



ThoughtSpot Deployment Guide

Release 6.0

December, 2019

© COPYRIGHT 2015, 2019 THOUGHTSPOT, INC. ALL RIGHTS RESERVED.

910 Hermosa Court, Sunnyvale, California 94085

This document may not, in whole or in part, be copied, photocopied, reproduced, translated, or reduced to any electronic medium or machine-readable form without prior consent in writing from ThoughtSpot, Inc.

All rights reserved. The ThoughtSpot products and related documentation are protected by U.S. and international copyright and intellectual property laws. ThoughtSpot and the ThoughtSpot logo are trademarks of ThoughtSpot, Inc. in the United States and certain other jurisdictions. ThoughtSpot, Inc. also uses numerous other registered and unregistered trademarks to identify its goods and services worldwide. All other marks used herein are the trademarks of their respective owners, and ThoughtSpot, Inc. claims no ownership in such marks.

Every effort was made to ensure the accuracy of this document. However, ThoughtSpot, Inc., makes no warranties with respect to this document and disclaims any implied warranties of merchantability and fitness for a particular purpose. ThoughtSpot, Inc. shall not be liable for any error or for incidental or consequential damages in connection with the furnishing, performance, or use of this document or examples herein. The information in this document is subject to change without notice.

Table of Contents

Overview	2
Hardware appliance	
Hardware appliance overview	3
Installing the ThoughtSpot Dell appliance	6
Supported hardware	18
Cable networking	21
Set up and start the appliance	26
Cloud	
Cloud overview	30
AWS	
Configuration options	33
Set up ThoughtSpot in AWS	36
GCP	
Configuration options	43
Set up ThoughtSpot in GCP	45

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 \[See page 3\]](#)
- [Amazon Web Services \(AWS\) EC2 \[See page 33\]](#)
- [Microsoft Azure \[See page 0\]](#)
- [Google Cloud Platform \(GCP\) \[See page 43\]](#)
- [VMware \[See page 0\]](#)

After you configure your nodes, you can contact [ThoughtSpot Support \[See page 0\]](#) 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 \[See page 18\]](#) 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 side²

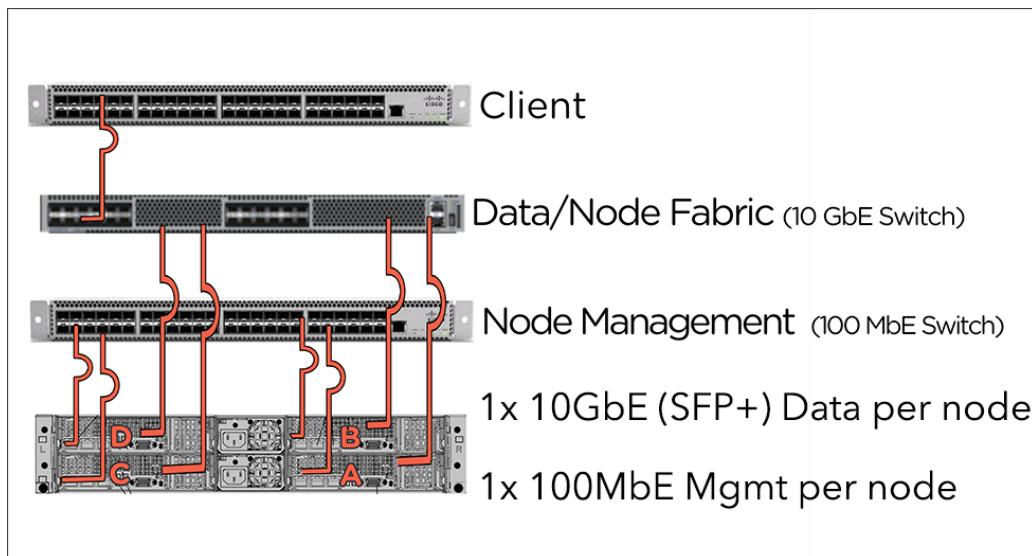
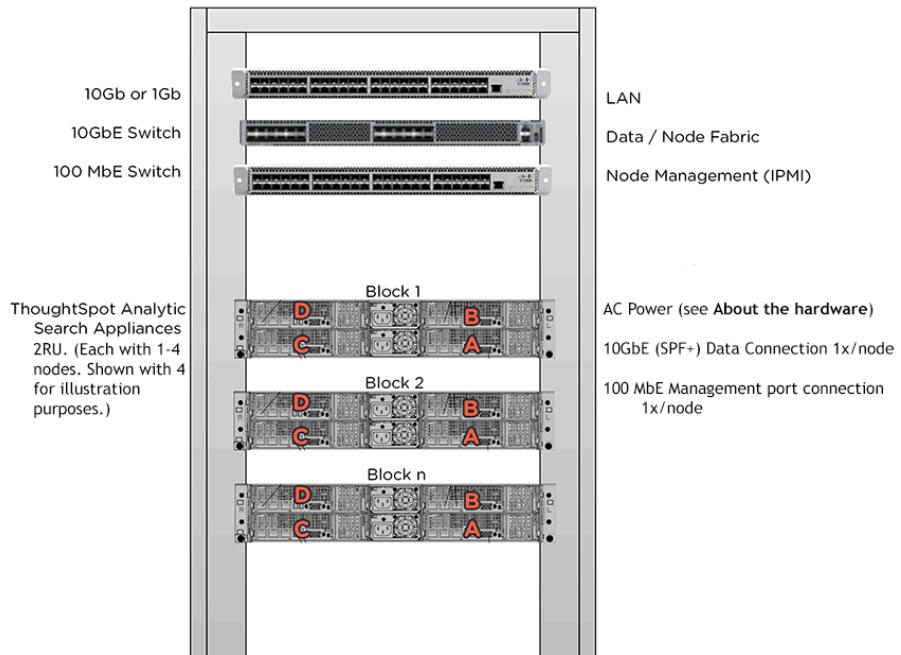
1. One 5m CAT 5e/6 network cable, per node, is provided with the appliance for management port connection. Customer supplied cable can be used if preferred.

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)



Installing the ThoughtSpot Dell appliance

Summary: Learn how to install the ThoughtSpot Dell appliance.

Installation Prerequisites

Ensure that you have the following items, information, and understanding of policies before you begin installing your Dell 6420 appliance:

- 10gbE switch with IPv6 broadcast and multicast enabled. You need one switch per node.
- Data center with proper cooling
- AC power
- 10G connection: SFP+ for switch side
- 10GbE network cables, either direct attach copper (DAC) or fiber. See [Cable Reference \[See page 0\]](#).
- 10bps switch for connection to the iDRAC (Out of Band Management) port
- Cat5 network cables
- Rack space (2U or 3.5 inches per appliance) and a power strip
- Monitor and keyboard
- Networking information, for data, management IPs, DNS, timezone, and default gateway IP. Contact your network administrator for this information, and fill out the ThoughtSpot site survey so that you have a quick reference.

