ThoughtSpot

ThoughtSpot Deployment Guide for Azure

Release 6.0

December, 2019

910 Hermosa Court, Sunnyvale, California 94085

Table of Contents

Installing Azure

Summary: Learn how to install a ThoughtSpot cluster on Azure.

Installation Prerequisites

Ensure the successful creation of the virtual machines (VMs) before you install the ThoughtSpot cluster on Azure.

1. **Review configuration options** Refer to Azure configuration options [See page 10] for detailed instance specs.
2. **Create the instance** Refer to Set up Azure for ThoughtSpot [See page 11] to create and launch your instance.
3. **Review required ports** Refer to Network Policies [See page 0] to view the required ports for successful operation of ThoughtSpot.

Configure Nodes

After creating the instance, you must configure the nodes.

Step 1: Log into your cluster

Use Terminal on a Mac or a terminal emulator on Windows to log into your cluster. Log in using the ssh private key provided by ThoughtSpot.

- If you do not have a private key, contact ThoughtSpot Support [See page 0] by email or through the support portal.

To log into your cluster, run `ssh -i <private-key> admin@<public-vm-ip>`.

```
$ ssh -i <private_key> admin@<public-vm-ip>
```

Step 2: Get a list of nodes to configure

Run the `tscli cluster get-config` command to get a list of the nodes that must be configured for the new cluster, and redirect it to the file `nodes.config`. You can find more information on this procedure in the nodes.config file reference [See page 0].

```
$ tscli cluster get-config |& tee nodes.config
```

Step 3: Configure the network of nodes

1. Refer to the autodiscovery of one node example [See page 0] before adding your specific network information.
2. Fill in the areas specified in Parameters of the `nodes.config` file [See page 0] with your specific network information.
    - If you have additional nodes, complete each node within the nodes.config file in the same way.

Make sure that you do not edit any part of the nodes.config file except the sections explained in Parameters of `nodes.config` [See page 0]. Deleting quotation marks, commas, or other parts of the code could cause setup to fail.

Step 4: Configure the nodes

Configure the nodes in the `nodes.config` file using the `set-config` command [See page 3].

1. Disable `Firewalld` by running `sudo systemctl stop firewalld` in your terminal. `Firewalld` is a Linux firewall that must be off for ThoughtSpot installation. When the cluster installer reboots the nodes, `Firewalld` automatically turns back on.
2. Run `cat nodes.config | tscli cluster set-config`.
    - If the command returns an error, refer to set-config error recovery [See page 4].

Set-config

```
$ sudo systemctl stop firewalld
$ cat nodes.config | tscli cluster set-config

Connecting to local node-scout
Setting up hostnames for all nodes
Setting up networking interfaces on all nodes
Setting up hosts file on all nodes
Setting up IPMI configuration
Setting up NTP Servers
Setting up Timezone
Done setting up ThoughtSpot
```

Set-config error recovery If the set-config fails with the following warning, restart the node-scout service by running `sudo systemctl restart node-scout`.

Restart node-scout service If you have this error, restart the node-scout:

```
Connecting to local node-scout WARNING: Detected 0 nodes, but f
ound configuration for only 1 nodes.
Continuing anyway. Error in cluster config validation: [] is no
t a valid link-local IPv6 address for node: 0e:86:e2:23:8f:76 C
onfiguration failed.
Please retry or contact support.
```

Restart node-scout with the following command, then retry the set-config command [See page 3].

```
$ sudo systemctl restart node-scout
```

The command output should no longer have a warning:

```
$ cat nodes.config | tscli cluster set-config

Connecting to local node-scout
Setting up hostnames for all nodes
Setting up networking interfaces on all nodes
Setting up hosts file on all nodes
Setting up IPMI configuration
Setting up NTP Servers
Setting up Timezone
Done setting up ThoughtSpot
```

Step 5: Confirm node configuration with the `get-config` command

Run `tscli cluster get-config` in your terminal to confirm node configuration.

