



Node Setup Guide

Version 5.2 May 17, 2019

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

ThoughtSpot, Inc. 1 Palo Alto Square
Building 1, Suite 200
Palo Alto, CA 94306

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 | 2 |
| | |
| Hardware appliance | |
| Hardware appliance overview | 3 |
| Supported hardware | 6 |
| Cable networking | 9 |
| Set up and start the appliance | 14 |
| | |
| Cloud | |
| Cloud overview | 17 |
| AWS | |
| AWS cluster overview | 20 |
| Configuration options | 21 |
| Set up AWS for ThoughtSpot | 23 |
| Deploy using CloudFormation CLI | 27 |
| Set up high availability | 29 |
| Azure | |
| Azure cluster overview | 31 |
| Configuration options | 32 |
| Set up Azure for ThoughtSpot | 33 |
| GCP | |
| GCP cluster overview | 43 |
| Set up GCP for ThoughtSpot | 44 |
| VMware | |
| Configuration overview | 52 |
| Set up VMware for ThoughtSpot | 54 |
| Network policies..... | 59 |
| Contact support..... | 70 |

Overview

Congratulations on purchasing ThoughtSpot. You can install a ThoughtSpot cluster on a hardware appliance, cloud service, or VMware appliance. Your ThoughtSpot installation cannot mix node types, your installation must be homogenous, so, for example either hardware or VMware nodes. You can, however, have an AWS 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](#)

Once you've configured 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 will need to supply the following items, as they will not be included with your ThoughtSpot appliance:

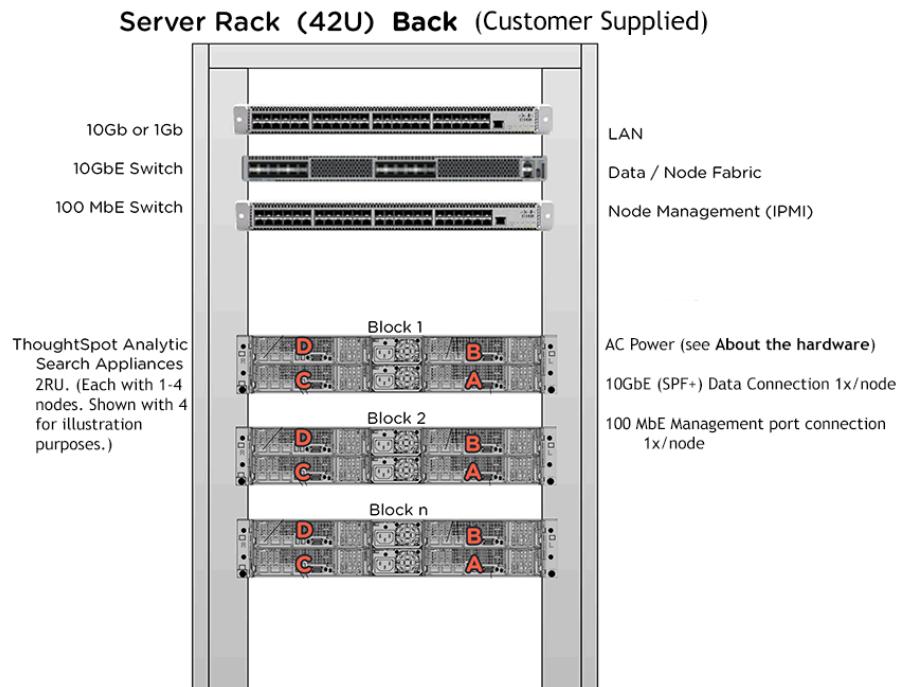
- Data center with proper cooling
- 2U of rack space per appliance (post depth 26.5" - 36.4")
- AC power **Attention:** Refer to [Hardware details](#) for power input requirements.
- 10GbE infrastructure (switch) - 1x port required / node
- 100MbE infrastructure (switch) - 1x port required /node
- Network cable Cat 5e/6 (node management)¹
- 10G connection: SFP+ for switch side²

