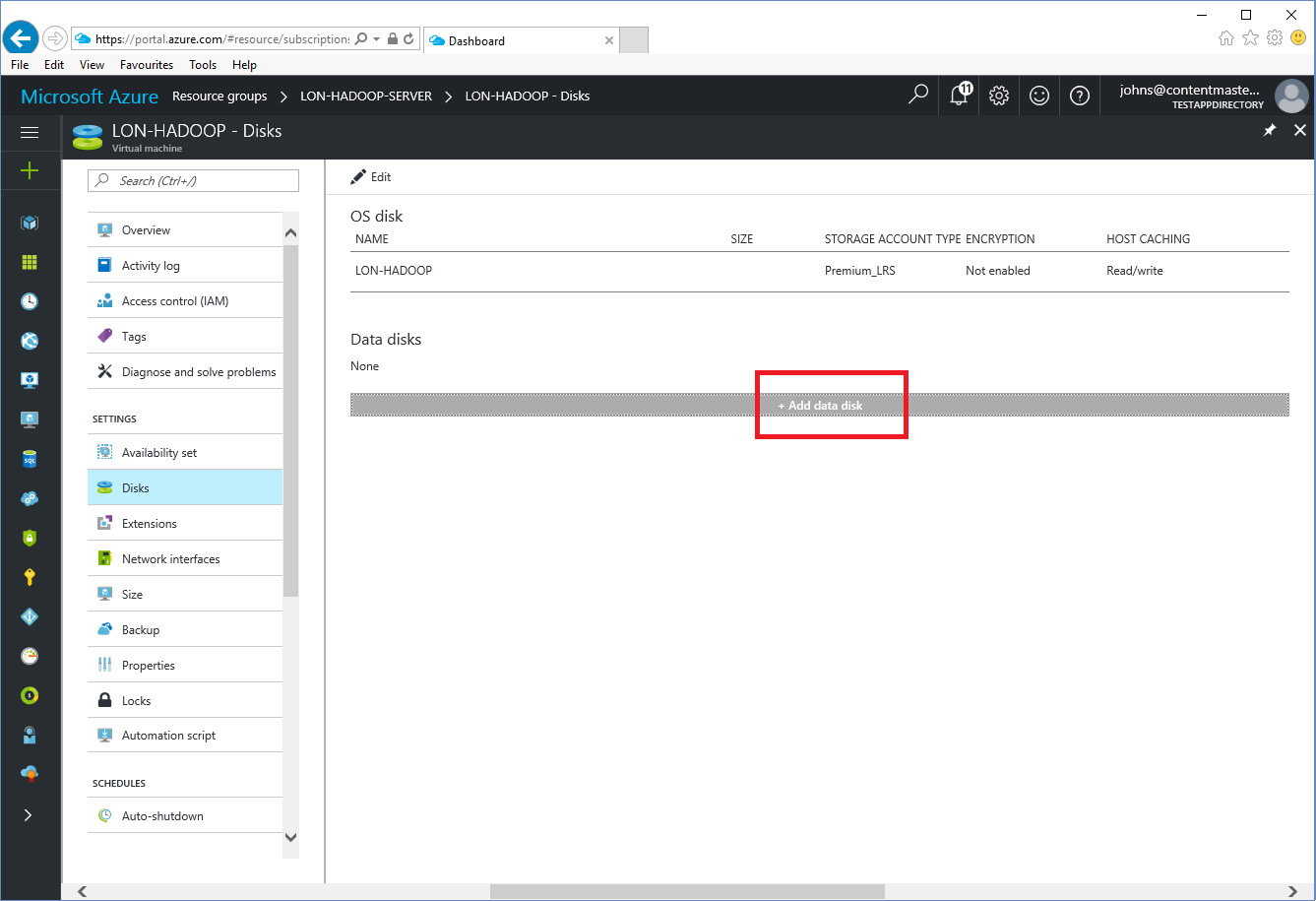
1. Click **Save**, wait for the configuration changes to be saved before continuing, and then close the **LON-HADOOP-ip - Configuration** blade.



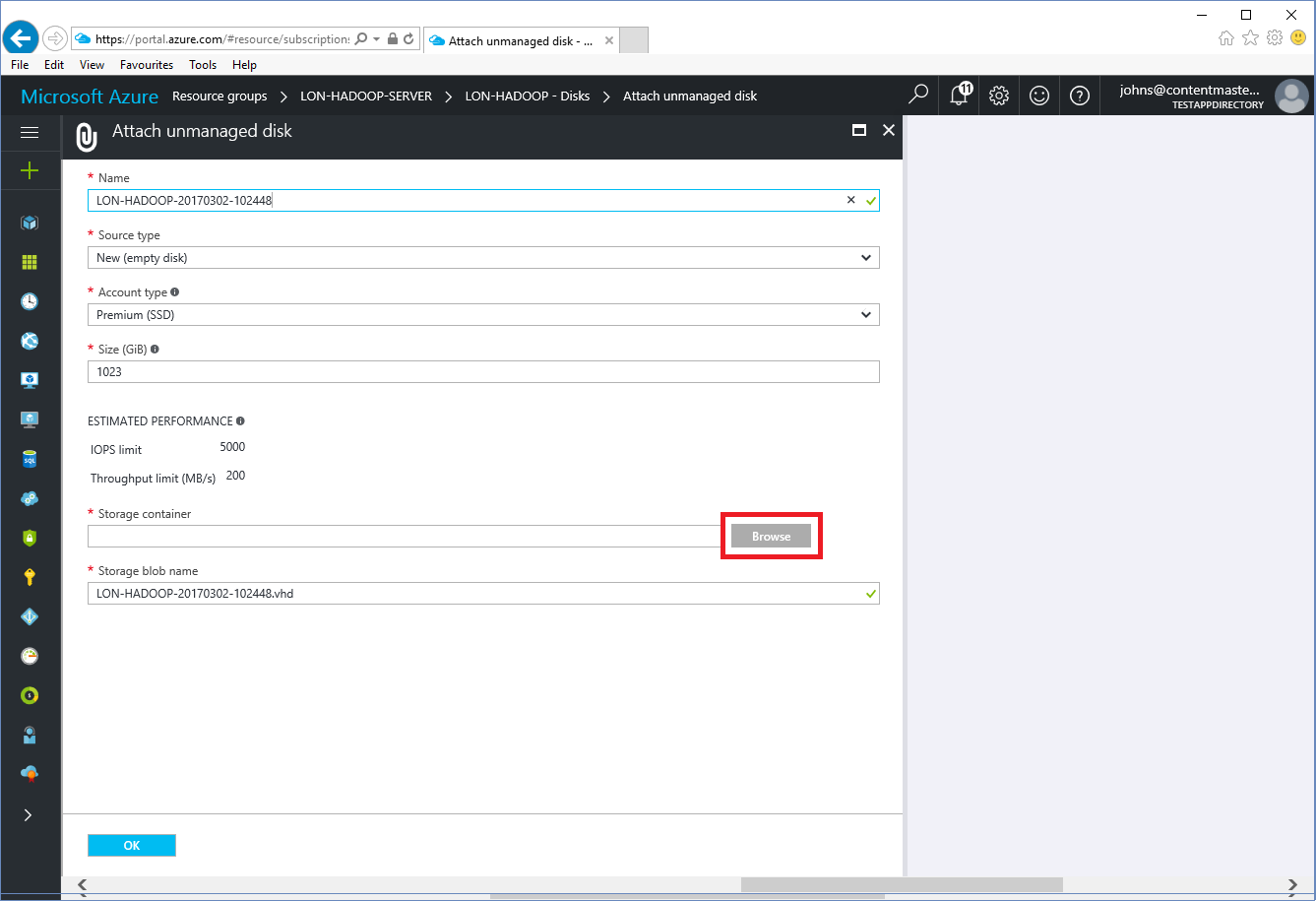
1. In the **LON-HADOOP** blade, under **Settings**, click **Disks**.



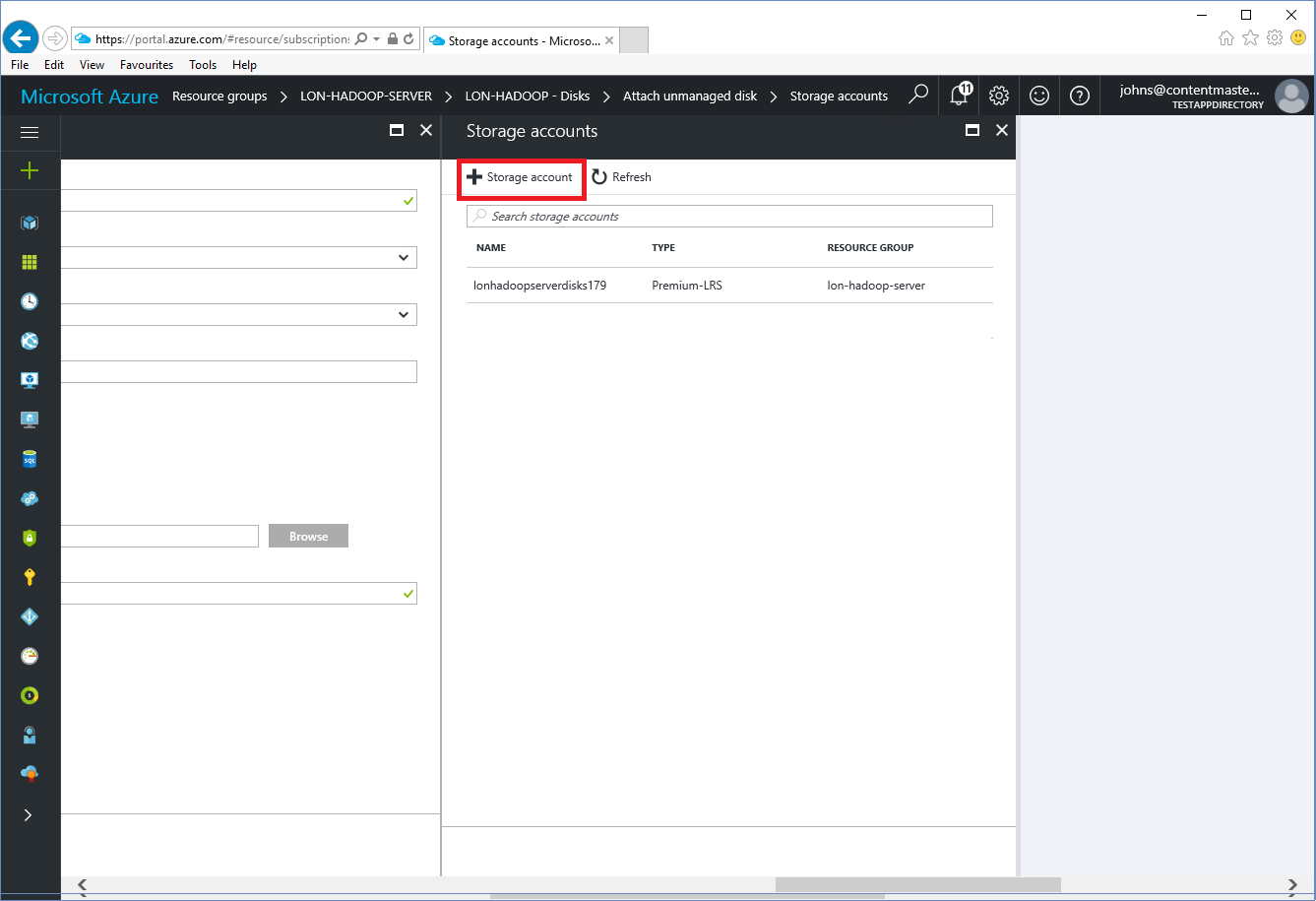
1. In the **LON-HADOOP - Disks** blade, click **Add data disk**.



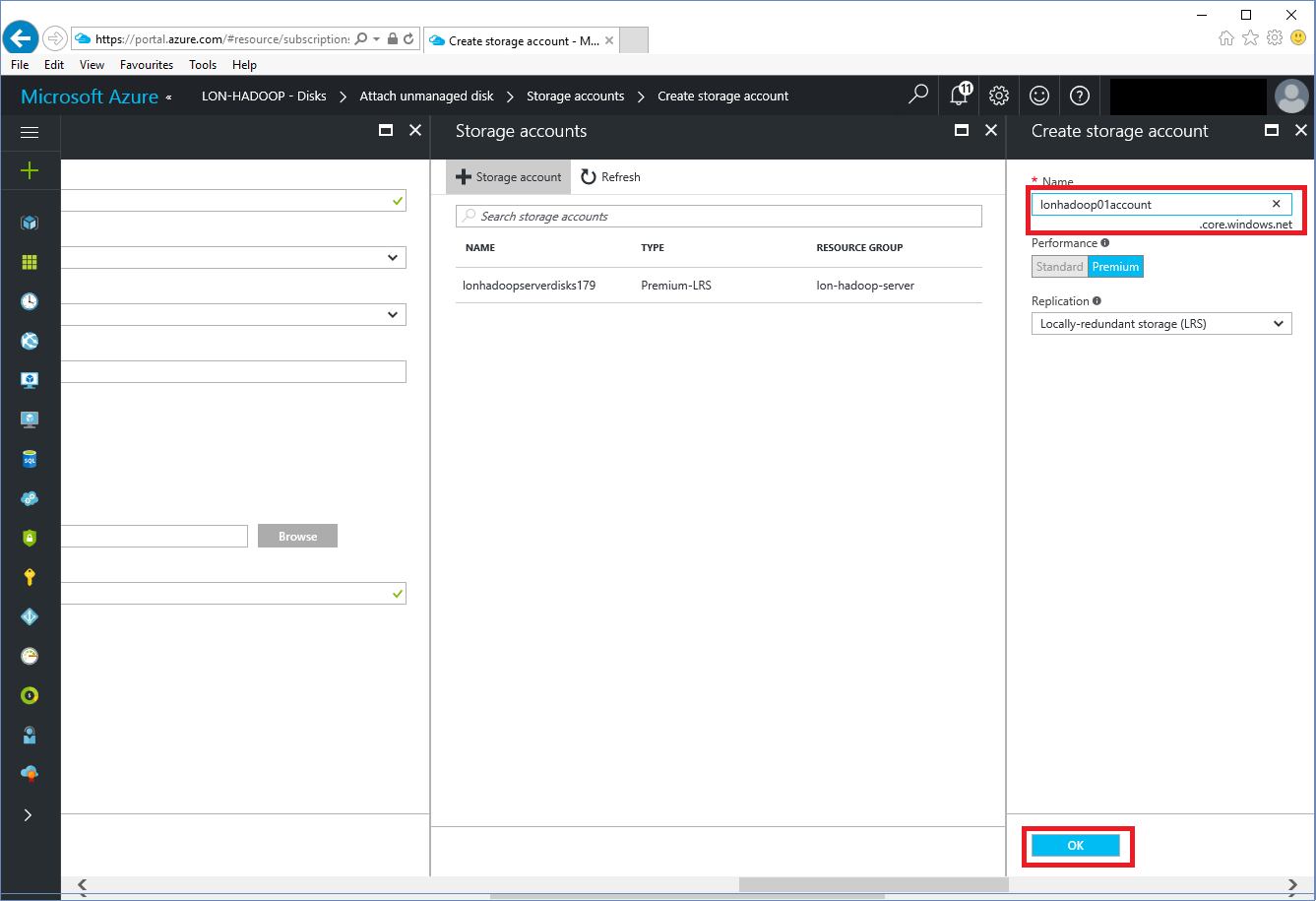
1. In the **Attach unmanaged disk** blade, next to the **Storage** container box, click **Browse**.



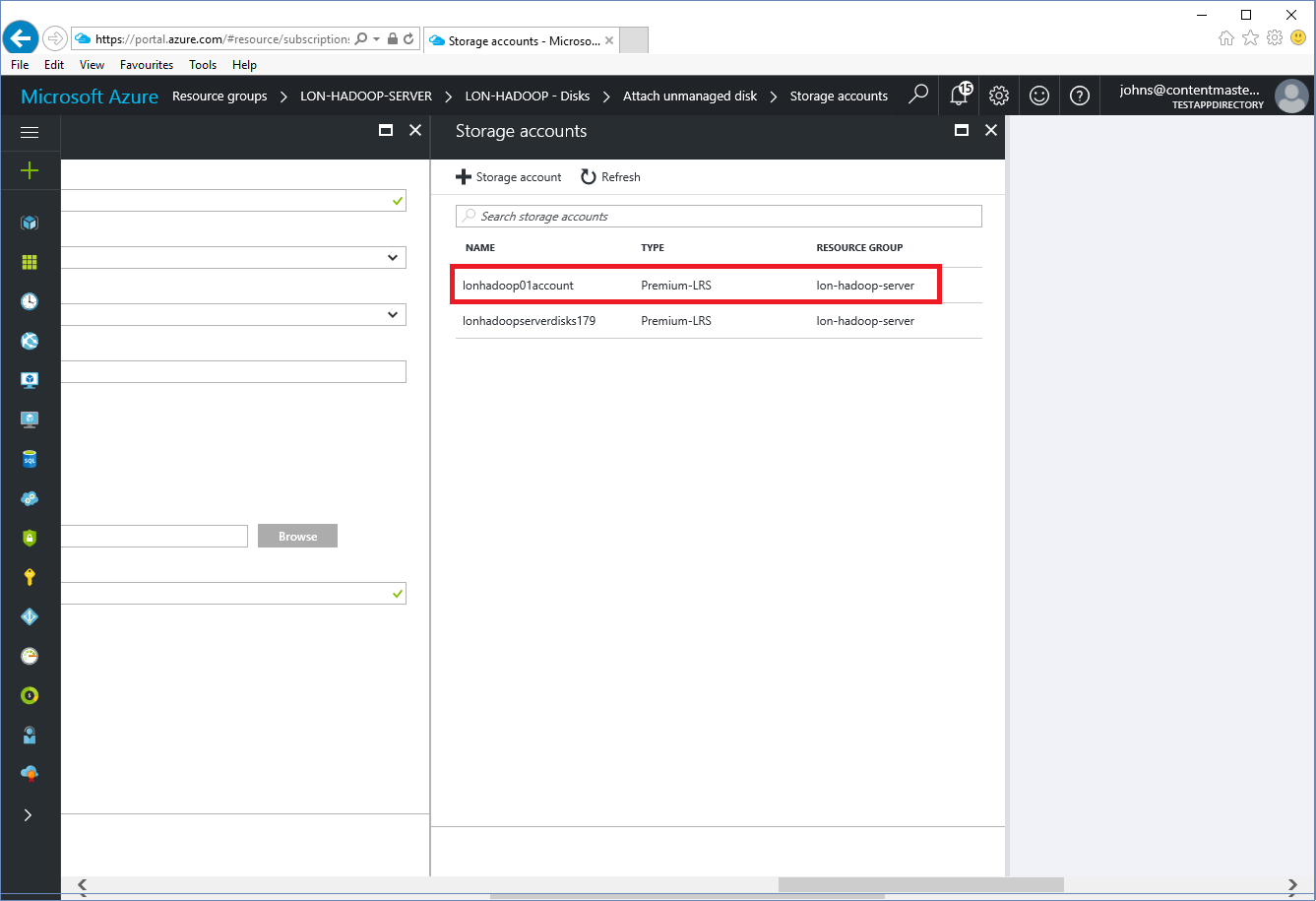
1. In the **Storage accounts** blade, click **+ Storage account**.