Confirm node configuration

```
$ tscli cluster get-config

{
  "ClusterId": "",
  "ClusterName": "",
  "DataNetmask": "255.255.252.0",
  "DataGateway": "192.168.4.1",
  "IPMINetmask": "255.255.252.0",
  "IPMIGateway": "192.168.4.1",
  "Timezone": "America/Los_Angeles",
  "NTPServers": "0.centos.pool.ntp.org,1.centos.pool.ntp.or
g,2.centos.pool.ntp.org,3.centos.pool.ntp.org",
  "DNS": "192.168.2.200,8.8.8.8",
  "SearchDomains": "example.company.com",
  "Nodes": {
        "ac:1f:6b:8a:77:f6": {
          "NodeId": "ac:1f:6b:8a:77:f6",
          "Hostname": "Thoughtspot-server1",
          "DataIface": {
            "Name": "eth2",
            "IPv4": "192.168.7.70"
          },
          "IPMI": {
            "IPv4": "192.168.5.70"
          }
        }
  }
}
```

Install Cluster

Next, install the cluster using the release tarball (est. time 1 hour).

If you do not have a link to download the release tarball, open a support ticket at ThoughtSpot Support [See page 0] to access the release tarball.

Step 1: Run the Installer

1. Copy the downloaded release tarball to `/home/admin` with the command `scp 0.0.tar.gz admin@hostname:/home/admin/file-name`. Replace '0.0' with your release number. Replace 'hostname' and 'file-name' with your specific hostname and the name of the tarball file.

```
$ scp 0.0.tar.gz admin@hostname:/home/admin/file-name
```

2. Run `tscli cluster create <release>`.

```
$ tscli cluster create 6.0.tar.gz
```

3. Edit the output with your specific cluster information. For more information on this process, refer to Using the `cluster create` command [See page 0] and Parameters of the `cluster create` command [See page 0].

The cluster installer automatically reboots all the nodes after the install. `Firewalld` automatically turns back on. Wait at least 15 minutes for the installation process to complete. The system is rebooting, which takes a few minutes. Log into any node to check the current cluster status, using the command `tscli cluster status`.

Step 2. Check Cluster Health

Once the cluster is installed, check its status with the `tscli cluster status` command.

Cluster Status

```
$ tscli cluster status
Cluster: RUNNING
Cluster name    : thoughtspot
Cluster id      : 1234X11111
Number of nodes : 3
Release         : 6.0
Last update     = Wed Oct 16 02:24:18 2019
Heterogeneous Cluster : False
Storage Type    : HDFS

Database: READY
Number of tables in READY state: 2185
Number of tables in OFFLINE state: 0
Number of tables in INPROGRESS state: 0
Number of tables in STALE state: 0
Number of tables in ERROR state: 0

Search Engine: READY
Has pending tables. Pending time = 1601679ms
Number of tables in KNOWN_TABLES state: 1934
Number of tables in READY state: 1928
Number of tables in WILL_REMOVE state: 0
Number of tables in BUILDING_AND_NOT_SERVING state: 0
Number of tables in BUILDING_AND_SERVING state: 128
Number of tables in WILL_NOT_INDEX state: 0
```

Step 3. Finalize Installation

After the cluster status changes to “Ready,” log into the ThoughtSpot application on your browser.

Follow these steps:

1. Start a browser from your computer.
2. Enter your secure IP information on the address line.

   ```
   https:<IP-address>
   ```

3. If you don’t have a security certificate for ThoughtSpot, you must bypass the security warning to proceed:
   - Click **Advanced**
   - Click **Proceed**

4. The ThoughtSpot login page appears.
5. In the ThoughtSpot login window [See page 9], enter admin credentials, and click **Sign in**. Ask your network administrator if you do not know the network credentials. ThoughtSpot recommends changing the default admin password.

References

Use these references for successful installation and administration of ThoughtSpot.