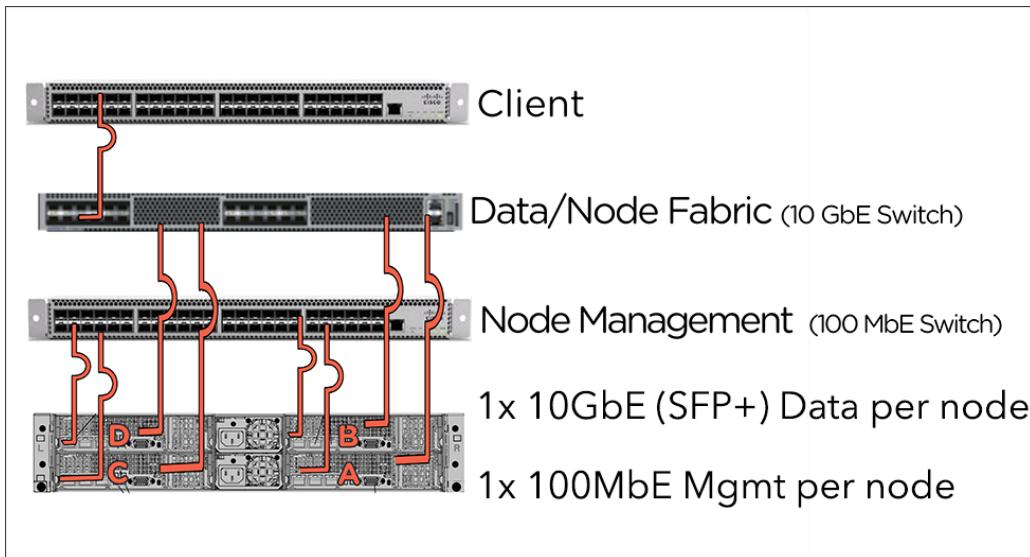
1. One 5m CAT 5e/6 network cable, per node, is provided with the appliance for management port connection. Customer supplied cable can be used if preferred.

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.





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 will 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 to each other via 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 specifications as detailed below.

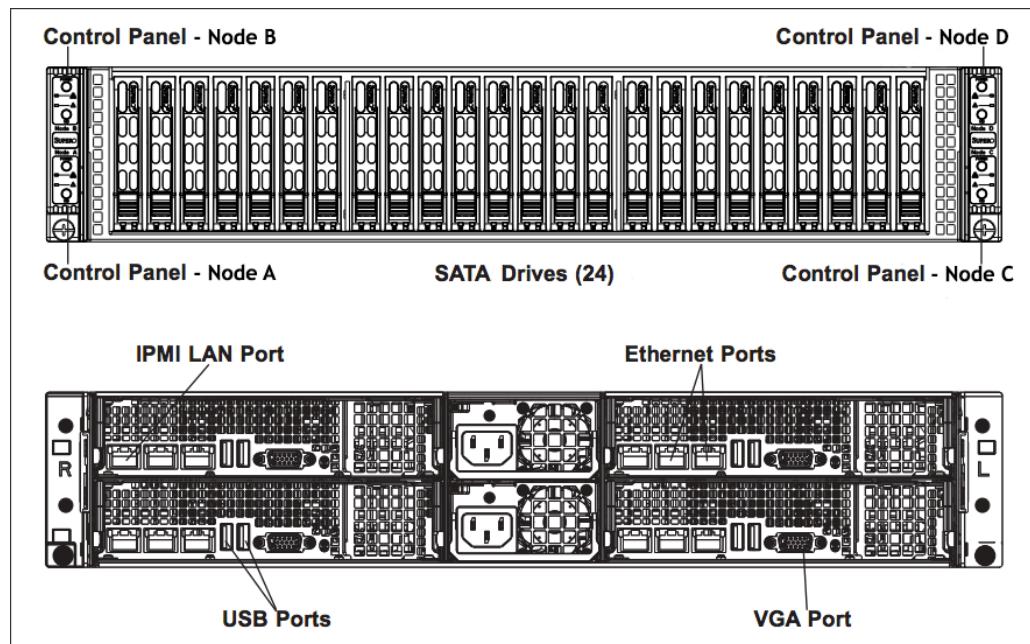
| Details | Haswell |
|-----------------------|---|
| Dimensions | 2 RU chassis (17.25 x 3.47 x 28.5 in.) |
| # of nodes | Populated with 1 to 4 nodes |
| Node specifications | Each node is independent and consists of a server board (removable from rear), 1x 200GB SSD, 3x 2TB HDD |
| Max power consumption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