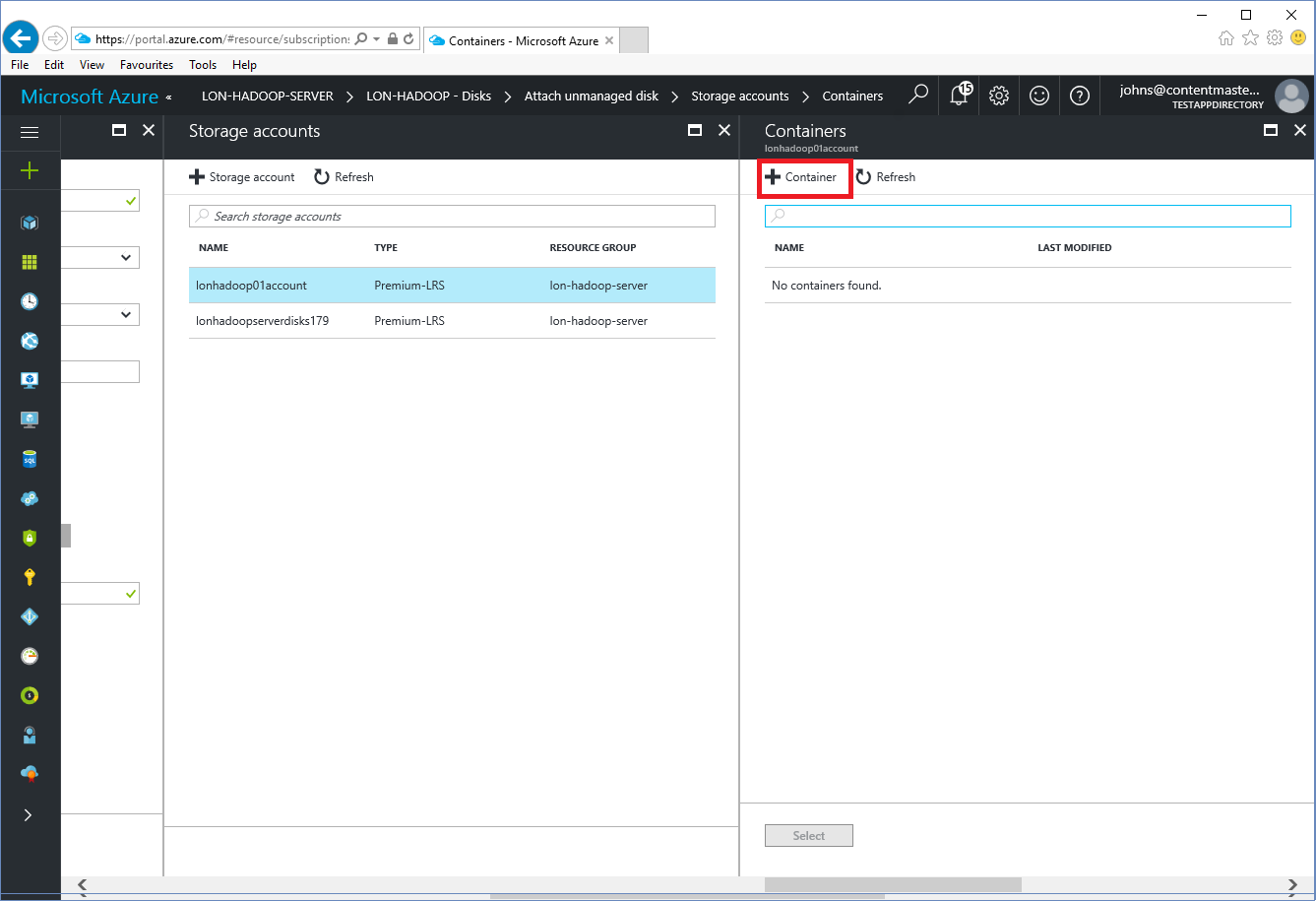
1. In the **Create storage account** blade, in the **Name** box, type **lonhadoop*nn*account** (without hyphens), and then click **OK**.



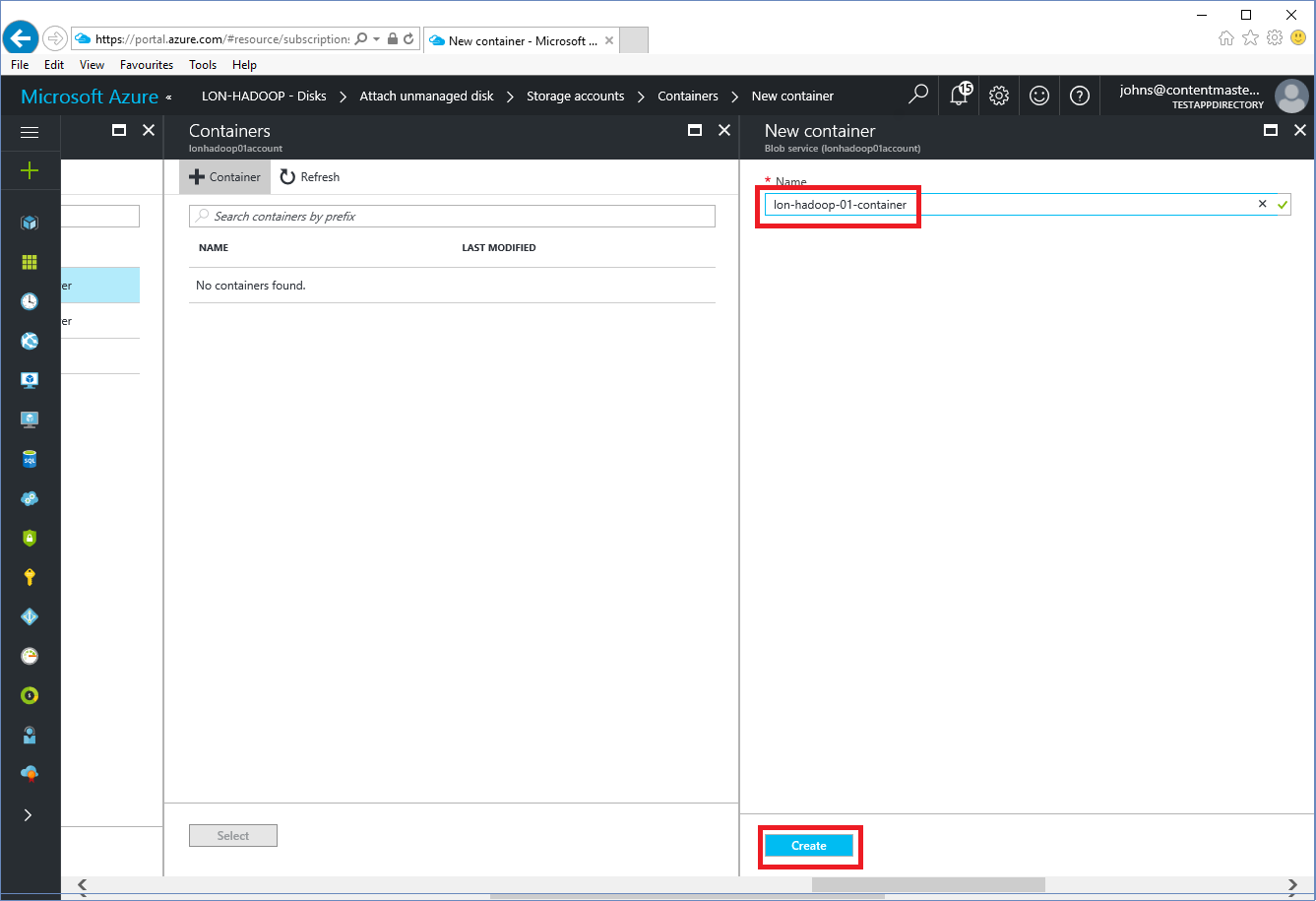
1. In the **Storage accounts** blade, click the **lonhadoop*nn*account**



1. In the **Containers** blade, click **+ Container**.



1. In the **New container** blade, in the **Name** box, type **lon-hadoop-*nn*-container**, and then click **Create**.



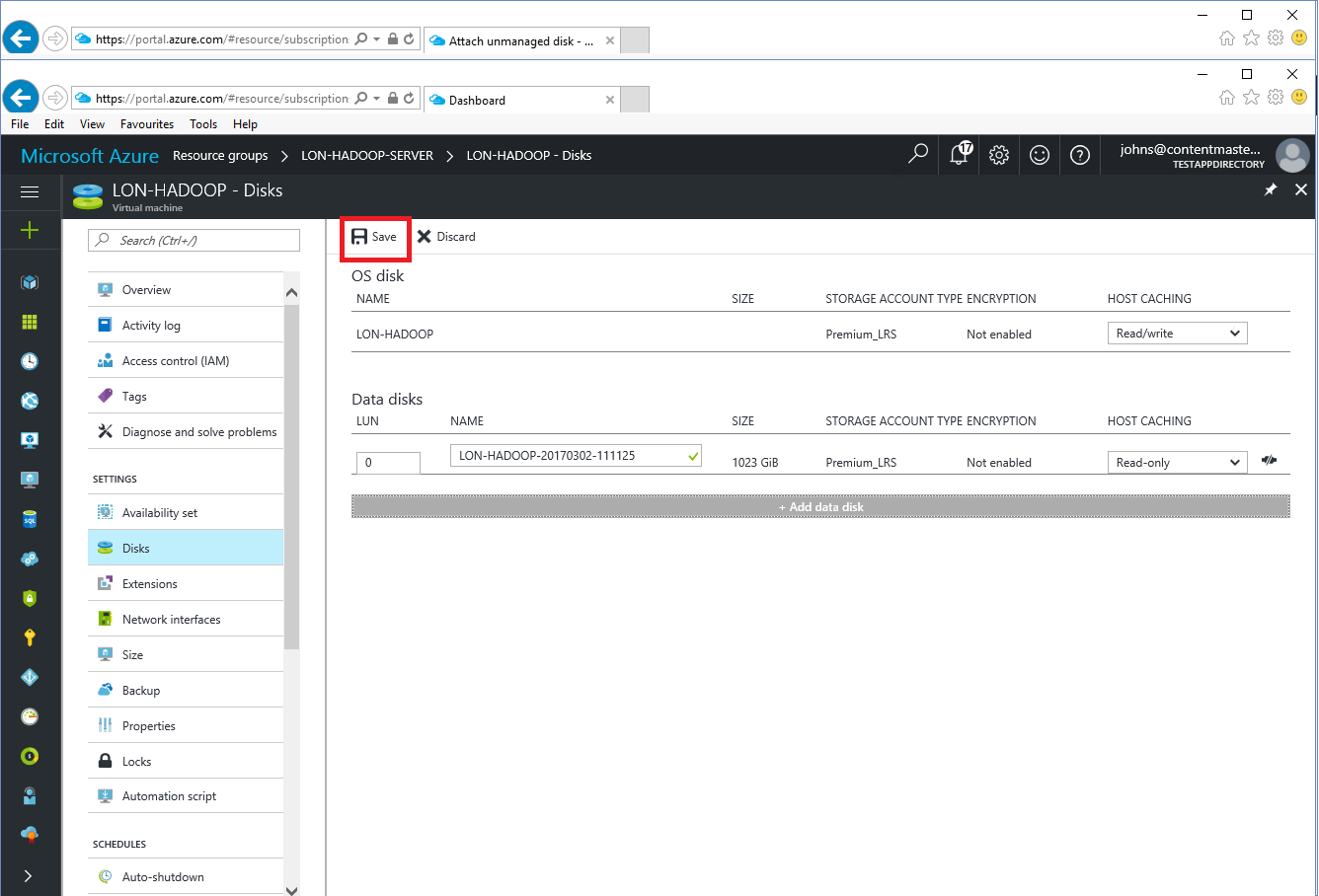
1. In the **Containers** blade, click **lon-hadoop-*nn*-container**, and then click **Select**.



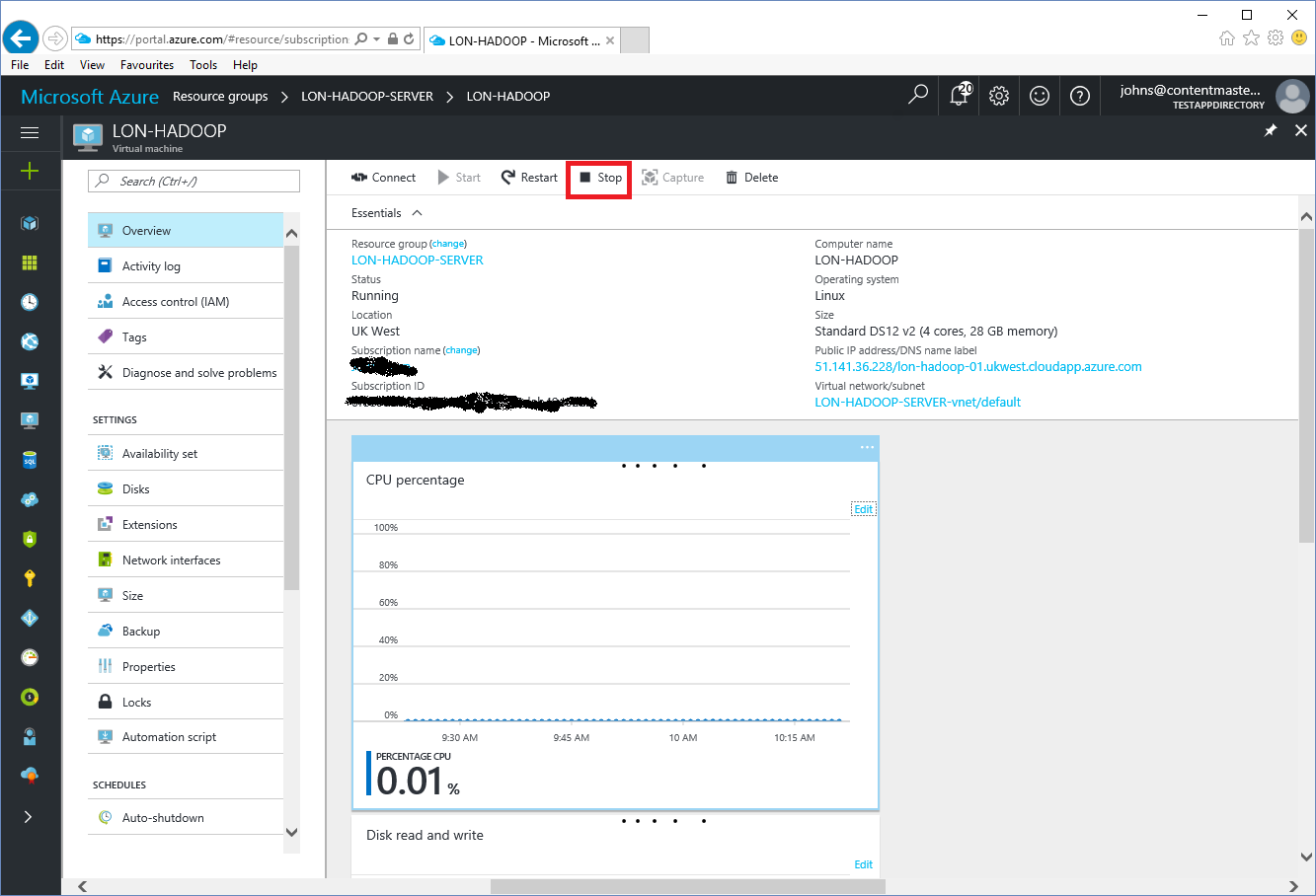
1. In the **Attach unmanaged disk** blade, click **OK**.



1. In the **LON-HADOOP - Disks** blade, click **Save**, and wait for the disk to be added to the virtual machine.



1. In the navigation blade on the left side of the portal, click **Resource groups**, and then click the **LON-HADOOP-SERVER** resource group.
2. In the **LON-HADOOP-SERVER** blade, click the **LON-HADOOP** virtual machine.
3. In the **LON-HADOOP** blade, click **Stop**.



Use the table below to record the public IP address and DNS name of the VM.

|  |  |  |
| --- | --- | --- |
| VM Name | Public IP Address | Public DNS Name  *For example:*  *lon-hadoop-01.ukwest.cloudapp.azure.com* |
| LON-HADOOP |  |  |

# Configure the LON-HADOOPVirtual Machine

You need to configure the VM before installing Hadoop. Specifically, you need to:

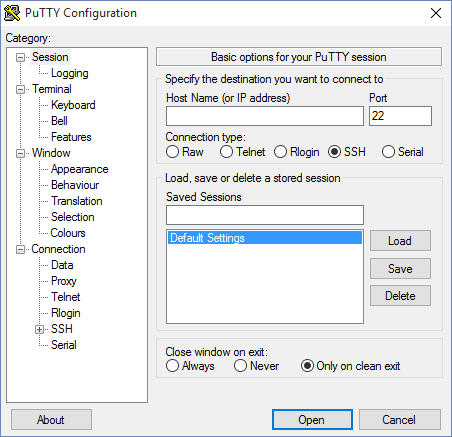
* Set a password for the root user.
* Configure password-less authentication (Hadoop requires this).
* Install the NTP service (for coordinating time across the network).
* Set the hostname configuration to include the fully qualified DNS name of the VM.
* Temporarily disable the Linux firewall.
* Configure and mount the disk for holding the HDFS filesystem.

Perform the following process:

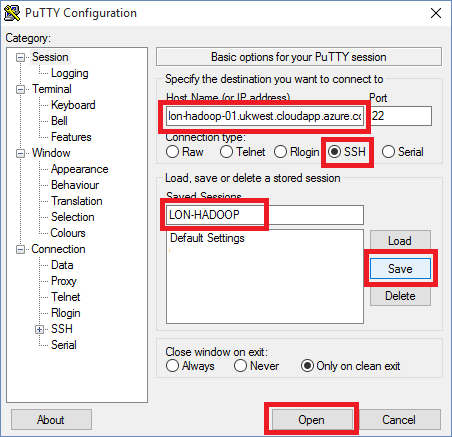
1. In the Azure portal, in the navigation blade on the left side of the portal, click **Resource groups**.
2. Click the **LON-HADOOP-SERVER** resource group.
3. In the **LON-HADOOP-SERVER** blade, click the **LON-HADOOP** virtual machine.
4. In the **LON-HADOOP** blade, click **Start**, and then wait for the VM to start running.



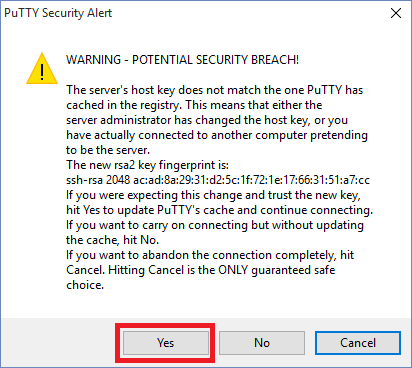
1. On the desktop machine, open a command prompt.
2. In the command prompt window, run the **putty** command. The putty utility should start and the **PuTTY Configuration** window should appear.



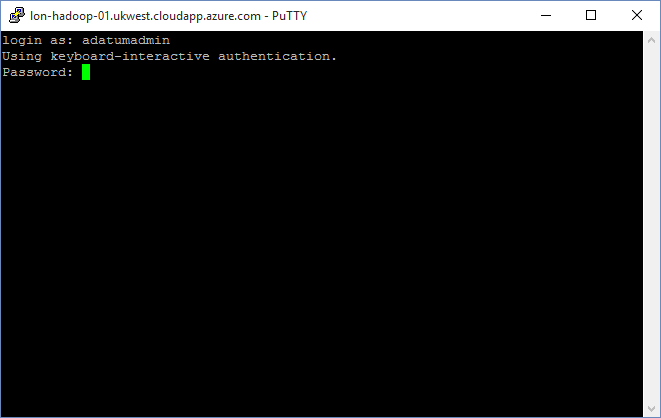
1. In the **Host Name (or IP address)**  box, type either the full public DNS name or the public IP address of the **LON-HADOOP** virtual machine, as listed in the table at the end of the previous procedure. Make sure the connection type is set to **SSH**. In the **Saved Sessions** box, type **LON-HADOOP**, click **Save**, and then click **Open**.



