ThoughtSpot

Node Setup Guide

Version THEVERSION December 02, 2019

Copyright for ThoughtSpot publications. © 2019 ThoughtSpot, Inc. All rights reserved.

ThoughtSpot, Inc. 910 Hermosa Ct
Sunnyvale, CA 94085

All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. ThoughtSpot is a trademark of ThoughtSpot, Inc. in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Table of Contents

Overview

You can install a ThoughtSpot cluster on a hardware appliance, cloud service, or VMware appliance. Your ThoughtSpot installation cannot mix node types. For example, you can have either hardware or VMware nodes, but not both. You can, however, have a cloud cluster for development and use an appliance for production.

This guide instructs you how to prepare each of the following:

- Hardware appliance
- Amazon Web Services (AWS) EC2
- Microsoft Azure
- Google Cloud Platform (GCP)
- VMware

After you configure your nodes, you can contact ThoughtSpot Support by phone, mail, email, or by filing a support ticket.

Hardware appliance overview

Summary: What is in the box.

The ThoughtSpot appliance hardware will be installed in a rack in your data center. This section describes the typical physical configuration.

Hardware provided by ThoughtSpot

When your ThoughtSpot appliance arrives, the following items will be included:

Item Name	UOM	Qty
Round Hole to Sq Hole Adapter Kit (For Slide Rail Management)	Each	1
Power Cord, C13 to C14, 6 feet	Each	2
Power Cord, C13 to NEMA 5-15, 6 feet[1] **This power cord is not included with the Haswell platform.**	Each	2
Document, Rack Rail Installation, TS-2000	Each	1
TS-2000 Quick Start Guide	Each	1
Bezel Assembly, TS-2000	Each	1
Slide Rail Kit	Each	1
Appliance (containing 1-4 nodes, depending on ordered configuration)	Each	1
SFP+ Connector per ordered node (data connection)	Each	1
5m Fiber cable per ordered node (data connection)	Each	1
5m Network cable per ordered node (management connection)	Each	1

1: The supply voltage, 120 VAC, available when using a NEMA-15 power cord is an insufficient input to achieve the full power output required by the Haswell power supply. Only the C13 to C14 power cord should be used with the Haswell platform.

Additional hardware requirements

You must supply the following items, as they will not be included with your ThoughtSpot appliance:

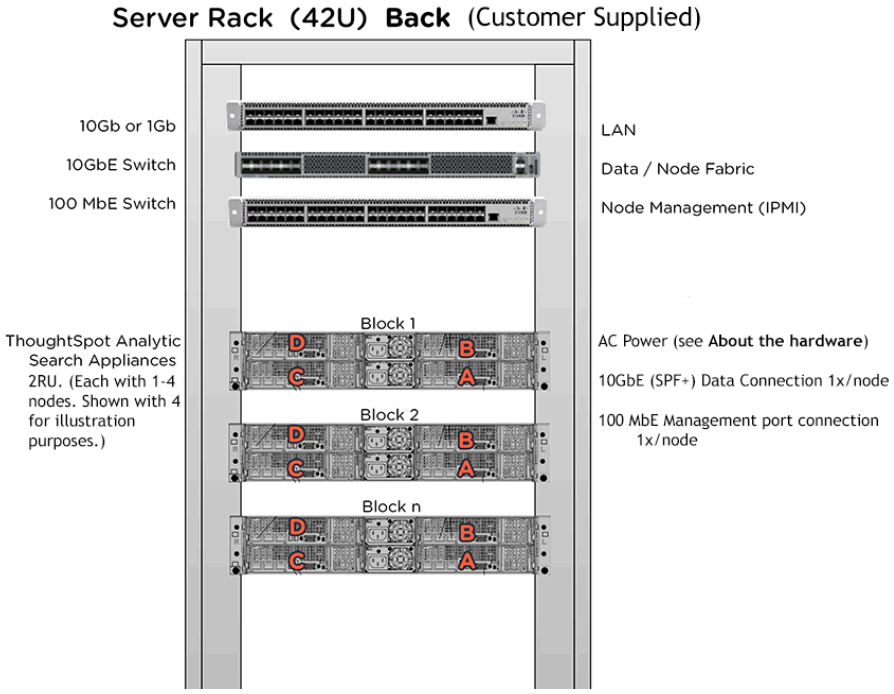
- Data center with proper cooling
- 2U of rack space per appliance (post depth 26.5” - 36.4”)
- AC power **Attention:** Refer to Hardware details for power input requirements.
- 10GbE infrastructure (switch) - 1x port required / node
- 100MbE infrastructure (switch) - 1x port required /node
- Network cable Cat 5e/6 (node management)[1]
- 10G connection: SFP+ for switch side[2]

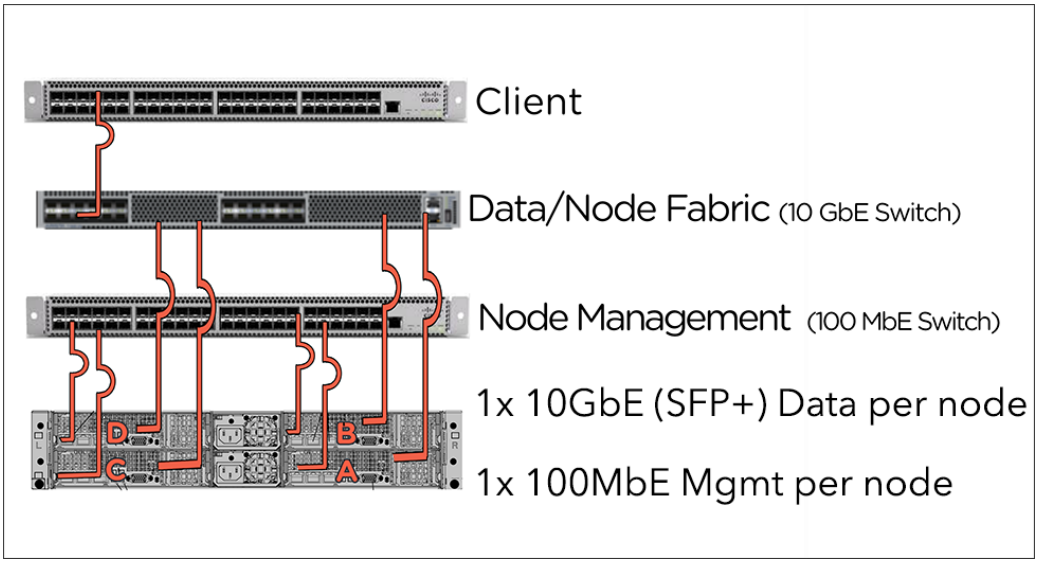
1. One 5m CAT 5e/6 network cable, per node, is provided with the appliance for management port connection. Customer supplied cable can be used if preferred.

2. One SFP+ connector is provided, per node, for the node side data connection. One 5m fiber cable is also provided. The customer must provide switch side SFP+ that is compatible with their switch. Customer supplied DAC cables or fiber cables can be used if preferred.

Typical physical deployment

These diagrams show a physical configuration with three blocks of four nodes each. Your appliance can have 1-4 nodes, depending on the ordered configuration.


Server Rack (42U) Back (Customer Supplied)
10Gb or 1Gb
10GbE Switch
100 MbE Switch
LAN
Data / Node Fabric
Node Management (IPMI)
Block 1
Block 2
Block n
ThoughtSpot Analytic Search Appliances 2RU. (Each with 1-4 nodes. Shown with 4 for illustration purposes.)
AC Power (see About the hardware)
10GbE (SPF+) Data Connection 1x/node
100 MbE Management port connection 1x/node


Client
Data/Node Fabric (10 GbE Switch)
Node Management (100 MbE Switch)
1x 10GbE (SFP+) Data per node
1x 100MbE Mgmt per node

Supported hardware

Summary: Required and provided installation hardware.

This section lists all required hardware that is needed to successfully install your ThoughtSpot appliance in your data center. Some hardware will be provided with your appliance, while the rest must be provided on-site.

The ThoughtSpot instance hardware is configured for fast data searching and reliability. This overview details the hardware specification and installation. The system is made up of compute nodes, which form a cluster. The 2U system includes up to 4 nodes and can hold up to 1TB of data. This can be scaled out.

Network connection

Before you can access ThoughtSpot, you need a network connection.

Refer to Network Ports in the Administrator's Guide to see which ports must remain open to outside traffic for handling certain network requests and for inter-cluster communication. The Administrator's Guide also provides information on network security and how to test your network connectivity between nodes.

Here are some more details on ports and node communication:

- Port redundancy (bonding) is not supported. Only one 10G port is active per node.
- Nodes communicate with each other through the 10G connection (data ports).
- All nodes should be on the same VLAN – ideally connected to the same top of rack switch.
- IPMI ports are used for management functions of the nodes.

Appliance hardware platforms

You can deploy the ThoughtSpot Analytical Search engine on Haswell appliance hardware platforms, with the following specifications:

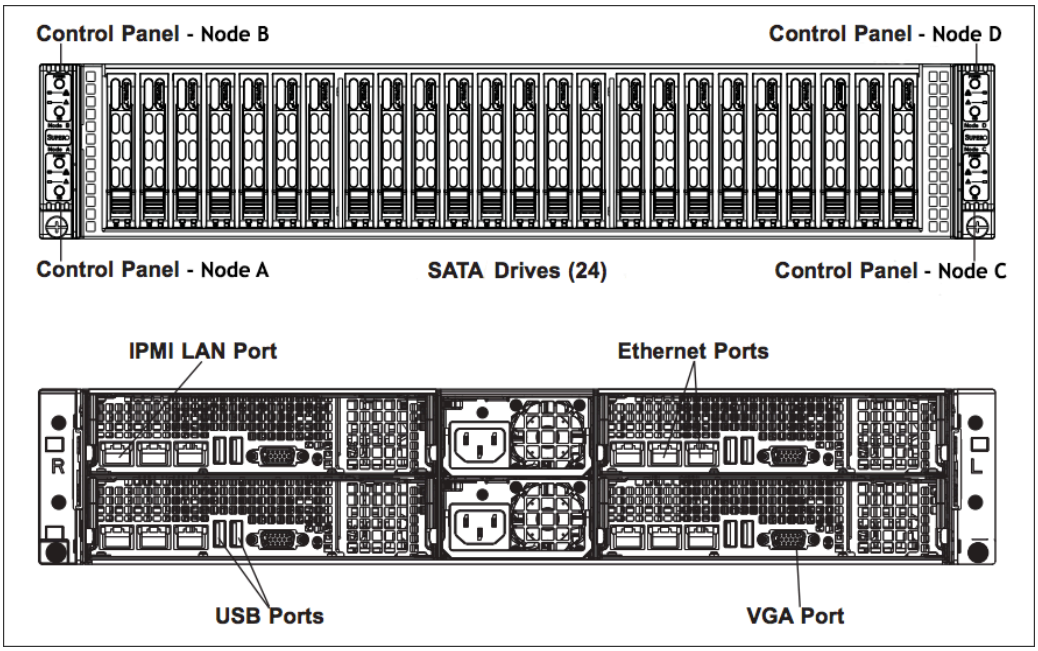
Details	Haswell
Dimensions	2 RU chassis (17.25 x 3.47 x 28.5 in.)
# of nodes	Populated with 1 to 4 nodes
Node specifications	Each node is independent and consists of a server board (removable from rear), 1x 200GB SSD, 3x 2TB HDD
Max power consumption	2000 W
Required power input	200-240 / 11.8 - 9.8A / 50-60Hz

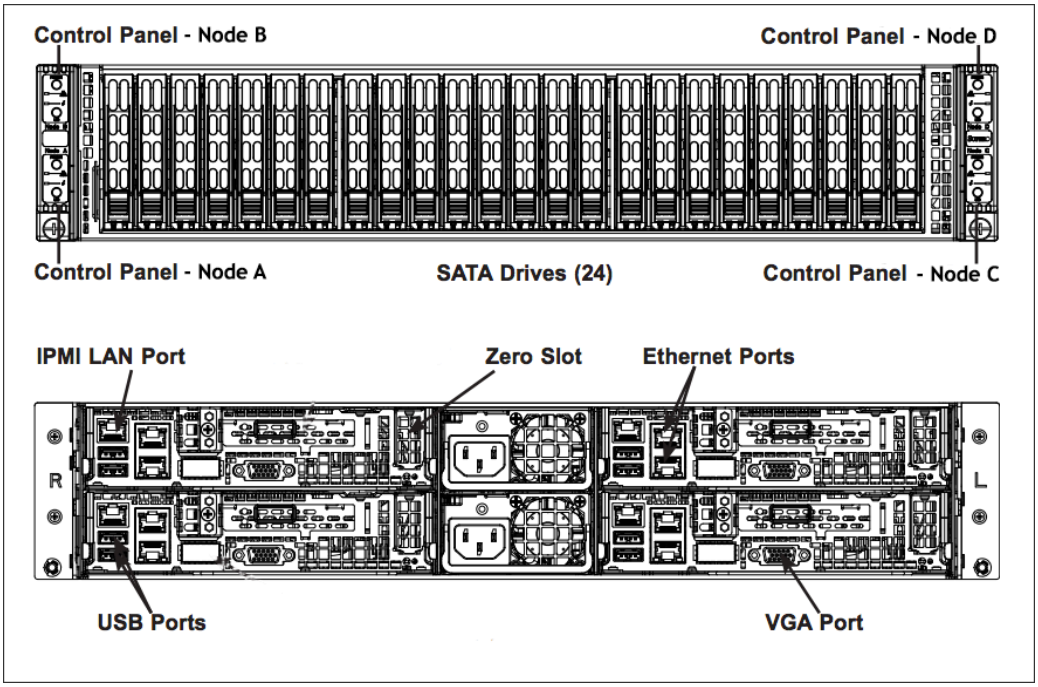
Note: ThoughtSpot deployments are no longer offered on Ivy Bridge platforms.

Chassis views

These diagrams show the front and rear chassis views. The marked features are present on all four nodes on the rear of the chassis even though they are only pointed out on one node in the diagrams.

The chassis appear fully populated (4-nodes). Your appliance may be populated with 1-4 nodes, depending on the ordered configuration. If less than 4-nodes were ordered, the empty slot will be filled with a filler panel.


Control Panel - Node B
Control Panel - Node D
Control Panel - Node A
SATA Drives (24)
Control Panel - Node C
IPMI LAN Port
Ethernet Ports
R
L
USB Ports
VGA Port


Control Panel - Node B
Control Panel - Node D
Control Panel - Node A
SATA Drives (24)
Control Panel - Node C
IPMI LAN Port
Zero Slot
Ethernet Ports
R
L
USB Ports
VGA Port

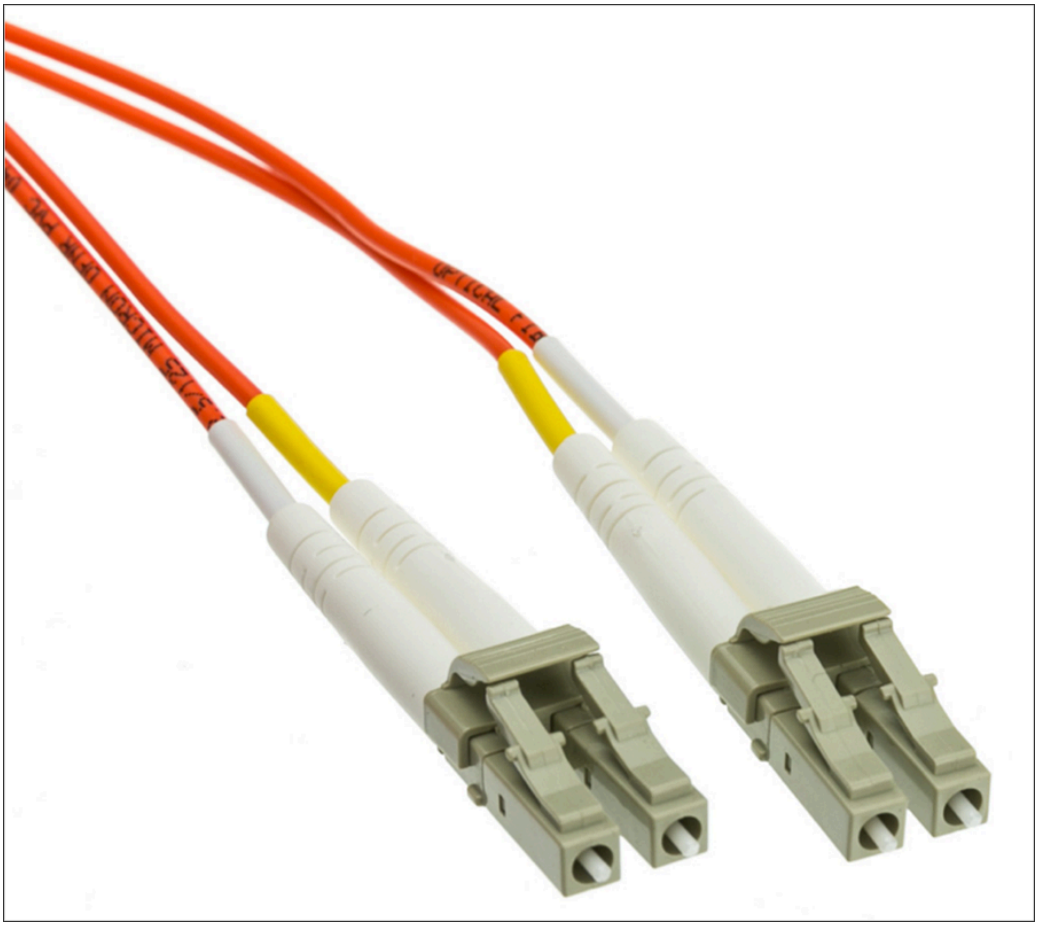
Cable networking

This section reviews the types of cables needed for 10GbE networking and how to plug them in. There are three types of cables to consider for 10GbE networking:

- Fiber
- Direct Attach Copper (DAC)
- Category 6a (not supported by ThoughtSpot)

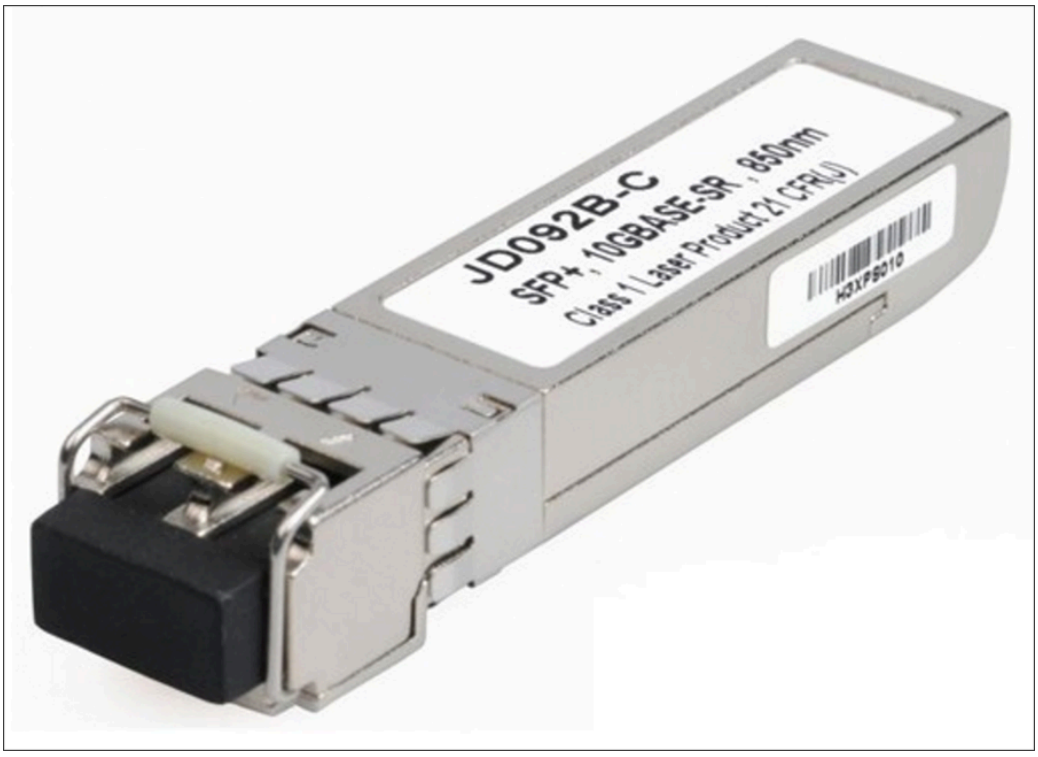
Option 1 - Fiber cables

Fiber can be run long distances to the switch.

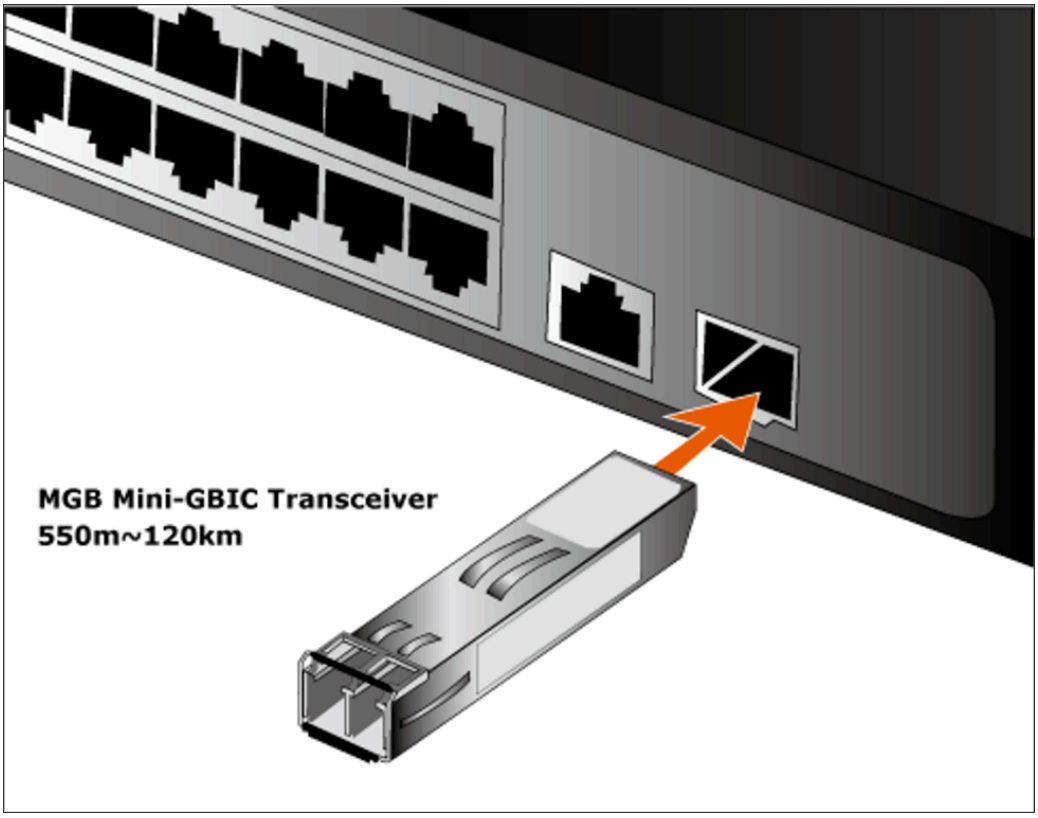


These cables require gigabit interface converters (GBICs), SFP+ form factor.

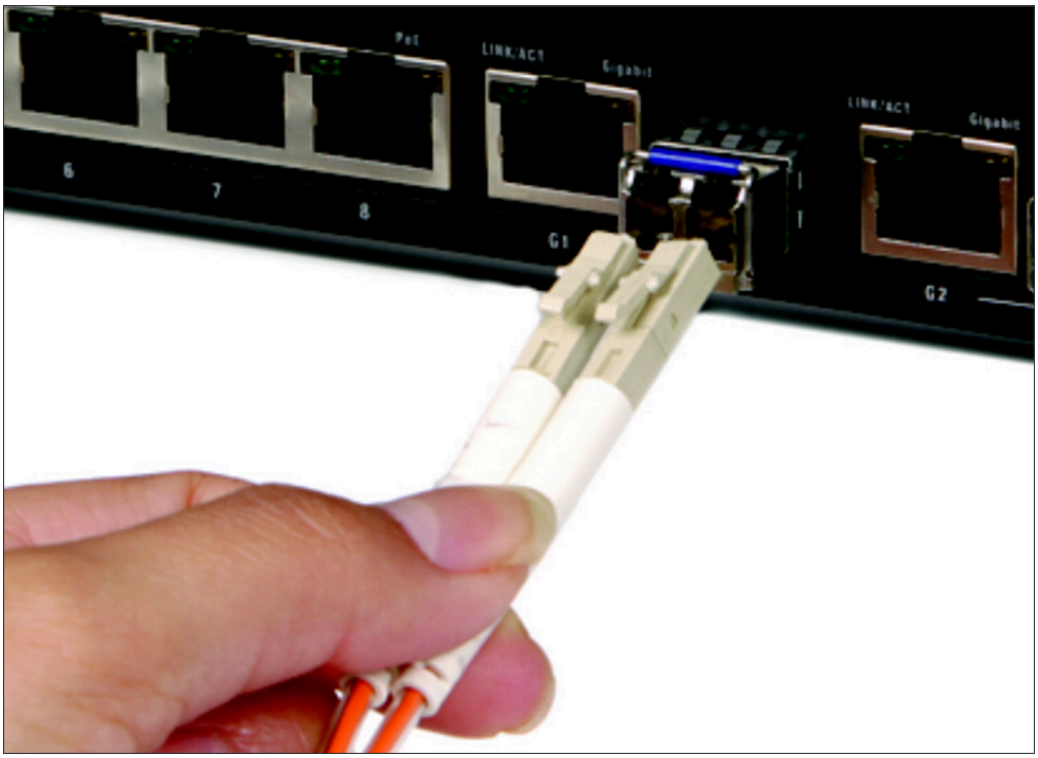
Remember: ThoughtSpot does not supply cables or GBICs



The GBIC must be plugged into a data port on the back of the appliance before plugging in the fiber cables.

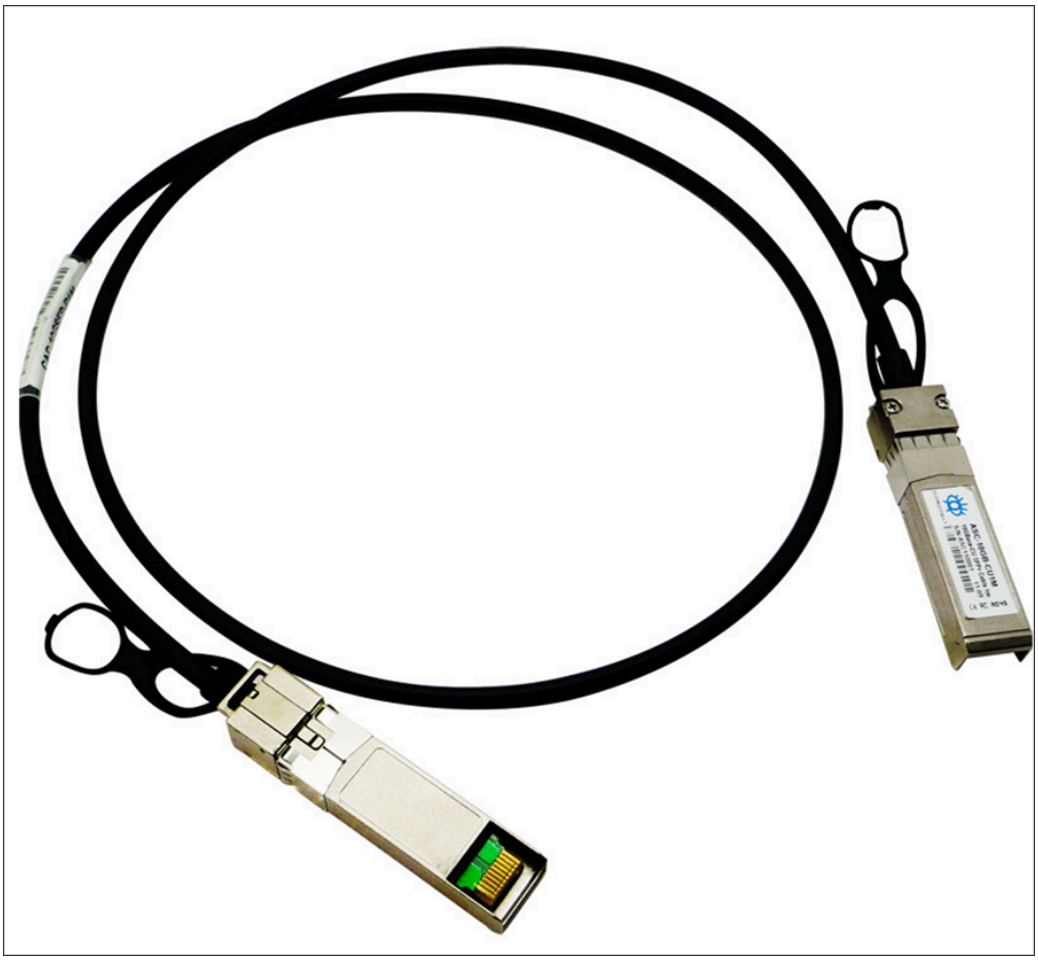


The fiber cables must then be plugged into the GBIC.

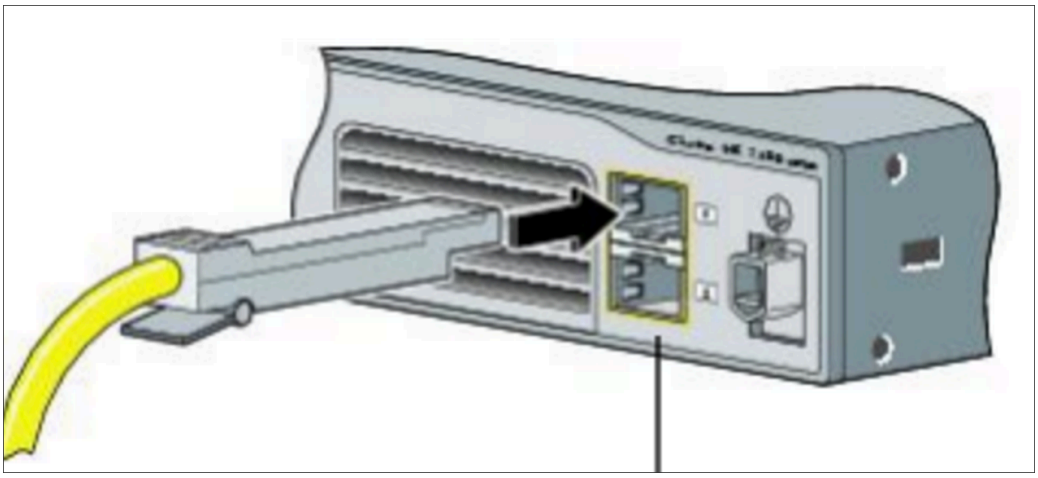


Option 2 - DAC/Twinax cables

Copper can only be run short distances to the switch. An SFP+ is attached to the cable.



Here is how you would plug in a DAC cable.



Non-option - Category 6a cables (not supported by ThoughtSpot)

There are no adapters for these cables. The 10GbE NIC (Network Interface Card) used on the ThoughtSpot appliance is not compatible with this type of cable/connection.

Setup and start the appliance

This section explains how to install and start the appliance.

Before you begin

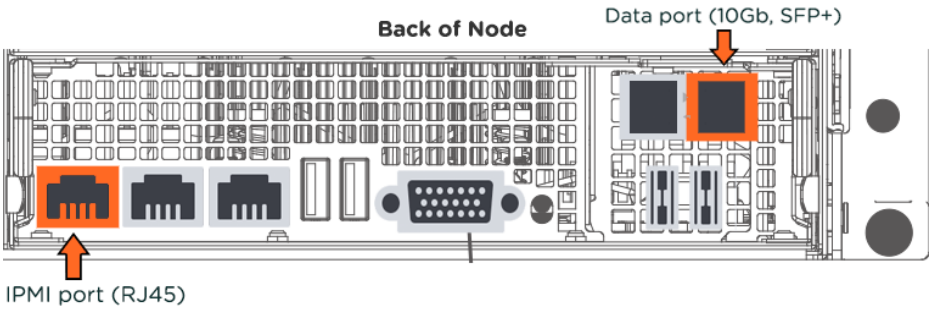
The ThoughtSpot appliance comes pre-installed with all the required software. Network settings on the appliance are required prior to using the appliance. Reference ThoughtSpot's site survey for the information specific to the customer's network environment that is required to configure the appliance.

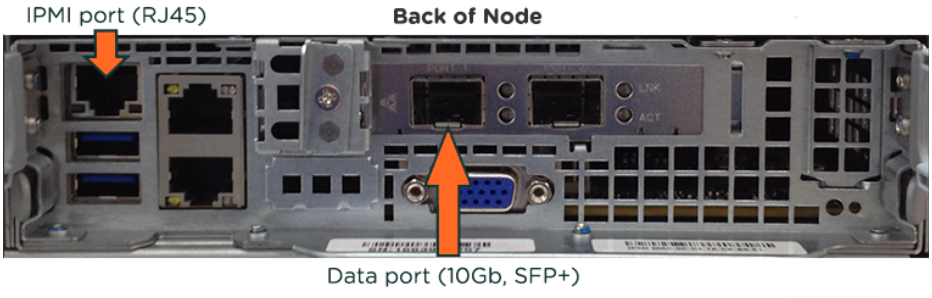
- If ThoughtSpot's site survey form was completed and returned to ThoughtSpot prior to the appliance being shipped, the appliance may be pre-configured for your network environment and ready to install and connect to your network.
- If the network configuration was not pre-set, then this step must be done as part of the installation process.
- If assistance is needed to determine the configuration status of the appliance, please contact ThoughtSpot Support.