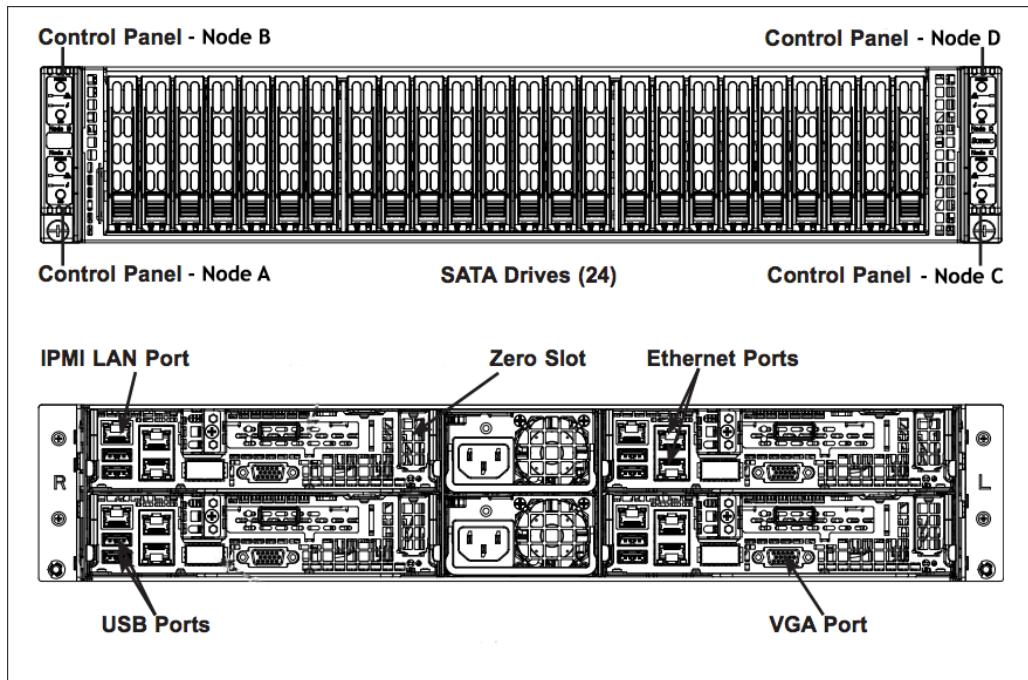
Note: ThoughtSpot deployments are no longer offered on Ivy Bridge platforms.

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 are shown 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.