1. In the **PuTTY Security Alert** dialog box, click **Yes**.



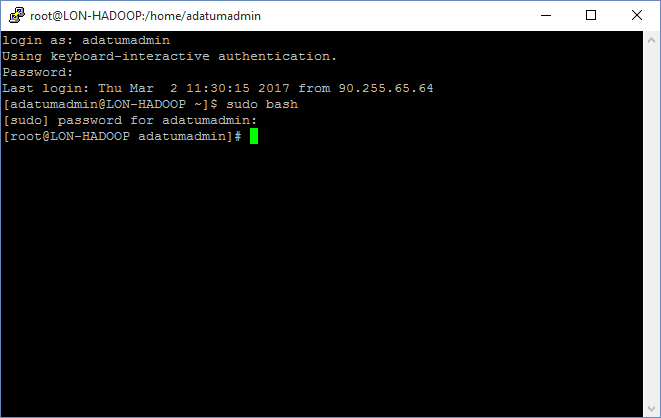
1. In the PuTTY terminal window that appears, at the **login as** prompt, log in as **adatumadmin** with password **Pa55w.rdPa55w.rd**. Note that Linux user names are case sensitive.



1. In the PuTTY terminal window, run the following command:

sudo bash

1. At the **password for adatumadmin** prompt, type **Pa55w.rdPa55w.rd**. You should now be running as the **root** user.



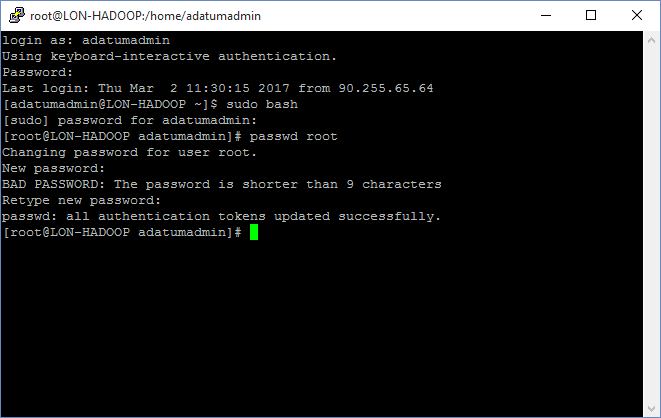
1. In the PuTTY terminal window, run the following command to set the root pasword:

passwd root

1. At the **New password** prompt, type **Pa55w.rd**.

Note that you might receive a warning about a bad password. Ignore this warning.

1. At the **Retype new password** prompt, type **Pa55w.rd**



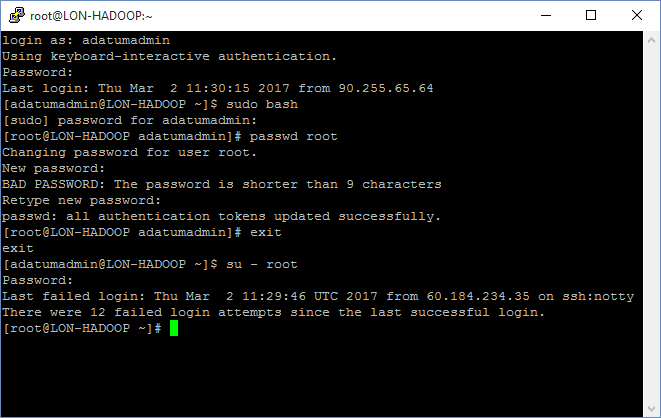
1. In the PuTTY terminal window, run the following command:

exit

1. In the PuTTY terminal window, run the following command to log in as root using the new password:

su - root

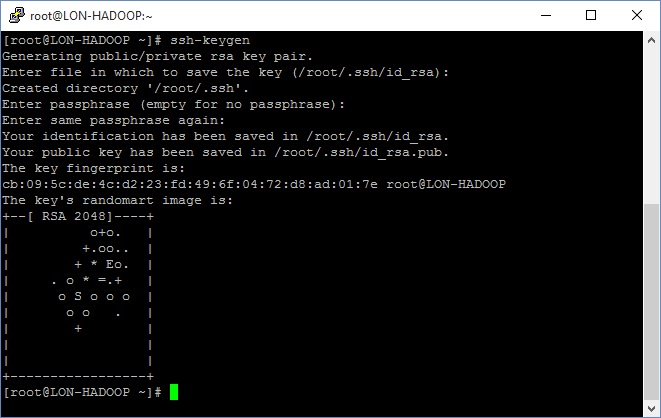
1. At the **Password** prompt, enter **Pa55w.rd**.



1. In the PuTTY terminal window, run the following command to create SSH keys for performing password-less authentication:

ssh-keygen

1. At the prompt **Enter file in which to save the key (/root/.ssh/id\_rsa)**, press Enter.
2. At the prompt **Enter passphrase (empty for no passphrase)**, press Enter.
3. At the prompt **Enter same passphrase again**, press Enter.



1. In the PuTTY terminal window, run the following command:

cat .ssh/id\_rsa.pub >> .ssh/authorized\_keys

1. In the PuTTY terminal window, run the following commands:

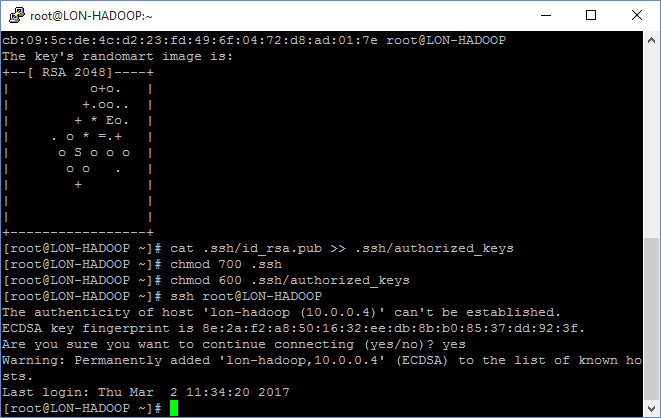
chmod 700 .ssh

chmod 600 .ssh/authorized\_keys

1. In the PuTTY terminal window, run the following command to test password-less authentication:

ssh root@LON-HADOOP

1. At the prompt **Are you sure you want to continue connecting (yes/no)**, enter **yes**.

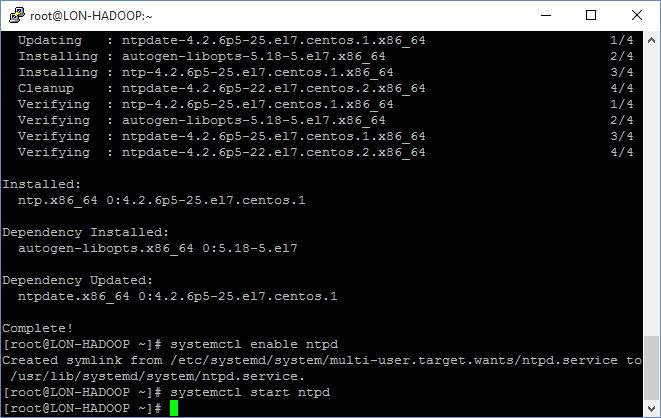


1. In the PuTTY terminal window, run the following commands to install the NTP (time) service:

yum install -y ntp

systemctl enable ntpd

systemctl start ntpd



If the **Is this ok** prompt appears, press **y**, and then press Enter.

1. In the PuTTY terminal window, run the following command to open the hosts configuration file:

nano /etc/hosts

1. In the editor, change the first line to:

127.0.0.1 localhost *fqdn*

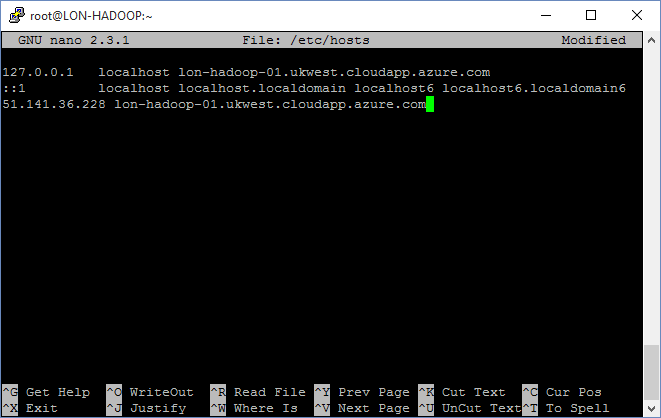
where ***fqdn*** is the fully qualified domain name of the VM (such as lon-hadoop-01.ukwest.cloudapp.azure.com). You recorded this information in the table at the end of the previous procedure.

1. Add the following line to the end of the file. Replace ***aa.bb.cc.dd*** with the public IP address of the VM, and replace ***fqdn*** with the fully qualified domain name of the VM:

*aa.bb.cc.dd fqdn*

For example:

51.140.225.246 lon-hadoop-01.ukwest.cloudapp.azure.com

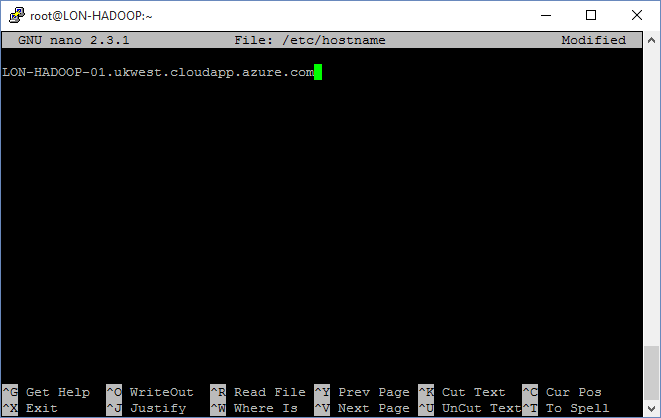


1. Press **Ctrl+O** to save the file. Save it as **/etc/hosts** when prompted, and then press Enter.
2. Press **Ctrl+X** to close the editor.
3. In the PuTTY terminal window, run the following command to open the hostname configuration file:

nano /etc/hostname

1. In the editor, append the domain name suffix to the VM name to specify the fully qualified DNS name of the VM. For example:

LON-HADOOP-01.ukwest.cloudapp.azure.com

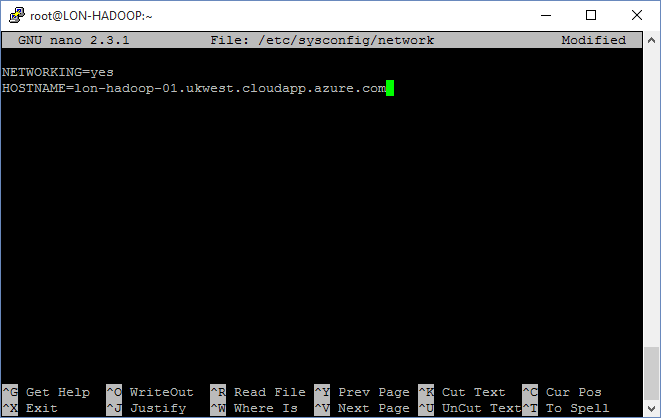


1. Press **Ctrl+O** to save the file. Save it as **/etc/hostname** when prompted, and then press Enter.
2. Press **Ctrl+X** to close the editor.
3. In the PuTTY terminal window, run the following command to open the network configuration file:

nano /etc/sysconfig/network

1. In the editor, change the **HOSTNAME** variable and specify the fully qualified DNS name of the VM. For example:

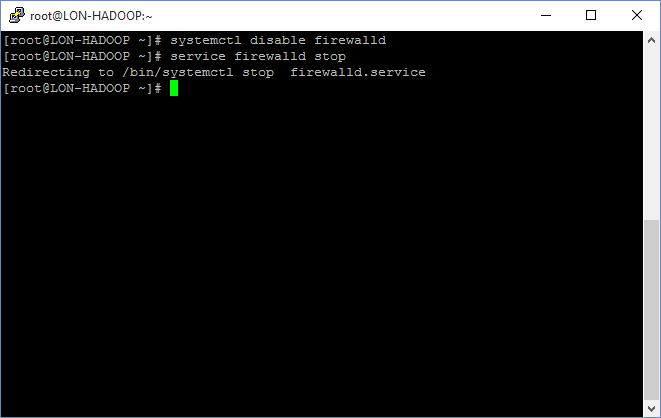
HOSTNAME=lon-hadoop-01.ukwest.cloudapp.azure.com



1. Press **Ctrl+O** to save the file. Save it as **/etc/sysconfig/network** when prompted, and then press Enter.
2. Press **Ctrl+X** to close the editor.
3. In the PuTTY terminal window, run the following commands to disable the Linux firewall:

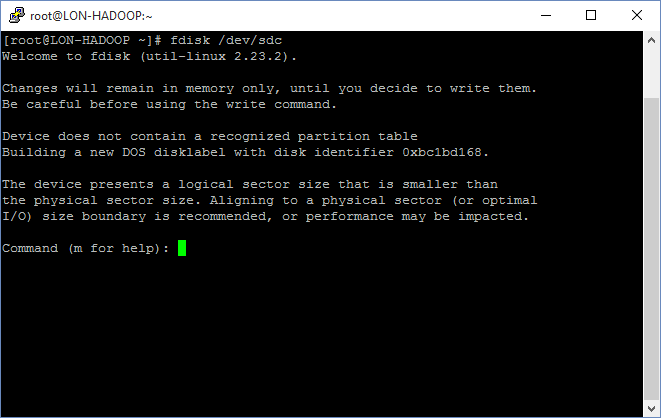
systemctl disable firewalld

service firewalld stop

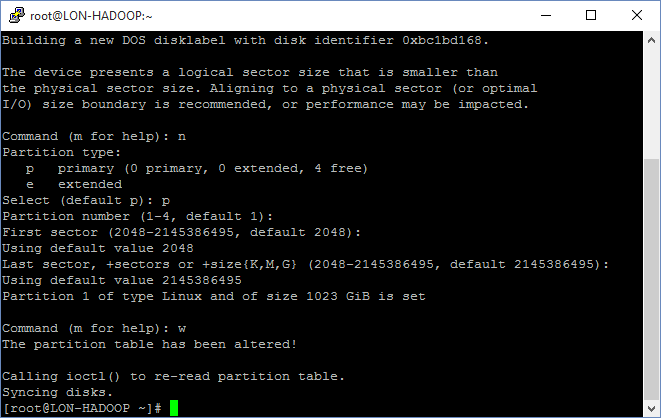


1. In the PuTTY terminal window, run the following command to create a new partition for holding the HDFS file system on the /dev/sdc device:

fdisk /dev/sdc

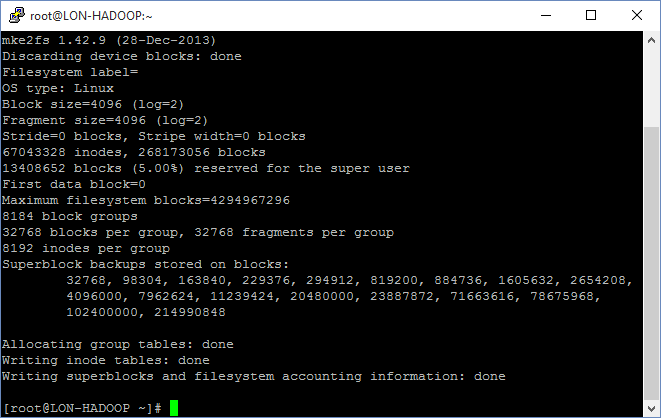


1. At the **Command** prompt in fdisk, type **n**, and then press Enter. This command creates a new partition on the device.
2. At the **Partition type**, type **p** and then press Enter.
3. At the **Partition number**, press Enter.
4. At the **First sector** prompt, press Enter.
5. At the **Last sector** prompt, press Enter.
6. At the **Command** prompt in fdisk, type **w**, and then press Enter. This command saves the new disk configuration and exits fdisk.



1. In the PuTTY terminal window, run the following command to create a file system on the new partition:

mkfs /dev/sdc1



1. In the PuTTY terminal window, run the following command to create a directory for mounting the HDFS disk:

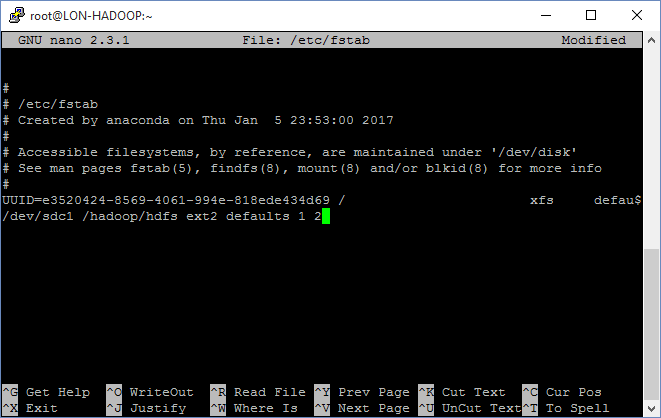
mkdir -p /hadoop/hdfs

1. In the PuTTY terminal window, run the following command to open the mount table configuration file:

nano /etc/fstab

1. In the editor, add the following line to the end of the file:

/dev/sdc1 /hadoop/hdfs ext2 defaults 1 2



1. Press **Ctrl+O** to save the file. Save it as **/etc/fstab** when prompted, and then press Enter.
2. Press **Ctrl+X** to close the editor.
3. In the PuTTY terminal window, run the following commands to mount the /dev/sdc1 disk on the /hadoop/hdfs directory and make it accessible by Hadoop:

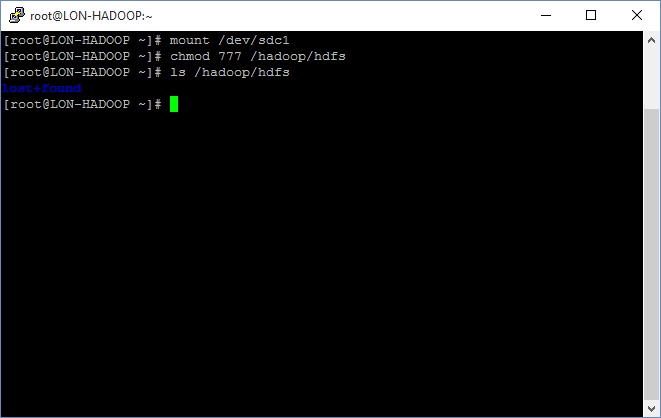
mount /dev/sdc1

chmod 777 /hadoop/hdfs

1. In the PuTTY terminal window, run the following command to list the contents of the /hadoop/hdfs directory:

ls /hadoop/hdfs

The /hadoop/hdfs directory should contain a single directory named lost+found.



1. In the PuTTY terminal window, run the following command to create an account for the course instructor:

useradd instructor

1. In the PuTTY terminal window, unlock the instructor account:

passwd instructor

1. At the **New password** prompt, enter **Pa55w.rd** (ignore the BAD PASSWORD warning)
2. At the **Retype new password** prompt, enter **Pa55w.rd**
3. In the PuTTY terminal window, create an account for each student. Name the student accounts student01, student02, etc (you might have more or fewer than 15 students; create accounts as necessary):

useradd student01

useradd student02

useradd student03

...

useradd student15

1. Unlock the student accounts using the **passwd** command, as described above for the instructor account.
2. In the PuTTY terminal window, run the following:

nano /etc/ssh/sshd\_config

1. In the editor, add the following line to the end of the file (enter this text on a single line):

AllowUsers root adatumadmin instructor student01 student02 student03 student04 student05 student06 student07 student08 student09 student10 student11 student12 student13 student14 student15

1. Press **Ctrl+O** to save the file. Save it as **/etc/ssh/sshd\_config** when prompted, and then press Enter.
2. Press **Ctrl+X** to close the editor.
3. In the PuTTY terminal window, run the following command to restart the sshd daemon:

service sshd restart

1. Close the PuTTY terminal window. If prompted, click **OK** to confirm that you wish to exit the session.

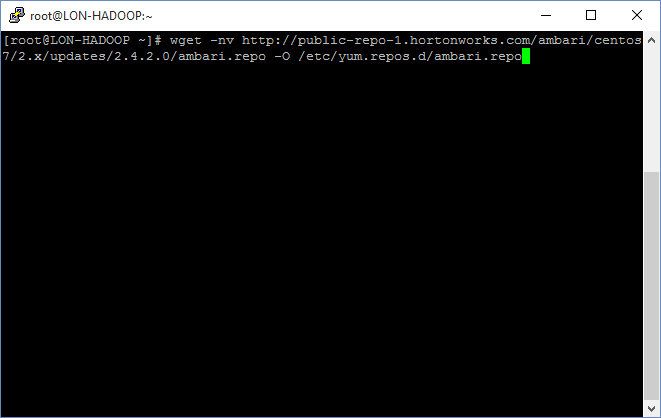
# Install Apache Ambari on the LON-HADOOP Virtual Machine

The simplest way to install and configure Hadoop is to use the Apache Ambari configuration and monitoring utility.