Installation procedure

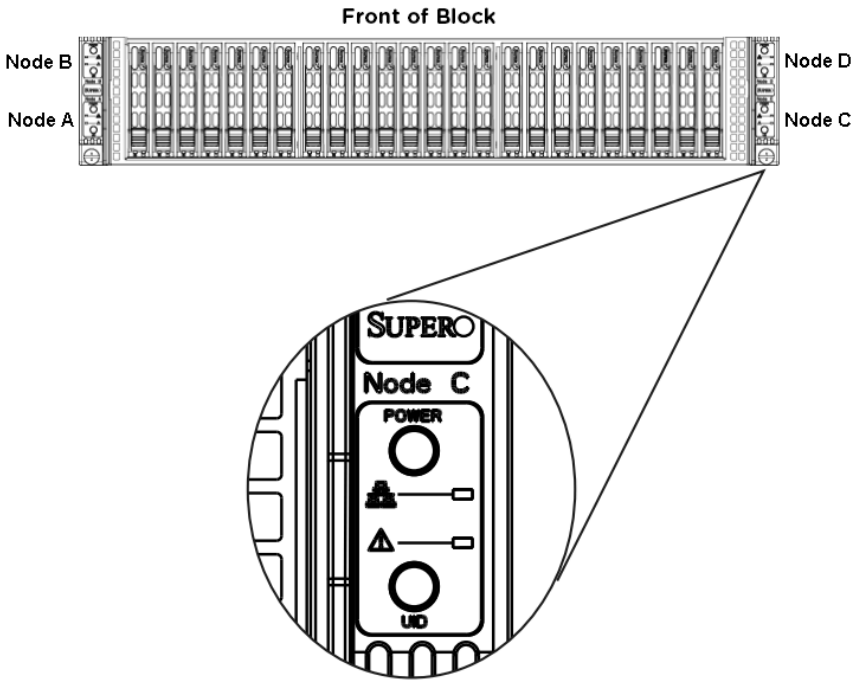
To install and start the appliance and connect to your network:

1. Refer to the Rack Install Guide to install the appliance securely in your data center.
2. Plug in the power cord, but do not turn the appliance on yet. See the figure of the **Location of the power and UID buttons on the control panel** for the power button location.
3. Connect the IPMI dedicated LAN port to a dedicated LAN for system management.
4. Connect the data port(s) on the back of the appliance to your 10GbE network switch. Only the one 10GbE port shown in the following **Back of Node** figures is active. Only one 10GbE port connection is needed.





5. Turn on the appliance by pressing and releasing the power button for each node and allow time for the nodes to boot up completely.



Each node has its own power and UID buttons. Turning the system off using the power button removes the main power, but keeps standby power supplied to the system. Therefore, you must unplug the AC power cord from any external power source before servicing. The power button for each node has a built-in LED which will turn green when the power is on.

There is also a UID button, which is used to turn on or off the blue light function of the LED. After the blue light is activated, the unit can be easily located in very large racks and server banks. A blue LED is also illuminated on the corresponding node, visible from the rear of the chassis.

6. After the appliance has been turned on, verify that both LEDs (IPMI and data NICs) on each network card are lit.
7. Connect a keyboard and monitor to each node in turn. You should see a login prompt on the screen. If you don't see one or the screen isn't responsive, press the key combination control, alt, and F2 on your keyboard, which should allow you to attempt to log in.
8. Log in as username admin, using the default password.
9. Run the following commands and capture the output at every stage:

```
 sudo ipmitool lan print 1
 sudo ipmitool lan set 1 ipsrc static
 sudo ipmitool lan set 1 defgw ipaddr <IPMI_GATEWAY_ADD
R>
 sudo ipmitool lan set 1 netmask <IPMI_VLAN_SUBNET_MASK>
 sudo ipmitool lan set 1 ipaddr <IPMI_NIC_IP_ADDR>
 sudo ipmitool lan print 1
```

10. Run and capture the output of the following commands as well:

```
ifconfig eth0
ifconfig eth1
ifconfig eth2
ifconfig eth3
sudo ethtool eth0
sudo ethtool eth1
sudo ethtool eth2
sudo ethtool eth3
```

11. Share the output of all commands with the ThoughtSpot team who will then determine the next steps.

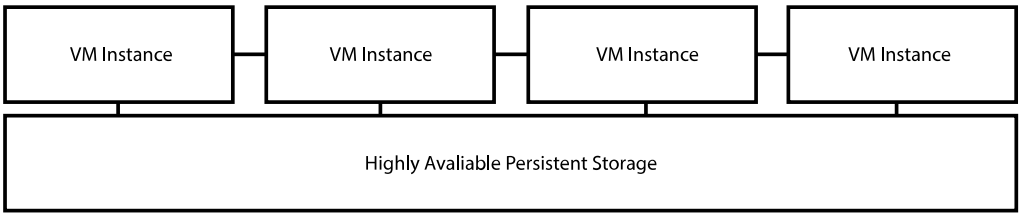
Cloud overview

ThoughtSpot can currently be deployed in the following cloud provider environments:

- Amazon Web Services (AWS) EC2
- Microsoft Azure
- Google Cloud Platform (GCP)

The ThoughtSpot cloud deployment consists of cloud compute (VM) instances and an underlying persistent storage layer. The number of instances required for a cloud deployment is based on the size of the data that needs to be analyzed in ThoughtSpot. The instances act as a distributed cluster of nodes to serve query responses.

	AWS	Azure	GCP
Compute	Virtual machines deployed in your AWS VPC	Virtual machines in your Azure VNET	Virtual machines in your GCP VPC
Persistent storage	Deployment options: 1. Elastic Block Storage 2. S3 + Elastic Block Storage	Premium SSD Managed Disks	Zonal SSD persistent disk



To determine the number of instances and the persistent storage requirements to provision your cluster, please refer to the available instance types for your cloud service provider in the next section.

ThoughtSpot cloud instance types

Refer to the following guidelines for how to set up ThoughtSpot on each cloud service:

- AWS instance types
- Azure instance types

- GCP instance types

Reducing your cloud infrastructure costs

ThoughtSpot recommends following these guidelines to help reduce the cost of your cloud deployment.

Use small and medium instance types when applicable

For ThoughtSpot customers who are deploying their instance with lower data sizes (<=100 GB), ThoughtSpot supports “small” (20 GB data) and “medium” (100 GB data) instance types, as provided at the links above, to help reduce the costs of cloud infrastructure. These are instances with lower CPU/RAM sizes (16/32 vCPU and 128 GB/256 RAM). Advanced lean configuration is required before any data can be loaded onto these instances.

Please contact ThoughtSpot support for assistance with this configuration.

Shut down and restart your cluster

If you do not need your ThoughtSpot cluster to be up and running 24/7, you can shut down your cluster and restart it during normal usage hours to save on the infrastructure costs of running ThoughtSpot instances in cloud provider environments.

To shut down and restart your cluster, do the following in the tscli:

1. Ensure there are no issues with the cluster by running: `$ tscli cluster check`

   The above command should return no failure messages.

2. Stop the cluster by running: `$ tscli cluster stop`

   Wait until you see the message: “Done stopping cluster”

3. Go to your cloud provider’s console and shut down all of the ThoughtSpot VMs in your cluster.

4. When you are ready to use ThoughtSpot again, start up your node VMs.

5. Restart your cluster by running: `$ tscli cluster start`

You should see the message: ”Started pre-existing cluster”

Depending on the size of your cluster, you may need to wait several minutes before the system is up and running. Make sure you budget for this startup time to ensure that the system is fully operational before you expect people to use it.

6. Ensure that your cluster is ready for use by running: `$ tscli cluster status`

   The following messages are displayed to indicate your cluster is up and running:

   `... Cluster: RUNNING`

   `Database: READY`

   `Search Engine: READY`

Automating your cloud deployment

You can automate your deployment, using the free tools in the ThoughtSpot Cloud Deployment GitHub repository.

For more information about automating your cloud deployment, read Deploying ThoughtSpot in the Cloud Using Terraform and Ansible.

Installing Amazon Web Services

Summary: Learn how to install Amazon Web Services (AWS).

Installation Prerequisites

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

1. **Review configuration options** Refer to AWS configuration options for detailed instance specs.
2. **Create the instance** Refer to Set up AWS for ThoughtSpot to create and launch your instance.
3. **Review required ports** Refer to Network Policies 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

Log into your cluster with admin credentials from Terminal on a Mac or a terminal emulator on Windows.

1. Run `ssh admin@clusterIP` or `ssh admin@hostname`, replacing ‘clusterIP’ or ‘hostname’ with your specific network information.
2. Enter your admin password.
    - Ask your network administrator if you don’t know the password.

```
$ ssh admin@clusterIP
```

Step 2: Get a list of nodes to configure

Run the `tscli cluster get-config` command to get a list of the nodes to configure for the new cluster, and redirect it to the file `nodes.config`. You can find more information on this process in the nodes.config file reference.

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

Step 3: Configure the network of nodes

1. Add your specific network information for the nodes in the `nodes.config` file, as demonstrated in the autodiscovery of one node example.
2. Fill in the areas specified in Parameters of the `nodes.config` file 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`. 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.

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.

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.

```
$ 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 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>`.
    - If you are using an s3 bucket for object storage, include the flag `--enable_cloud_storage s3`.