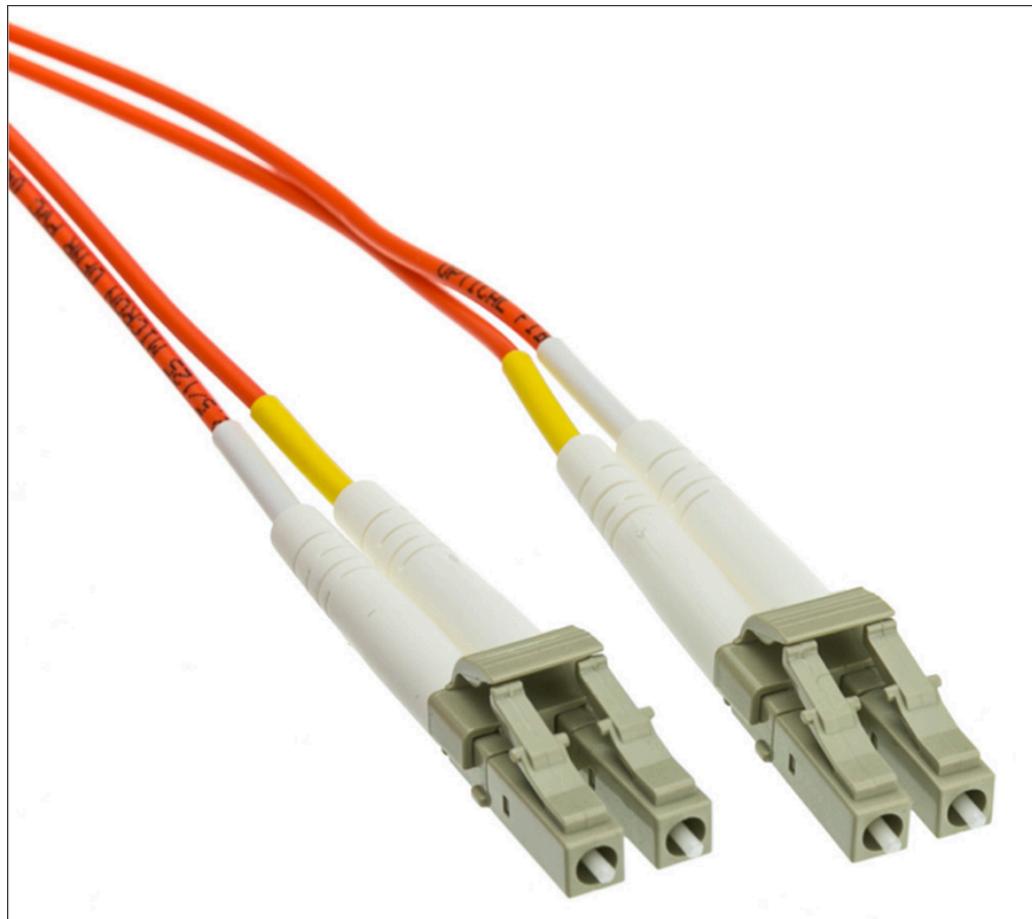
Cable networking

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.

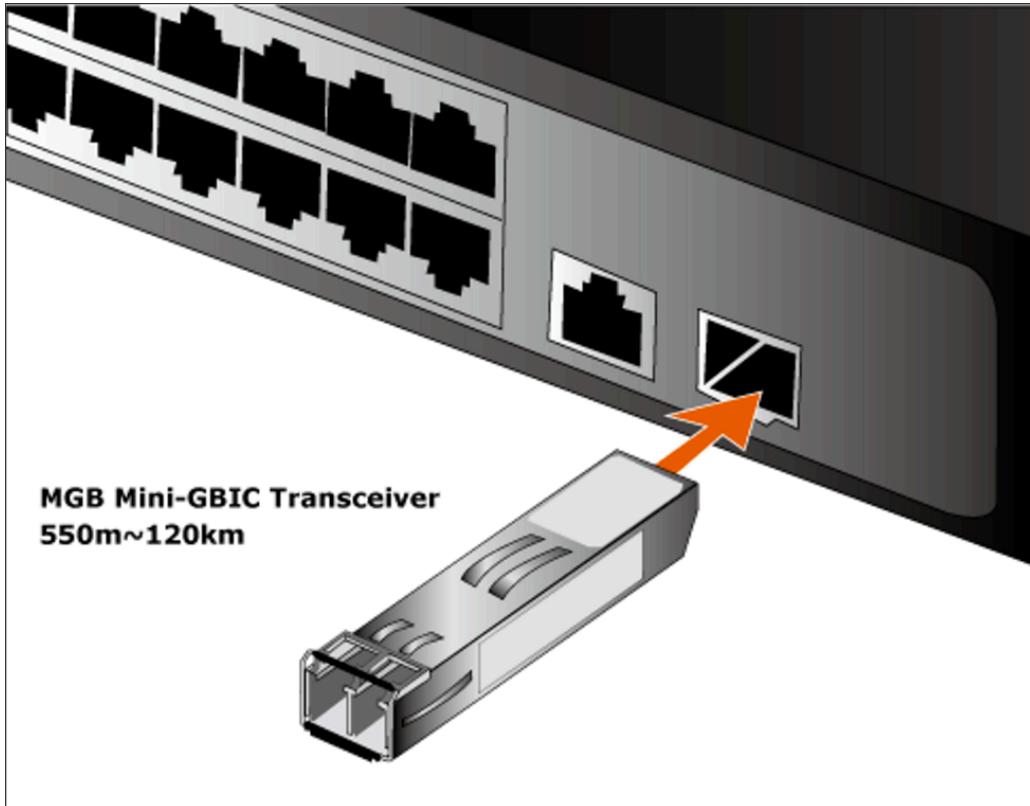


These cables require gigabit interface converters (GBICs), SFP+ form factor.