1. On the desktop machine, open a command prompt if one is not already available.
2. In the command prompt window, run the **putty** command.
3. In the **PuTTY Configuration** window, in the list of saved sessions, click **LON-HADOOP**, click **Load**, and then click **Open**.
4. In the PuTTY terminal window that appears, at the **login as** prompt, log in as **root** with password **Pa55w.rd**. Note that password-less authentication has not been enabled for the desktop computer, so you still need to specify a password when logging in.
5. In the PuTTY terminal window, run the following command to download Ambari:

wget -nv http://public-repo-1.hortonworks.com/ambari/centos7/2.x/updates/2.4.2.0/ambari.repo -O /etc/yum.repos.d/ambari.repo

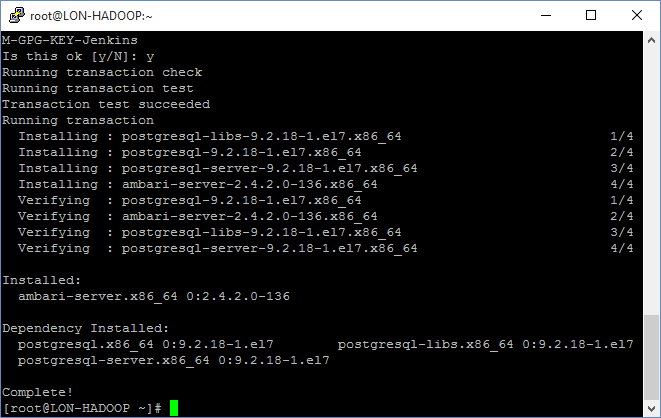
Note that this is a single command that you should enter on one line.



1. In the PuTTY terminal window, run the following command to install Ambari:

yum install ambari-server

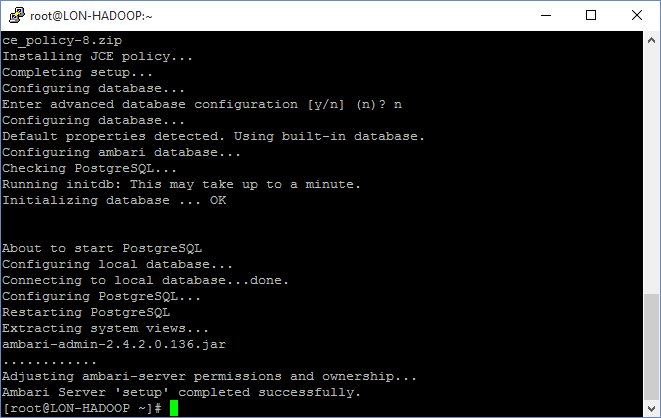
1. At each **Is this ok** prompt, press **y**, and then press Enter.



1. In the PuTTY terminal window, run the following command to set up the Ambari server:

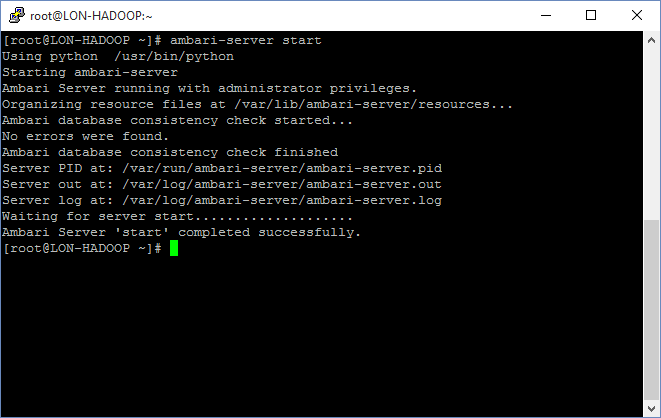
ambari-server setup

1. At the **OK to continue** prompt, press **y**, and then press Enter.
2. At the **Customize user account for ambari-server daemon** prompt, press **n**, and then press Enter.
3. At the **Checking JDK … Enter choice** prompt, press **1**, and then press Enter.
4. At the **Do you accept the Oracle Binary Code License Agreement** prompt, press **y**, and then press Enter.
5. At the **Enter advanced database configuration** prompt, press **n**, and then press Enter.



1. In the PuTTY terminal window, run the following command to start the Ambari server:

ambari-server start



1. Close the PuTTY terminal window. When prompted, click **OK** to confirm that you wish to exit the session.

# Install Hadoop on the LON-HADOOPVirtual Machine

The Ambari server runs as a web application on each VM. You install Hadoop by connecting to the Ambari server from the desktop machine using Internet Explorer.

1. On the desktop machine, open a command prompt window, and run the following commands to copy the private key used to connect to the VM to the C:\TEMP folder on desktop:

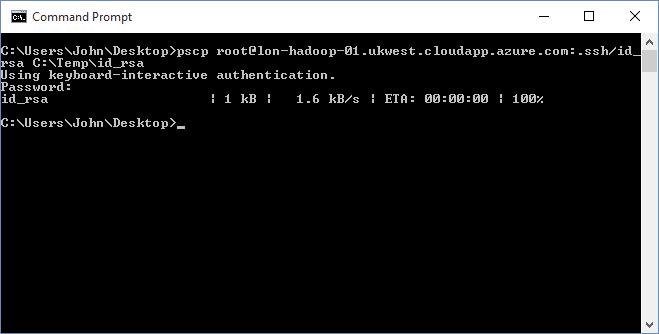
mkdir C:\Temp

pscp root@*fqdn*:.ssh/id\_rsa C:\Temp\id\_rsa

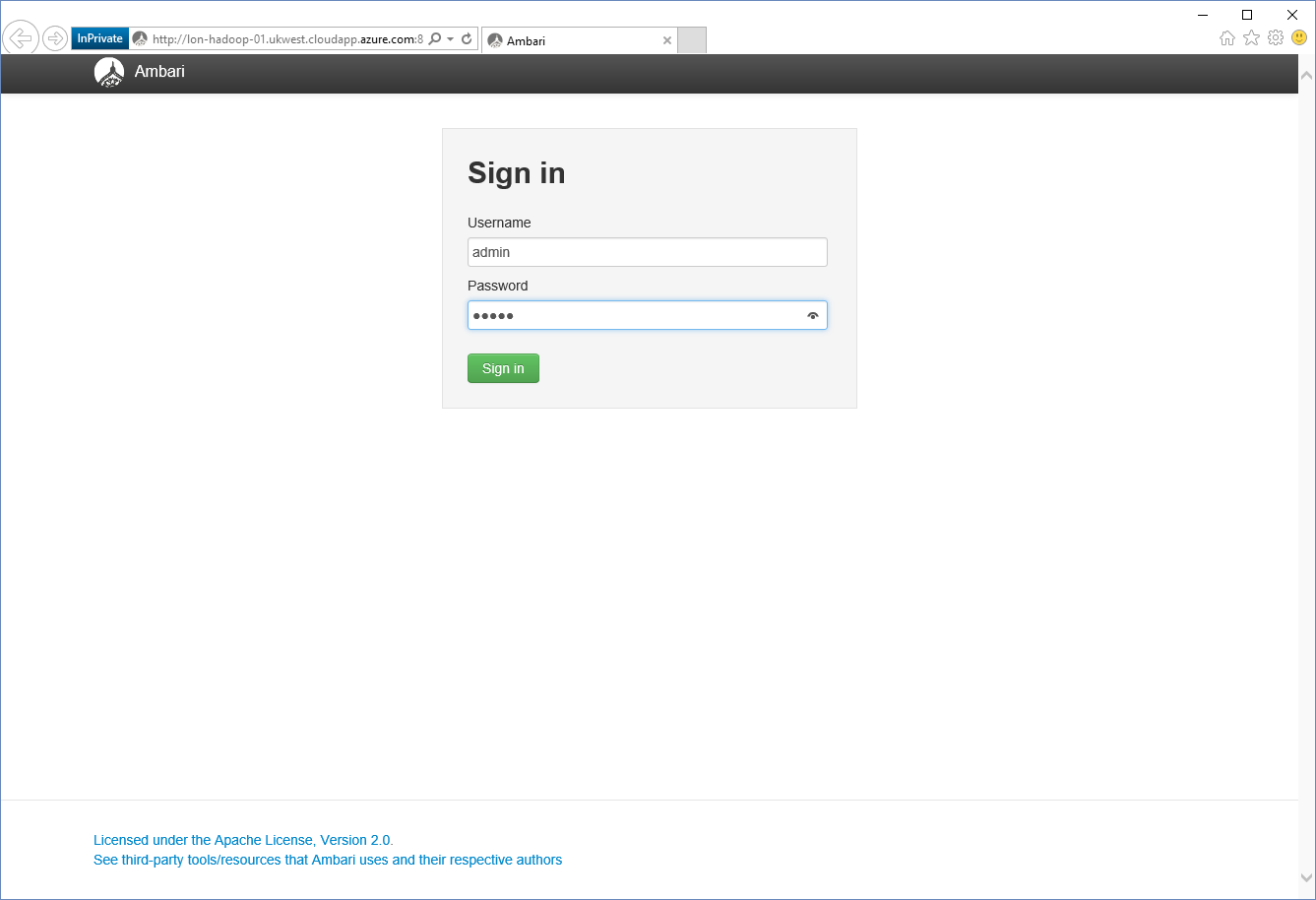
Replace ***fqdn*** with the fully qualified DNS name of the LON-HADOOP VM. For example:

pscp root@lon-hadoop-01.ukwest.cloudapp.azure.com:.ssh/id\_rsa C:\Temp\id\_rsa

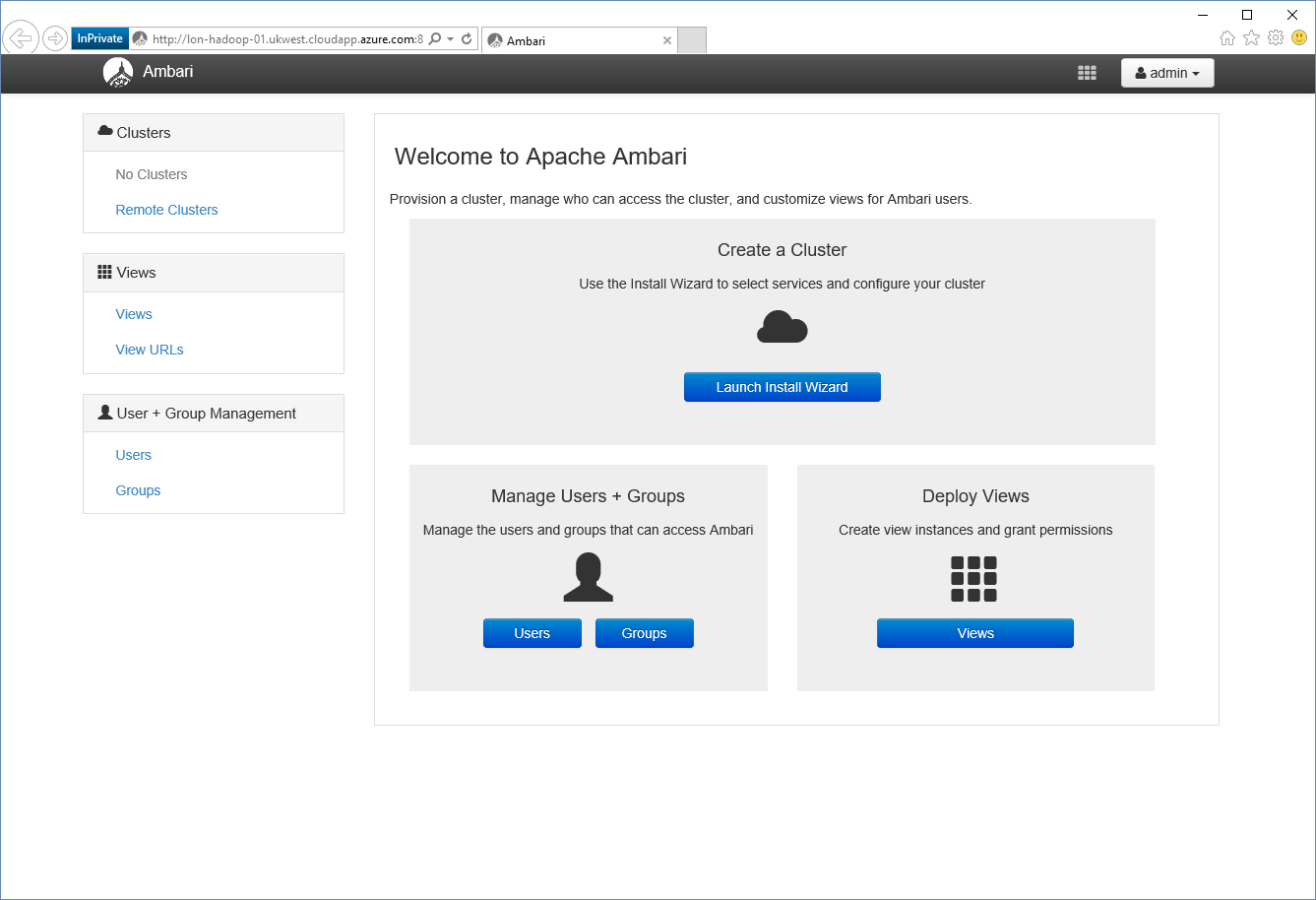
1. If the **Store key in cache** prompt appears, type **y**, and then press Enter.
2. At the **Password** prompt, enter **Pa55w.rd**, and then press Enter.



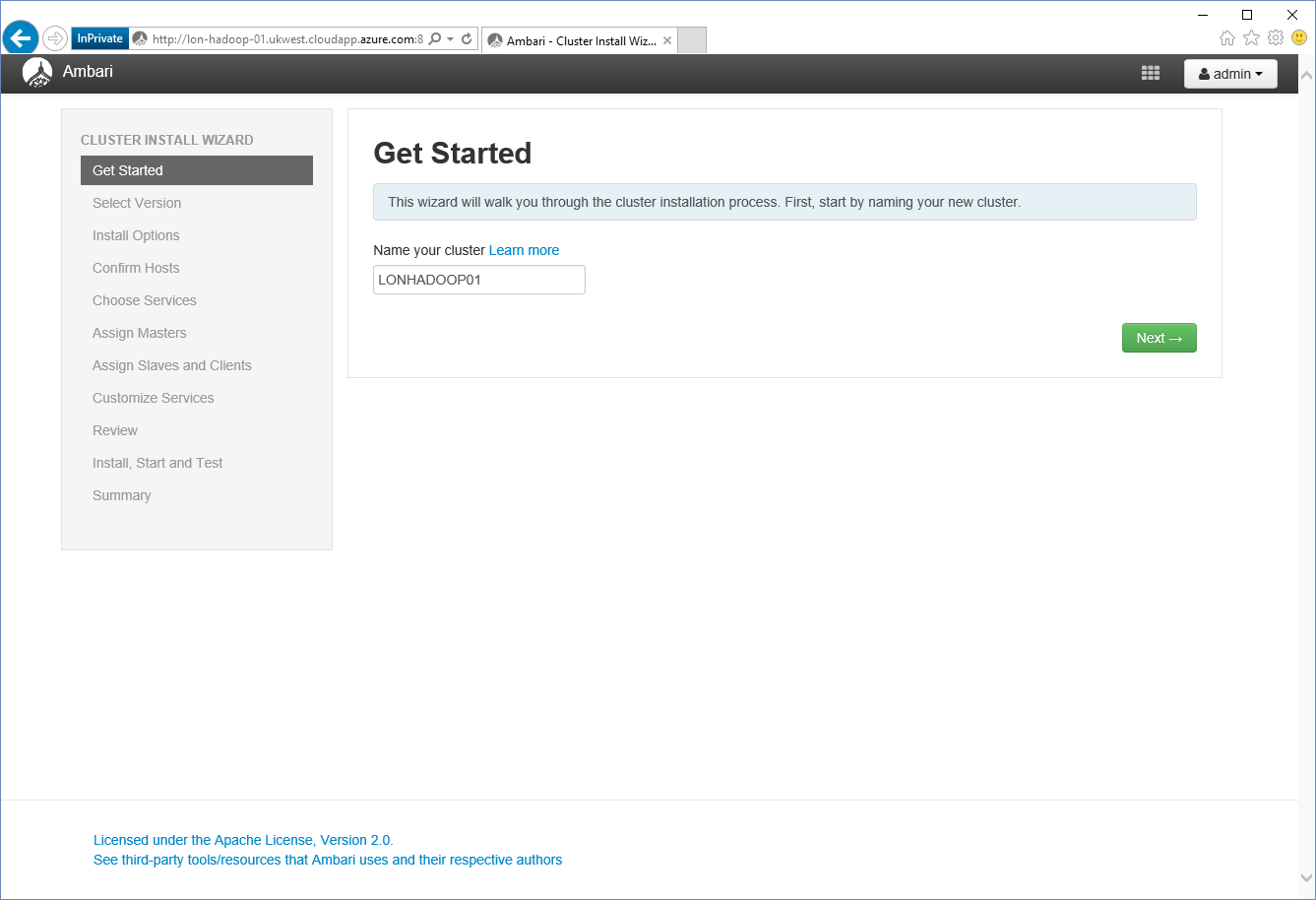
1. Open Internet Explorer, and navigate to [http://*fqdn*:8080](http://fqdn:8080), where fqdn is the fully qualified domain name of the LON-HADOOP VM. For example, http://lon-hadoop-01.ukwest.cloudapp.azure.com:8080
2. On the **Sign in** page, for the **Username** and **Password**, type **admin**, and then click **Sign in**.