About the Hardware

These pictures show the front and back view of the Dell C6420 appliance.

Appliance Front View



Appliance Back View



Connect the Appliance

After you rack and stack the appliance, you can begin to configure it.

Step 1: Connect switches to 10GbE ports

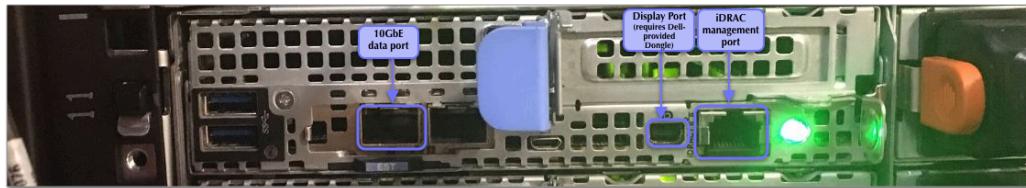
Connect the 10GbE port of each node, as illustrated in [Appliance Port Location \[See page 7\]](#), to the 10GbE switches on your own rack using either fiber or DAC cables.

Refer to the [Cable reference \[See page 0\]](#) for information on the cable types:

- Fiber Cables [See page 0]
 - DAC Cables [See page 0]

Note: Ask your hardware vendor for more details about what they supply and what you need to buy.

Appliance Port Location



- Connect to switches **only** the nodes you want in your cluster. Power off or disconnect any other nodes or appliances, to prevent accidental configuration of incorrect nodes.

Note: You need at least three nodes for high availability (HA). Each appliance can have up to four nodes.

- You must connect all nodes, even if using only one node, to a 10GbE switch.

Step 2: Connect iDRAC ports

Connect the iDRAC management ports of each node to the management switch. If you need help finding the ports, see [Appliance Port Location \[See page 7\]](#).

Step 3: Connect a keyboard and monitor

Connect a keyboard and monitor to the appliance. You need these to initially configure the appliance, and you can disconnect them later. Use the adapter Dell provides. Plug it into the Display Port shown in [Appliance Port Location \[See page 7\]](#), and plug the monitor in on the other side of the adapter.

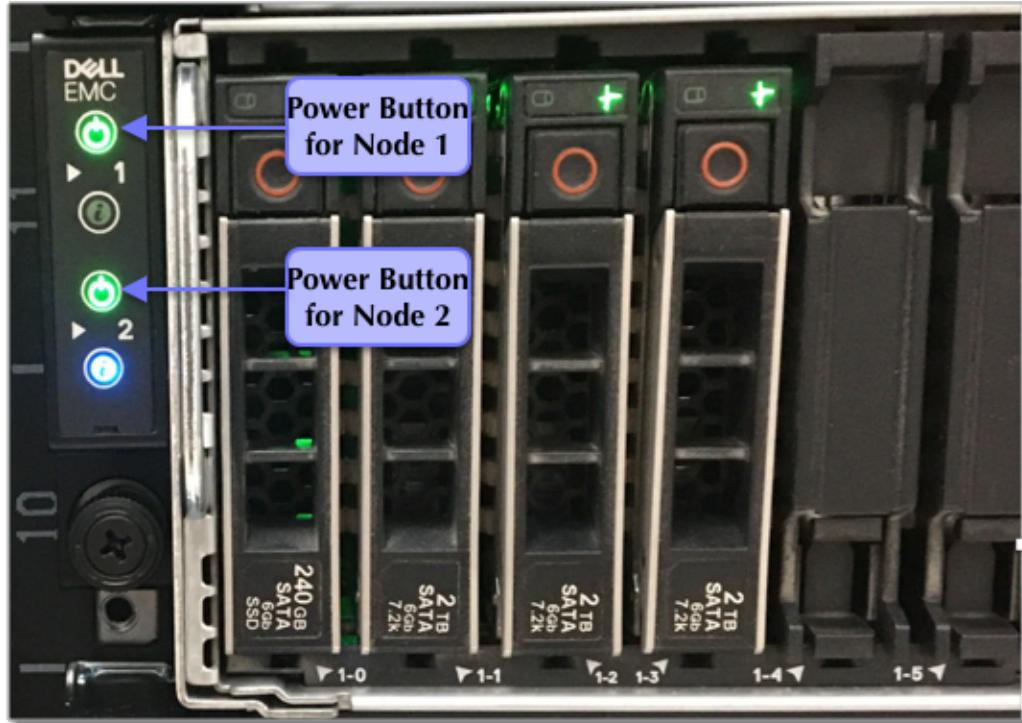
Dell-provided display to VGA adapter



Step 4: Turn on nodes

Turn on power for the nodes by pressing the power button for each one; see [Appliance Power Button \[See page 8\]](#).

Appliance Power Button



Note: There is one power button for each node.

Configure the management settings

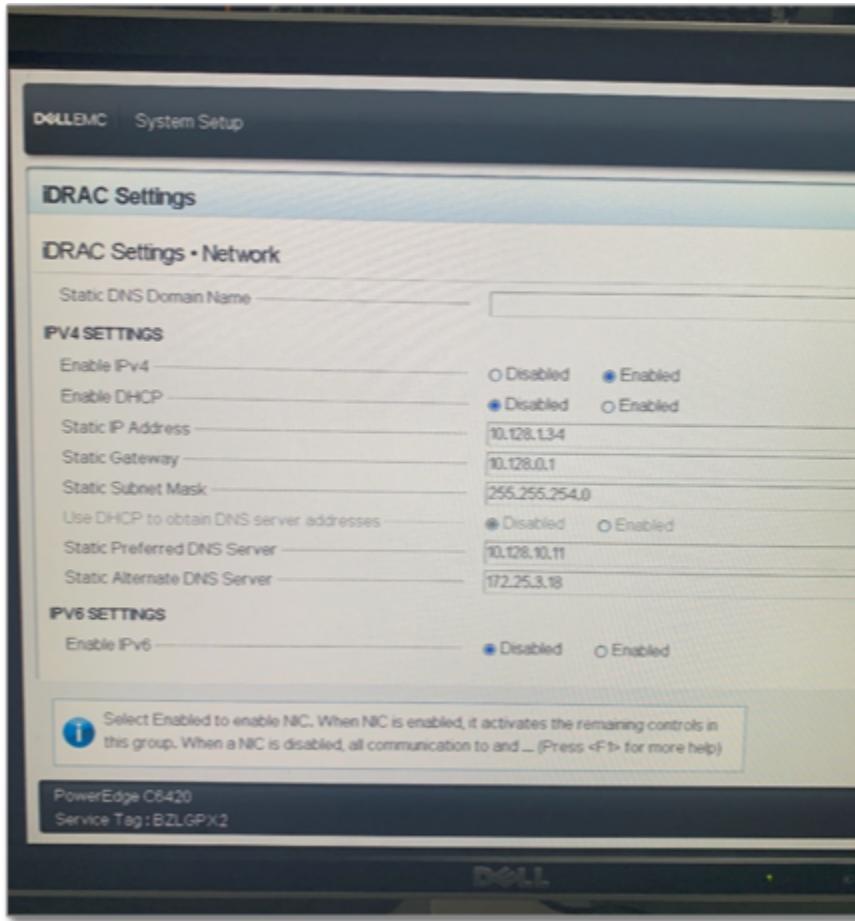
Next, input your specific network information to configure the management settings. Refer to [Dell Management Configuration \[See page 10\]](#). If you need additional guidance, view [Dell Support \[See page 0\]](#) for this product.