```
 $ tscli cluster create 6.0.tar.gz --enable_clou
d_storage s3
```

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

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).

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, enter admin credentials, and click **Sign in**. ThoughtSpot recommends changing the default admin password.

References

Use these references for successful installation and administration of ThoughtSpot.

- The `nodes.config` file
- Parameters of the `nodes.config` file
- Using the `cluster create` command
- Parameters of the `cluster create` command
- ThoughtSpot Documentation
- Contact Support

AWS configuration options

Summary: Your specific instances require specific configurations of memory, CPU, storage, and networking capacity.

ThoughtSpot can be deployed in your AWS environment by deploying compute (VM) instances in your Amazon VPC as well as an underlying persistent storage infrastructure. Currently two configuration modes are supported by ThoughtSpot:

- Mode 1: Compute VMs + EBS-only persistent storage
- Mode 2: Compute VMs + EBS and S3 persistent storage

The cost of infrastructure for deploying ThoughtSpot is cheaper when using S3. However, there are differences in where data is loaded, as well as in the backup and restore procedure. For assistance in choosing the best mode for your organization, contact your ThoughtSpot representative. For more information on purchasing ThoughtSpot on AWS, see: ThoughtSpot Pricing.

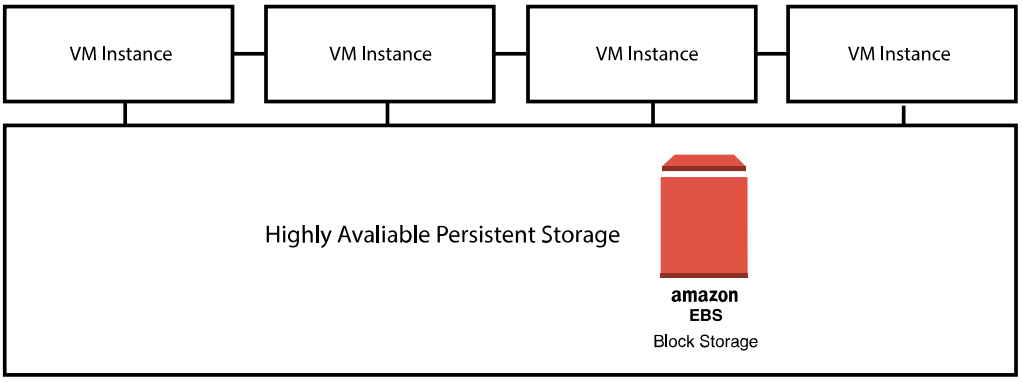
All AWS VMs in a ThoughtSpot cluster must be in the same availability zone (and therefore, also in the same region). ThoughtSpot does not support deploying VMs in the same cluster across availability zones. For more information, see Regions and Availability Zones in Amazon's AWS documentation.

ThoughtSpot AWS instance types

The following sections contain the supported and recommended instance types for a ThoughtSpot AWS deployment. When setting up your cluster in AWS, use the information here to select an instance type, configure the number of instances required for the storage you need, and add data volumes to your cluster.

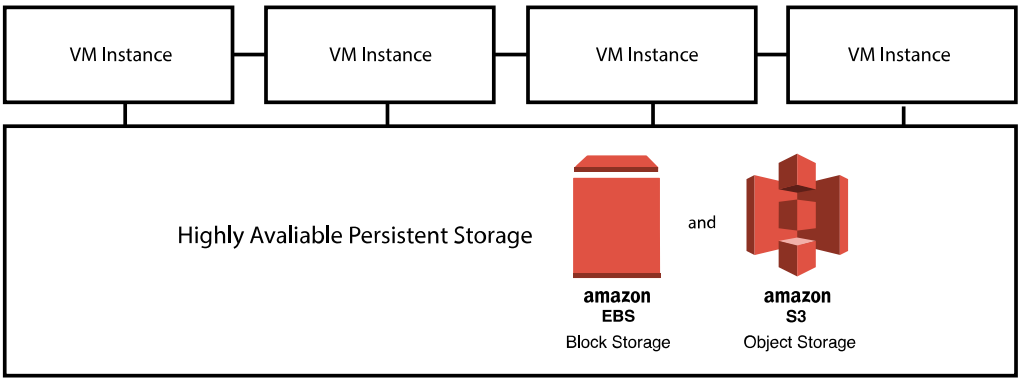
For example: If you were deploying a total cluster data size of 1 TB using the standard r5.16xlarge instance type, you would need 4 instances (VMs), because the per-VM user data capacity of that instance type is 250 GB. If you were deploying EBS-only data volumes, you would need 2x1 TB data volumes per VM.

VMs with EBS-only persistent storage



Per VM user data capacity	Instance type	CPU/RAM	Recommended per-VM EBS volume
20 GB	r4.4xlarge, r5.4xlarge	16/122, 16/128	2X 400 GB
100 GB	r4.8xlarge, r5.8xlarge	32/244, 32/256	2X 400 GB
192 GB	m5.24xlarge	96/384	2X 1 TB
250 GB	r4.16xlarge, r5.16xlarge	64/488, 64/512	2x 1 TB
384 GB	r5.24xlarge	96/768	2X 1.5 TB

VMs with EBS and S3 persistent storage



Per VM user data capacity	Instance type	CPU/RAM	Recommended per-VM EBS volume
20 GB	r4.4xlarge, r5.4xlarge	16/122, 16/128	1x 500 GB
100 GB	r4.8xlarge, r5.8xlarge	32/244, 32/256	1x 500 GB
192 GB	m5.24xlarge	96/384	1x 500 GB
250 GB	r4.16xlarge, r5.16xlarge	64/488, 64/512	1x 500 GB
384 GB	r5.24xlarge	96/768	1x 500 GB

Note: The S3 bucket size is approximately equal to the size of the user data.

Related information

- EC2 instance types
- EC2 pricing
- EBS pricing
- Placement groups

Set up AWS for ThoughtSpot

Summary: After you determine your configuration options, you must set up your virtual machines (VMs) on AWS using a ThoughtSpot Amazon Machine Image (AMI).

Overview of ThoughtSpot setup in AWS

The high-level process for setting up ThoughtSpot in AWS involves these steps:

1. Gain access to ThoughtSpot AMIs.
2. Choose a VM instance configuration recommended by ThoughtSpot.
3. Set up your Amazon S3 bucket (optional).
4. Set up your ThoughtSpot cluster in AWS.
5. Contact ThoughtSpot to finish setting up your cluster.
6. Open the required network ports for communication for the nodes in your cluster and end users.

About the ThoughtSpot AMI

An Amazon Machine image (AMI) is a preconfigured template that provides the information required to launch an instance. You must specify an AMI when you launch an instance in AWS.

To make deployment easy, the ThoughtSpot AMI includes a custom ThoughtSpot image, with the following components:

- A template for the root volume for the instance, such as an operating system, an appliance server, and applications.
- Launch permissions that control which AWS accounts can use the AMI to launch instances.
- A block device mapping that specifies the volumes to attach to the instance when it launches.

The ThoughtSpot AMI has specific applications on a CentOS base image. The AMI includes the EBS volumes necessary to install ThoughtSpot in AWS. When you launch an EC2 instance from this image, it automatically sizes and provisiones the EBS volumes. The base AMI includes 200 GB (xvda), 2X400 GB (xvdb), and SSD (gp2). It contains the maximum number of disks to handle a fully loaded VM.

Prerequisites

To install and launch ThoughtSpot, you must have the following:

- Familiarity with Linux administration, and a general understanding of cloud deployment models.
- The necessary AWS Identity and Access Management (IAM) users and roles assigned to you to access and deploy the various AWS resources and services as defined in the Required AWS components section that follows.

  For more information about IAM, see: What Is IAM? in Amazon's AWS documentation.

Required AWS components

- An AWS VPC. For details, see VPC and Subnets in Amazon's AWS documentation.
- A ThoughtSpot AMI. For details, see the next section.
- AWS security groups. For required open ports, see network policies.
- AWS VM instances. For instance type recommendations, see ThoughtSpot AWS instance types.
- EBS volumes.
- (Optional) If deploying with S3 persistent storage, one S3 bucket dedicated to each ThoughtSpot cluster.

Guidelines for setting up your EC2 instances

- Sign in to your AWS account.
- Copy the following ThoughtSpot public AMI which has been made available in N. California region to your AWS region:

  AMI Name: thoughtspot-image-20190718-dda1cc60a58-prod

  AMI ID: ami-0b23846e4761375f1

  Region: N. California

  > ℹ **Note:** The AMI is backward-compatible with ThoughtSpot releases 5.1.x - 5.2.x.

- Choose the appropriate EC2 instance type: See ThoughtSpot AWS instance types for supported instance types.
- Networking requirements: 10 GbE network bandwidth is needed between the VMs. This is the

default for the VM type recommended by ThoughtSpot.

- Security: The VMs that are part of a cluster need to be accessible by each other, which means they need to be on the same Amazon Virtual Private Cloud (VPC) and subnetwork. Additional external access may be required to bring data in/out of the VMs to your network.
- Number of EC2 instances needed: Based on the datasets, this number will vary. Please check ThoughtSpot AWS instance types for recommended nodes for a given data size.
- Staging larger datasets (> 50 GB per VM), may require provisioning additional attached EBS volumes that are SSD (gp2).

Setting up your Amazon S3 bucket

If you are going to deploy your cluster using the S3-storage option, you must set up that bucket before you set up your cluster. Contact ThoughtSpot Support to find out if your specific cluster size will benefit from the S3 storage option.

To set up an Amazon S3 bucket in AWS, do the following:

1. In AWS, navigate to the S3 service dashboard by clicking **Services**, then **S3**.
2. Make sure the selected region in the top-right corner of the dashboard is the same region in which you plan to set up your cluster.
3. Click **Create bucket**.
4. In the **Name and region** page, enter a name for your bucket, select the region where to set up the cluster, and click **Next**.
5. On the **Properties** page, click **Next**.
6. On the Configure options page, make sure **Block *all* public access** is selected and click **Next**.
7. On the Set permissions page, click **Create bucket**.

Encrypting your data at rest on Amazon EBS or S3 in AWS

ThoughtSpot makes use of EBS for the data volumes to store persistent data (in the EBS deployment model) and the boot volume (in the EBS and S3 deployment models). ThoughtSpot recommends that you encrypt your data volumes prior to setting up your ThoughtSpot cluster. If you are using the S3 persistent storage model, you can encrypt the S3 buckets using SSE-S3. ThoughtSpot does not currently support AWS KMS encryption for AWS S3.

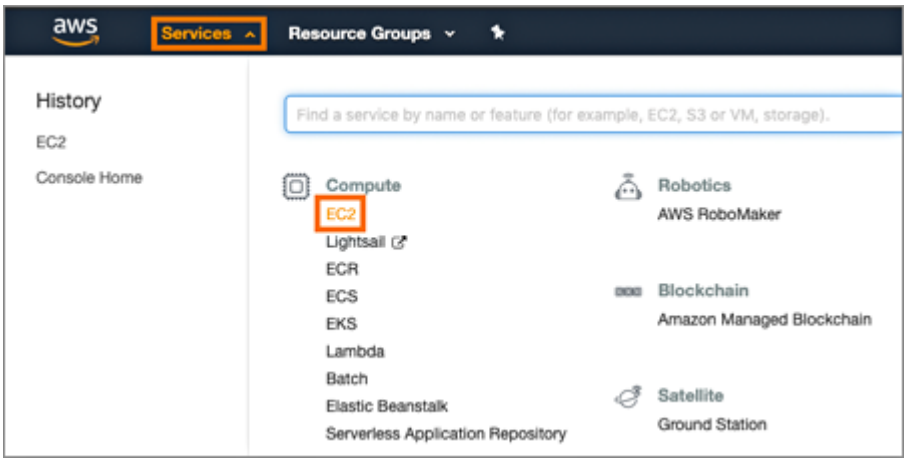
For more information on encryption supported with AWS:

- For EBS, see Amazon EBS Encryption in Amazon's AWS documentation.
- For S3, see Amazon S3 Default Encryption for S3 Buckets in Amazon's AWS documentation.

Setting up your ThoughtSpot cluster in AWS

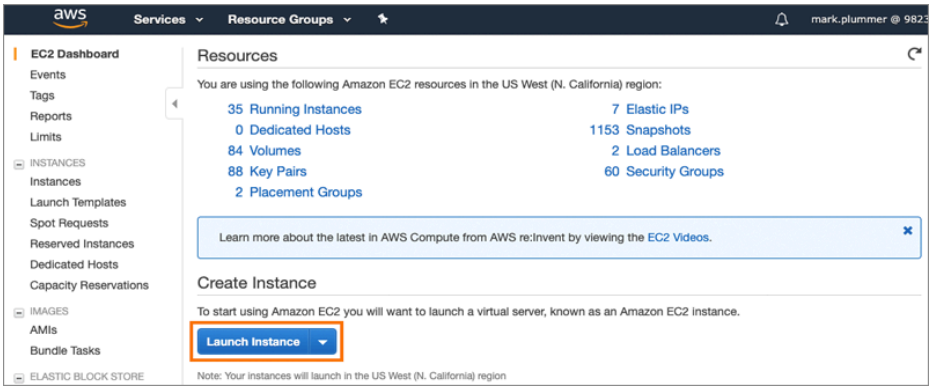
To set up a ThoughtSpot cluster in AWS, do the following:

1. In AWS, navigate to the EC2 service dashboard by clicking **Services**, then **EC2**.



2. Make sure your selected region is correct in the top-right corner of the dashboard. If not, select a different region you would like to launch your instance in. Let ThoughtSpot support know if you change your region.

3. Start the process of launching a VM by clicking **Launch Instance**.



4. Click the **My AMIs** tab, find the ThoughtSpot AMI from the list, and click **Select**.

5. On the Choose an Instance Type page, select a ThoughtSpot-supported instance type. (See ThoughtSpot AWS instance types.)

6. Click **Next: Configure Instance Details**.

7. Configure the instances by choosing the number of EC2 instances you need. The instances must be on the same VPC and subnetwork. ThoughtSpot will set up the instances to be in the same ThoughtSpot cluster.

   S3 storage setting: If you are going to use the S3 storage option, you must go to the **IAM role** menu and select **ec2rolewithfulls3access**. This setting gives your instance access to all S3 buckets in your account's region. If you want to restrict the access to a specific bucket, you must create a new IAM role that provides access to the specific bucket, and select it instead. For details on that, click **Create new IAM role**.

8. Click **Next: Add Storage**. Add the required storage based on the storage requirements of the instance type you have selected, and the amount of data you are deploying. For specific storage requirements, refer to ThoughtSpot AWS instance types.

9. When you are done modifying the storage size, Click **Next: Add Tags**.

10. Set a name for tagging your instances and click **Next: Configure Security Group**.

11. Select an existing security group to attach new security groups to so that it meets the security requirements for ThoughtSpot.

    ☑ **Tip: Security setting for ThoughtSpot**

    - The VMs need intragroup security, i.e. every VM in a cluster must be accessible from one another. For easier configuration, ThoughtSpot recommends that you enable full access between VMs in a cluster.
    - Additionally, more ports must be opened on the VM to provide data staging capabilities to your network. Check Network policies to determine the minimum required ports that must be opened for your ThoughtSpot appliance.

12. Click **Review and Launch**. After you have reviewed your instance launch details, click **Launch**.

13. Choose a key pair. A key pair consists of a public and private key used to encrypt and decrypt login information. If you don't have a key pair, you must create one, otherwise you won't be able to SSH into the AWS instance later on.

14. Click **Launch Instances**. Wait a few minutes for it to fully start up. After it starts, it will appear on the EC2 console.

Prepare the VMs (ThoughtSpot Systems Reliability Team)

⚠ **Important:** This procedure is typically done by a ThoughtSpot Systems Reliability Engineer (SRE). Please consult with your ThoughtSpot Customer Service or Support Engineer on these steps.

Before we can install a ThoughtSpot cluster, an administrator must log into each VM through SSH as user "admin", and complete the following preparation steps:

1. Run `sudo /usr/local/scaligent/bin/prepare_disks.sh` on every machine.
2. Configure each VM based on the site-survey.

When complete, your storage is mounted and ready for use with your cluster.

Launch the cluster

Upload the TS tarball to one of the VMs and proceed with the normal cluster creation process, using tscli cluster create.

If you are going to use S3 as your persistent storage, you must enable it when running this command, using the **enable_cloud_storage** flag. Example: `tscli cluster create 6.0-167.tar.gz --enable_cloud_storage=s3a`

When the setup is complete, you can load data into ThoughtSpot for search analytics.

Open the required network ports

To determine which network ports to open for a functional ThoughtSpot cluster, see Network policies.

Related information

EC2 Best Practices

Loading data from an AWS S3 bucket

Set up high availability for AWS

Summary: This article explains how to set up High Availability (HA) for your ThoughtSpot cluster using the AWS Elastic File System (EFS)

Configure high availabiity

To set up High Availability (HA) for your ThoughtSpot cluster using the AWS Elastic File System (EFS), follow these steps:

1. Create an EFS File System that spans across different availability zones, and across different subnets.

2. Create two ThoughtSpot clusters in each availability zone and in the subnets, where the file system was created.

3. Change the IP addresses of the cluster, if necessary.

4. Create an EFS directory in the `/home/admin` path, and issue the following command to mount the previously created file system.

   Modify the fields as necessary for your installation.

   ```
   sudo mount -t nfs -o nfsvers=4.1,rsize=1048576,wsize=104
   8576,hard,timeo=600,retrans=2,
                        noresvport fs-f756f1ee.efs.us-wes
   t-1.amazonaws.com:/ /home/admin/efs/
   ```

   To ensure that all clusters with EFS mount points have read and write permissions, modify permissions:

   ```
   chmod 777 /home/admin/efs
   ```

5. On the first cluster, create a snapshot on to the EFS mount point, and backup it.

```
tscli snapshot create EfsTest HA 2
tscli backup create --mode full --type full
                    --storage_type local EfsTest /home/a
dmin/efs/Efs-backup
```

6. Ensure that the backup is successful, and that it can be accessed from all clusters where EFS is mounted.

7. Take down the first cluster instances.

8. On the second cluster, delete the existing cluster, and create a new one by restoring from the first cluster backup. This is accessible from the EFS mount point.