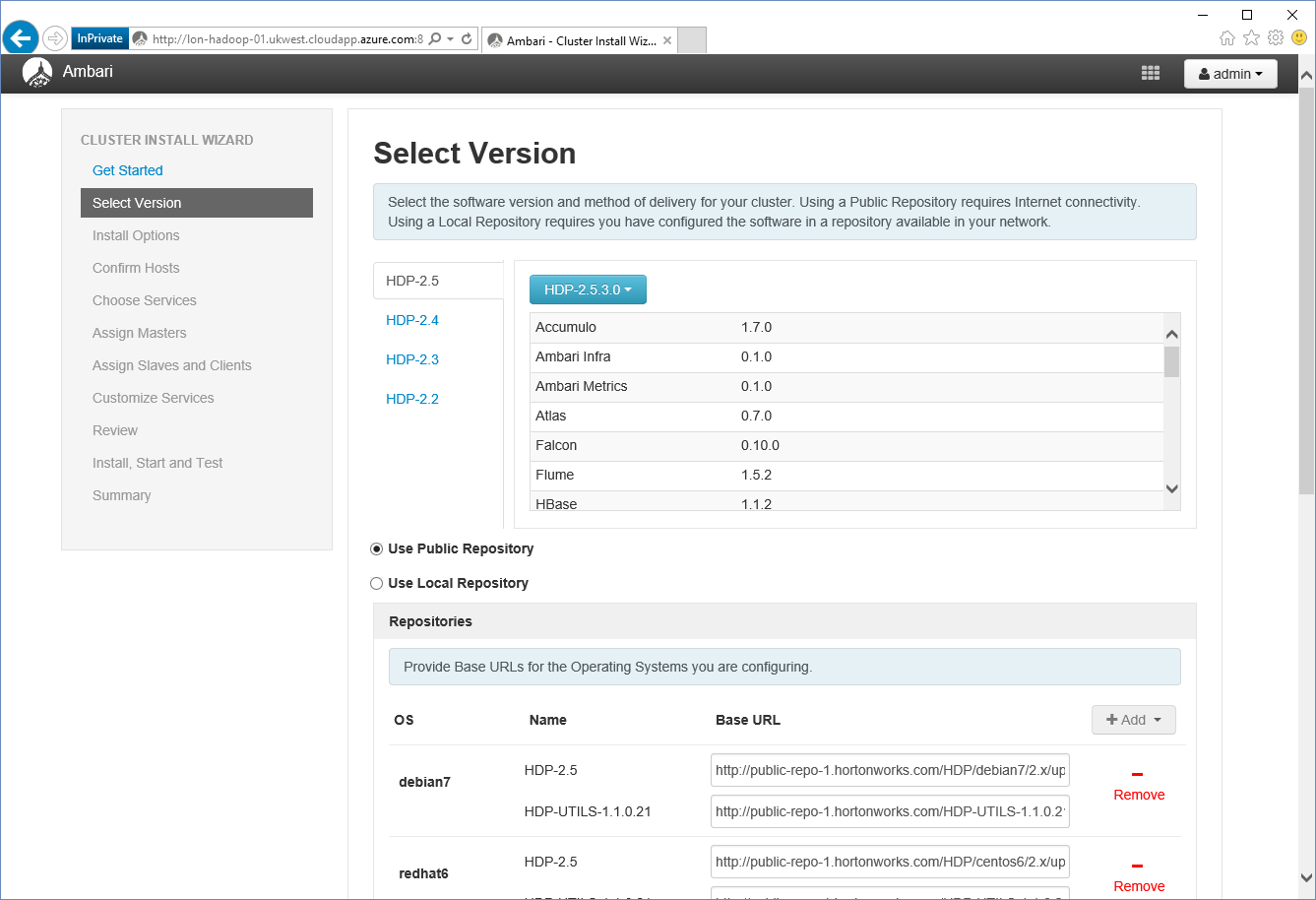
1. On the **Welcome to Apache Ambari** page, click **Launch Install Wizard**.



1. On the **Get Started** page, in the **cluster name** box, type **LONHADOOP*nn*** (without the hyphens) and then click **Next**.



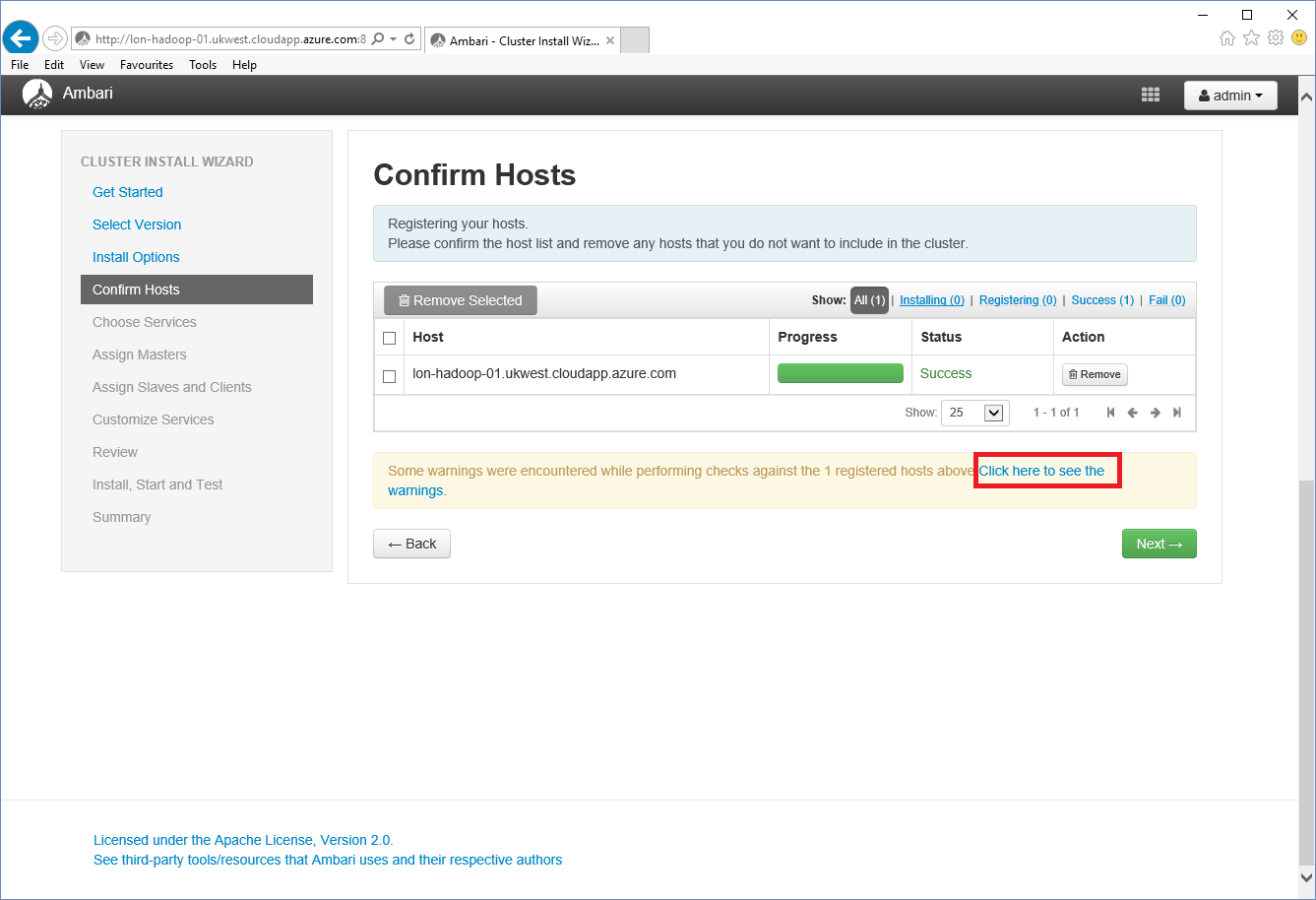
1. On the **Select Version** page, accept the default values, scroll to the bottom of the page, and then click **Next**.



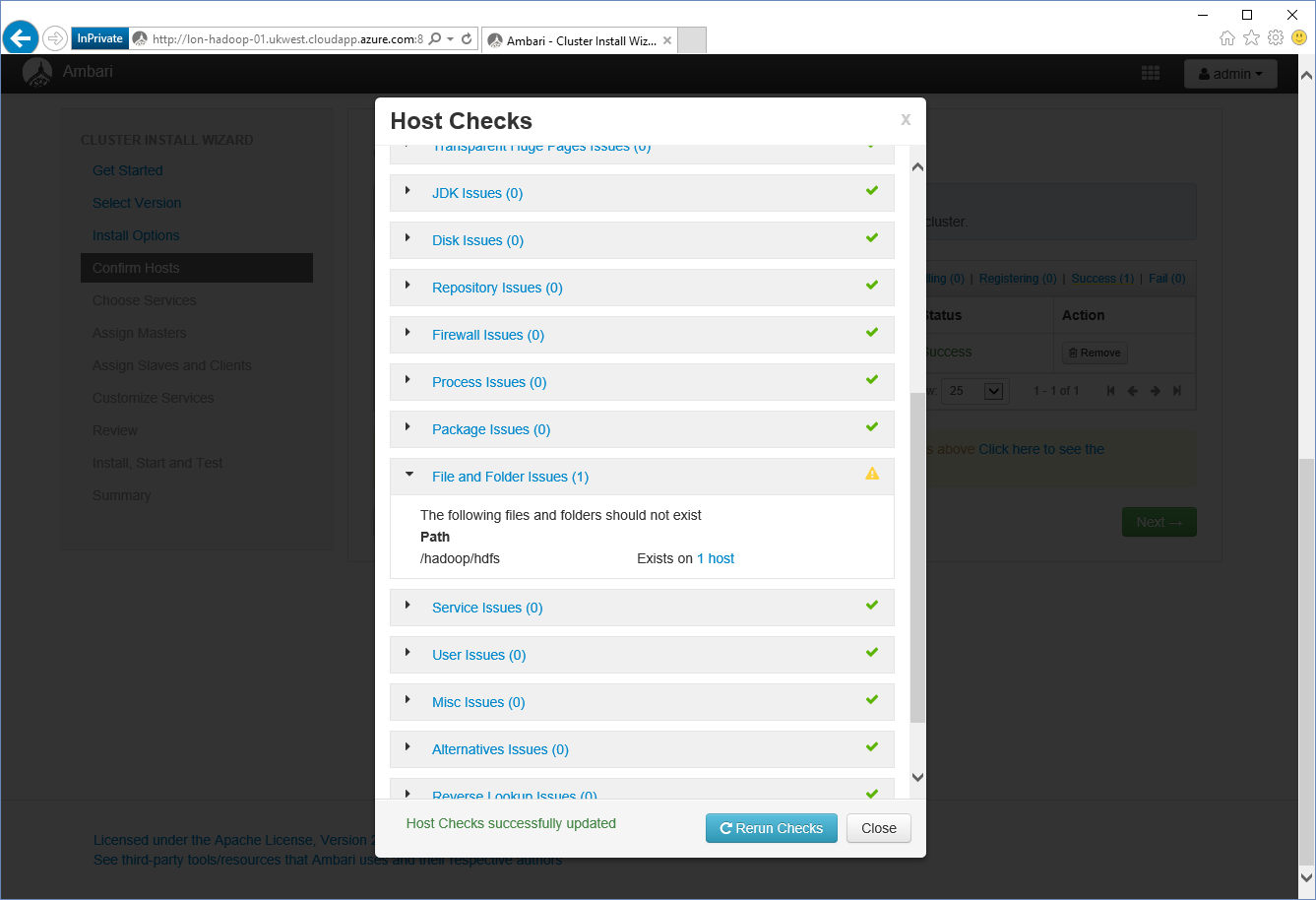
1. On the **Install Options** page, in the **Target Hosts** box, enter the fully qualified domain name (for example, lon-hadoop-01.ukwest.cloudapp.azure.com).
2. In the **Host Registration Information** section click **Browse**, move to the **C:\Temp** folder, select the **id\_rsa** file, click **Open**, and then click **Register and Confirm**.



1. On the **Confirm** Hosts page, you will see a warning generated, click **Click here to see the warnings**.

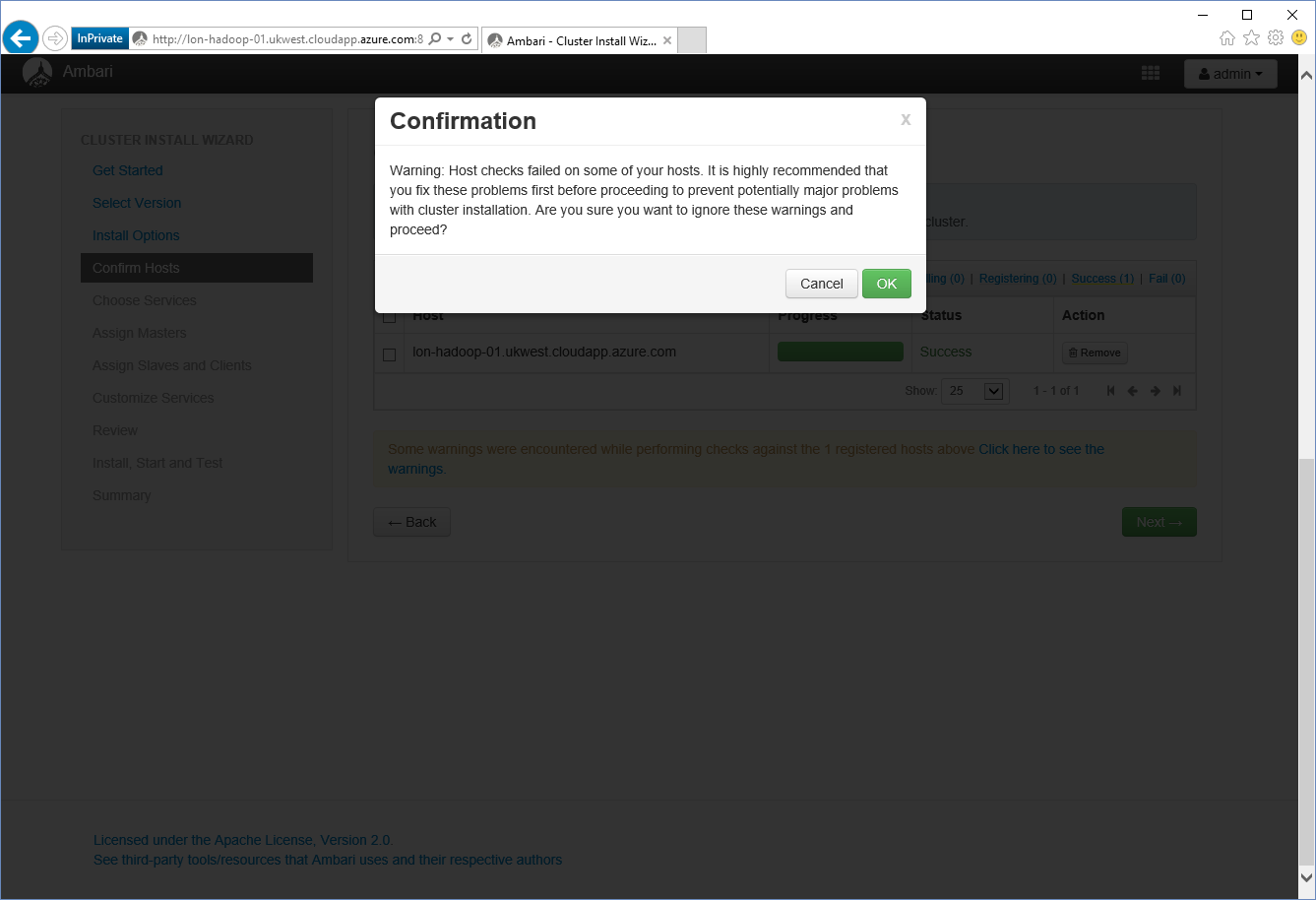


1. On the **Host Checks** page, all checks should be successful apart from a file and folder issue:

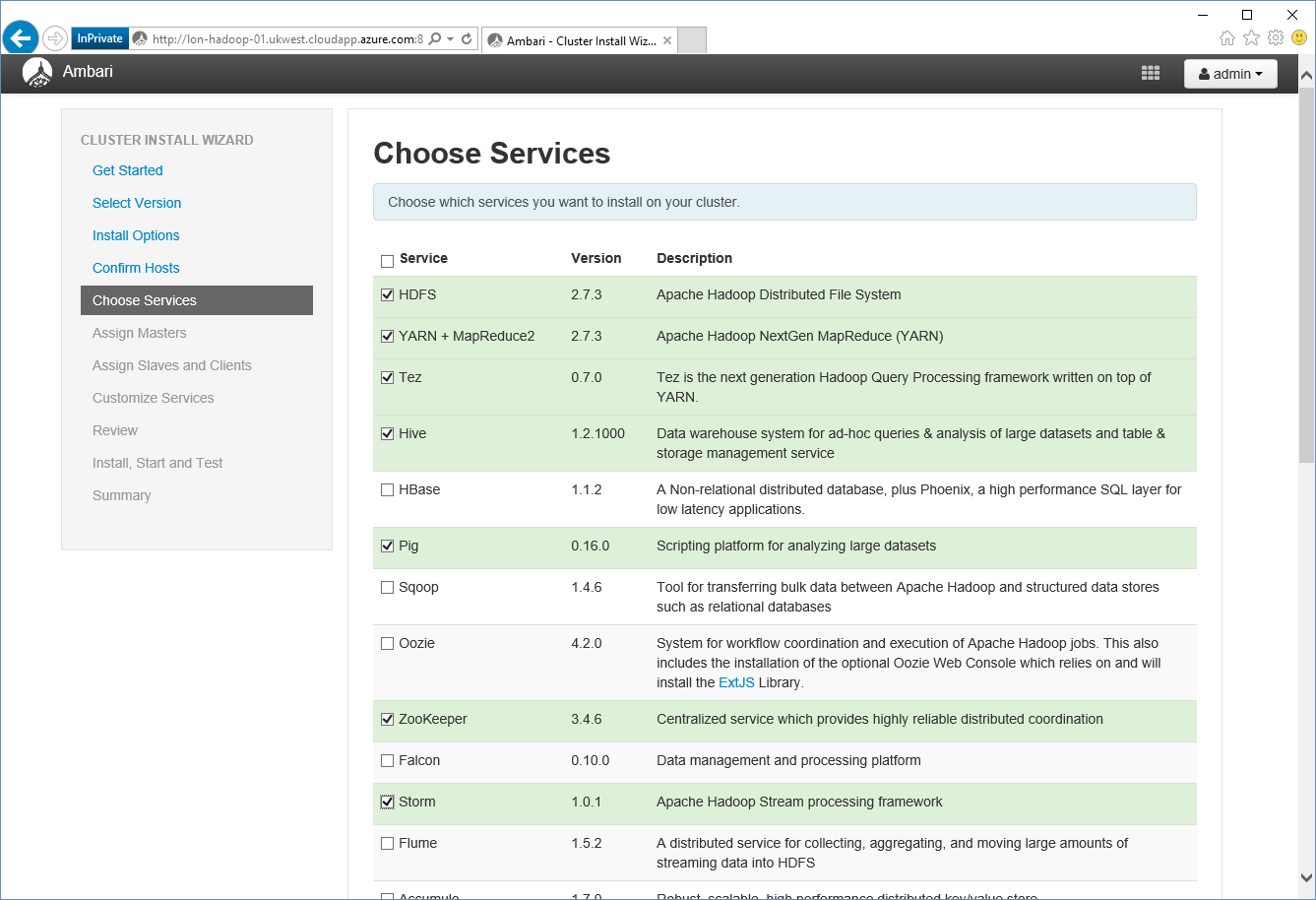


This warning occurs because the wizard assumes it will create the HDFS file system in a new directory on the Hadoop server, but we are using a separate partition that we have already mounted. You can ignore this warning, and click **Close**.