1. **Open the iDRAC settings modal** Before the node boots, a screen appears on your monitor with several options. Click F11 to enter the Boot Manager.
2. **Press F2** Click F2 when the option to do so appears on your screen.
3. **Select iDRAC** In the Bios setup screen, there are several options. Select **iDRAC** to configure your iDRAC settings.
4. **Select network configuration** From the iDRAC settings options, select **network**.
5. **Fill out the iDRAC settings form** Add your specific network information for the IP address, Gateway, and Netmask in the empty boxes. DNS information is optional. Refer to your ThoughtSpot site survey for a quick reference, and ask your network administrator for help if you have not filled out the site survey yet.

- For **Enable IPv4**, select **enabled**.
- For **Enable DHCP**, select **disabled**.
- For **Enable IPv6**, select **disabled**.

6. **Save changes and reboot** Follow the prompts on the monitor to save changes to the management settings form, exit, and reboot the system.
7. **Log into ThoughtSpot** When the system reboots, the login page appears. Log in as an administrator. Ask your network administrator if you do not know the admin credentials.

Dell Management Configuration



Configure Nodes on the Command Line

Once you have connected the appliance, a command line appears on your console. Configure the nodes on this command line.

Step 1: Get a list of nodes to configure

Make sure you have logged into your cluster. If you have not, use admin credentials to log into your cluster. Then, run the `tscli cluster get-config` command to get a list of the nodes to configure for the new cluster. Redirect it to the file `nodes.config`. You can find more information on this process in the [nodes.config file reference \[See page 0\]](#).

```
$ 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 \[See page 0\]](#).
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 [\[See page 11\]](#). Run `$ cat nodes.config | tscli cluster set-config`.

- If the command returns an error, refer to [set-config error recovery \[See page 12\]](#).

Set-config

```
$ 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 \[See page 11\]](#).

```
$ 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` on the command line 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). Make sure you can connect to ThoughtSpot remotely. If you can, you can run the installer on your local computer.

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.

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’ with your specific hostname.
- Replace ‘file-name’ with 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 ouput 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. 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`.

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
```

3. Finalize Installation

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

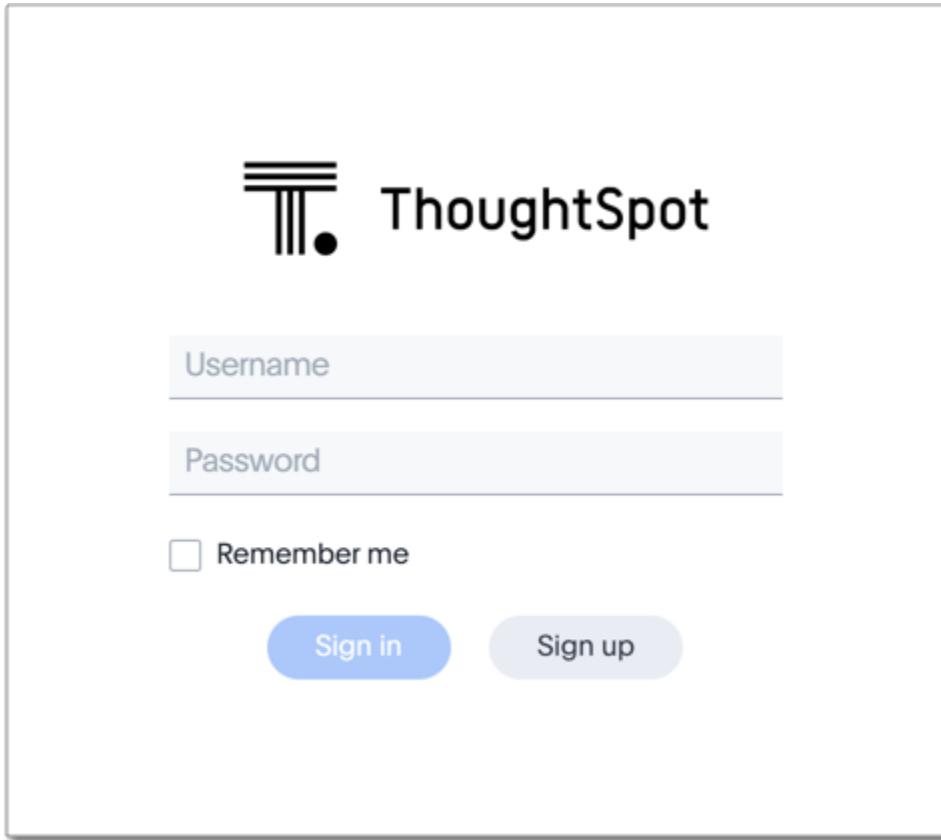
Follow these steps:

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 \[See page 17\]](#), 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 \[See page 0\]](#)
- [Parameters of the `nodes.config` file \[See page 0\]](#)
- [Using the `cluster create` command \[See page 0\]](#)
- [Parameters of the `cluster create` command \[See page 0\]](#)
- [Cable Reference \[See page 0\]](#)
- [ThoughtSpot Documentation \[See page 0\]](#)
- [Contact Support \[See page 0\]](#)

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 \[See page 0\]](#) 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 \[See page 0\]](#) 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:

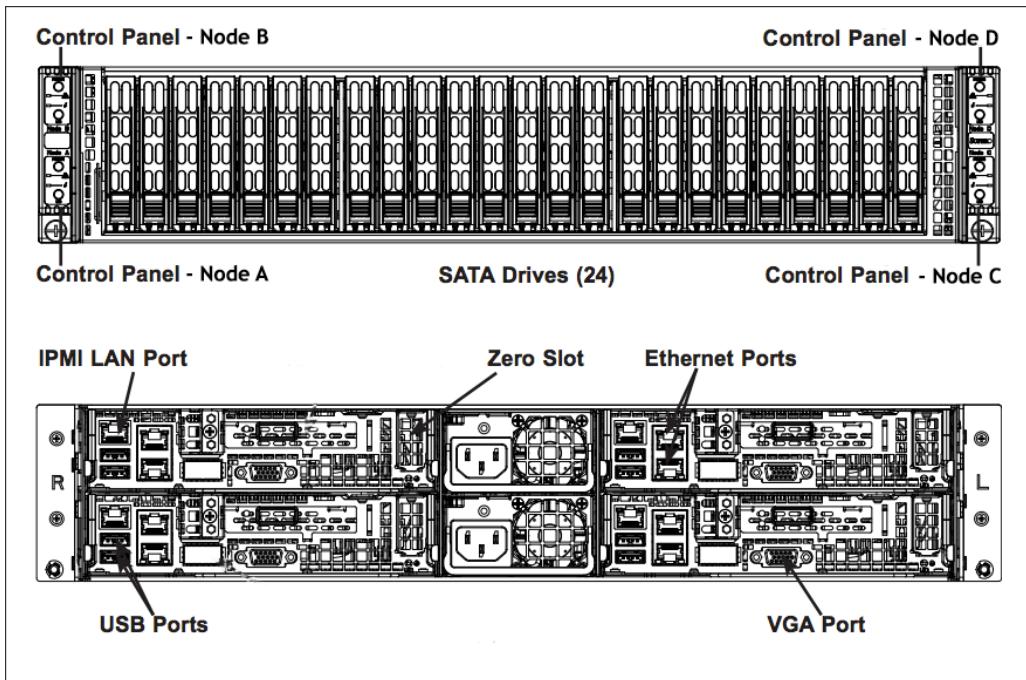
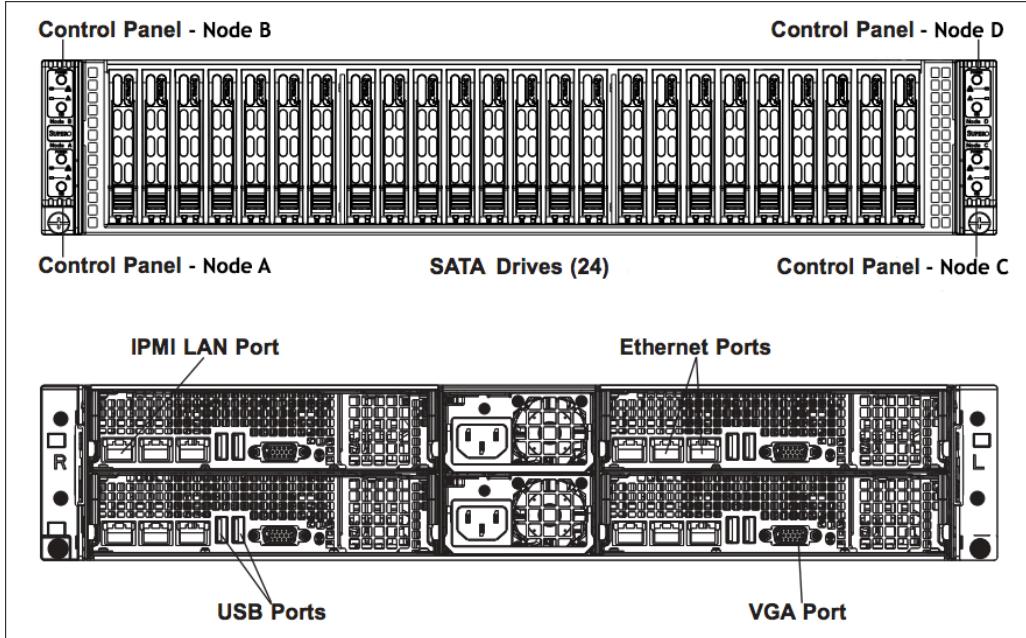
Details	Haswell
Dimensions	2 RU chassis (17.25 x 3.47 x 28.5 in.)
# of nodes	Populated with 1 to 4 nodes
Node specifica-tions	Each node is independent and consists of a server board (removable from rear), 1x 200GB SSD, 3x 2TB HDD
Max power con-sumption	2000 W
Required power input	200-240 / 11.8 - 9.8A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



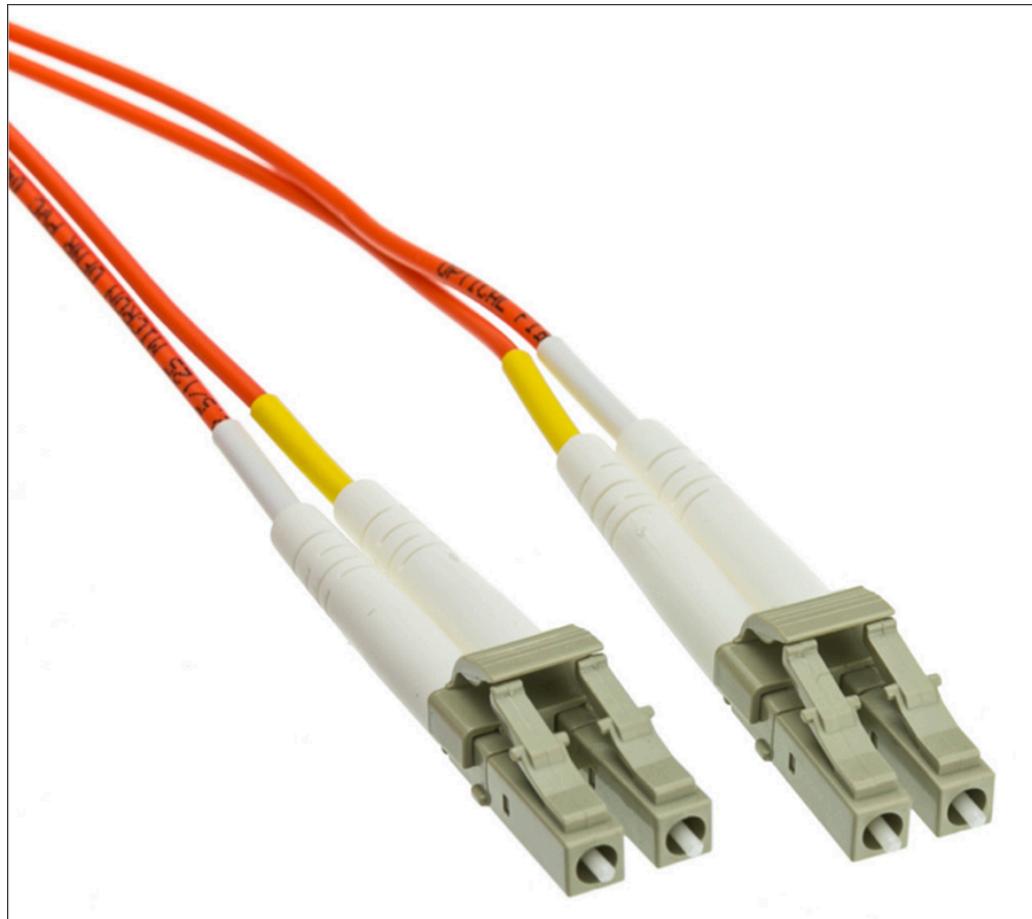
Cable networking

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.