```
tscli cluster restore /home/admin/EFS/Efs-backup
```

Your cluster should now be successfully restored to the second cluster from the backup on the EFS, achieving HA for ThoughtSpot clusters.

Replace a cluster

For information on how to recover from infrastructure failure scenarios, see: Cluster replacement.

Note: At this time, ThoughtSpot does not support AWS Auto Scaling or deployment across AWS availability zones or regions.

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 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? 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 Azure for ThoughtSpot

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 on 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.

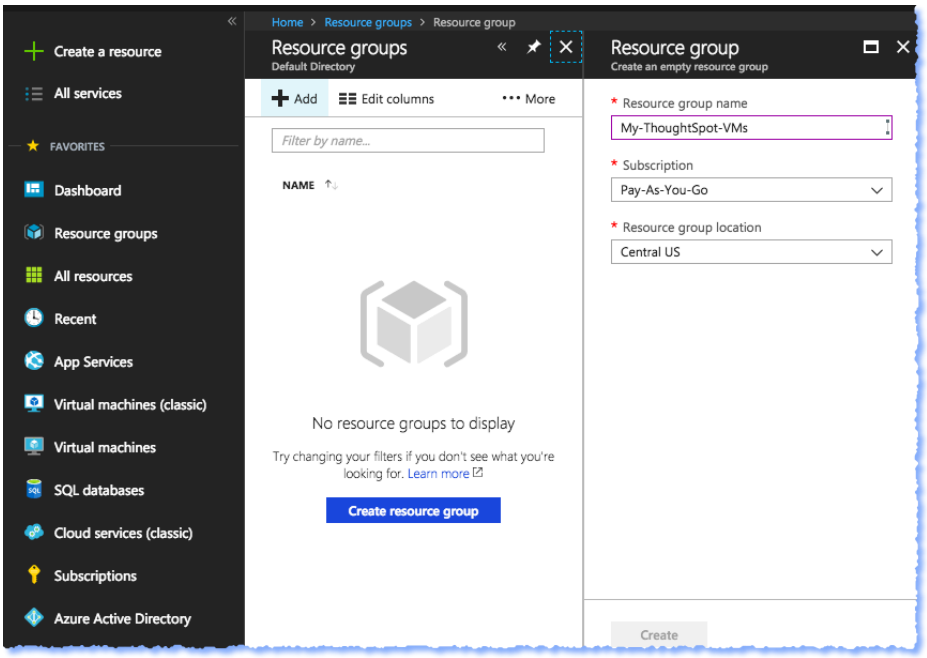
Create an instance

To get started, you need to log into the Azure portal, create a resource group, get the ThoughtSpot Virtual Machine on the Azure Marketplace, 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.

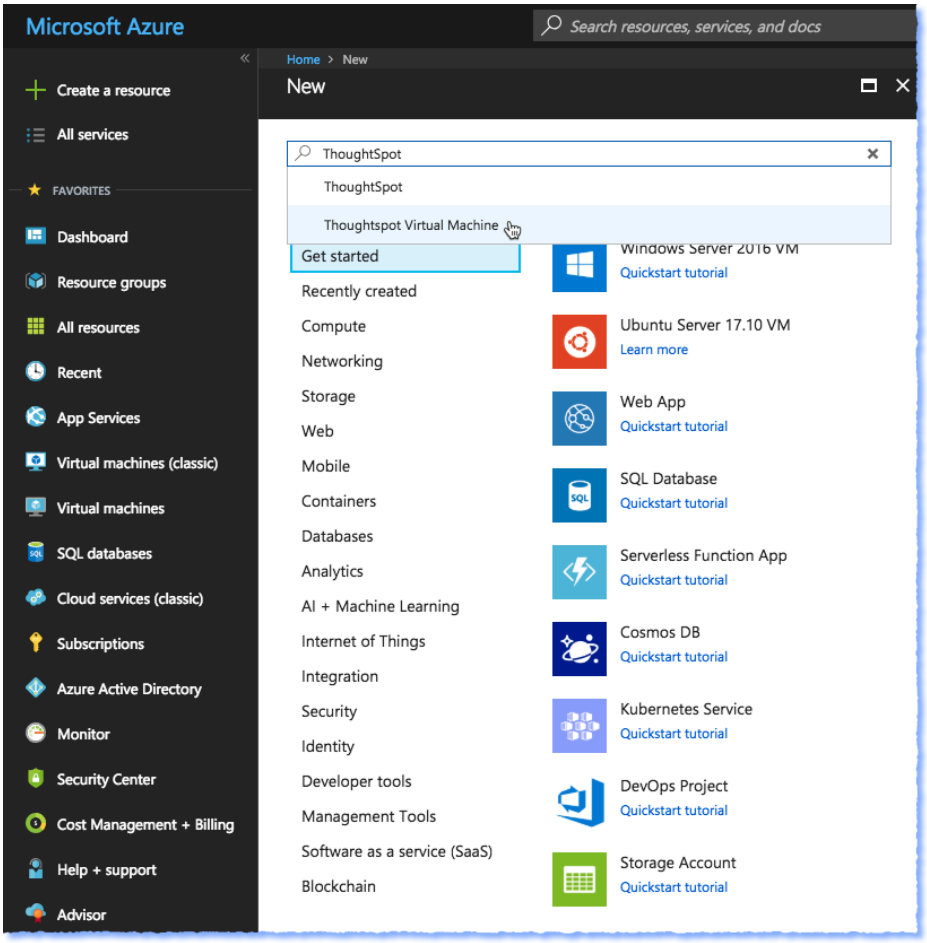
   In a browser, go to http://azure.microsoft.com, and log in to your Azure account.

2. Create a Resource Group.

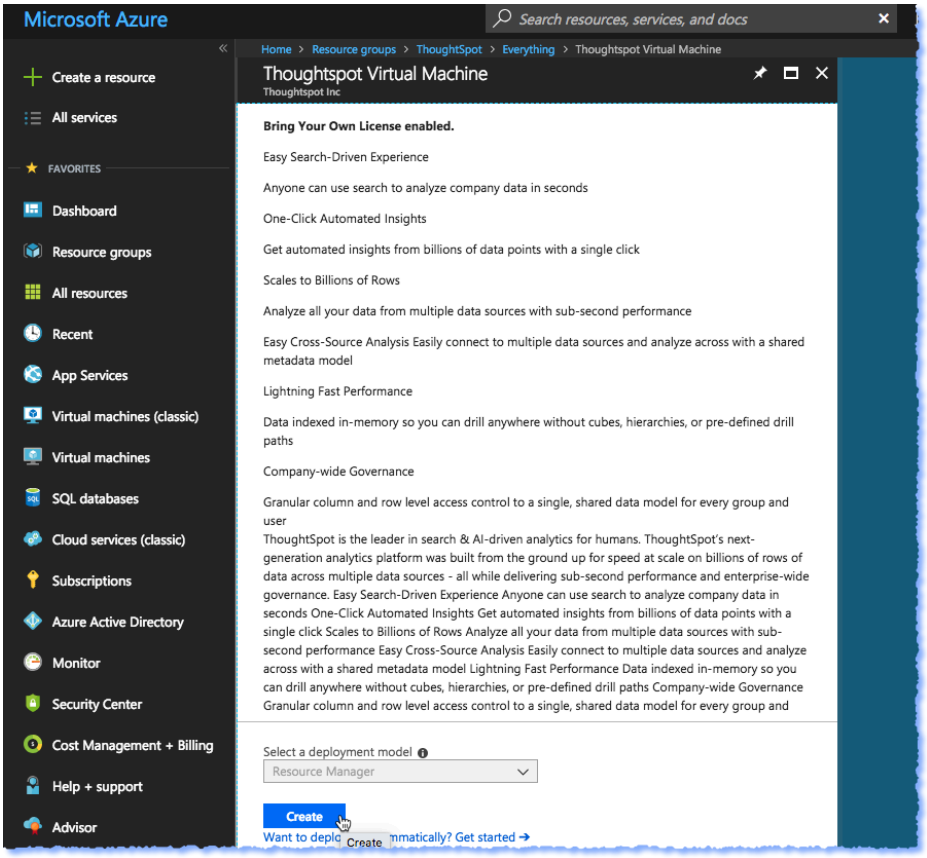


3. Next, create a resource based on the ThoughtSpot Virtual Machine.

   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.

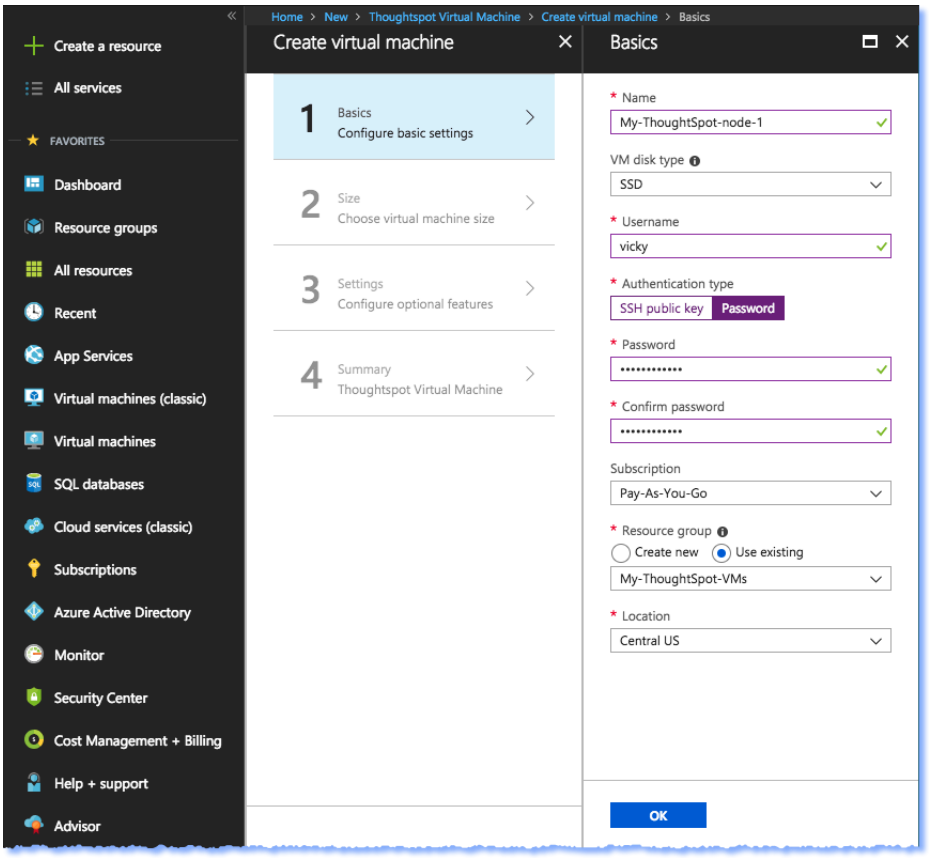
   > ☑ **Tip:** the new SSD disk types are currently available for only particular regions, so if you choose this disk type, make sure it's supported on the region you chose for your VM.

   See Standard SSD Disks for Virtual Machine workloads for more on SSD disks.

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

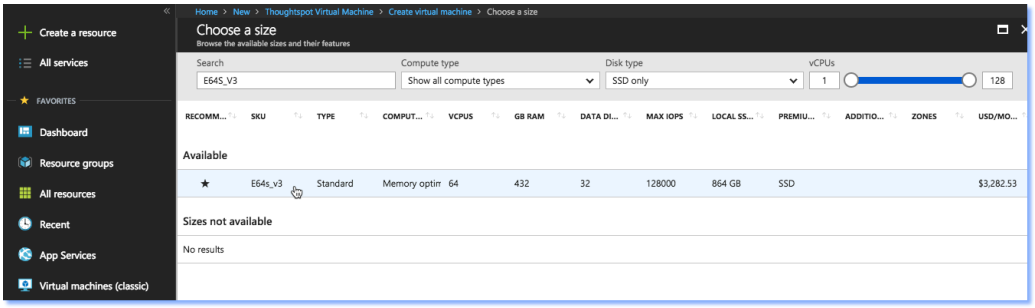
4. Select a location.

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.

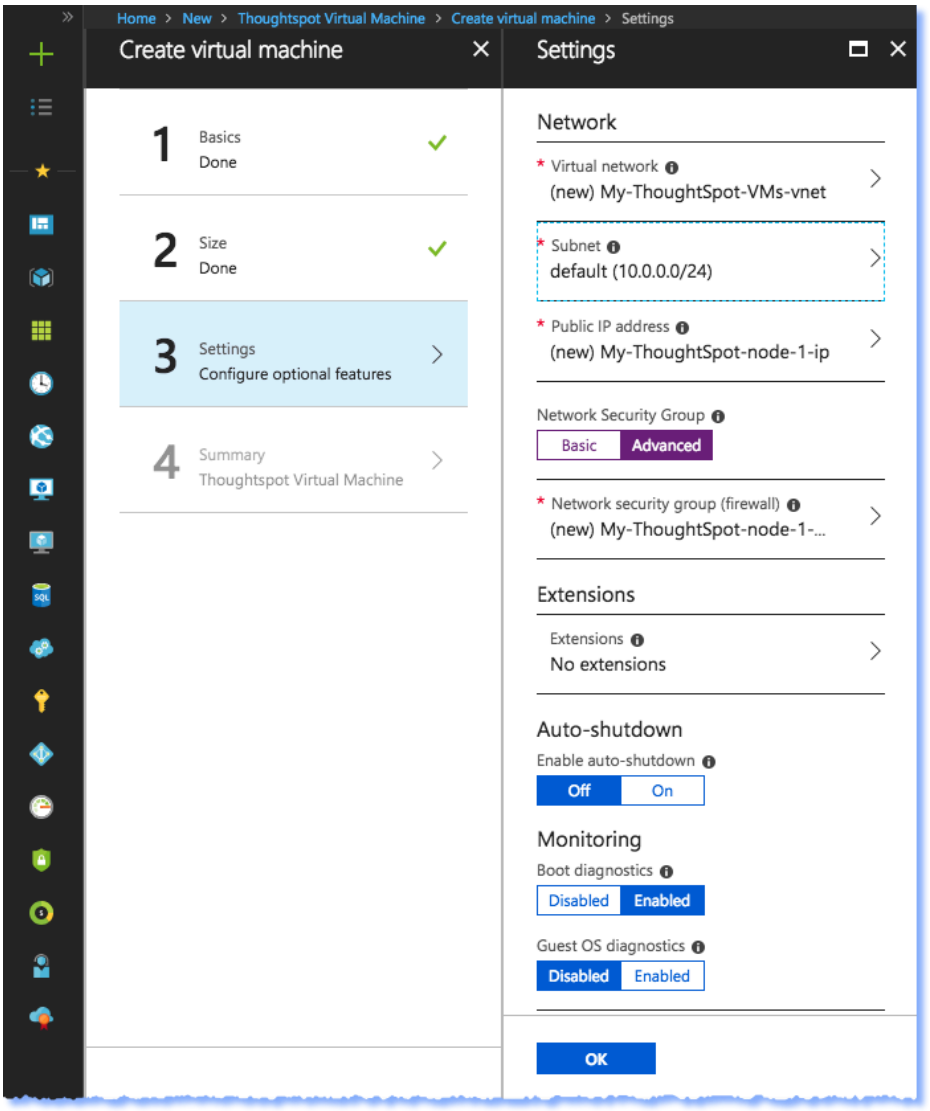
3. Open the necessary Inbound and Outbound ports to ensure that the ThoughtSpot processes do not get blocked.

   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.

GCP configuration options

ThoughtSpot can be deployed in your GCP environment by deploying compute (VM) instances in your VPC as well as an underlying persistent storage infrastructure. Currently two configuration modes are supported by ThoughtSpot:

- Mode 1: Compute VMs + SSD Persistent Disk storage-only
- Mode 2: Compute VMs + SSD Persistent Disk and Google Cloud Storage (GCS).

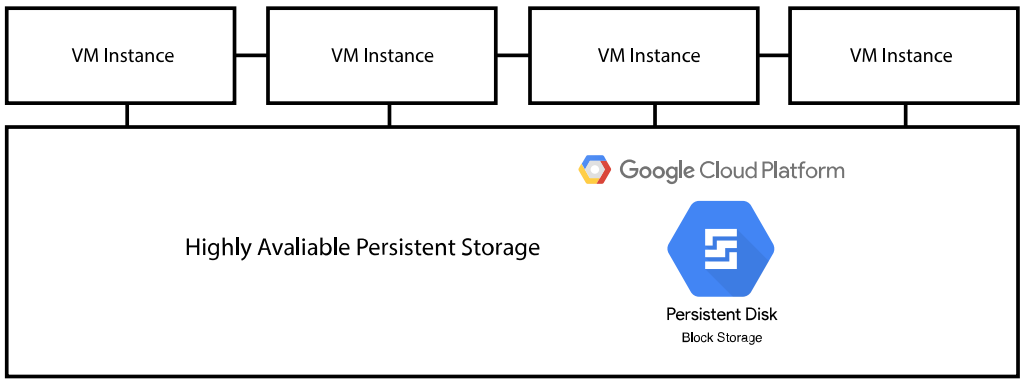
For more information about Persistant Storage, see Zonal Persistent SSD disks in Google's Cloud documentation.

For more information about Google Cloud Storage, see Cloud Storage Buckets in Google's Cloud documentation.

All GCP VMs (nodes) in a ThoughtSpot cluster must be in the same zone (and, therefore, also in the same region). ThoughtSpot does not support deploying VMs (nodes) of the same cluster across zones. For more information, see Regions and Zones in Google's Cloud documentation.

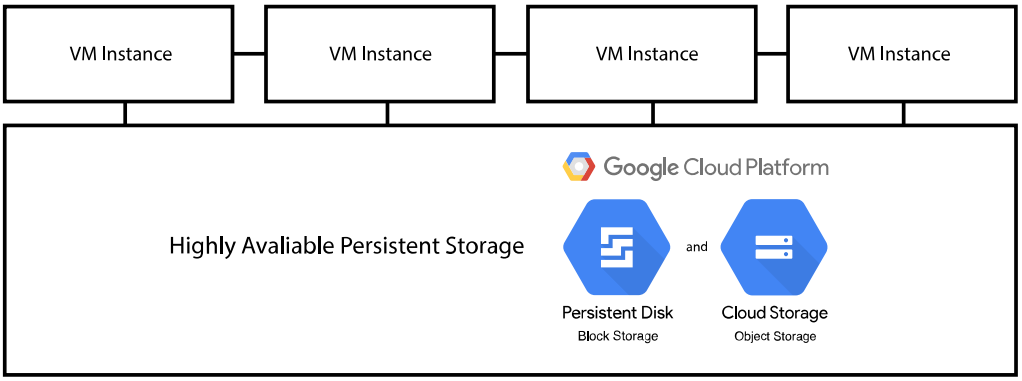
ThoughtSpot GCP instance types

VMs with Persistent Disk-only storage



Per VM user data capacity	Instance type	CPU/RAM	Recommended per-VM Zonal Persistent SSD Disk volume
208 GB	n1-highmem-64	64/416	2x 1 TB
312 GB	n1-highmem-96	96/624	2x 1.5 TB
100 GB	n1-highmem-32	32/208	2X 400 GB
20 GB	n1-highmem-16	16/122	2X 400 GB
180 GB	n1-standard-96	96/330	2X 1 TB

VMs with Persistent Disk and Google Cloud storage



Per VM user data capacity	Instance type	CPU/RAM	Recommended per-VM Zonal Persistent SSD Disk volume
208 GB	n1-highmem-64	64/416	1X 500 GB
312 GB	n1-highmem-96	96/624	1X 500 GB
100 GB	n1-highmem-32	32/208	1X 500 GB
20 GB	n1-highmem-16	16/122	1X 500 GB
180 GB	n1-standard-96	96/330	1X 500 GB

Set up GCP for ThoughtSpot

After you've determined your configuration options, set up your virtual machines (VMs). The ThoughtSpot base image for booting the VMs and some other aspects of system setup will be shared with you on GCP by ThoughtSpot.

About the ThoughtSpot and Google Cloud Platform

ThoughtSpot uses a custom image to populate VMs on GCP. The base image is a Centos derived image, which will be available to you in your Google Compute Engine project for Boot disk options under Custom Images.

Ask your ThoughtSpot liaison for access to this image. We need the Google account/email ID of the individual who will be signed into your organization's GCP console. We will share ThoughtSpot's GCP project with them so they can use the contained boot disk image for creating ThoughtSpot VMs.

Overview

Before you can create a ThoughtSpot cluster, you must provision VMs. You use the Google Compute Engine (GCP) platform for creating and running VMs.

The following topics walk you through this process.

Prerequisites

1. Ensure that **Network Service Tier** is set to **Premium** for all VMs to be used in your ThoughtSpot cluster.
2. A ThoughtSpot cluster requires 10 Gb/s bandwidth (or better) between any two nodes. This must be established before creating a new cluster.

Setting up your Google Cloud Storage (GCS) bucket

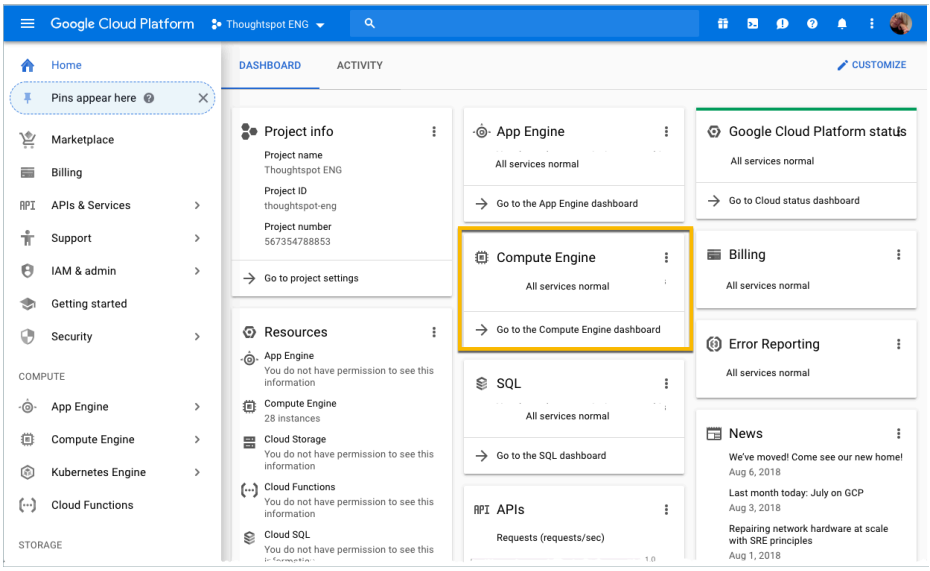
If you are going to deploy your cluster using the GCS-storage option, you must set up that bucket before you set up your cluster. Contact ThoughtSpot Support to find out if your specific cluster size will benefit from the GCS storage option.

1. Sign in to the Google Cloud Console.
2. Go to the Storage dashboard.
3. Click **CREATE BUCKET**.
4. Enter a name for your bucket, and click **CONTINUE**.
5. For location type, select **Region** and use the Location drop-down menu to select the region where you are going to set up your instance, and click **CONTINUE**.