- The `nodes.config` file [See page 0]
- Parameters of the `nodes.config` file [See page 0]
- Using the `cluster create` command [See page 0]
- Parameters of the `cluster create` command [See page 0]
- ThoughtSpot Documentation [See page 0]
- Contact Support [See page 0]

Azure configuration options

ThoughtSpot can be deployed in your Azure environment by deploying compute (VM) instances in your VNET as well as an underlying persistent storage infrastructure. Currently we support Premium SSD Managed Disks for persistent storage. For more information, see Managed Disks pricing [See page 0] in Microsoft's Azure documentation.

All Azure VMs (nodes) in a ThoughtSpot cluster must be in the same availability zone (and, therefore, also in the same region). ThoughtSpot does not support deploying VMs(nodes) of the same cluster across availability zones. For more information, see What are Availability Zones in Azure? [See page 0] in Microsoft's Azure documentation.

ThoughtSpot Azure instance types

Per VM user data capacity	Instance type	CPU/ RAM	Recommended per-VM Premium SSD Managed Disk volume
200 GB	E64sv3	64/432	2x1 TB
100 GB	E32sv3	32/256	2X 400 GB
20 GB	E16sv3	16/128	2X 400 GB
120 GB	D64v3	64/256	2X 1 TB

Set up ThoughtSpot in Azure

Summary: After you determine your configuration options, you must set up your virtual machines using a ThoughtSpot image for Azure.

About the ThoughtSpot image

To provision ThoughtSpot in the Azure portal, you'll need to access the ThoughtSpot Virtual Machine in the Azure Marketplace.

The ThoughtSpot Virtual Machine comes provisioned with the custom ThoughtSpot image to make hosting simple. A virtual machine is a preconfigured template that provides the information required to launch an instance of ThoughtSpot. It includes the following:

- A template for the root volume for the instance (for example, an operating system, an appliance server, and applications).

The ThoughtSpot Virtual Machine has the ThoughtSpot software installed and configured, on an CentOS base image. Check with your ThoughtSpot contact to learn about the latest version of the ThoughtSpot Virtual Machine.

Due to security restrictions, the ThoughtSpot Virtual Machine does not have default passwords for the administrator users. When you are ready to obtain the password, contact ThoughtSpot Support.

Set up ThoughtSpot in Azure

Follow these steps to provision and set up the VMs and launch ThoughtSpot.

Prerequisites

Complete these steps before launching your ThoughtSpot Virtual Machine:

1. Obtain an Azure login account.
2. Set up usage payment details with Microsoft Azure.
3. Set up a Resource Group.

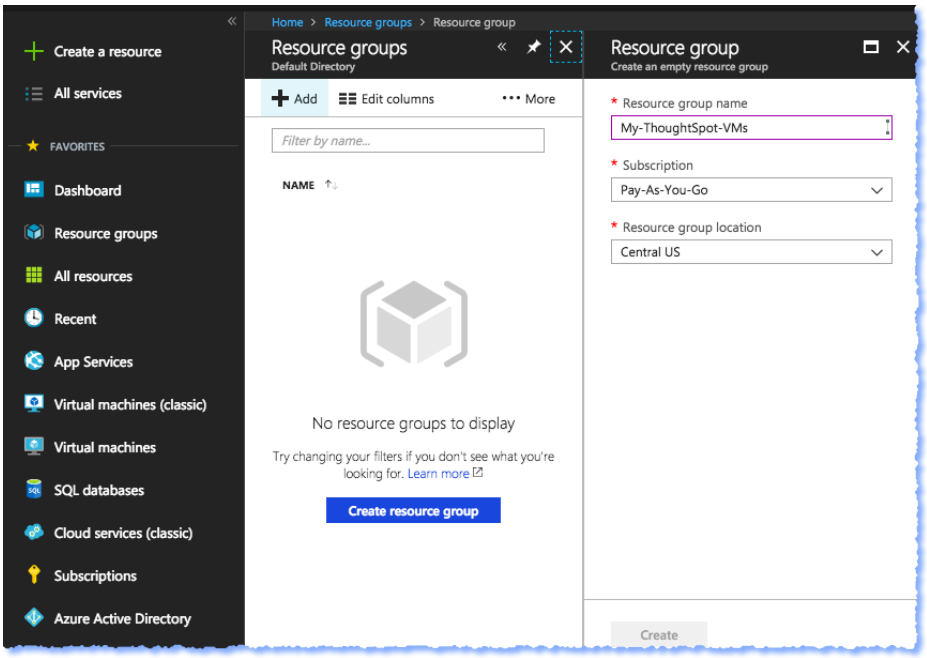
Create an instance

To get started, you need to log into the Azure portal, create a resource group, get the ThoughtSpot Virtual Machine [See page 0] on the Azure Marketplace [See page 0], create a resource based on the VM, and complete initial setup. You can either start at the Marketplace or from within the resource group you just created, as described here.

1. Log in to the Azure portal.

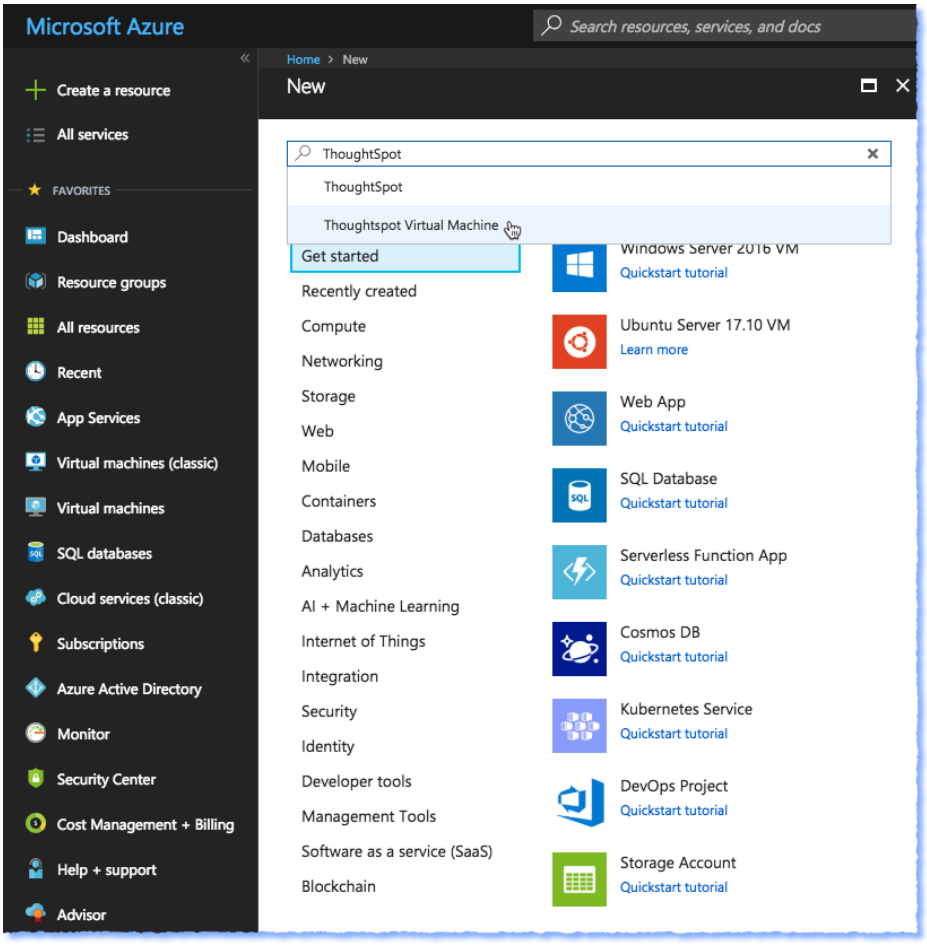
   In a browser, go to http://azure.microsoft.com [See page 0], and log in to your Azure account.