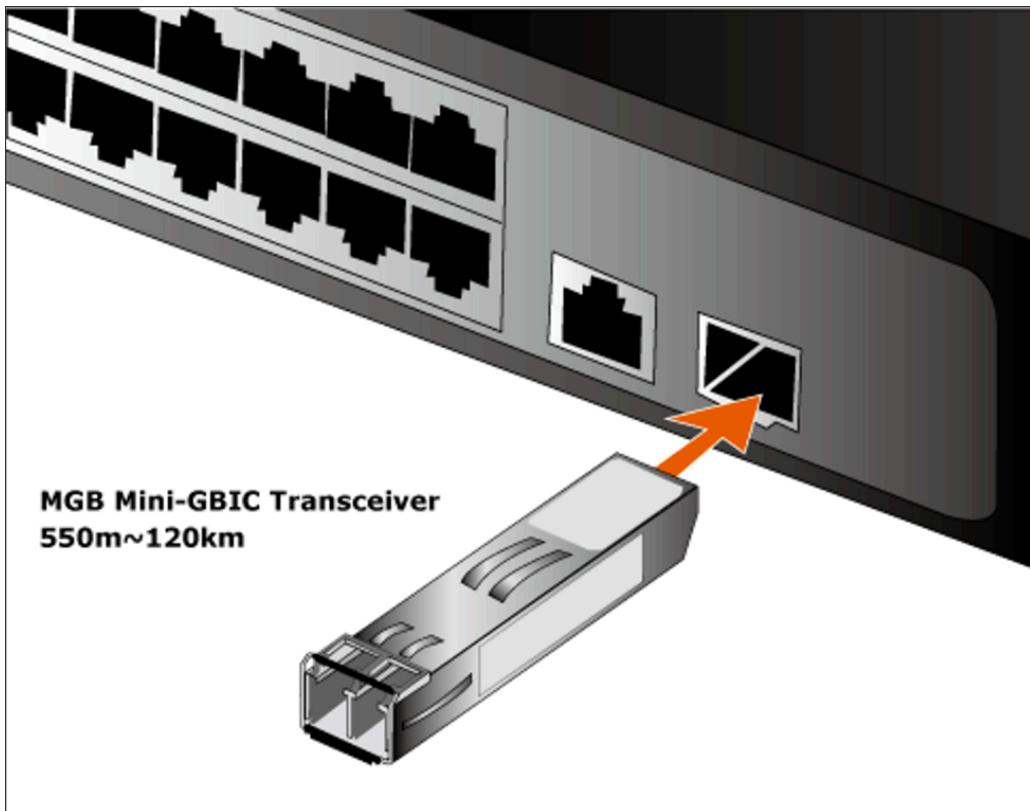


These cables require gigabit interface converters (GBICs), SFP+ form factor.

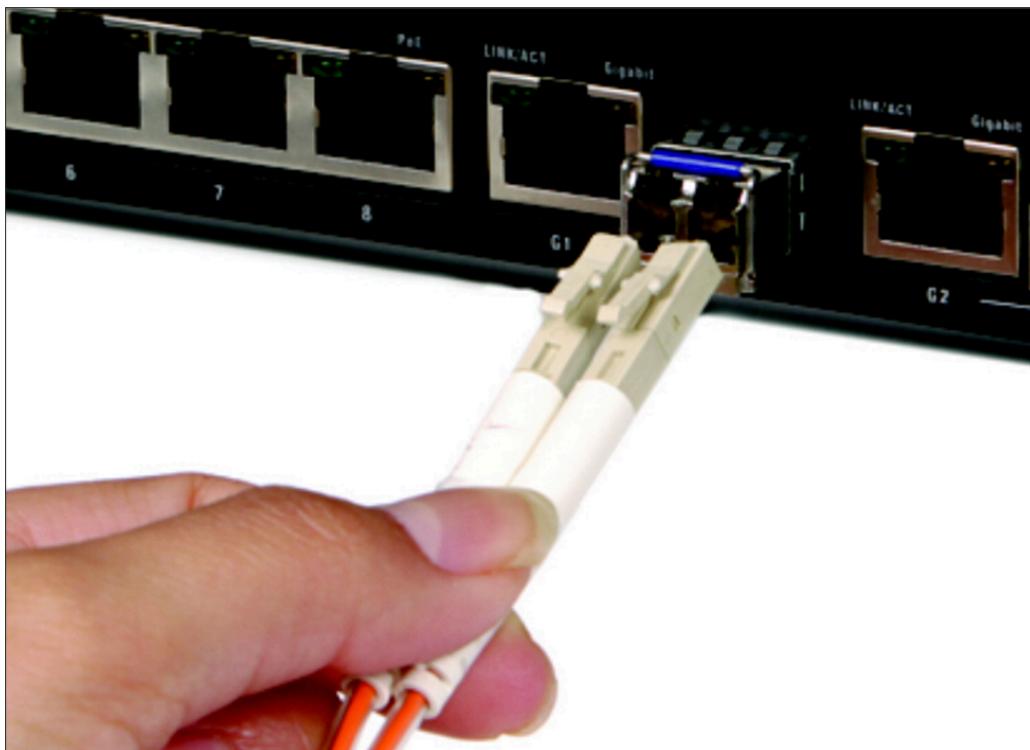
Remember: ThoughtSpot does not supply cables or GBICs



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.