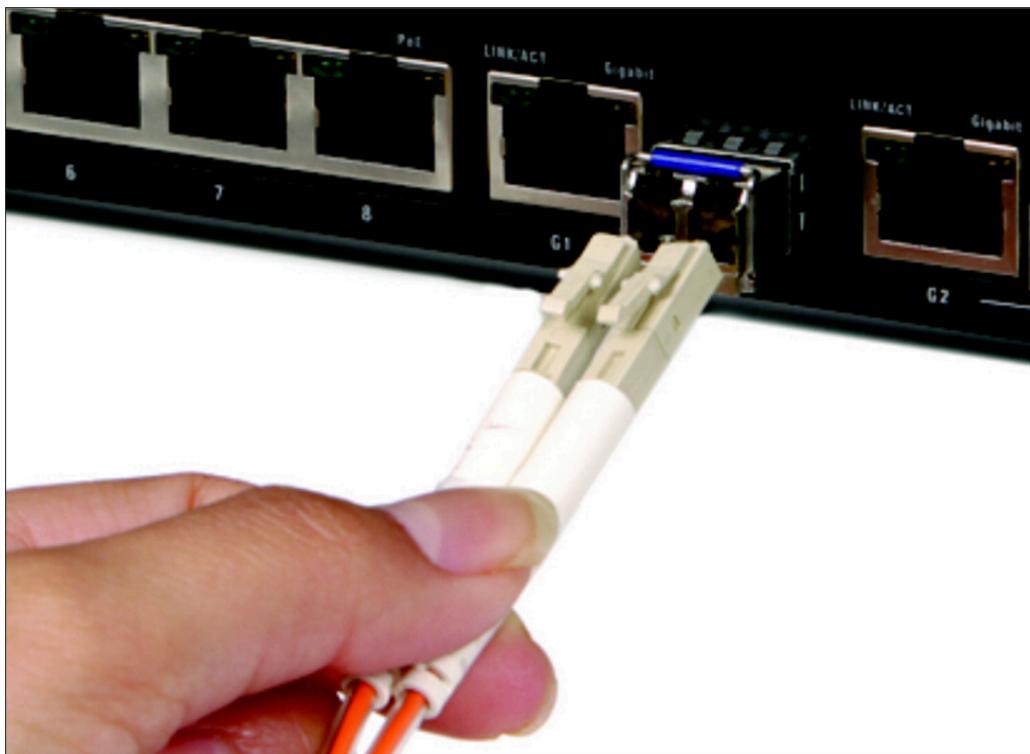
Remember: ThoughtSpot does not supply cables or GBICs



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.