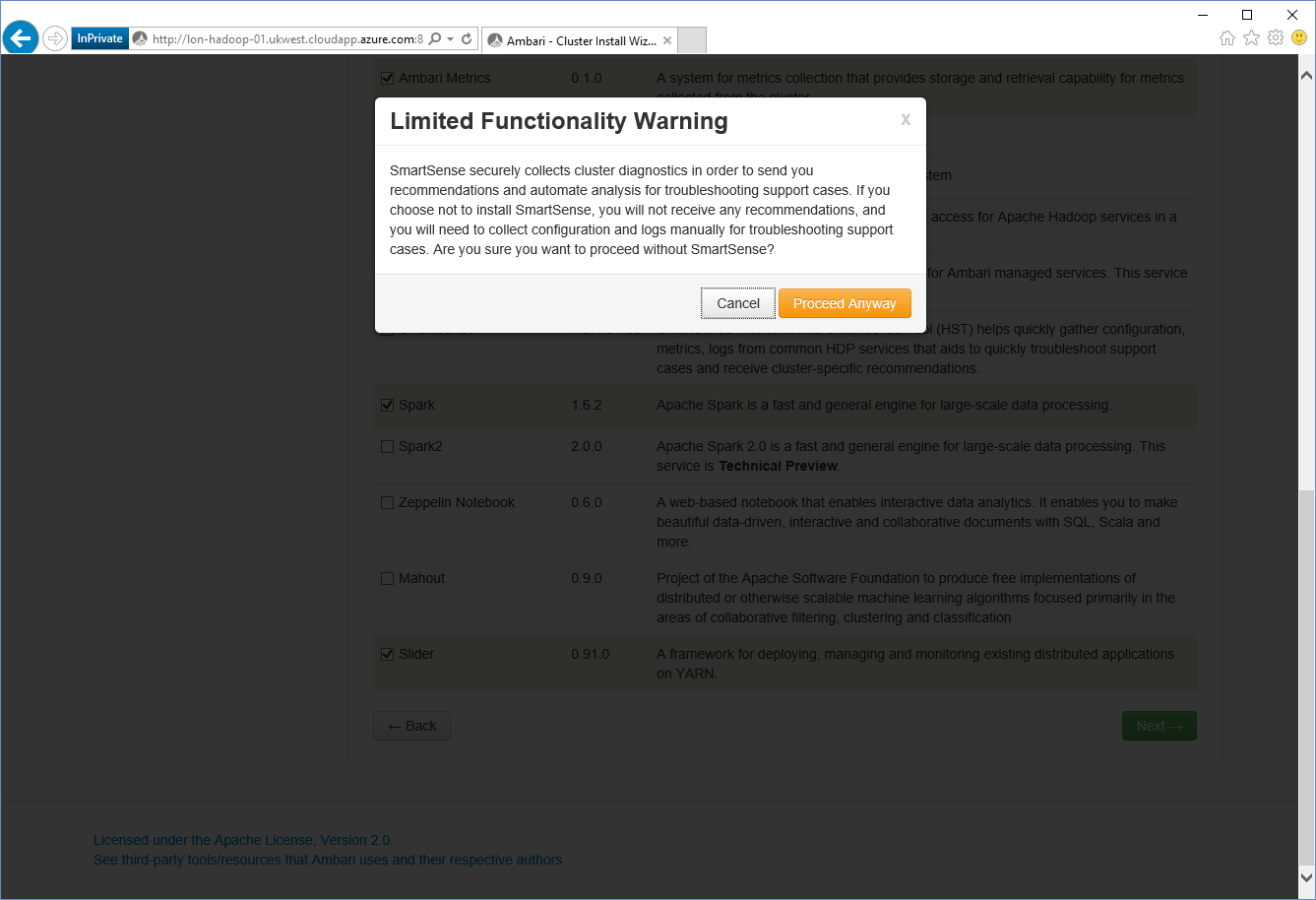
1. On the **Confirm Hosts** page, click **Next**. In the **Confirmation** dialog box, click **OK**.



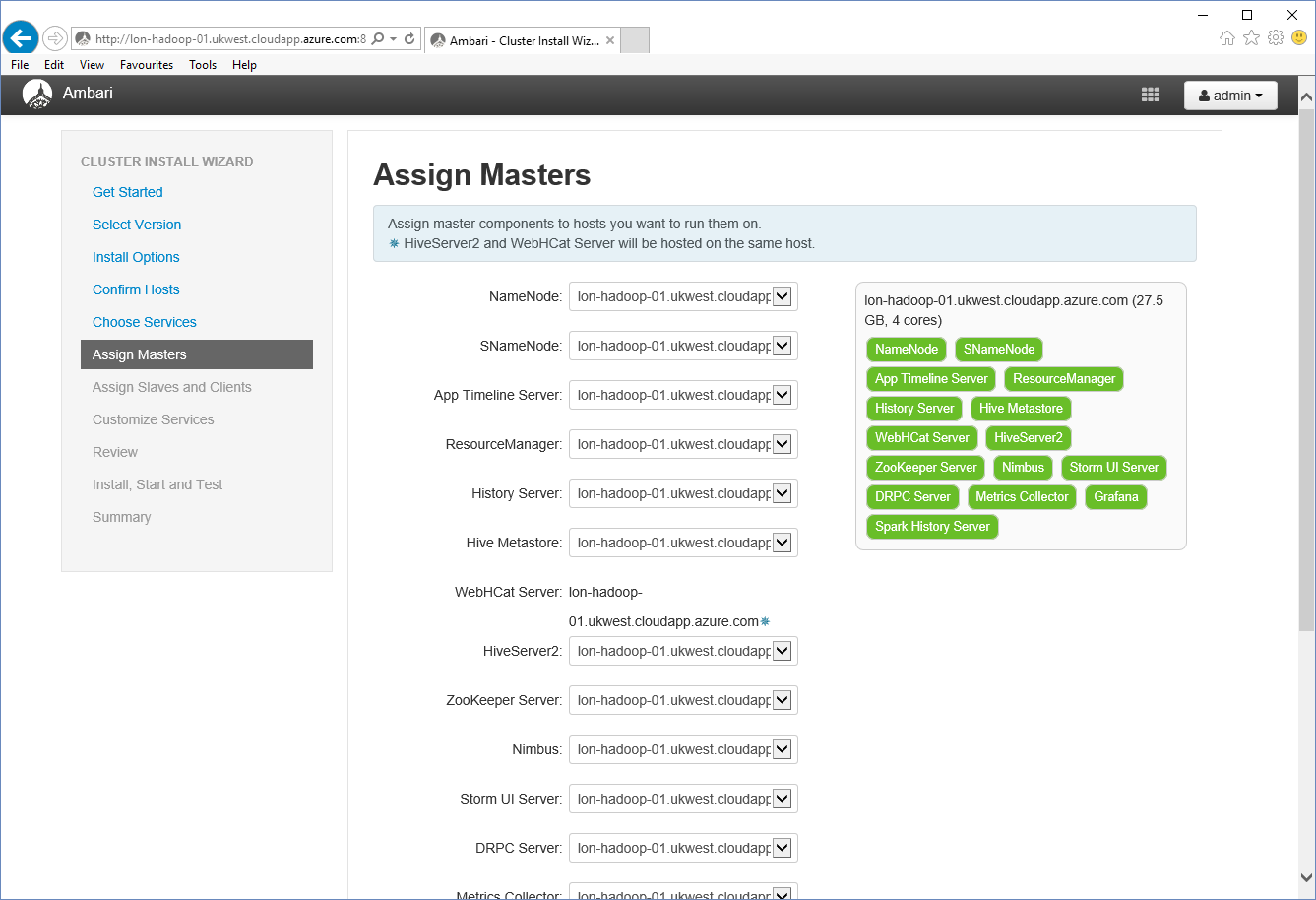
1. On the **Choose Services** page, select the following services, and then click **Next**.
   * HDFS
   * YARN + MapReduce2
   * Tez
   * Hive
   * Pig
   * Zookeeper
   * Storm
   * Ambari Metrics
   * Spark
   * Slider



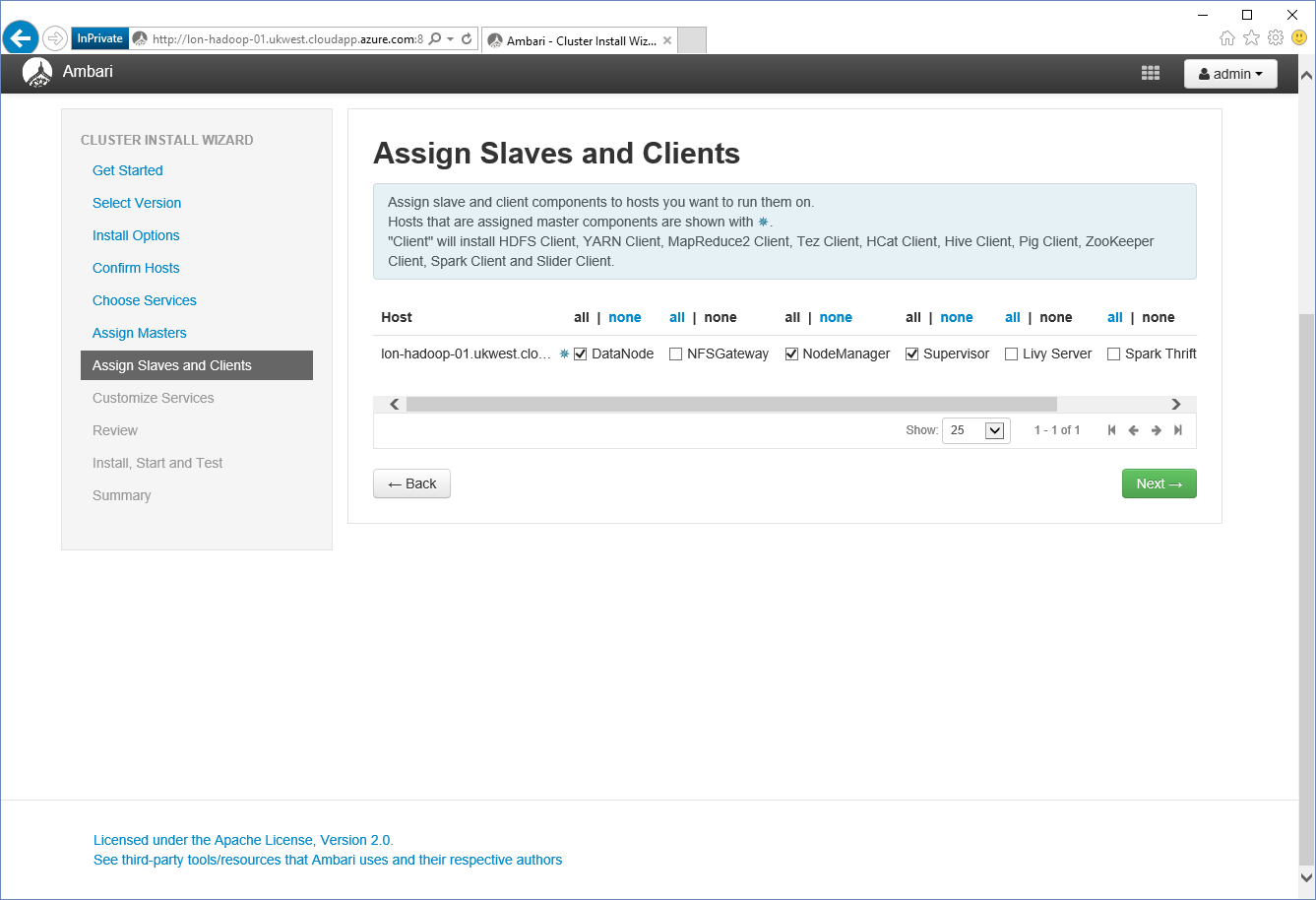
1. In the **Limited Functionality Warning** message box, click **Proceed Anyway** (this cluster does not require SmartSense).



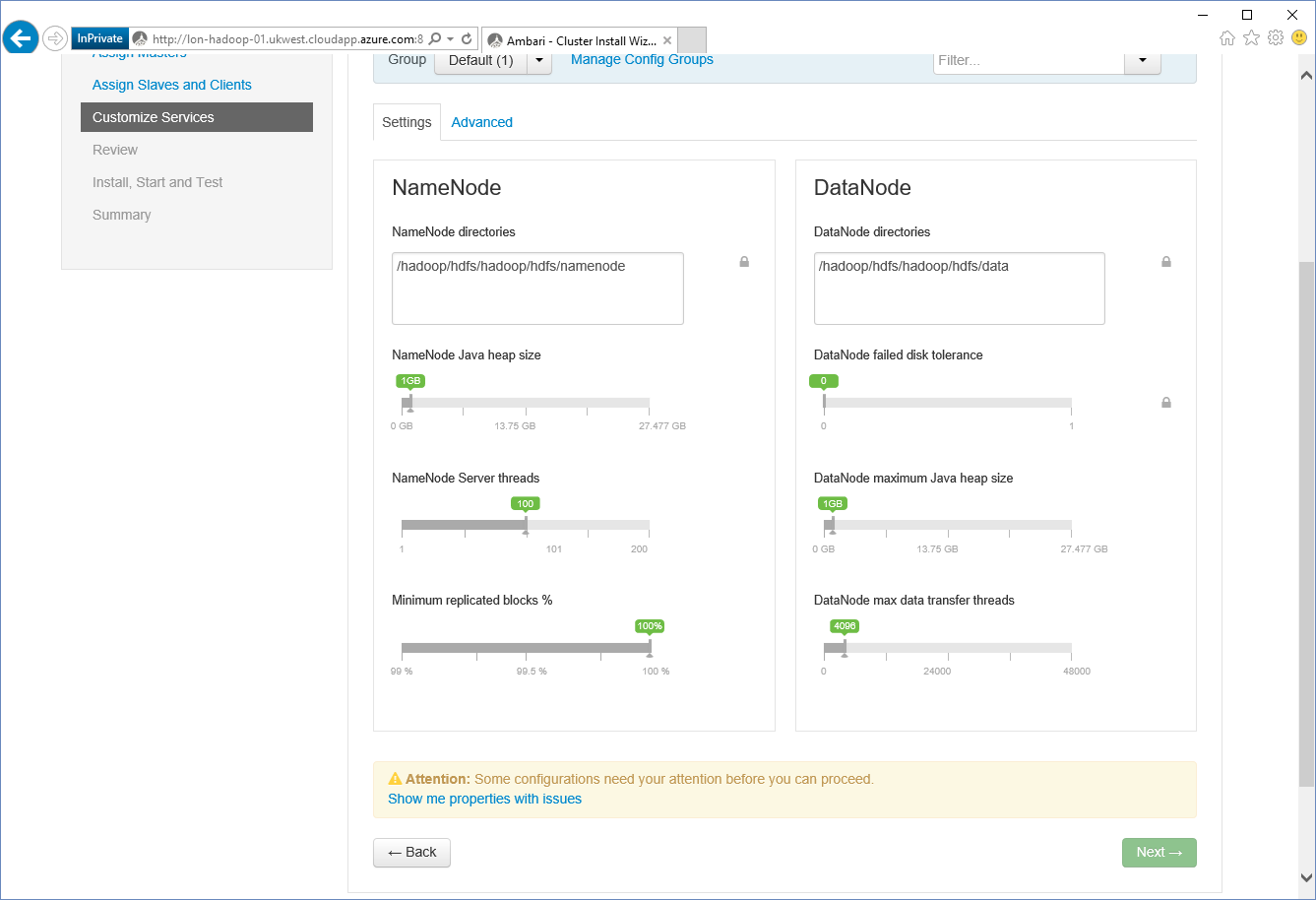
1. On the **Assign Masters** page, accept the default values and click **Next**.



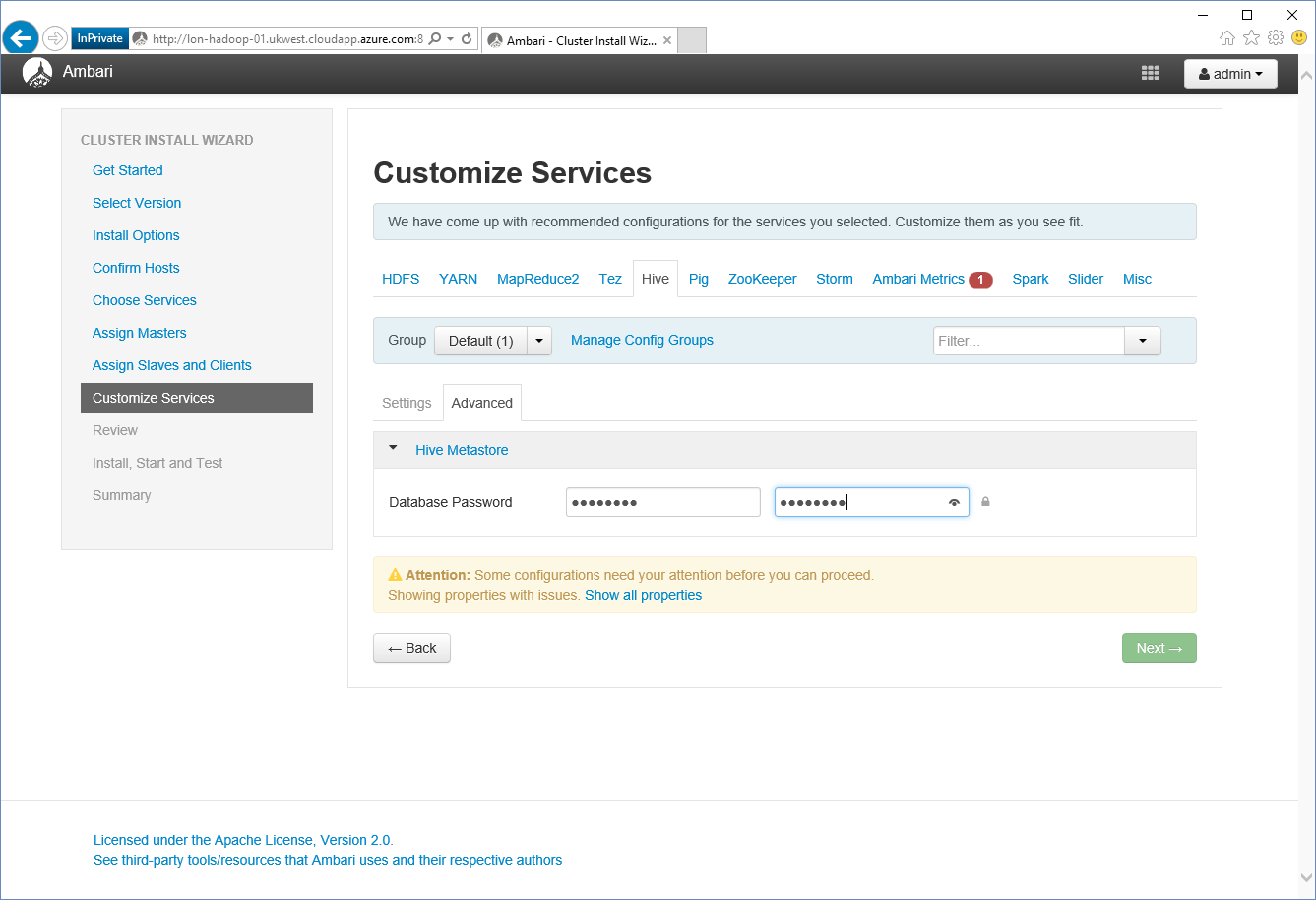
1. On the **Assign Slaves and Clients** page, accept the default values and click **Next**.