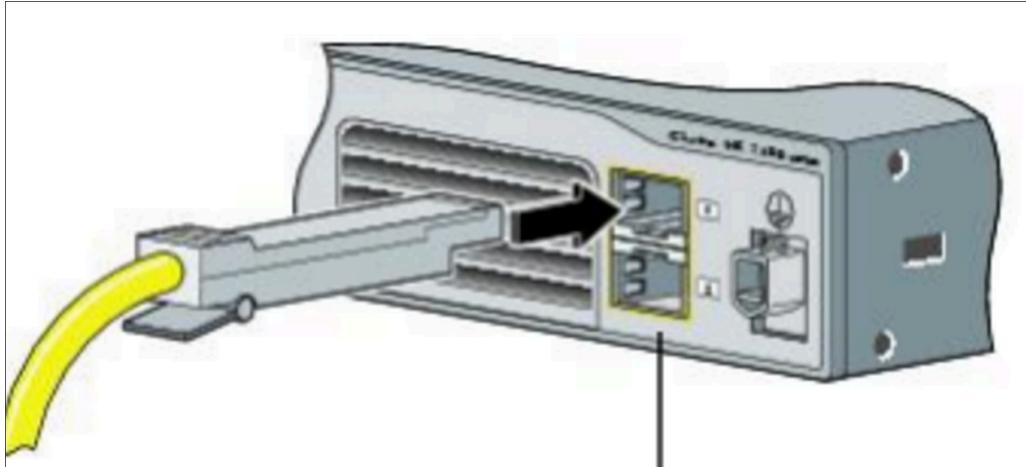


Option 2 - DAC/Twinax cables

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

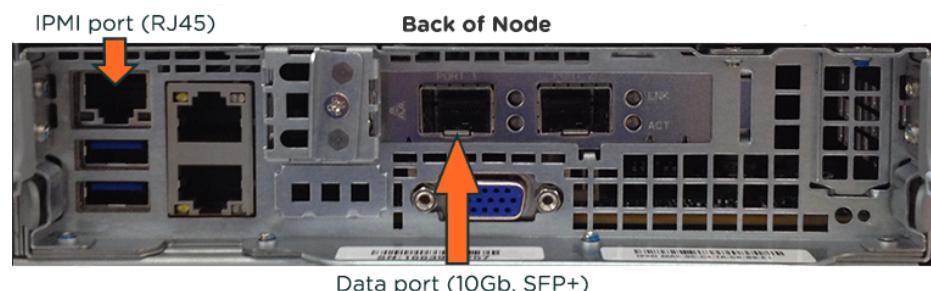
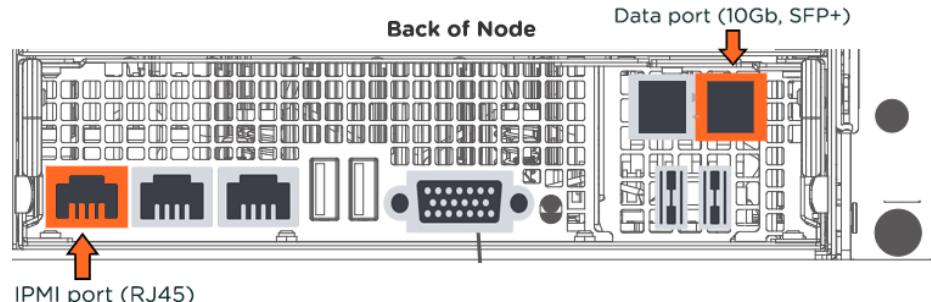
The ThoughtSpot appliance comes pre-installed with all the required software. Network settings on the appliance are required prior to using the appliance. Reference ThoughtSpot's site survey for the information specific to the customer's network environment that is required to configure the appliance.

- 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.

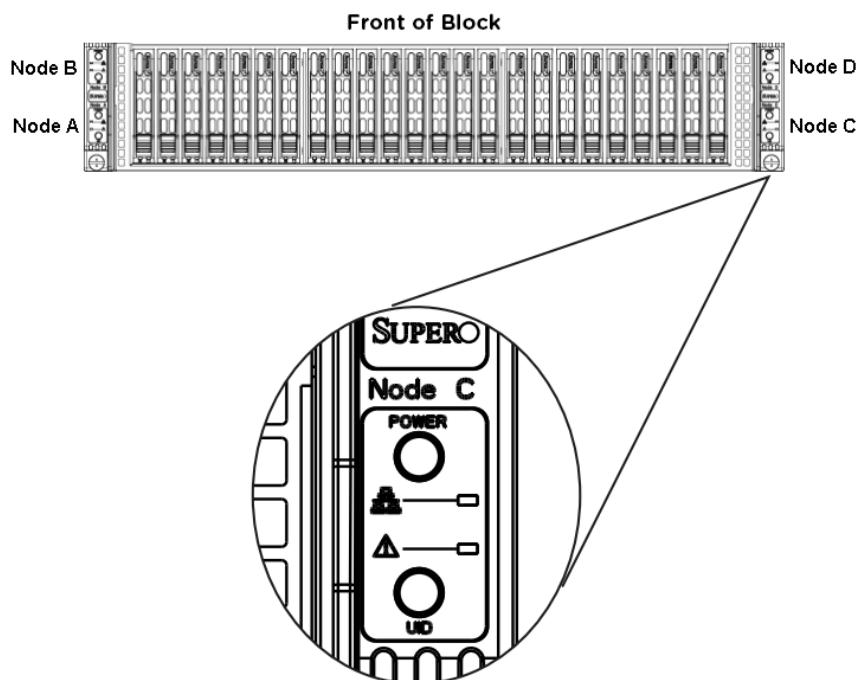
Installation procedure

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.

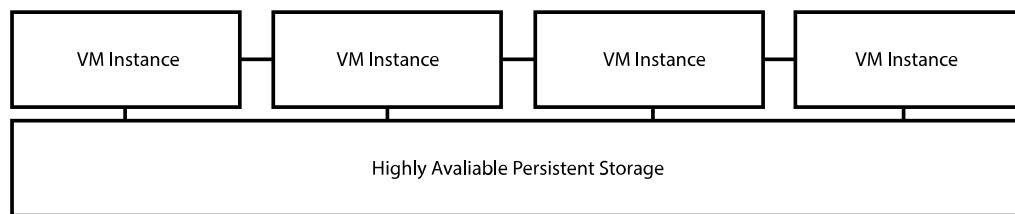
Cloud overview

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

- [Amazon Web Services \(AWS\) EC2 \[See page 33\]](#)
- [Microsoft Azure \[See page 0\]](#)
- [Google Cloud Platform \(GCP\) \[See page 43\]](#)

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 \[See page 0\]](#)
- [Azure instance types \[See page 0\]](#)

- GCP instance types [See page 43]

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](#) [See page 0].

For more information about automating your cloud deployment, read [Deploying ThoughtSpot in the Cloud Using Terraform and Ansible](#) [See page 0].

AWS configuration options

Summary: Your 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 \[See page 0\]](#).

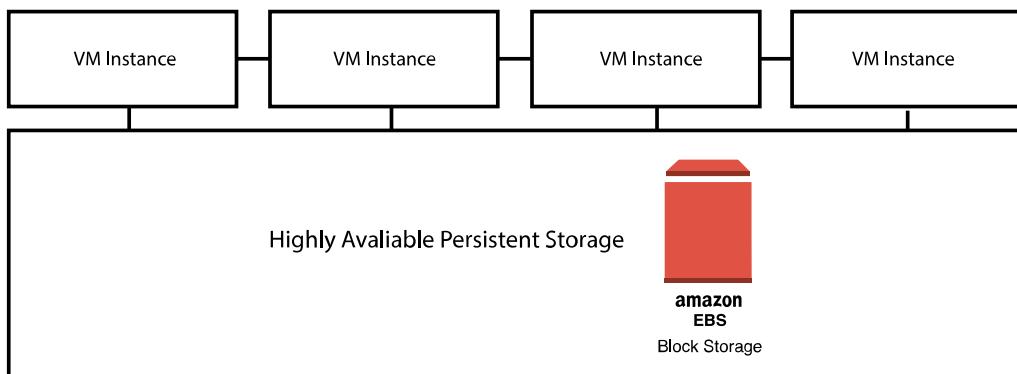
All AWS VMs in a ThoughtSpot cluster must be in the same availability zone (and therefore, also in the same region). ThoughtSpot does not support deploying VMs in the same cluster across availability zones. For more information, see [Regions and Availability Zones \[See page 0\]](#) in Amazon's AWS documentation.