2. Create a Resource Group.

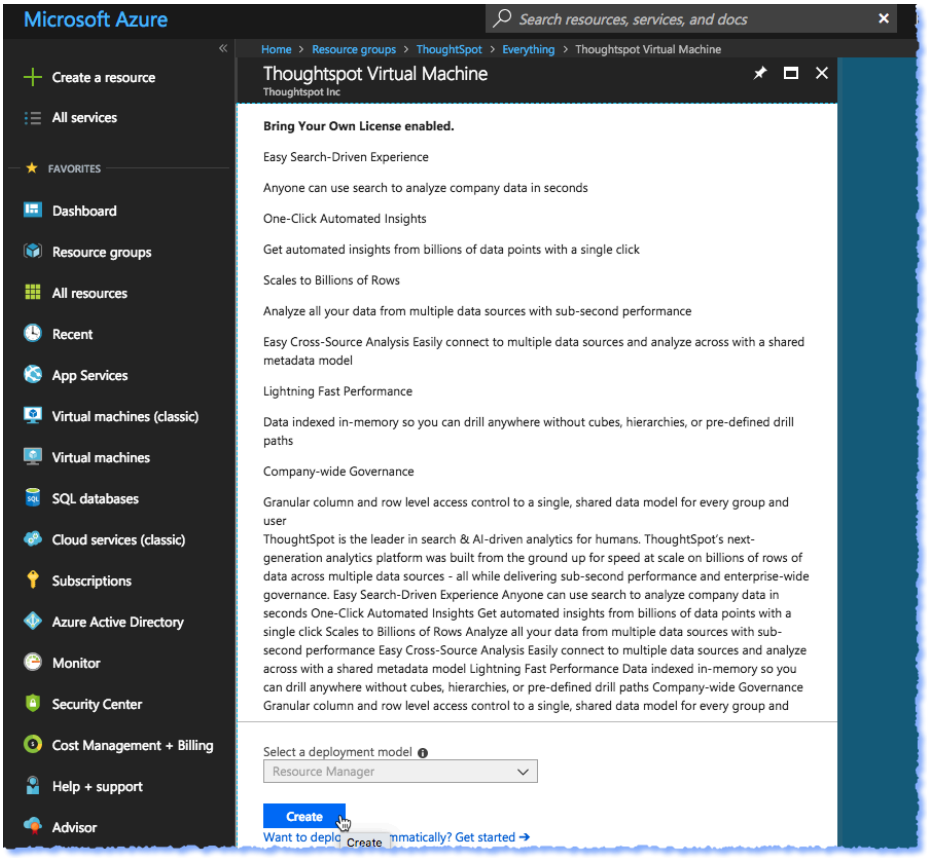


3. Next, create a resource based on the ThoughtSpot Virtual Machine.

   a. Click **Create a resource**, search the Marketplace for the ThoughtSpot Virtual Machine, and select it.



b. On the ThoughtSpot Virtual Machine page, click **Create**.