6. For default storage class, make sure **Standard** is selected, and click **CONTINUE**.
7. For access control model, make sure **Set permissions uniformly at bucket-level** is selected, and click **CONTINUE**.
8. For advanced settings, leave Encryption set to **Google-managed key**, do not set a retention policy, and click **CREATE**.

When you create your instance, make sure you set Storage to **Read Write** access.

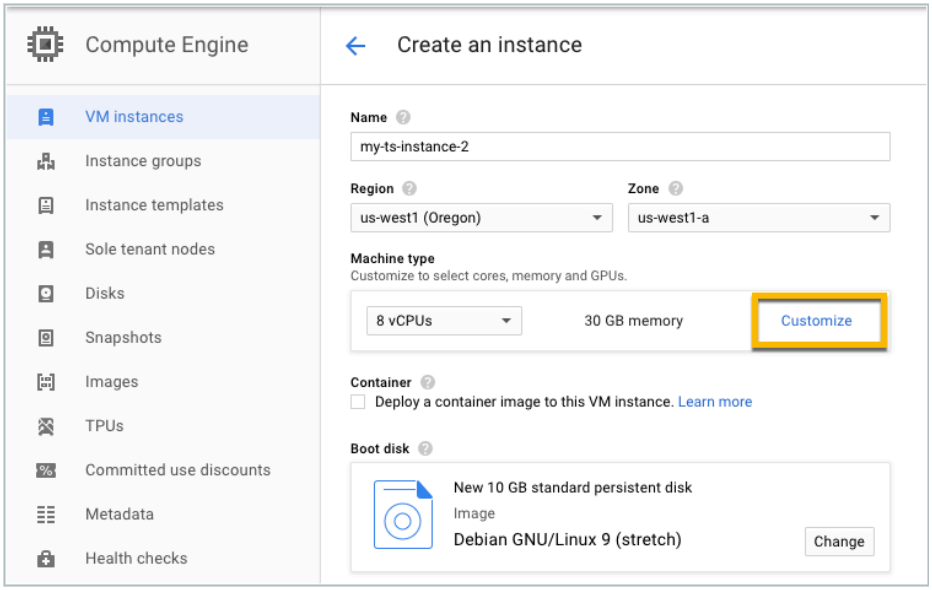
Create an instance

1. Sign in to the Google Cloud Console.
2. Go to the Compute Engine dashboard, and select the associated ThoughtSpot project.



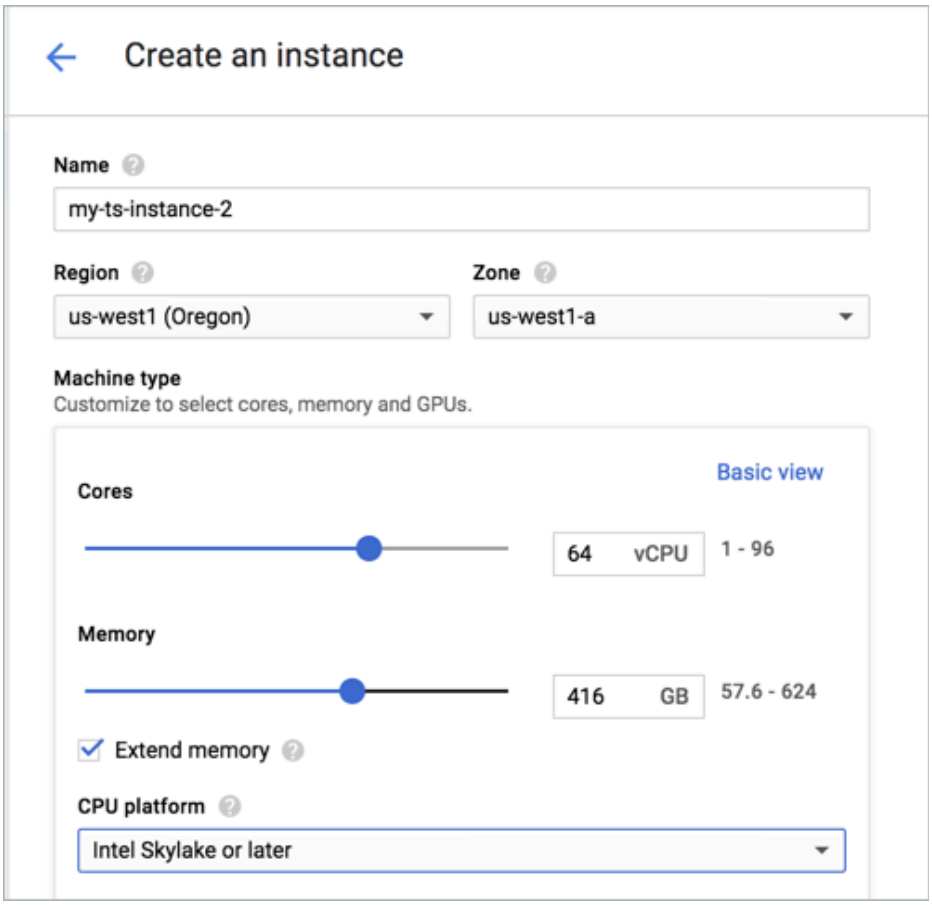
3. Select **VM instances** on the left panel and click **CREATE INSTANCE**.

4. Provide a name for the instance, choose a region, choose number of CPUs (e.g., 8 vCPUs for a cluster), and click **Customize** to further configure CPUs and memory.



5. For **Machine type** set the following configuration:

Setting	Value
Cores	`64 vCPU`
Memory	`416 GB`
Extend memory	Enabled (checkmark)
CPU platform	`Automatic` (or select one of the preferred CPU platforms, `Intel Skylake` or `Intel Broadwell`, if available)

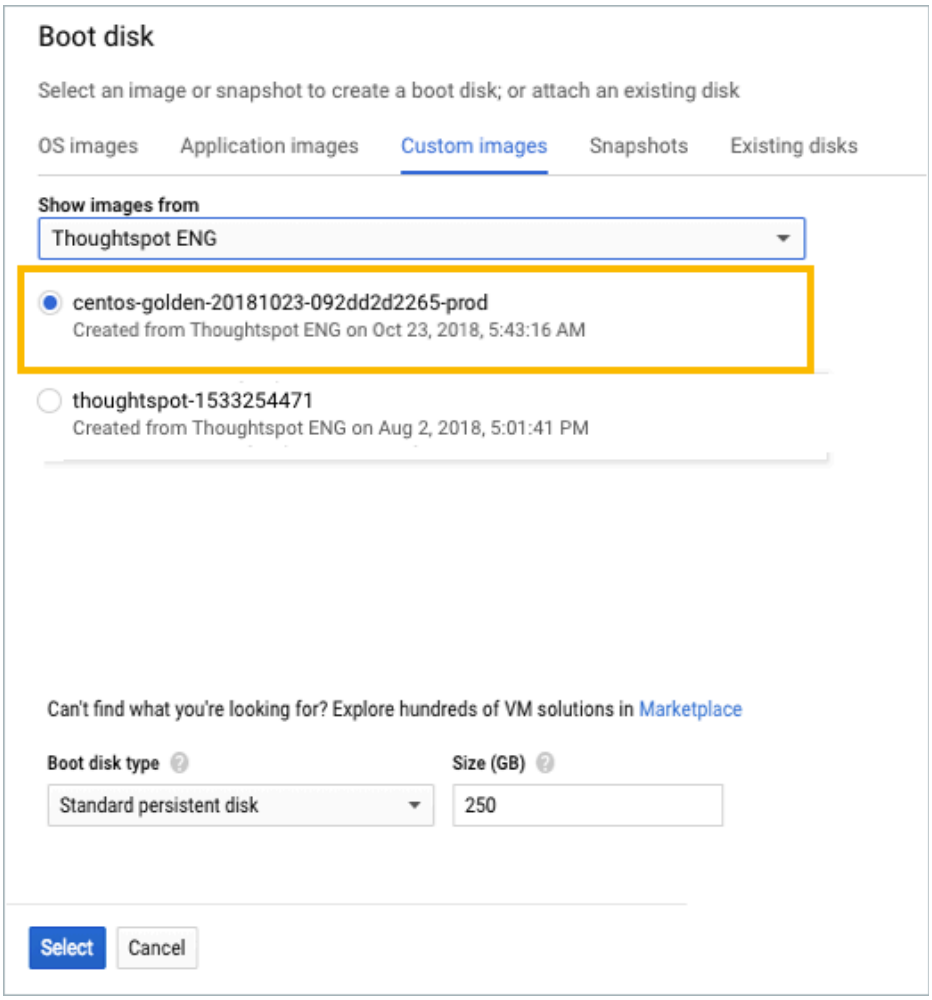


6. Configure the Boot disk.

a. Scroll down to the find the **Boot disk** section and click **Change**.

b. Click **Custom Images** on the tabs at the top, select a ThoughtSpot base image and configure the boot disk as follows:

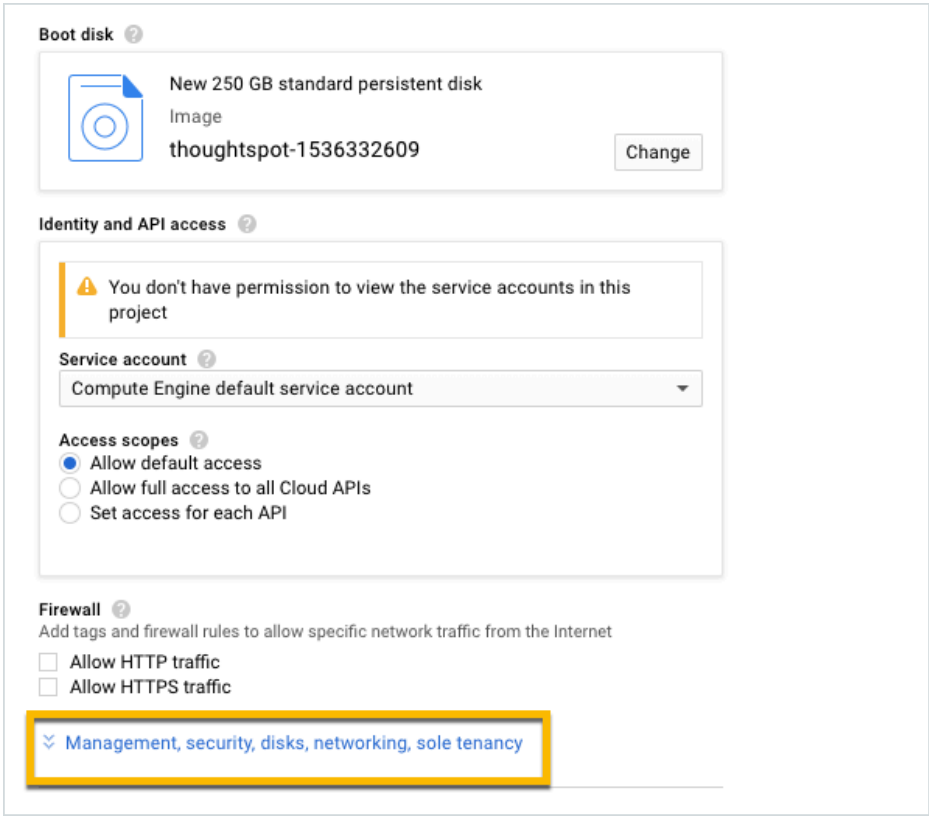
Setting	Value
Image	`ThoughtSpot`
Boot disk type	`Standard persistent disk`
Size (GB)	`250`



> ℹ **Note:** ThoughtSpot updates these base images with patches and enhancements. If more than one image is available, the latest one is always at the top of the list. Both will work, but we recommend using the latest image because it typically contains the latest security and maintenance patches.

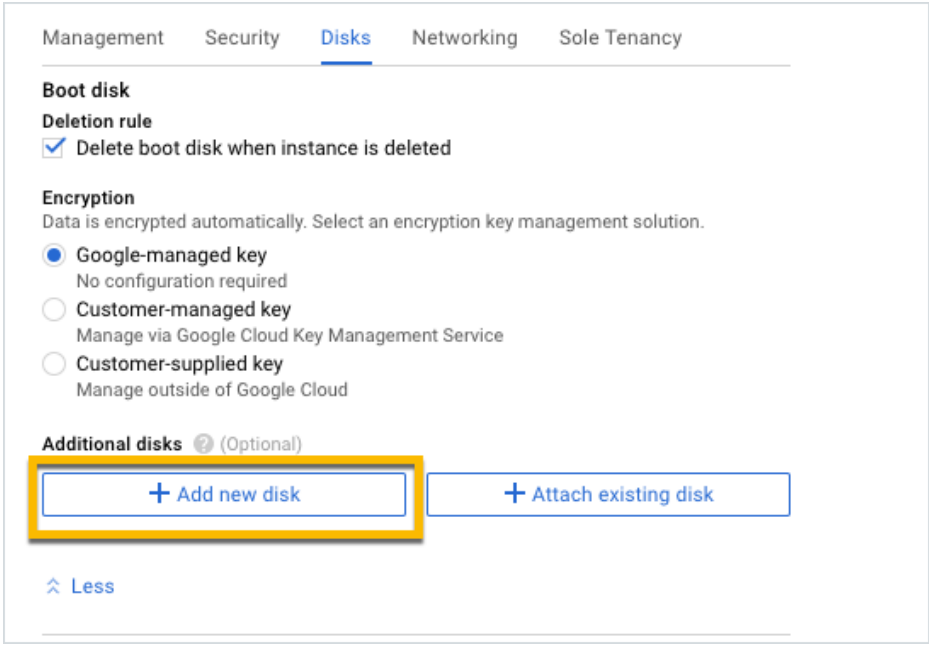
c. Click **Select** to save the boot disk configuration.

7. Back on the main configuration page, click to expand the advanced configuration options (**Management, security, disks, networking, sole tenancy**).



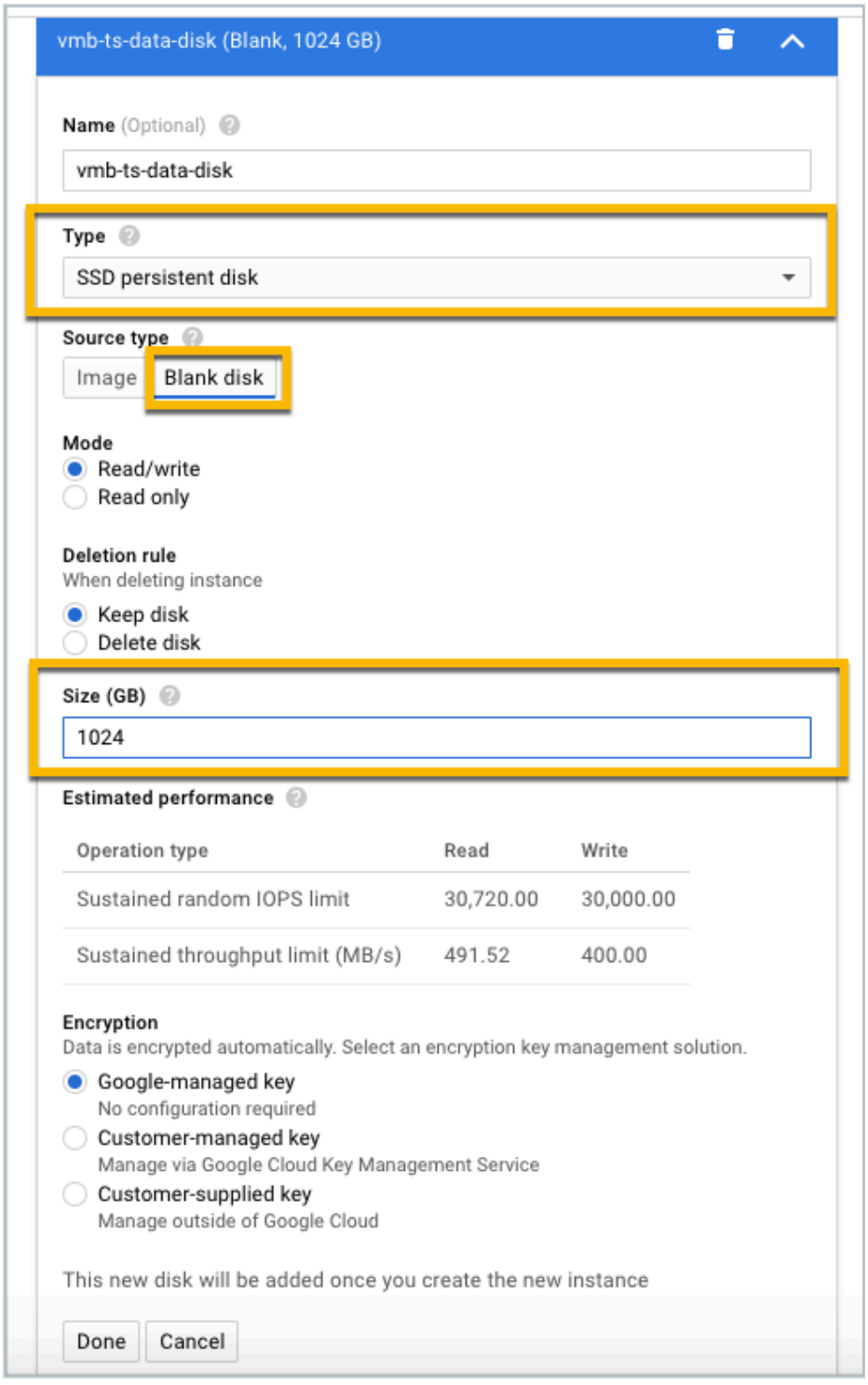
8. Attach two 1 TB SSD drives. These drives will be used for the data storage.

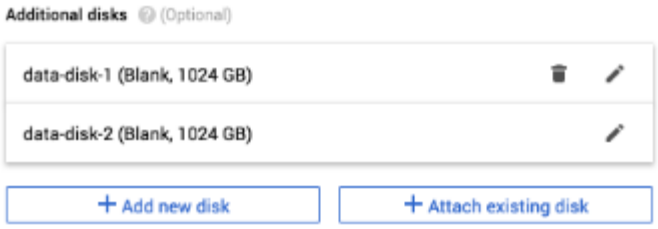
   a. Click the **Disks** tab, and click **Add new disk**.



b. Configure the following settings for each disk.

Setting	Value
Type	`SSD persistent disk`
Source type	`Blank disk`
Size (GB)	`1024`


vmb-ts-data-disk (Blank, 1024 GB)
Name (Optional)
vmb-ts-data-disk
Type
SSD persistent disk
Source type
Image
Blank disk
Mode
Read/write
Read only
Deletion rule
When deleting instance
Keep disk
Delete disk
Size (GB)
1024
Estimated performance
Operation type
Read
Write
Sustained random IOPS limit
30,720.00
30,000.00
Sustained throughput limit (MB/s)
491.52
400.00
Encryption
Data is encrypted automatically. Select an encryption key management solution.
Google-managed key
No configuration required
Customer-managed key
Manage via Google Cloud Key Management Service
Customer-supplied key
Manage outside of Google Cloud
This new disk will be added once you create the new instance
Done
Cancel



9. (For use with GCS only) In the Identity and API access section, make sure Service account is set to **Compute Engine default service account**, and under Access scopes, select **Set access for each API**.
10. (For use with GCS only) Scroll down to the Storage setting, and set it to one of the following options:
    - To use Google Cloud Storage (GCS) as persistent storage for your instance, select **Read Write**.
    - To only use GCS to load data into ThoughtSpot, select **Read Only**.
11. Customize the network settings as needed, preferably use your default VPC settings.
12. Repeat these steps to create the necessary number of such VMs.

Prepare the VMs (ThoughtSpot Systems Reliability Team)

> ⚠ **Important:** This procedure is typically done by a ThoughtSpot Systems Reliability Engineer (SRE). Please consult with your ThoughtSpot Customer Service or Support Engineer on these steps.

Before we can install a ThoughtSpot cluster, an administrator must log into each VM through SSH as user “admin”, and complete the following preparation steps:

1. Run `sudo /usr/local/scaligent/bin/prepare_disks.sh` on every machine.
2. Configure each VM based on the site-survey.

Launch the cluster

Upload the TS tarball to one of the VMs and proceed with the normal cluster creation process, using tscli cluster create.

If you are going to use GCS as your persistent storage, you must enable it when running this command, using the **enable_cloud_storage** flag. Example: `tscli cluster create 6.0-167.tar.gz --enable_cloud_storage=gcs`

Open the required network ports

To determine which network ports to open for a functional ThoughtSpot cluster, see Network policies.

Related information

Connecting to Google Cloud Storage buckets

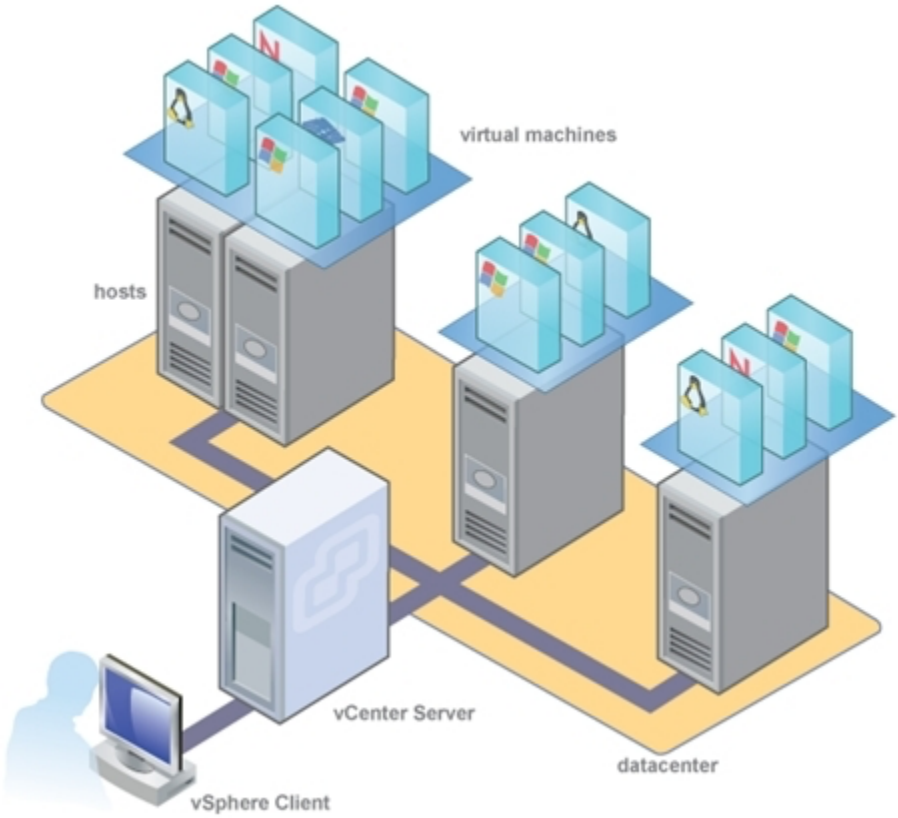
Loading data from a GCP GCS bucket

VMware configuration overview

Congratulations on purchasing the ThoughtSpot instance. This section is an overview of the ThoughtSpot AI-Driven analytics platform hosted on the VMware vSphere Hypervisor (ESXi) 6.5 environment.

About ThoughtSpot on VMware

The VMware virtualization platform provides highly scalable and efficient memory and CPU resources management that can be used by ThoughtSpot instances. Additionally, the VMware virtualization environment is an easy transition between development and production environments. The following diagram shows the components of a VMware and ThoughtSpot architecture:



Note: This is a generic representation; Only CentOS-based virtual machines are supported with ThoughtSpot.

Your database capacity will determine the number of ThoughtSpot instances and the instance network/storage requirements. In addition, you can scale your ThoughtSpot VMs as your dataset size grows.

Supported configurations

ThoughtSpot Engineering has performed extensive testing of the ThoughtSpot platform on VMware for the best performance, load balancing, scalability, and reliability. Based on this testing, ThoughtSpot recommends the following *minimum specifications* for an individual VMware ESXi host machine:

Per VM user data capacity	CPU/RAM	Data disk
20 GB	16/128 GB	800 GB
100 GB	32/256 GB	800 GB
256 GB	72/512 GB	6 TB
Note: All cores must be hyperthreaded. 200GB SSD boot disk required for all configurations.		

Locally attached storage provides the best performance.

SAN can be used, but must comply with the following requirements:

- 136 MBps minimum random read bandwidth
- 240 random IOPS (~4ms seek latency)

NAS/NFS is not supported since its latency is so high that it tends to be unreliable.

All virtualization hosts should have VMware vSphere Hypervisor (ESXi) 6.5 installed.

ThoughtSpot provides a VMware template (OVF) together with a VMDK (Virtual Machine Disk) file for configuring a VM. VMDK is a file format that describes containers for virtual hard disk drives to be used in virtual machines like VMware Workstation or VirtualBox. OVF is a platform-independent, efficient, extensible, and open packaging distribution format for virtual machines.

The ThoughtSpot VM configuration uses thin provisioning and sets the recommended reserved memory, among other important specifications. You can obtain these files from your ThoughtSpot Customer Success Engineer.

Questions or comments?

We hope your experience with ThoughtSpot is excellent. Please let us know how it goes, and what we can do to make it better. You can contact ThoughtSpot by filing a support ticket, email or phone.

Set up VMware for ThoughtSpot

Summary: Learn how to install a ThoughtSpot cluster in a VMware environment.

This page explains how to install a ThoughtSpot cluster in a VMware VSphere Hypervisor (ESXi) 6.5 environment. For each hardware node, you must:

- Complete the prerequisites
- Use the ThoughtSpot Open Virtualization Format (OVF) file to create a virtual machine (VM)
- Add hard disks to the VM

Prerequisites

This installation process assumes you have already acquired your host machines. You can install on a cluster with any number of nodes. A one node cluster is suitable for a sandbox environment but is insufficient for a production environment. You need at least three nodes for high availability (HA), but there is no limit on the number of nodes.

1. Make sure you have installed the Hypervisor on each of your nodes.

   The VM template, by default, captures a 72-core configuration. If your physical host has more than 72 cores, you may want to edit VM to have (`n–2`) cores (for a physical host with n cores) to fully take advantage of computing power of the physical host. Extra cores help performance.

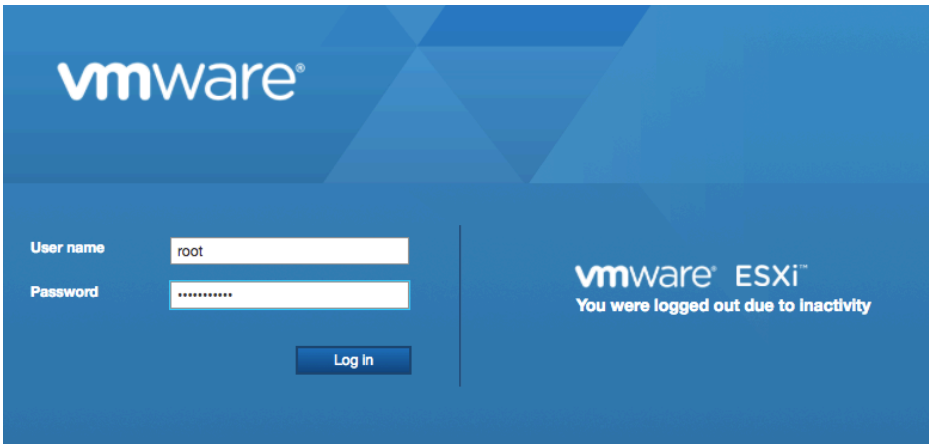
   You should aim to allocate 490 GB or more RAM.

2. Create datastores for all solid-state drive (SSD) and hard drive devices.

Use the OVF to Create a VM

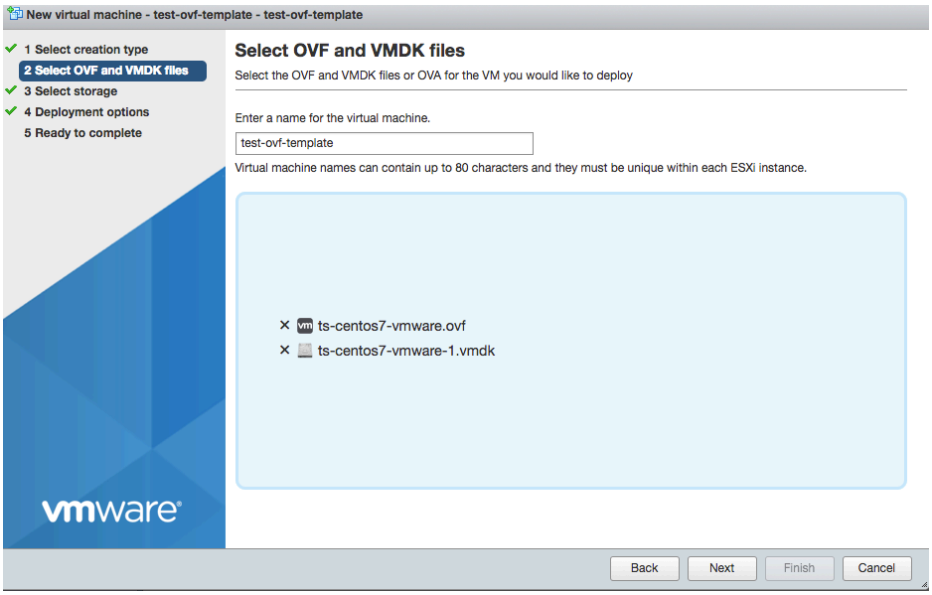
1. **Download** the `ThoughtSpot OVF` to a location on an accessible disk.

2. Log into the ESXi web portal.



3. Select **Virtual Machines > Create/Register VM**.

   The system displays the dialog for selecting an OVF template.



4. Choose the OVF template and click **Next**.

   The system prompts you to select a storage.

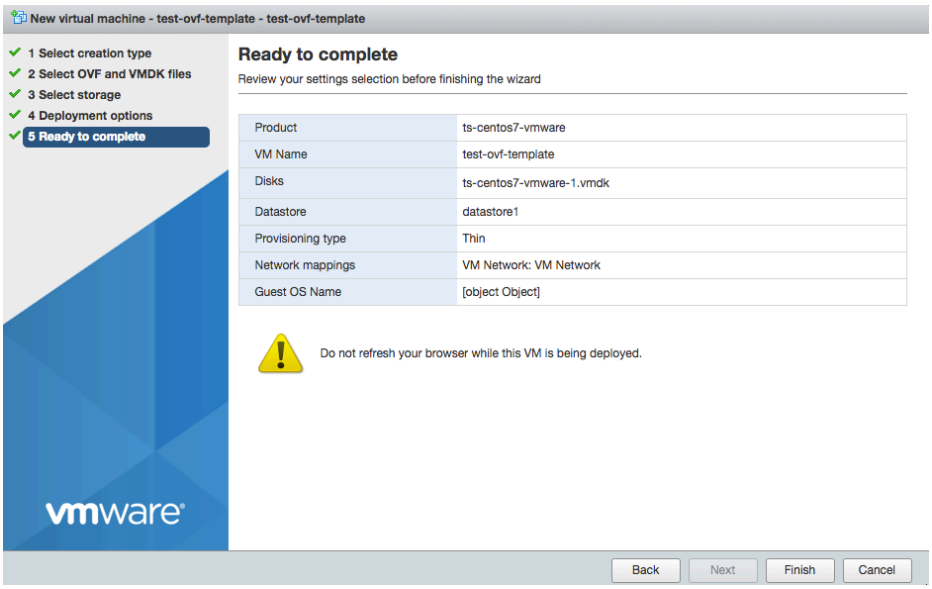
5. Choose the SSD as the destination and click **Next**.

   The system displays the **Deployment Options** dialog.

6. Enter the options and click **Next**.

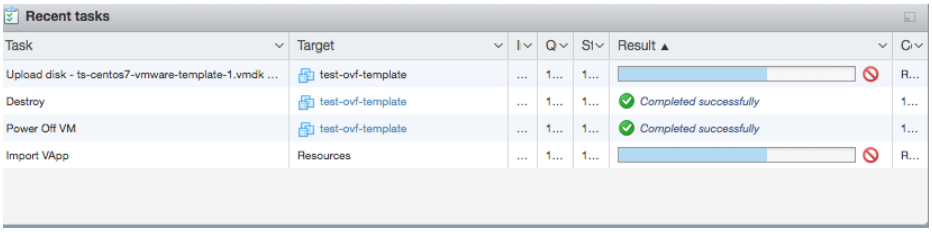
Setting	Value
Network mappings	Select the correct network for your installation.
Disk provisioning	Choose Thin.
Power on automatically	Check this box.

7. Review your selection and click **Finish**.



8. Wait for the template to be loaded.

   Depending on your network speed, loading can take several minutes.



9. Make sure that VM is powered off.

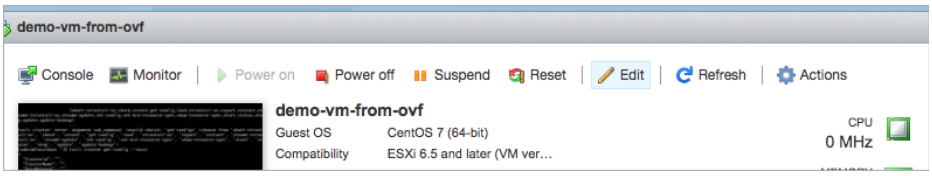
Add hard disks to the VM

Use Case	HDFS Disk Requirements
POC	2 x 1 TB on HDD
Production	3 x 2 TB on HDD

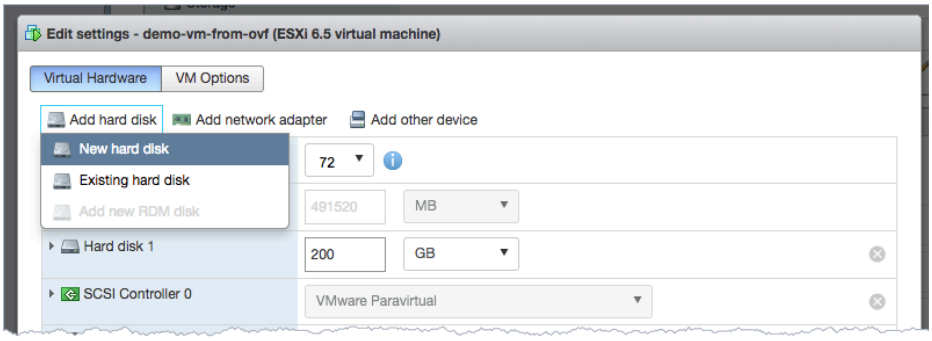
For a proof of concept (POC), follow these steps to create two 1TB HDFS disks on HDD storage, as shown here (2 x 1TB).

For production deployments, ThoughtSpot requires you to have three 2TB HDFS disks on HDD (3 x 2TB). For this use case, follow these same steps to create the additional, larger capacity disks.

1. Edit the VM you just created.



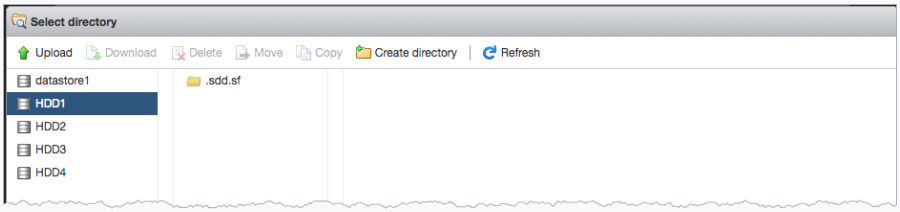
2. Select **Add hard disk > New hard disk**.



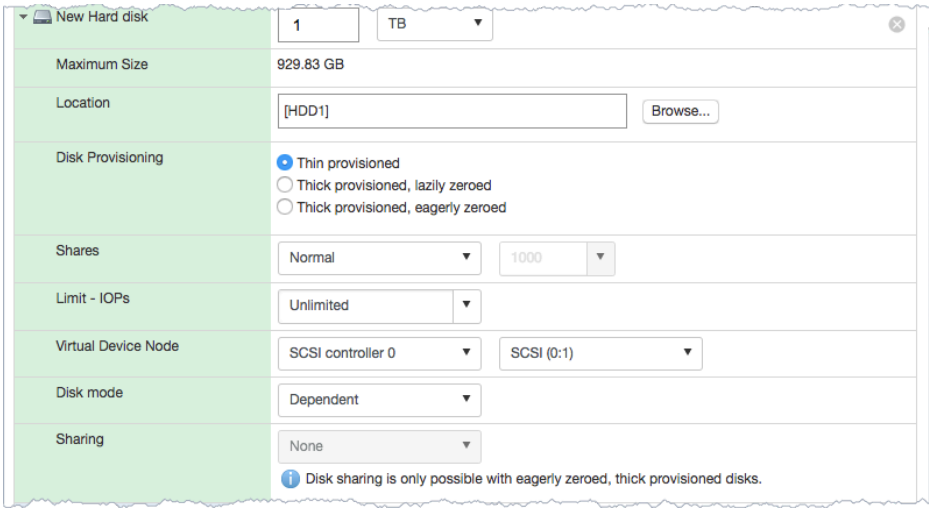
   The new disk appears as a new row under the only existing SSD row.

3. Click the **New Hard disk** to expand the detailed configuration options.

4. For a proof of concept, set the options as follows. (For production deployments, set the size to 2TB.)

Setting	Value
size	1 TB
Location	Use the **Browse** button to select the hard disk store. 
Thin provisioned	Check this box.

You should see something similar to the following:



5. Save your changes.
6. Repeat steps 1-5 to create more hard disks.
7. Power on the VM
8. After the VM is online, run the following command to prepare the HDFS disks:

```
sudo /usr/local/scaligent/bin/prepare_disks.sh
```

Next steps

There is no network at this point on your VMs. As a prerequisite:

1. Verify that Network Adapter type is set to VMware vmxnet3 (Recommended).
2. Verify that all ESXi hosts in your VMware farm for ThoughtSpot have been trunked to the VLAN assigned to your ThoughtSpot VMs.
3. Verify that the console of all ThoughtSpot VMs is accessible in VMware vCenter Server.

After you finish, go to the ThoughtSpot Support website and use the support ticket for installation tasks. If necessary, create a new ticket.

Network policies

Summary: Lists the required and optional ports for an installation.

For regular operations and for debugging, there are some ports you must keep open to network traffic from end users. Another, larger list of ports must be kept open for network traffic between the nodes in the cluster.

Required ports for operations and debugging

The following ports must be open for requests from your user population. There are two main categories: operations and debugging.

Port	Protocol	Service Name	Direction	Source	Destination	Description
22	SSH	SSH	bidirectional	Administrators IP addresses	All nodes	Secure shell access. Also used for scp (secure copy).
80	HTTP	HTTP	bidirectional	All users IP addresses	All nodes	Hypertext Transfer Protocol for website traffic.
443	HTTPS	HTTPS	bidirectional	All users IP addresses	All nodes	Secure HTTP.
12345	TCP	Simba	bidirectional	Administrators IP addresses	All nodes	Port used by ODBC and JDBC drivers when connecting to ThoughtSpot.
2201	HTTP	Orion master HTTP	bidirectional	Administrator IP addresses	All nodes	Port used to debug the cluster manager.
2101	HTTP	Oreo HTTP	bidirectional	Administrator IP addresses	All nodes	Port used to debug the node daemon.
4001	HTTP	Falcon worker HTTP	bidirectional	Administrator IP addresses	All nodes	Port used to debug the data cache.

Port	Protocol	Service Name	Direction	Source	Destination	Description
4251	HTTP	Sage master HTTP	bidirectional	Administrator IP addresses	All nodes	Port used to debug the search engine.

Network Ports

This reference lists the potential ports to open when setting up your security group.

Required ports for intracluster operation

Static ports are used for communication between services within the cluster. ThoughtSpot recommends that you open all ports within a cluster. This not required, but it will ensure that cluster communication works properly if additional ports are used in a future software release.

If your organization does not allow you to open all ports, make sure you open the required intracluster ports listed in the following table. In addition, a number of ports are dynamically assigned to services, which change between runs. The dynamic ports come from the range of ports that are dynamically allocated by Linux (20K+).

Port	Protocol	Service Name	Direction	Source	Dest.	Description
80	TCP	nginx	inbound	All nodes	All nodes	Primary app HTTP port (nginx)
443	TCP	Secure nginx	inbound	All nodes	All nodes	Primary app HTTPS port (nginx)
2100	RPC	Oreo RPC port	bidirectional	All nodes	All nodes	Node daemon RPC
2101	HTTP	Oreo HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Node daemon HTTP

Port	Protocol	Service Name	Direction	Source	Dest.	Description
2181	RPC	Zookeeper servers listen on this port for client connections	bidirectional	All nodes	All nodes	Zookeeper servers listen on this port for client connections
3181	RPC	Zookeeper servers listen on this port for client connections	bidirectional	All nodes	All nodes	Zookeeper servers listen on this port for client connections
4181	RPC	Zookeeper servers listen on this port for client connections	bidirectional	All nodes	All nodes	Zookeeper servers listen on this port for client connections
2200	RPC	Orion master RPC port	bidirectional	All nodes	All nodes	Internal communication with the cluster manager
2201	HTTP	Orion master HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Port used to debug the cluster manager
2205	TCP	Cluster update service TCP port	bidirectional	All nodes	All nodes	Internal communication with the cluster manager
2210	RPC	Cluster stats service RPC port	bidirectional	All nodes	All nodes	Internal communication with the stats collector
2211	HTTP	Cluster stats service HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Port used to debug the stats collector
2230	RPC	Callosum stats collector RPC port	bidirectional	All nodes	All nodes	Internal communication with the BI stats collector

Port	Protocol	Service Name	Direction	Source	Dest.	Description
2231	HTTP	Callosum stats collector HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Port used to debug the BI stats collector
2240	RPC	Alert manager	bidirectional	All nodes	All nodes	Port where alerting service receives alert events
2241	RPC	Alert manager	bidirectional	All nodes	All nodes	Port where alerting service receives alert events
2888	RPC	Ports used by Zookeeper servers for communication between themselves	bidirectional	All nodes	All nodes	Ports used by Zookeeper servers for communication between themselves
3181	RPC	Ports used by Zookeeper servers for communication between themselves	bidirectional	All nodes	All nodes	Ports used by Zookeeper servers for communication between themselves
3888	RPC	Ports used by Zookeeper servers for communication between themselves	bidirectional	All nodes	All nodes	Ports used by Zookeeper servers for communication between themselves
4000	RPC	Falcon worker RPC port	bidirectional	All nodes	All nodes	Port used by data cache for communication between themselves
4001	HTTP	Falcon worker HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Port used to debug the data cache
4002	HTTP	Falcon worker HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Port used to debug the data cache

Port	Protocol	Service Name	Direction	Source	Dest.	Description
4003	RPC	Falcon worker RPC port	bidirectional	All nodes	All nodes	Port used by data cache for communication between themselves
4004	RPC	Falcon worker RPC port	bidirectional	All nodes	All nodes	Port used by data cache for communication between themselves
4021	RPC	Sage metadata service port (exported by Tomcat), Callosum services like meta-data services, medata-dependency service, scheduling service, session-less service, spotiq service	bidirectional	All nodes	All nodes	Port where search service contacts metadata service for metadata
4181	RPC	Ports used by Zookeeper servers for communication between themselves	bidirectional	All nodes	All nodes	Ports used by Zookeeper servers for communication between themselves
4201	HTTP	Sage auto complete server HTTP interface port	bidirectional	Admin IP addresses and all nodes	All nodes	Port used to debug the search service
4231	HTTP	Sage index server HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Port used to debug the search service
4232	RPC	Sage index server metadata subscriber port	bidirectional	All nodes	All nodes	Port used for search service internal communication
4233	RPC	Sage index server RPC port	bidirectional	All nodes	All nodes	Port used for search service internal communication

Port	Protocol	Service Name	Direction	Source	Dest.	Description
4241	HTTP	Sage auto complete server HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Port used to debug the search service
4242	RPC	Sage auto complete server RPC port	bidirectional	All nodes	All nodes	Port used for search service internal communication
4243	RPC	Sage auto complete server metadata subscriber port	bidirectional	All nodes	All nodes	Port used for search internal communication
4244	RPC	Sage auto complete server metadata subscriber port	bidirectional	All nodes	All nodes	Port used for search internal communication
4245	RPC	Sage auto complete server metadata subscriber port	bidirectional	All nodes	All nodes	Port used for search internal communication
4243	RPC	Sage auto complete server metadata subscriber port	bidirectional	All nodes	All nodes	Port used for search internal communication
4251	RPC	Sage master RPC port	bidirectional	All nodes	All nodes	Port used for search service internal communication
4405	RPC	Diamond (graphite) port	bidirectional	All nodes	All nodes	Port used for communication with monitoring service
4406	RPC	Diamond (graphite) port	bidirectional	All nodes	All nodes	Port used for communication with monitoring service
4500	RPC	Trace vault service RPC port	bidirectional	All nodes	All nodes	Trace collection for ThoughtSpot services
4501	HTTP	Trace vault service HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Debug trace collection

Port	Protocol	Service Name	Direction	Source	Dest.	Description
4851	RPC	Graphite manager RPC port	bidirectional	All nodes	All nodes	Communication with graphite manager
4852	HTTP	Graphite manager HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Debug graphite manager
4853	RPC	Elastic search stack (ELK) manager RPC port	bidirectional	All nodes	All nodes	Communication with log search service
4853	HTTP	Elastic search stack (ELK) manager HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Debug log search service
9200	RPC	Elastic search (ELK)	bidirectional	All nodes	All nodes	Communication with log search service
5021	RPC	Callosum services like meta-data services, medata-dependency service, scheduling service, session-less service, spotiq service	bidirectional	All nodes	All nodes	Port where search service contacts meta-data service for metadata
5432	Postgres	Postgres database server port	bidirectional	All nodes	All nodes	Communication with Postgres database
6021	RPC	Callosum services like meta-data services, medata-dependency service, scheduling service, session-less service, spotiq service	bidirectional	All nodes	All nodes	Port where search service contacts meta-data service for metadata
7021	RPC	Callosum services like meta-data services, medata-dependency service, scheduling service, session-less service, spotiq service	bidirectional	All nodes	All nodes	Port where search service contacts meta-data service for metadata

Port	Protocol	Service Name	Direction	Source	Dest.	Description
8020	RPC	HDFS namenode server RPC port	bidirectional	All nodes	All nodes	Distributed file system (DFS) communication with clients
8021	RPC	Callosum services like meta-data services, medata-dependency service, scheduling service, session-less service, spotiq service	bidirectional	All nodes	All nodes	Port where search service contacts metadata service for metadata
8080	HTTP	Tomcat	bidirectional	All nodes	All nodes	BI engine communication with clients
8081	HTTP	Callosum/Tomcat status	bidirectional	All nodes	All nodes	BI engine communication with clients
8787	HTTP	Periscope (UI) service HTTP port	bidirectional	All nodes	All nodes	Administration UI back end
8888	HTTP	HTTP proxy server (tinyproxy)	bidirectional	All nodes	All nodes	Reverse SSH tunnel
11211	Memcached	Memcached server port	bidirectional	All nodes	All nodes	BI engine cache
12345	ODBC	Simba server port	bidirectional	All nodes	All nodes	Port used for ETL (extract, transform, load)
8480	HTTP	HDFS journalnode server HTTP port	bidirectional	All nodes	All nodes	Debug DFS metadata
8485	HTTP	HDFS journalnode server HTTP port	bidirectional	All nodes	All nodes	Debug DFS metadata
50070	HTTP	HDFS namenode server HTTP port	bidirectional	All nodes	All nodes	Debug DFS metadata
50090	HTTP	HDFS secondary namenode server HTTP port	bidirectional	All nodes	All nodes	Debug DFS metadata
50075	HTTP	HDFS datanode server HTTP port	bidirectional	All nodes	All nodes	Debug DFS data

Port	Protocol	Service Name	Direction	Source	Dest.	Description
50010	HTTP	HDFS datanode server HTTP port	bidirectional	All nodes	All nodes	Debug DFS data
50020	HTTP	HDFS datanode server HTTP port	bidirectional	All nodes	All nodes	Debug DFS data
7000	TCP	Cassandra KV store database	bidirectional	All nodes	All nodes	Debug DFS data
7001	TCP	Cassandra	bidirectional	All nodes	All nodes	Debug DFS data
9042	HTTP	Munshi server impression service, Cassandra	bidirectional	All nodes	All nodes	Debug DFS data
9160	TCP	Cassandra	bidirectional	All nodes	All nodes	Debug DFS data
4010	HTTP	Falcon moderator	bidirectional	All nodes	All nodes	Debug DFS data
4011	HTTP	Falcon moderator	bidirectional	All nodes	All nodes	Debug DFS data
20123 - 32768	TCP (dynamic)	Dynamic port in this range used for various services and anciliary services like atlas, caffeine, callhome, callosum, falcon, monitoring, munshi server, nlp, object_search, postgres, sage UBR, spotiq snapshot, timely	All nodes	Services		
5270	TCP	Cluster monitoring service (ELK)	bidirectional	All nodes	All nodes	Services
5271	TCP	Cluster monitoring service (ELK)	bidirectional	All nodes	All nodes	Services
5601	TCP	Kibana UI (ELK)	bidirectional	All nodes	All nodes	Services
6311	TCP	R service	bidirectional	All nodes	All nodes	Services
8008	TCP	Video recorder	bidirectional	All nodes	All nodes	Services

Port	Protocol	Service Name	Direction	Source	Dest.	Description
9090	TCP	Timely	bidirectional	All nodes	All nodes	Services
	ICMPv4	Used for health check of cluster nodes	bidirectional	All nodes	All nodes	Services

Required ports for inbound and outbound cluster access

ThoughtSpot uses static ports for inbound and outbound access to the cluster.

Port	Protocol	Service Name	Direction	Source	Dest.	Description
22	SCP	SSH	bidirectional	ThoughtSpot Support	All nodes	Secure shell access.
80	HTTP	HTTP	bidirectional	ThoughtSpot Support	All nodes	Hypertext Transfer Protocol for website traffic.
443	HTTPS	HTTPS	bidirectional	ThoughtSpot Support	All nodes	Secure HTTP.
12345	TCP	Simba	bidirectional	ThoughtSpot Support	All nodes	Port used by ODBC and JDBC drivers when connecting to ThoughtSpot.
2049	TCP	NFS: In case one needs to mount NFS share on TS node.	bidirectional	ThoughtSpot Support	All nodes	Port used by NFS.
123	UDP	NTP service	bidirectional	ThoughtSpot Support	All nodes	Port used by NTP service.

Port	Protocol	Service Name	Direction	Source	Destination	Description
443	TCP	HTTPS	outbound	All nodes	208.83.110.20	For transferring files to thoughtspot.egnyte.com.

Port	Protocol	Service Name	Direction	Source	Destination	Description
443	TCP	HTTPS	outbound	All nodes	For transferring product usage data to mixpanel cloud.	outbound
443	TCP	HTTPS	outbound	All nodes	je8b47jfif.execute-api.us-east-2.amazonaws.com s3.us-west-1.amazonaws.com s3-us-west-1.amazonaws.com s3.dualstack.us-west-1.amazonaws.com	For transferring monitoring data to InfluxCloud. (Given address will resolve to point to AWS instances).
25 or 587	SMTP	SMTP or Secure SMTP	outbound	All nodes and SMTP relay (provided by customer)	All nodes	Allow outbound access for the IP address of whichever email relay server is in use. This is for sending alerts to ThoughtSpot Support.
389 or 636	TCP	LDAP or LDAPS	outbound	All nodes and LDAP server (provided by customer)	All nodes	Allow outbound access for the IP address of the LDAP server in use.

Required ports for IPMI (Intelligent Platform Management Interface)

ThoughtSpot uses static ports for out-of-band IPMI communications between the cluster and ThoughtSpot support.

Port	Protocol	Service Name	Direction	Source	Dest.	Description
80	HTTP	HTTP	bidirectional	ThoughtSpot Support	All nodes	Hypertext Transfer Protocol for website traffic.
443	TCP	S-HTTP	bidirectional	ThoughtSpot Support	All nodes	IPMI GUI and for HTML5-based IPMI console access.

Port	Protocol	Service Name	Direction	Source	Dest.	Description
623	UDP	Serial-over-LAN	bidirectional	ThoughtSpot Support	All nodes	IPMI GUI and for HTML5-based IPMI console access.

Contact support

You can contact ThoughtSpot by phone, mail, email, or by filing a support ticket.

File a support ticket

If you encounter a technical issue, file a support ticket using the Support Portal ticket filing system at:

http://support.thoughtspot.com/

Please provide as much detail as possible about your issue, to help us resolve it quickly.

You need a Support Portal login to file a ticket. Please contact ThoughtSpot to get an account, if necessary.

Address

ThoughtSpot, Inc.

910 Hermosa Ct

Sunnyvale, CA 94085

Phone numbers

Phone Number	Description
1-800-508-7008 ext 1	ThoughtSpot Support
1-800-508-7008	Toll free number for ThoughtSpot headquarters.

Email

Reason for contacting	Email
For sales inquiries.	sales@thoughtspot.com
For customer support and software update inquiries.	support@thoughtspot.com
For other inquiries.	hello@thoughtspot.com