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.

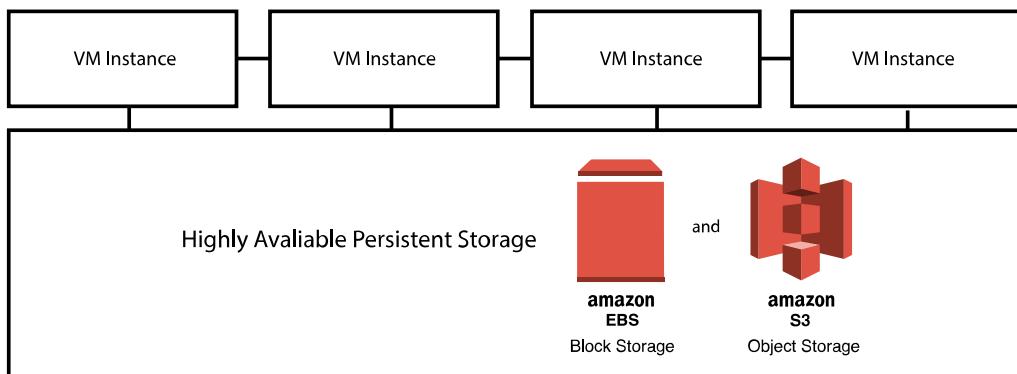
For example: If you were deploying a total cluster data size of 1 TB using the standard r5.16xlarge instance type, you would need 4 instances (VMs), because the per-VM user data capacity of that instance type is 250 GB. If you were deploying EBS-only data volumes, you would need 2x1 TB data volumes per 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 \[See page 0\]](#)
- [EC2 pricing \[See page 0\]](#)
- [EBS pricing \[See page 0\]](#)
- [Placement groups \[See page 0\]](#)

Set up ThoughtSpot in AWS

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 provisions 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? \[See page 0\]](#) in Amazon's AWS documentation.

Required AWS components

- An AWS VPC. For details, see [VPC and Subnets \[See page 0\]](#) in Amazon's AWS documentation.
- A ThoughtSpot AMI. For details, see the next section.
- AWS security groups. For required open ports, see [network policies \[See page 0\]](#).
- AWS VM instances. For instance type recommendations, see [ThoughtSpot AWS instance types \[See page 0\]](#).
- 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 \[See page 0\]](#).
- 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

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 \[See page 0\]](#)

[0\]](#) 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 varies. Please check [ThoughtSpot AWS instance types \[See page 0\]](#) 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 \[See page 0\]](#) to find out if your specific cluster size can 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

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.

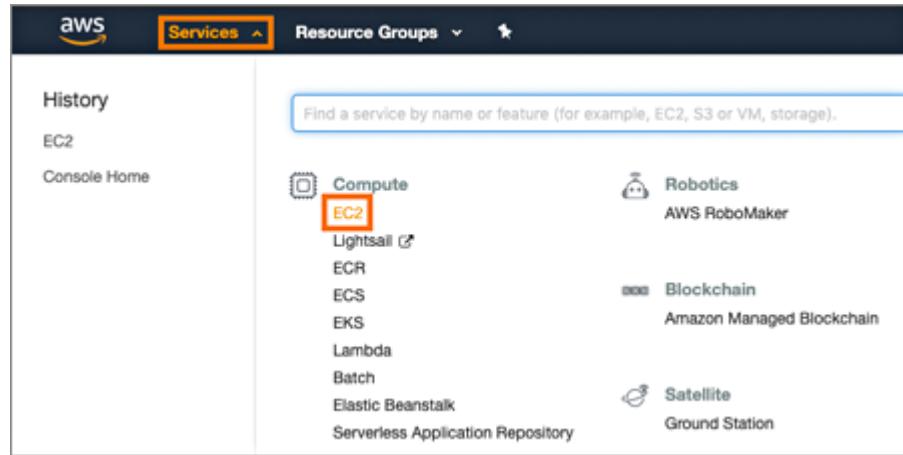
For more information on encryption supported with AWS:

- For EBS, see [Amazon EBS Encryption \[See page 0\]](#) in Amazon's AWS documentation.
- For S3, see [Amazon S3 Default Encryption for S3 Buckets \[See page 0\]](#) in Amazon's AWS documentation.

Setting up your ThoughtSpot cluster

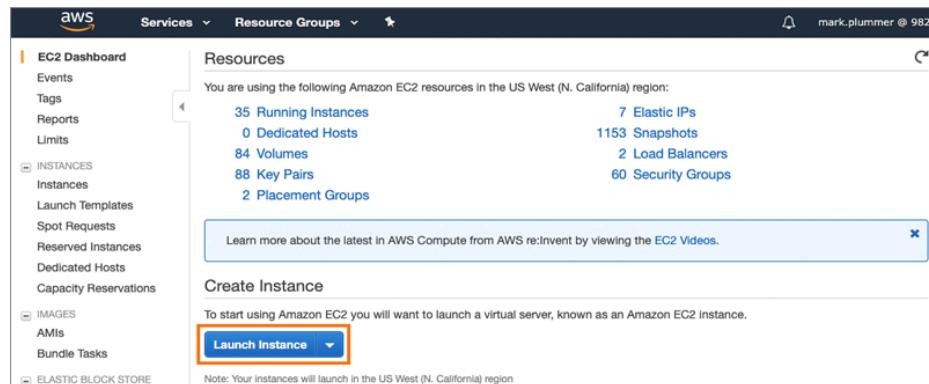
To set up a ThoughtSpot cluster in AWS, do the following:

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 \[See page 0\]](#).)
 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 sets 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.
- 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 \[See page 0\]](#).
 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 \[See page 0\]](#) 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 appears 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 \[See page 0\]](#).

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 \[See page 0\]](#).

Related information

[EC2 Best Practices \[See page 0\]](#)

[Loading data from an AWS S3 bucket \[See page 0\]](#)

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 Persistent Storage, see [Zonal Persistent SSD disks \[See page 0\]](#) in Google's Cloud documentation.