Configure basic settings

1. Provide a name and password for your new virtual machine.

2. Choose a disk type.

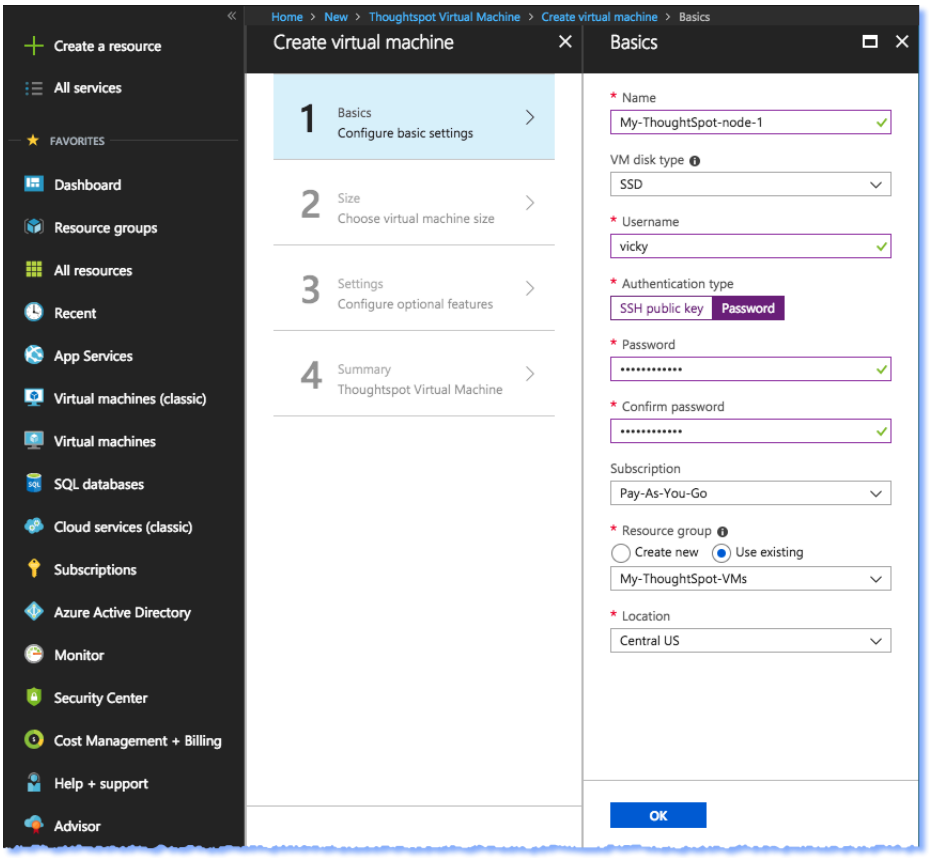
   > ☑ **Tip:** the new SSD disk types are currently available for only particular regions, so if you choose this disk type, make sure it's supported on the region you chose for your VM.

   See Standard SSD Disks for Virtual Machine workloads [See page 0] for more on SSD disks.

3. Provide a Resource Group, by clicking `existing` and selecting one.

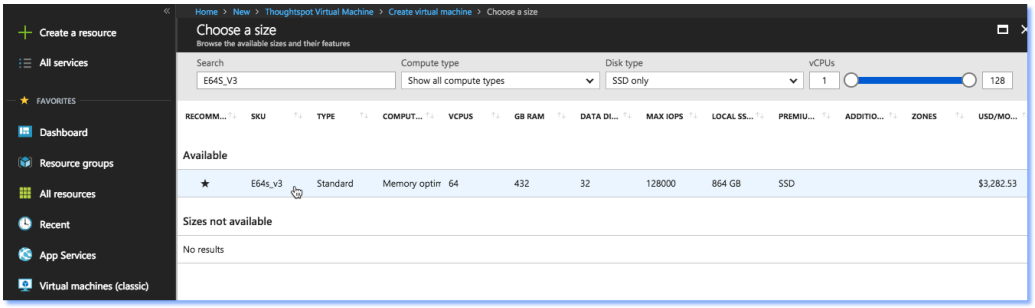
4. Select a location.

5. Click **OK** to save the Basics, which should look similar to the following example.



Choose a machine size

For **Choose a size**, select `E64S_V3 standard`.



Configure network settings, storage, and other options

Prerequisite: Get the details needed for setting up the Virtual Network, Subnet, and Network Security Group from your Azure support team.

1. For storage, select **Yes** to **use managed disks**.

2. Under **Network**, select **Virtual network**, then **Subnet**, then **Public IP addresses**, and set those names, addresses, and ranges approriately for your network.