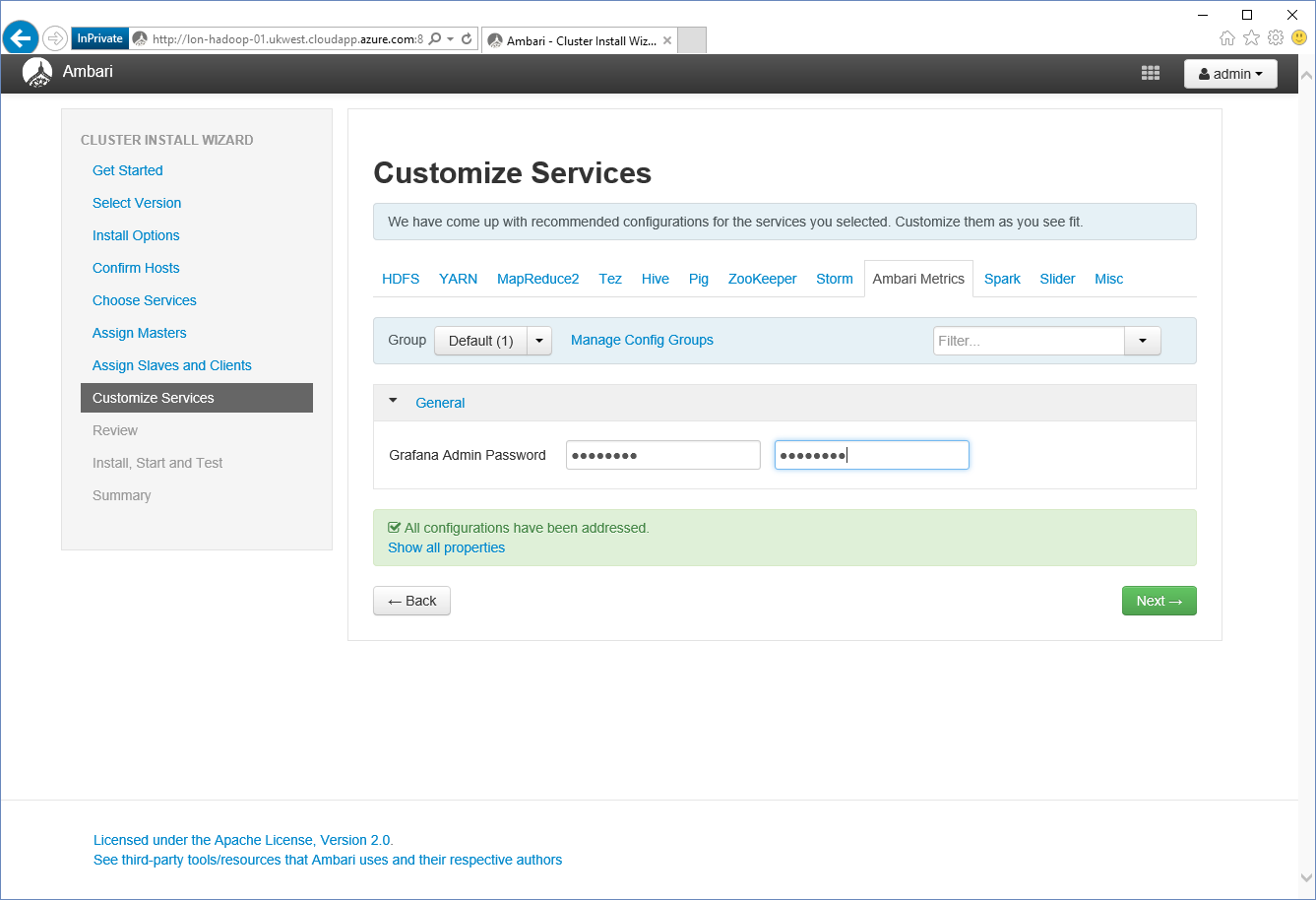
1. On the **Customize Services** page, click **Show me properties with issues** at the bottom of the page.



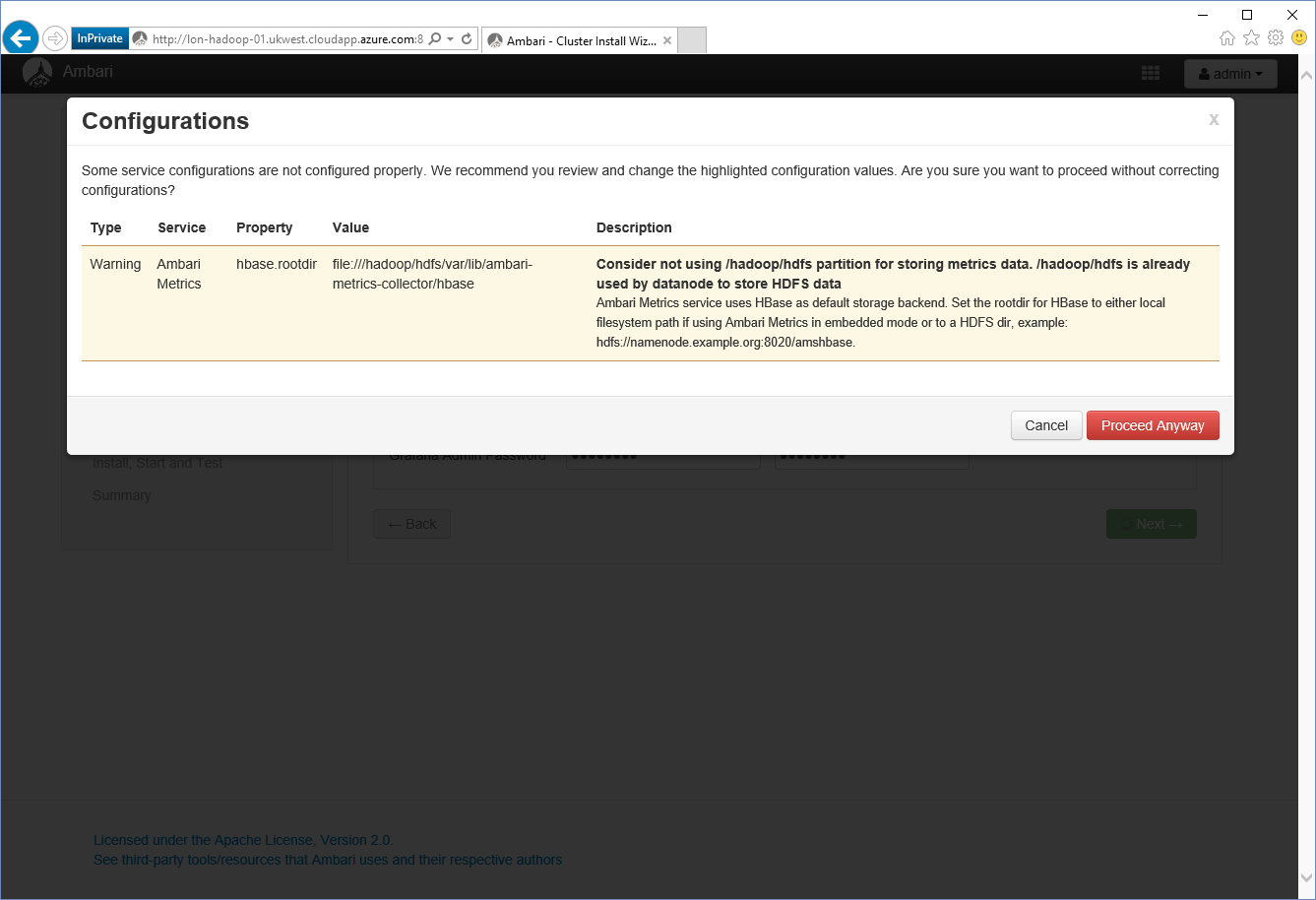
1. On the **Customize Services** page, on the **Hive** tab, in the **Hive Metastore** box, set the database password to **Pa55w.rd**, and then click the **Ambari Metrics** tab**.**



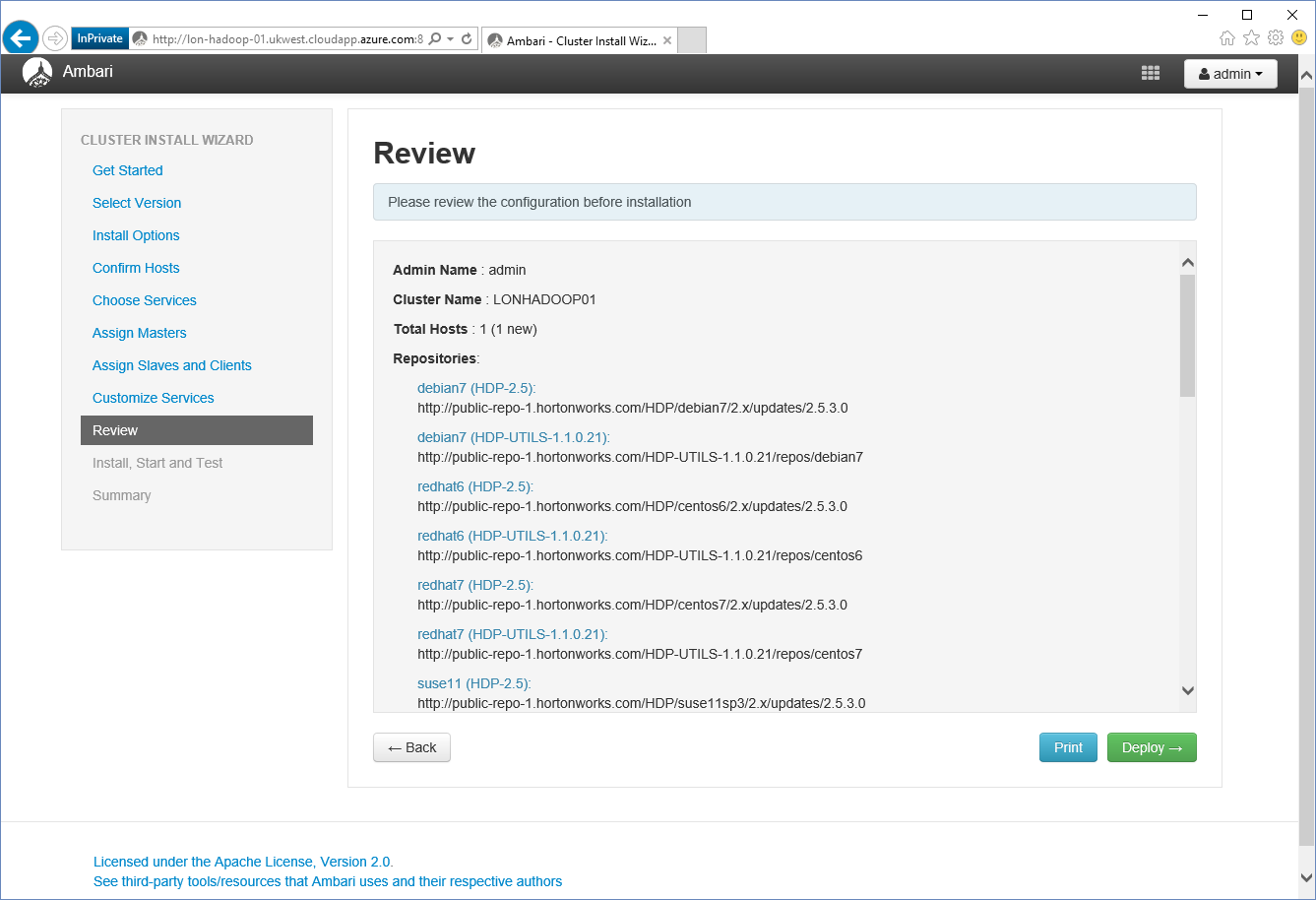
1. On the **Ambari Metrics** tab, in the **General** box, set the Grafana admin password to **Pa55w.rd**, and then click **Next.**



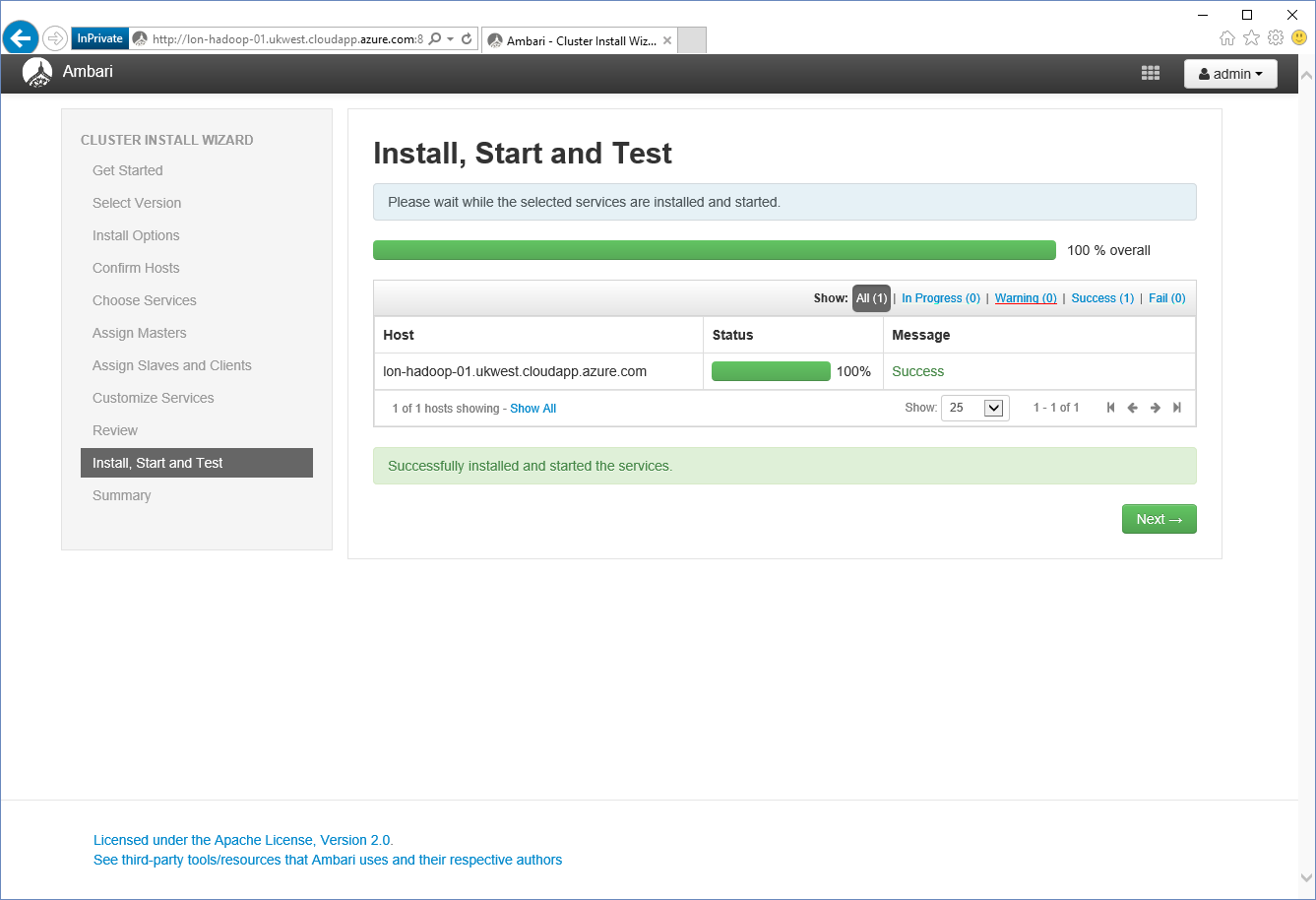
1. On the **Configurations** warning page, click **Proceed Anyway**.



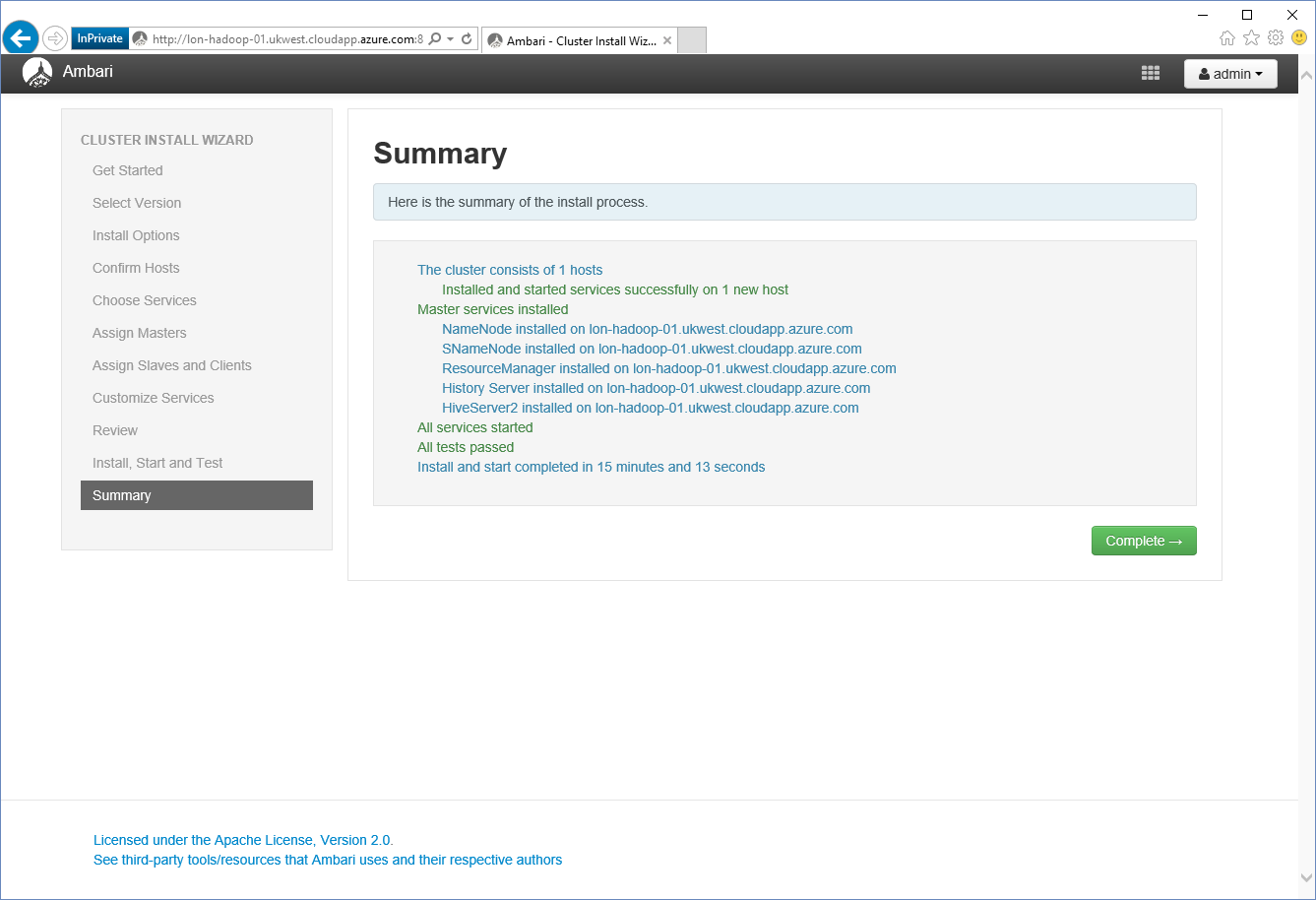
1. On the **Review** page, click **Deploy**.



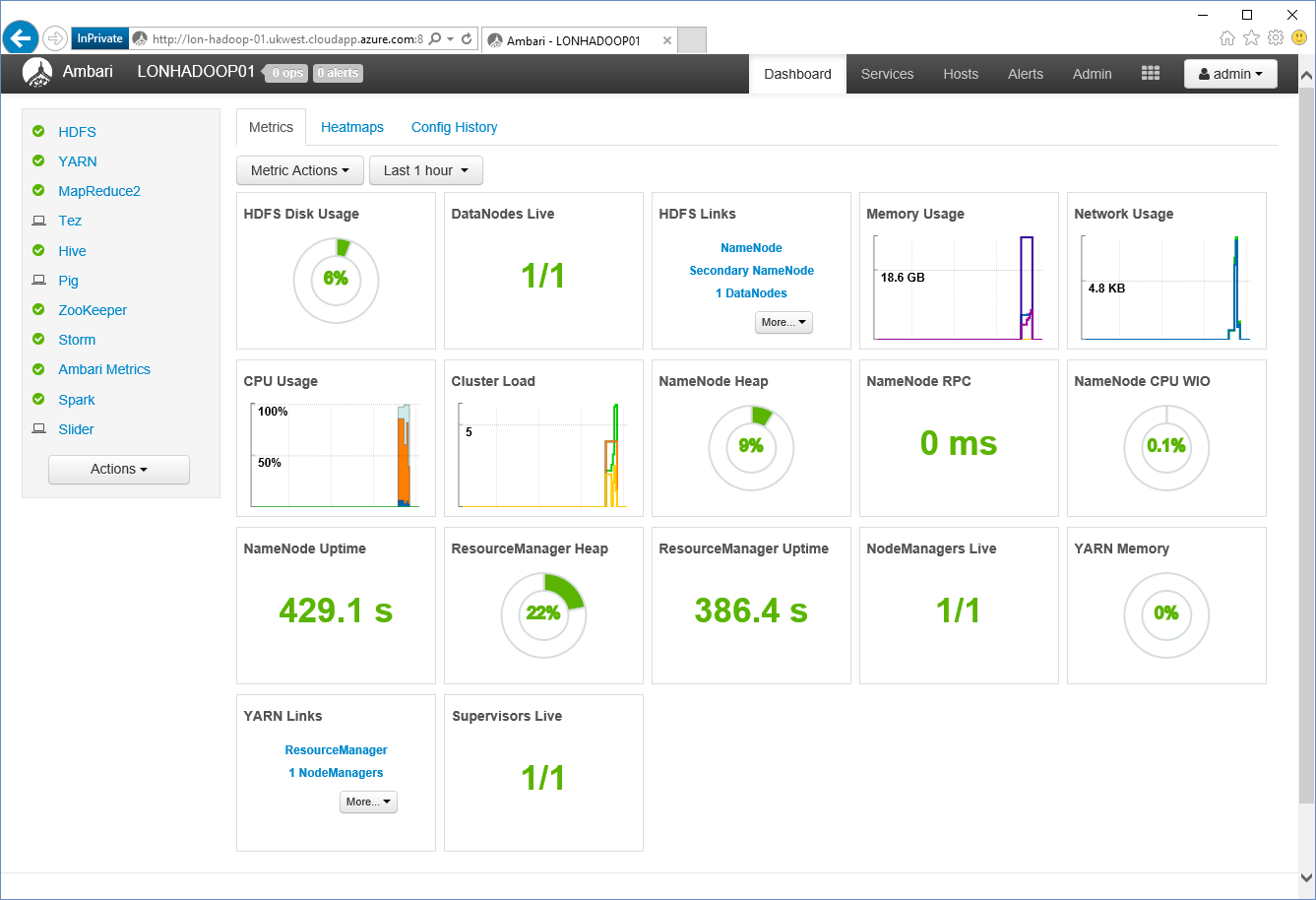
1. On the **Install, Start**, **and Test** page, wait while all services are deployed and started, and then click **Next**.