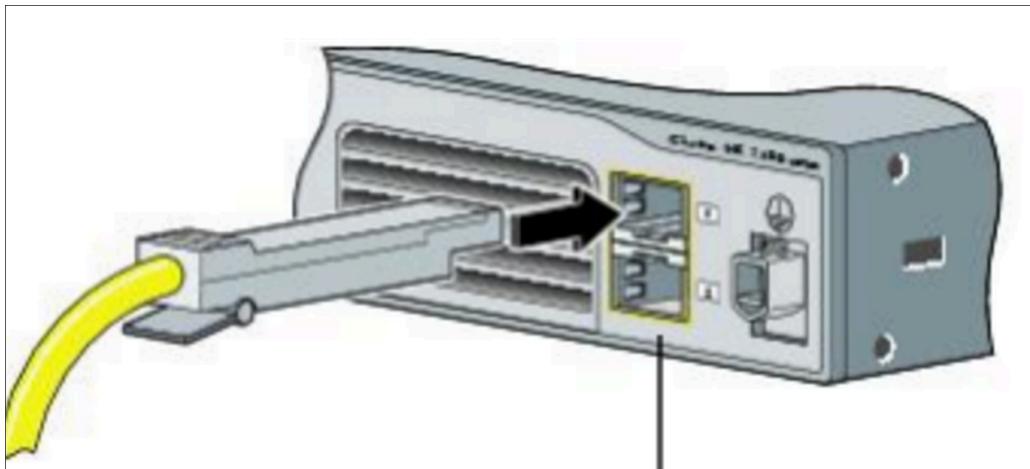


Option 2 - DAC/Twinax cables

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

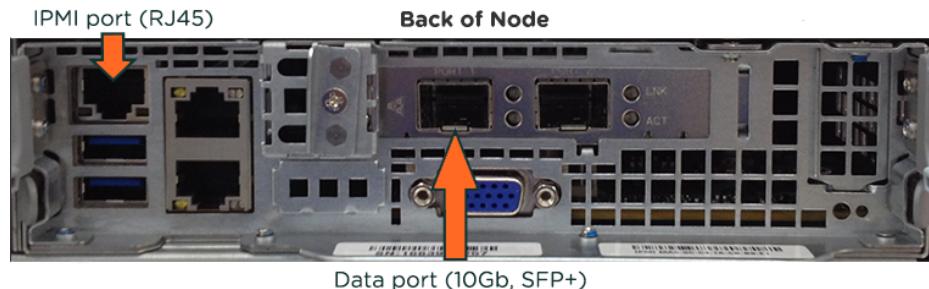
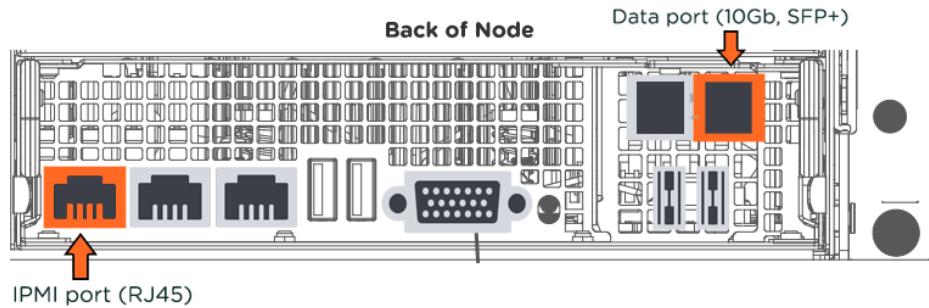
The ThoughtSpot appliance comes pre-installed with all the required software. Network settings on the appliance are required prior to using the appliance. Reference ThoughtSpot's site survey for the information specific to the customer's network environment that is required to configure the appliance.

- 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.

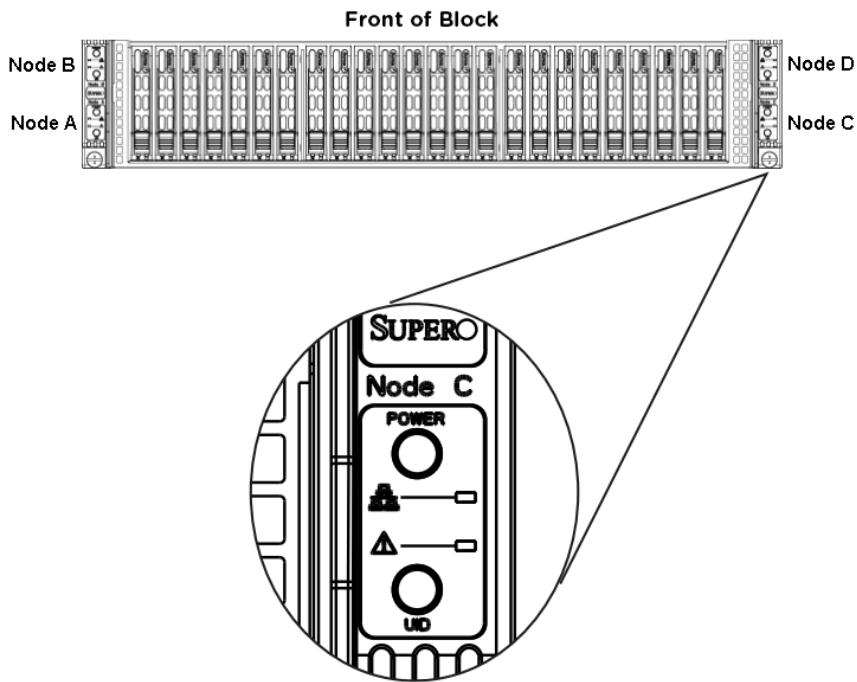
Installation procedure

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 **Back of Node** figures below is active. Only one 10GbE port connection is needed.