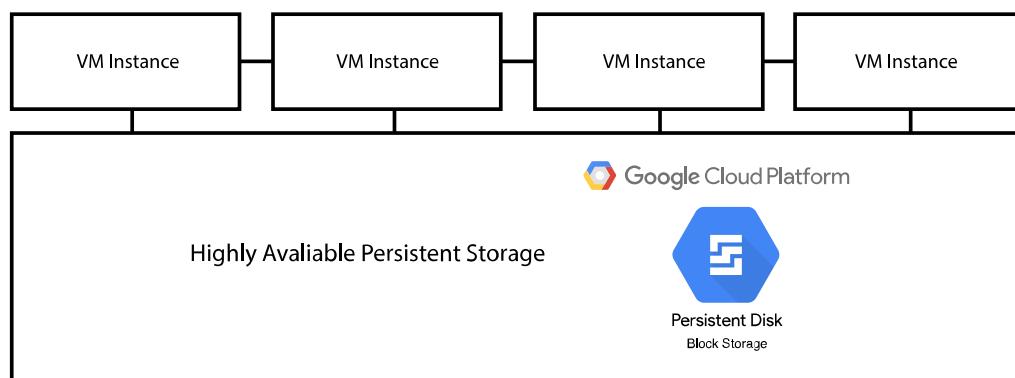
For more information about Google Cloud Storage, see [Cloud Storage Buckets \[See page 0\]](#) in Google's Cloud documentation.

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 \[See page 0\]](#) in Google's Cloud documentation.

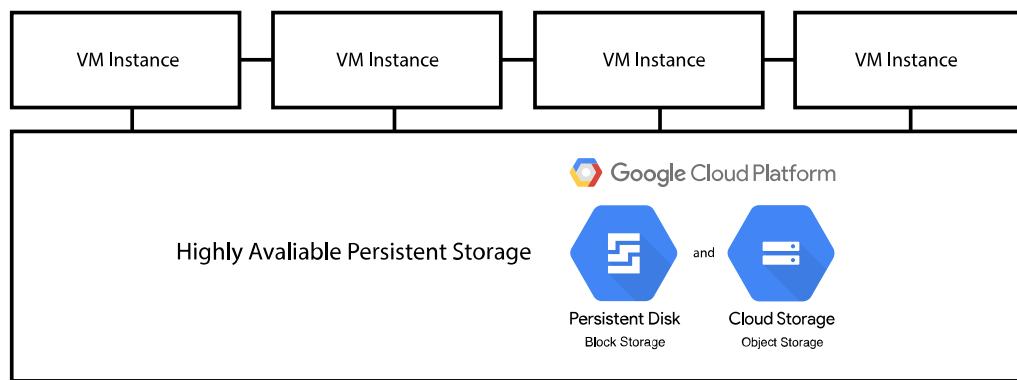
ThoughtSpot GCP instance types

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 ThoughtSpot in GCP

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 \[See page 0\]](#).

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 \[See page 0\]](#) to find out if your specific cluster size will benefit from the GCS storage option.

1. Sign in to the [Google Cloud Console \[See page 0\]](#).
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 \[See page 0\]](#).
2. Go to the Compute Engine dashboard, and select the associated ThoughtSpot project.

The screenshot shows the Google Cloud Platform dashboard for the project 'Thoughtspot ENG'. The left sidebar lists various services: Marketplace, Billing, APIs & Services, Support, IAM & admin, Getting started, Security, COMPUTE (App Engine, Compute Engine, Kubernetes Engine, Cloud Functions), and STORAGE. The 'Compute Engine' section is highlighted with a yellow box. The main content area displays 'Project info' (Project name: Thoughtspot ENG, Project ID: thoughtspot-eng, Project number: 567354788853) and a list of services: App Engine (All services normal), Compute Engine (28 instances, All services normal), Cloud Storage (All services normal), Cloud Functions (All services normal), Cloud SQL (All services normal), and API APIs (Requests (requests/sec)). A 'Compute Engine' card is also highlighted with a yellow box.

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.

The screenshot shows the 'Create an instance' dialog for the Compute Engine. The left sidebar lists options: VM instances (selected), Instance groups, Instance templates, Sole tenant nodes, Disks, Snapshots, Images, TPUs, Committed use discounts, Metadata, and Health checks. The main form has fields for Name (my-ts-instance-2), Region (us-west1 (Oregon)), Zone (us-west1-a), Machine type (8 vCPUs, 30 GB memory), Container (checkbox for deploying a container image), and Boot disk (New 10 GB standard persistent disk, Image: Debian GNU/Linux 9 (stretch)). A 'Customize' button is highlighted with a yellow box.

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)

[←](#) Create an instance

Name [?](#)
my-ts-instance-2

Region [?](#)
us-west1 (Oregon) ▾

Zone [?](#)
us-west1-a ▾

Machine type
Customize to select cores, memory and GPUs.

[Basic view](#)

Cores
64 vCPU 1 - 96

Memory
416 GB 57.6 - 624

Extend memory [?](#)

CPU platform [?](#)
Intel Skylake or later ▾

Automatic
Intel Skylake or later
Intel Broadwell or later

Either of these are preferred platforms

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

Boot disk

Select an image or snapshot to create a boot disk; or attach an existing disk

OS images Application images **Custom images** Snapshots Existing disks

Show images from

centos-golden-20181023-092dd2d2265-prod
Created from Thoughtspot ENG on Oct 23, 2018, 5:43:16 AM

thoughtspot-1533254471
Created from Thoughtspot ENG on Aug 2, 2018, 5:01:41 PM

Can't find what you're looking for? Explore hundreds of VM solutions in [Marketplace](#)

Boot disk type Size (GB)

Select **Cancel**

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**).

Boot disk

New 250 GB standard persistent disk
Image
thoughtspot-1536332609 Change

Identity and API access

⚠ You don't have permission to view the service accounts in this project

Service account Compute Engine default service account

Access scopes

- Allow default access
- Allow full access to all Cloud APIs
- Set access for each API

Firewall

Add tags and firewall rules to allow specific network traffic from the Internet

Allow HTTP traffic
 Allow HTTPS traffic

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**.

Management	Security	Disks	Networking	Sole Tenancy
------------	----------	--------------	------------	--------------

Boot disk

Deletion rule

Delete boot disk when instance is deleted

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

Additional disks (Optional)

+ Add new disk + Attach existing disk

Less

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)

Type

Source type

Mode
 Read/write
 Read only

Deletion rule
When deleting instance
 Keep disk
 Delete disk

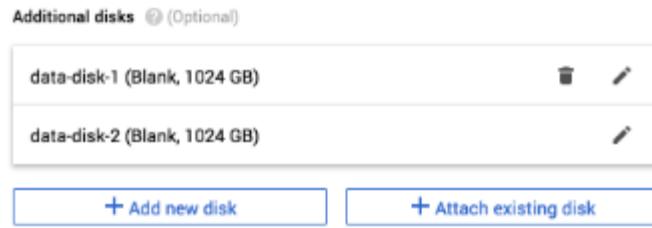
Size (GB)

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



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 \[See page 0\]](#).

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 \[See page 0\]](#).

Related information

[Connecting to Google Cloud Storage buckets \[See page 0\]](#)

[Loading data from a GCP GCS bucket \[See page 0\]](#)