1. On the **Summary** page, click **Complete**.



1. The Metrics page for the Hadoop cluster should appear:



# Install Microsoft R Server for Hadoop on the LON-HADOOPVirtual Machine

Microsoft R Server for Hadoop enables R clients to connect to the Hadoop cluster and perform data analysis operations. Some R packages also require you to install various Linux development tools, libraries, and programming language compilers.

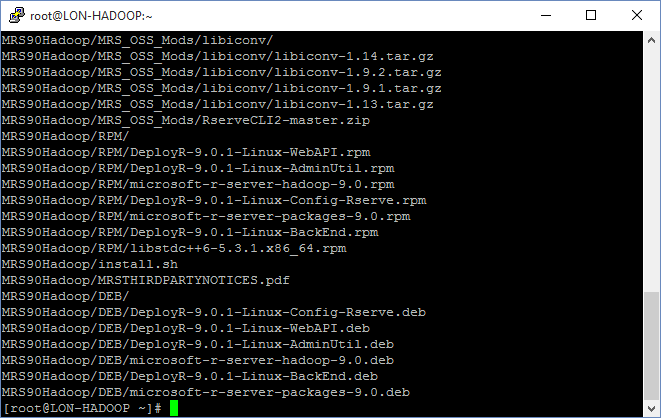
1. On the desktop machine, go to [**https://msdn.microsoft.com/en-us/microsoft-r/rserver-install-windows#download-r-server-installer**](https://msdn.microsoft.com/en-us/microsoft-r/rserver-install-windows#download-r-server-installer).
2. In the **Download R Server installer** section, click [**Visual Studio Dev Essentials**](https://na01.safelinks.protection.outlook.com/?url=http%3A%2F%2Fgo.microsoft.com%2Ffwlink%2F%3FLinkId%3D717968%26clcid%3D0x409&data=02%7C01%7CJim.Cochran%40microsoft.com%7C6aa25fa5114842a3ead008d4864a3b9c%7C72f988bf86f141af91ab2d7cd011db47%7C1%7C0%7C636281100775223390&sdata=0N92SJIT%2FHonn8GrenAFtOU%2Bs6mTf7kUIDG831rUrho%3D&reserved=0).
3. On the **Visual Studio Dev Essentials** page, click **Join or access now**, and enter your account information.
4. Sign-In with your credentials (either Hotmail or Outlook account).
5. Make sure you're in the right place: my.visualstudio.com.
6. In the **To access the following Visual Studio Dev Essentials benefits, please accept the terms of the program** message box, click **Accept**.
7. On the **Welcome to Visual Studio Dev Essentials** window, click **Confirm**.
8. At the top of the page, click on the **Downloads** tab.
9. In the search box, type **Microsoft R** **Server**, and the press Enter.
10. Under **Download Results**, click **Microsoft R** **Server 9.1.0**.
11. Next to **Microsoft R Server 9.1.0 for Hadoop (x64)**, click the **Direct download** icon.
12. In the message box, click **Save**.
13. Open a command prompt if one is not already available.
14. In the command prompt window, move to the folder containing the gzip file containing the Microsoft R Server 9.1.0 for Hadoop installer (en\_microsoft\_r\_server\_910\_for\_hadoop\_x64\_10323951.tar.gz in the example below), and run the following command to copy the installer to the VM, where fqdn is the fully qualified domain name of the LON-HADOOP VM:

pscp en\_microsoft\_r\_server\_910\_for\_hadoop\_x64\_10323951.tar.gz root@*fqdn*:r\_server\_910.gz

1. Enter **Pa55w.rd** for the root password when prompted.
2. On the desktop machine, run the **putty** command.
3. In the **PuTTY Configuration** window, in the list of saved sessions, click **LON-HADOOP**, click **Load**, and then click **Open**.
4. In the PuTTY terminal window that appears, at the **login as** prompt, log in as **root** with password **Pa55w.rd**. Note that password-less authentication has not been enabled for the desktop computer, so you still need to specify a password when logging in.
5. In the PuTTY terminal window, run the following commands to unpack the R Server installer file:

gunzip r\_server\_910.gz

tar xvf r\_server\_910

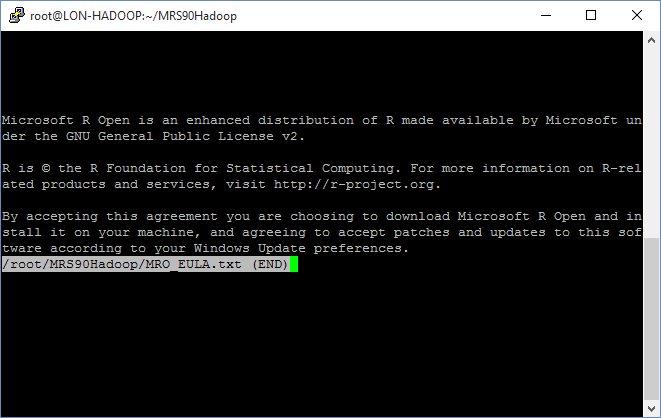


1. In the PuTTY terminal window, run the following commands to install R Server:

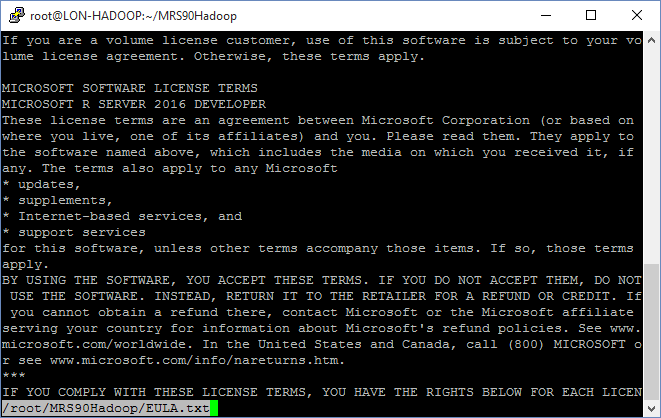
cd MRS91Hadoop

./install.sh -p

1. At the prompt, **Press [Enter] key to display the Microsoft R Open license. When finished reading, press q to continue**, press Enter.
2. When the license file is displayed, type **q**.



1. At the prompt **Do you agree to the terms of the previously displayed license. Choose [y]es | [n]o**, type **y**.
2. At the prompt **Press [Enter] key to display the Microsoft R Server license. When finished reading, press q to continue**, press Enter.
3. When the license file is displayed, type **q**.



1. At the prompt **Do you agree to the terms of the previously displayed license. Choose [y]es | [n]o**, type **y**.
2. In the PuTTY terminal window, run the following commands to connect as the **hdfs** user (this user has admin privileges over the HDFS file system):

su - hdfs

1. In the PuTTY terminal window, run the following commands to create the HDFS folders required by R server for the root user, together with the instructor and each of the students:

hadoop fs -mkdir /user/RevoShare/root

hadoop fs -mkdir /user/RevoShare/instructor

hadoop fs -mkdir /user/RevoShare/student01

hadoop fs -mkdir /user/RevoShare/student02

hadoop fs -mkdir /user/RevoShare/student03

...

hadoop fs -mkdir /user/RevoShare/student15

hadoop fs -chmod 777 /user/RevoShare/root

hadoop fs -chmod 777 /user/RevoShare/instructor

hadoop fs -chmod 777 /user/RevoShare/student01

hadoop fs -chmod 777 /user/RevoShare/student02

hadoop fs -chmod 777 /user/RevoShare/student03

...

hadoop fs -chmod 777 /user/RevoShare/student15

hadoop fs -mkdir /user/root

hadoop fs -mkdir /user/instructor

hadoop fs -mkdir /user/student01

hadoop fs -mkdir /user/student02

hadoop fs -mkdir /user/student03

...

hadoop fs -mkdir /user/student15

hadoop fs -chmod 777 /user/root

hadoop fs -chmod 777 /user/instructor

hadoop fs -chmod 777 /user/student01

hadoop fs -chmod 777 /user/student02

hadoop fs -chmod 777 /user/student03

...

hadoop fs -chmod 777 /user/student15

1. In the PuTTY terminal window, run the following command to return to running as the root user:

exit

1. In the PuTTY terminal window, run the following commands to create the file system folders required by R server for each user:

mkdir -p /var/RevoShare/root

mkdir -p /var/RevoShare/instructor

mkdir -p /var/RevoShare/student01

mkdir -p /var/RevoShare/student02

mkdir -p /var/RevoShare/student03

...

mkdir -p /var/RevoShare/student15

chmod 777 /var/RevoShare/root

chmod 777 /var/RevoShare/instructor

chmod 777 /var/RevoShare/student01

chmod 777 /var/RevoShare/student02

chmod 777 /var/RevoShare/student03

...

chmod 777 /var/RevoShare/student15

Ignore any warnings that occur if these directories already exist

1. In the PuTTY terminal window, run the following commands:

cd /usr/lib64

ln -s libpcre.so.1 libpcre.so.0

ln -s libicui18n.so.50 libicui18n.so.36

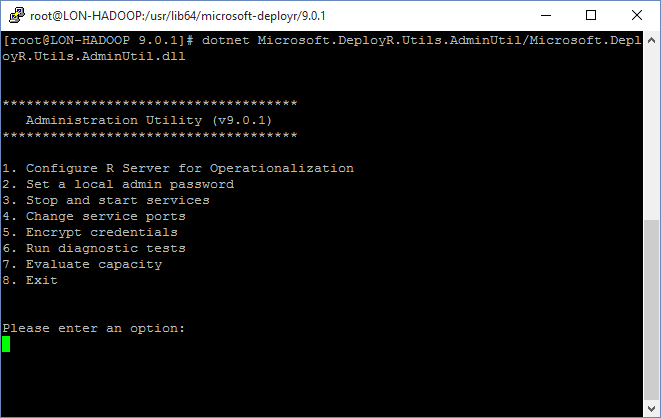
ln -s libicuuc.so.50 libicuuc.so.36

ln -s libicudata.so.50 libicudata.so.36

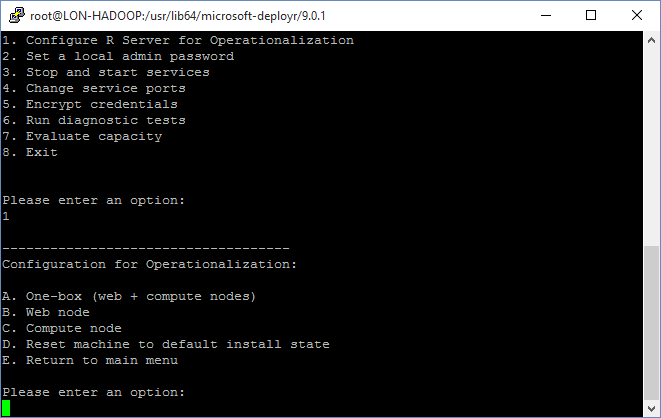
1. In the PuTTY terminal window, run the following commands to start the Microsoft R Administrator Utility:

cd /usr/lib64/microsoft-r/rserver/o16n/9.1.0

dotnet Microsoft.RServer.Utils.AdminUtil/Microsoft.RServer.Utils.AdminUtil.dll

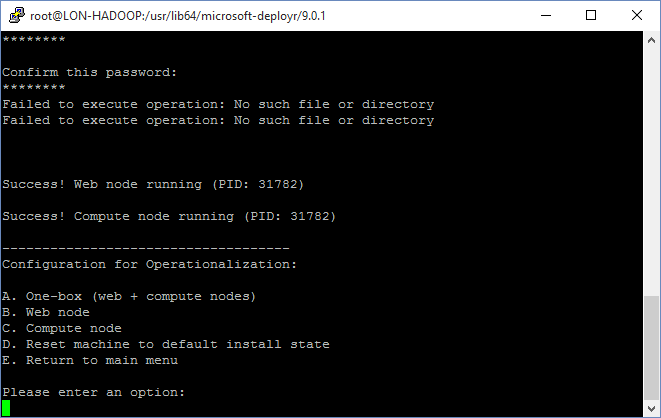


1. In the **Administration Utility** menu, type **1**, and then press Enter.
2. In the **Configuration for Operationalization** menu, type **A**, and then press Enter.

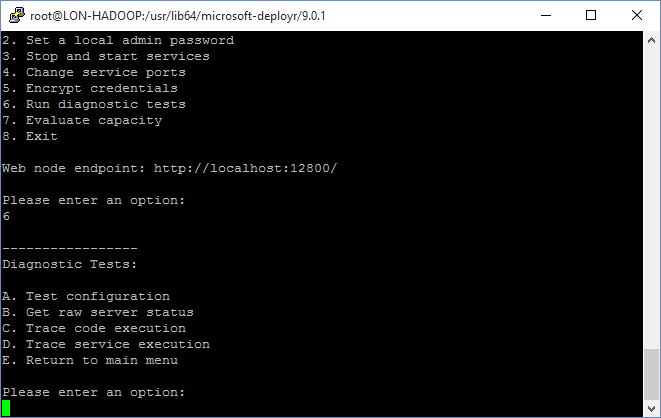


1. At the **Set the admin password** prompt, type **Pa55w.rd**, and then press Enter.
2. At the **Confirm this password** prompt, type **Pa55w.rd**, and then press Enter.

Note that the administration utility might report one or more warnings, **Failed to execute operation**. You can ignore these warnings.



1. In the **Configuration for Operationalization** menu, type **E**, and then press Enter.
2. In the **Administration Utility** menu, type **6**, and then press Enter.
3. In the **Diagnostic Tests** menu, type **A**, and then press Enter.



1. At the **Username** prompt, type **admin**, and then press Enter.
2. At the **Password** prompt, type **Pa55w.rd**, and then press Enter.
3. Verify that the diagnostic results show that the server is healthy:



1. In the **Diagnostic Tests** menu, type **E**, and then press Enter.
2. In the **Administration Utility** menu, type **8**, and then press Enter.
3. In the PuTTY terminal window, run the following commands to install the Linux development tools and compilers:

yum groupinstall "Development Tools"

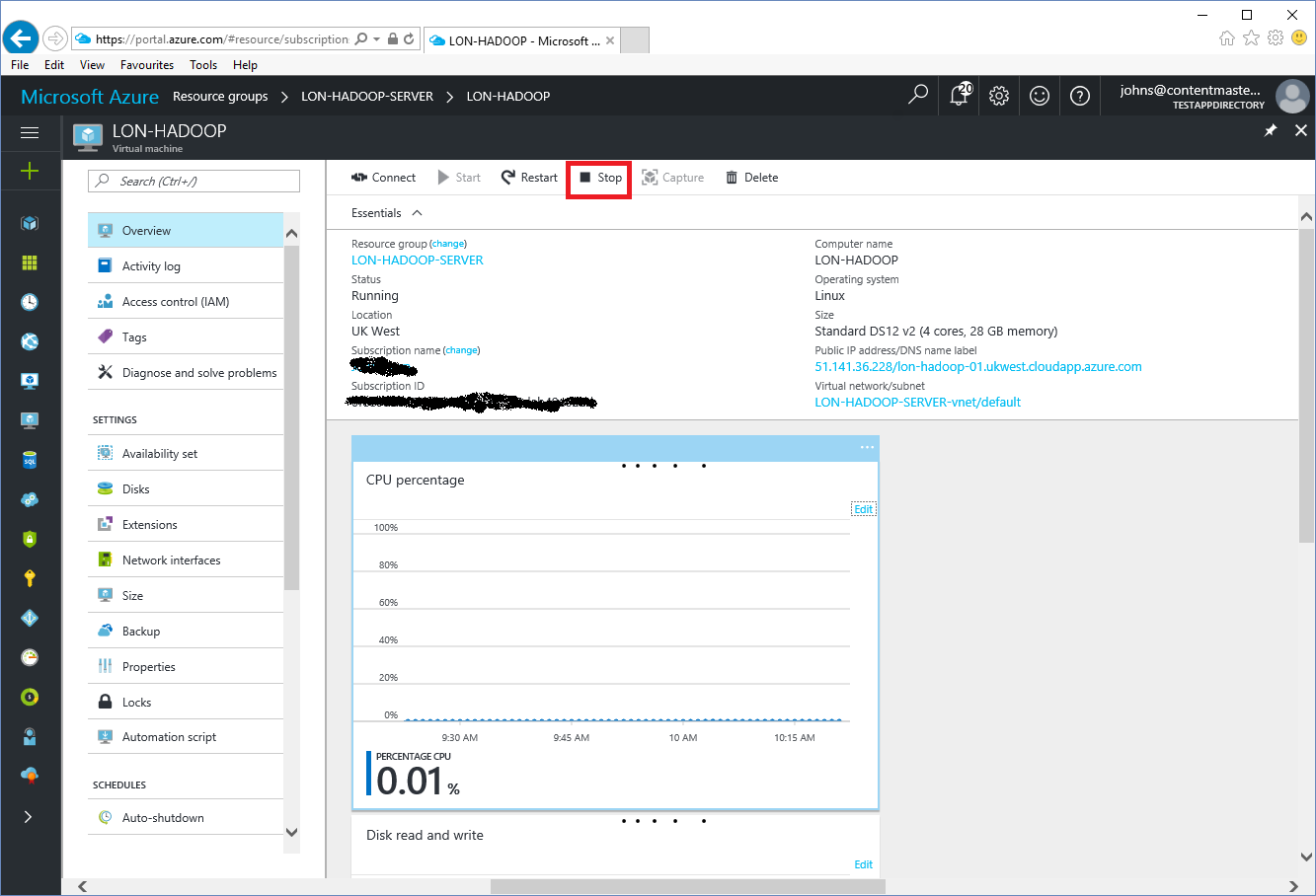
yum install libjpeg\*

At each **Is this ok** prompt, press **y**, and then press Enter.

1. Close the PuTTY terminal window. When prompted, click **OK** to confirm that you wish to exit the session.

# Shutdown the LON-HADOOP Virtual Machine

1. Using Internet Explorer, navigate to **portal.azure.com**.
2. Enter your Microsoft account credentials to log in.
3. In the navigation blade on the left side of the portal, click **Resource groups**, and then click the **LON-HADOOP-SERVER** resource group.
4. In the **LON-HADOOP-SERVER** blade, click the **LON-HADOOP** virtual machine.
5. In the **LON-HADOOP** blade, click **Stop**.



1. In the **Stop this virtual machine** dialog box, click **Yes**.
2. Close the **LON-HADOOP** blade.
3. Close Internet Explorer.