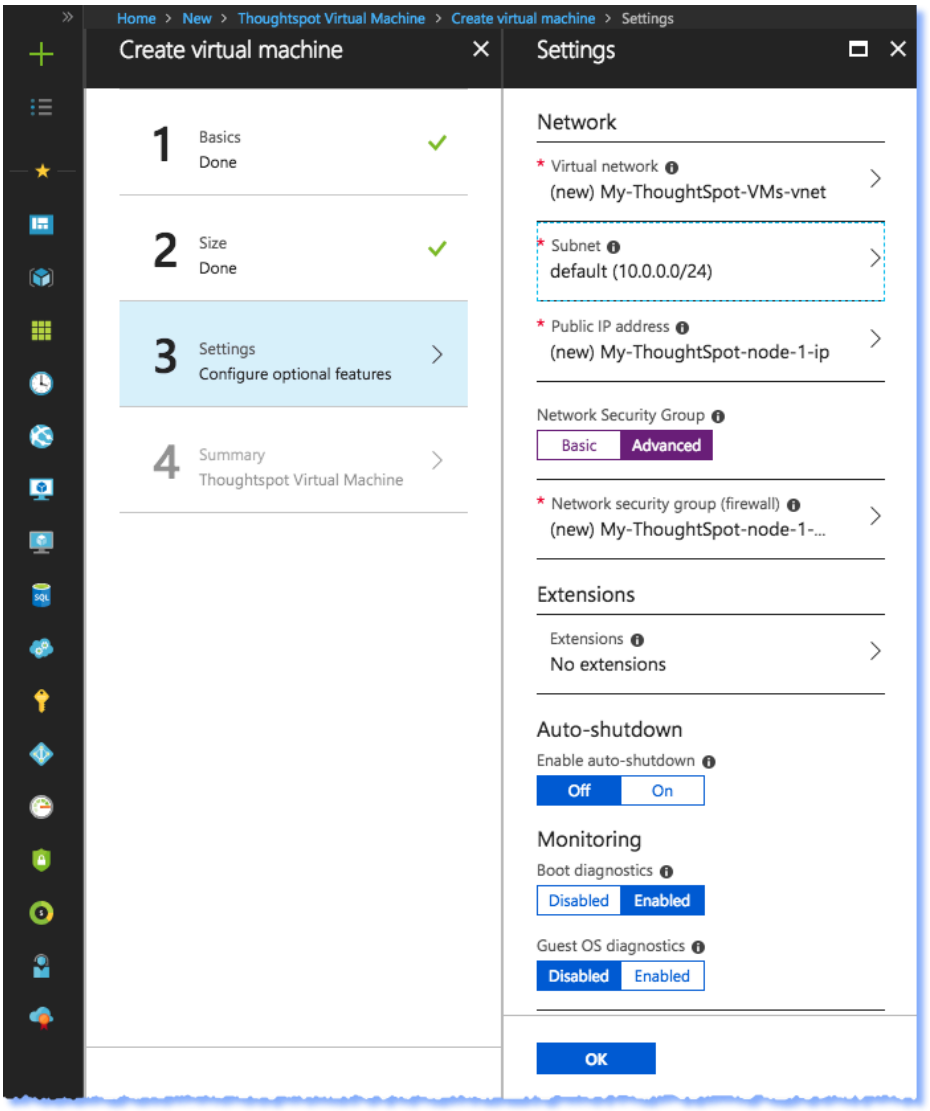
3. Open the necessary Inbound and Outbound ports to ensure that the ThoughtSpot processes do not get blocked.

   The minimum ports needed are:

Port	Protocol	Service
22	SSH	Secure Shell access
80	HTTP	Web access
443	HTTPS	Secure Web access
12345	TCP	ODBC and JDBC drivers access
2201	HTTP	Cluster Debugging
2101	HTTP	Node daemon Debugging
4001	HTTP	Data Cache Debugging

   ℹ **Note:** ThoughtSpot requires that nodes purchased from Azure must be reachable to each other so that they can communicate and form a distributed environment. ThoughtSpot only requires that those ports be accessible between nodes within a cluster. Use your discretion about whether to restrict public access or not for all nodes/ all ports

4. Leave other configurations such as `auto shutdown` and `monitoring` on their default settings.



5. Click **OK**.

   Azure will do the final validation check.

Review the Summary

Verify that the validation check succeeded and that summary of information shown is correct. If you find errors, reconfigure as needed.

When you are satisfied with the virtual machine setup, click **Create**.