- 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. Once 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.

- After the appliance has been turned on, verify that both LEDs (IPMI and data NICs) on each network card are lit.
- 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.
- Log in as username admin, using the default password.
- Once logged in, 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 currently supports the following cloud deployments:

- [Amazon Web Services \(AWS\) EC2](#)
- [Microsoft Azure](#)
- [Google Cloud Platform \(GCP\)](#)

This page provides recommendations and best practices for ThoughtSpot cloud deployments, including how to set up your ThoughtSpot instance and how to reduce infrastructure costs.

ThoughtSpot cloud instance types

This table provides guidelines on how to set up ThoughtSpot on each cloud service.

| Cloud Provider | Data shape | Use case | | |
|----------------|-------------------------------|-------------------------|----------------------|---------------------------|
| | | Total cluster data size | Per VM Data capacity | Recommended Instance type |
| AWS | Standard (1KB/row) | Up to 2 TB | 250 GB | r4.16xlargea |
| | | >2 TB | 384 GB (Large) | r5.24xlargeb |
| Azure | Thin rows (<300 bytes/row) | Up to 100 GB | 100 GB | r4.8xlargec |
| | | Up to 20 GB | 20 GB | r4.4xlargec |
| Azure | Standard (1KB/row) | Any | 192 GB | m5.24xlargeb |
| | | >100 GB | 200 GB | E64sv3 |
| Azure | | Up to 100 GB | 100 GB | E32sv3c |

| | | Up to 20 GB | 20 GB | E16sv3c |
|-----|-------------------------------|-------------------------|--------|----------------|
| | Thin rows (<300 bytes/row) | Any | 120 GB | D64v3b |
| GCP | Standard | Up to 3 TB (1KB/row) | 208 GB | n1-highmem-64 |
| | | >3 TB | 312 GB | n1-highmem-96 |
| | | Up to 100 GB | 100 GB | n1-highmem-32c |
| | | Up to 20 GB | 20 GB | n1-highmem-16c |
| | Thin rows (<300 bytes/row) | Any | 180 GB | n1-standard-96 |

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

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 shown in the table above to help reduce cost of cloud infrastructure. These are instances with lower CPU/RAM sizes (16/32 vCPU and 128 GB/256 RAM). Additional 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, 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.