Prepare for starting up ThoughtSpot

Prerequisite: To log in to the VM, you need the private key that is available in the image. You can obtain this from your ThoughtSpot contact.

1. Obtain the VM's public and private IP addresses.
    - To see the public IP, click the VM name link. This will show the public IP of the VM.
    - To see the private IP click Networking (under SETTINGS on the left side of the screen).
2. Connect to the VM through SSH, using the private key provided for the admin user.
    - You must file a support ticket to obtain this private key; it is necessary for the first login.
    - This key is different from the credentials, or the private keys supplied in earlier steps, which do not work in this context.
3. Update the password for both the `admin` and the `thoughtspot` users.

```
$ sudo passwd admin Changing password for user admin
$ sudo passwd thoughtspot Changing password for user tho
ughtspot
```

4. Update the file `/etc/hosts` with all the node IP addresses for the other VMs that will be part of the ThoughtSpot cluster.

Add Storage Disks

1. Go back to the VM and click it.
2. Add 2 SSD disks of 1TB each.
3. Click **Add data disk** and choose **Create disk from the menu**.
4. Create one mode data disk (demo-disk2) and save them both.
5. Click **Save** to add the disks to the VM.
6. Verify that the disks were added by issuing this command in the shell on the VM:

```
$ lsblk
```

Which returns results like:

```
NAME    MAJ:MIN RM  SIZE RO TYPE MOUNTPOINT
fd0       2:0    1    4K  0 disk
sda       8:0    0  200G  0 disk
├─sda1    8:1    0    1G  0 part /mntboot
├─sda2    8:2    0   20G  0 part /
├─sda3    8:3    0   20G  0 part /update
└─sda4    8:4    0  159G  0 part /export
sdb       8:16   0    1T  0 disk
└─sb1     8:17   0    1T  0 part /mnt/resource
sdc       8:32   0    1T  0 disk
sdd       8:48   0    1T  0 disk
sr0      11:0    1  628K  0 rom
```

7. Unmount the temporary disk by issuing:

```
$ sudo umount /mnt/resource
```

8. Prepare the disks /dev/sdc and /dev/sdd for ThoughtSpot by issuing the command:

```
$ sudo /usr/local/scaligent/bin/prepare_disks.sh /dev/sd
c /dev/sdd
```

Warning: Do not use the disk /dev/sdb. This is reserved for ThoughtSpot use.

9. Check the disks status by issuing:

```
$ df –h
```

10. Repeat the steps in this section for each node in your cluster.

Make network support settings

> ☑ **Tip:** All changes in this section must be re-applied each time after a cluster is created or updated. If these changes are not present, a reboot of the VMs will not have network access. So when updating these files, keep a backup to copy after any subsequent cluster creation or update.

1. Update hostnames for all the nodes by issuing:

   ```
   $ sudo hostnamectl set-hostname <HOSTNAME>
   ```

   If you are using a static name, you can issue:

   ```
   sudo hostnamectl set-hostname <HOSTNAME> --static
   ```

2. Update `/etc/hosts` with the IP and hostname:

   ```
   $ sudo vi /etc/sysconfig/network-scripts/ifcfg-eth0

   DEVICE=eth0 ONBOOT=yes BOOTPROTO=dhcp HWADDR=<Add eth0 M
   AC> TYPE=Ethernet USERCTL=no PEERDNS=yes IPV6INIT=no
   ```

3. Do not reboot any of the nodes, until these changes are made to each node:

   a. Open the grub file /update/etc/default/grub in an editor:

   ```
   $ sudo vi /update/etc/default/grub
   ```

   b. Change the line:

```
   GRUB_CMDLINE_LINUX="console=tty0 console=ttyS1,115200
n8"
```

to:

```
   GRUB_CMDLINE_LINUX="console=tty0 console=ttyS1,115200
n8 net.ifnames=0"
```

c. Save your changes.

4. Issue these commands:

```
$ sudo cp /update/etc/default/grub /etc/default/
$ rm /usr/local/scaligent/bin/setup-net-devices.sh
```

5. Reboot the nodes.