5. 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
```

AWS cluster overview

AWS can provide lots of memory and CPU for your ThoughtSpot instance, and it can be easily updated from development instances.

Your database capacity will determine the number of instances you'll need and the instance network/storage requirements. In addition, you can go with multiple VMs based on your dataset size.

The security group setting of your ThoughtSpot instance on AWS is up to you. You can find more information about which ports to open in the [network ports reference](#).

A little about AWS

AWS is a secure cloud services platform offered by Amazon. Using ThoughtSpot on AWS allows you to easily add instances as your dataset grows. You can do everything you'd normally want to do in a traditional data center with AWS. It features an on-demand delivery of IT resources and applications via the Internet with pay-as-you-go pricing.

Amazon EC2 is based on instance types and the region in which your instances are running. When you are connected to an instance, you can use it just like you use any other server. There is no minimum fee and you only pay for what you use.

Using Amazon EC2 lets you develop and deploy applications faster since there is no need to manage hardware. Therefore, it is easy to scale and manage computing capacity.

As persistent block level storage volumes, Amazon EBS helps with scaling your EC2 instances. Each EBS volume is automatically replicated to protect you from component failure, and offers low-latency performance.

AWS configuration options

ThoughtSpot engineering has performed extensive testing of the ThoughtSpot appliance on various Amazon Elastic Compute Cloud (EC2) and Amazon Elastic Block Store (EBS) configurations for best performance, load balancing, scalability, and reliability.

You can find information here on which configuration of memory, CPU, storage, and networking capacity you should be running for your instances. There are also details on how to configure your placement groups.

Hardware configurations

There is only one available hardware configuration for deploying ThoughtSpot on Amazon:

r4.16xlarge

Below are charts depicting the specifications for the configuration for EC2 and EBS requirements. Both EC2 and EBS requirements must be fulfilled to deploy on Amazon.

| Instance name | Data capacity | vCPUs | DRAM |
|---------------|---------------|-------|--------|
| r4.16xlarge | Up to 250 GB | 64 | 488 GB |

| Instance name | Data capacity | Root volume (SSD) | Data volume (SSD or HDD) |
|---------------|---------------|-------------------|--------------------------|
| r4.16xlarge | Up to 250 GB | 1 vol 200 GB | 2 vols 1 TB each |

ThoughtSpot software license sizes

ThoughtSpot only sells software licenses in multiples of 250 GB of data. So you can start with 250 GB, and add increments of 250 GB each time your data capacity needs increase. You can also choose to start off with more than 250 GB of data, as long as you know the best fit configuration for your data volume.

Lego blocks

If you aren't sure what kind of configuration you need, it might help to think of the hardware configurations in terms of simple Lego blocks. The r4.16xlarge size can be seen as a 250 GB block.

Note: ThoughtSpot does not support sizes other than r4.16xlarge.

Since the minimum data volume offered is 250 GB, you would need one r4.16xlarge block to match the data capacity. This scales linearly. So, 500 GB would require two r4.16xlarge blocks.

Regions, availability zones, and placement groups

AWS instances are configured to a location with regard to where the computing resources are physically located. You must specify a region, an availability zone, and below that, a placement group.

AWS nodes in a ThoughtSpot cluster must be in the same [availability zone](#) (and, therefore, also in the same region).

A [placement group](#) is a logical grouping of instances *within* a single availability zone. Placement groups are recommended for applications that benefit from low network latency, high network throughput, or both.

ThoughtSpot relies on high connectivity between nodes of a cluster, which is why creating a placement group is recommended. Being in same placement group will give you the best shot at the highest bandwidth across AWS EC2 instances and the lowest latencies. This will make the node-node network reach the closest AWS promised specs. Our default recommendation for a multi-instance setup requires a placement group since it works best for our application performance. Also, AWS will provide jumbo frames (9000 MTU) support in such situations, and they don't charge extra for being in the same placement group.

Having said that, ThoughtSpot will still work with EC2s in the cluster across placement groups in an availability zone.

Related information

- [EC2 instance types](#)
- [EC2 pricing](#)
- [EBS pricing](#)
- [Placement groups](#)

Set up AWS for ThoughtSpot

After you've determined your configuration options, you must set up your virtual machines (VMs) on AWS using a ThoughtSpot Amazon Machine Image (AMI). ThoughtSpot will share this AMI with you and work with you to set up the VMs in the AWS environment.

Overview of ThoughtSpot setup in AWS

The high-level process for setting up ThoughtSpot in AWS is the following:

1. Contact ThoughtSpot support to get access to ThoughtSpot AMIs.
2. Choose a VM instance configuration recommended by ThoughtSpot.
3. Work with ThoughtSpot to set up your ThoughtSpot cluster.
4. Open the required network ports for communication for the nodes in your cluster and end users.
5. [Deploy your AWS instance using CloudFormation CLI](#).

About the ThoughtSpot AMI

The ThoughtSpot AMI comes provisioned with the custom ThoughtSpot image to make deploying simple. 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. An AMI includes the following:

- A template for the root volume for the instance (for example, 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's launch.

To get the latest version of the ThoughtSpot AMI, contact ThoughtSpot and provide your AWS account ID and region where the VMs will be hosted. ThoughtSpot will then provide the current ThoughtSpot base AMI to you.

The ThoughtSpot AMI has specific applications on a CentOS base image. The EBS volumes required to install ThoughtSpot in AWS are included in the AMI. When you launch an EC2 instance from this image, the EBS volumes are automatically sized and provisioned. The base AMI includes 200 GB (xvda), 2X400 GB (xvdb), and SSD (gp2). It contains the maximum number of disks so that it handle the full load of the VM.

Prerequisites

ThoughtSpot instances on AWS need AWS EC2 instances to be provisioned in the AWS account before ThoughtSpot can be installed and launched. Please make sure you follow the guidelines below for setting up your EC2 details:

- Contact ThoughtSpot support to get access to ThoughtSpot AMIs
- Choose the appropriate EC2 instance type: See [ThoughtSpot cloud instance types](#) for supported instance types.
- Networking requirements: 10GbE 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

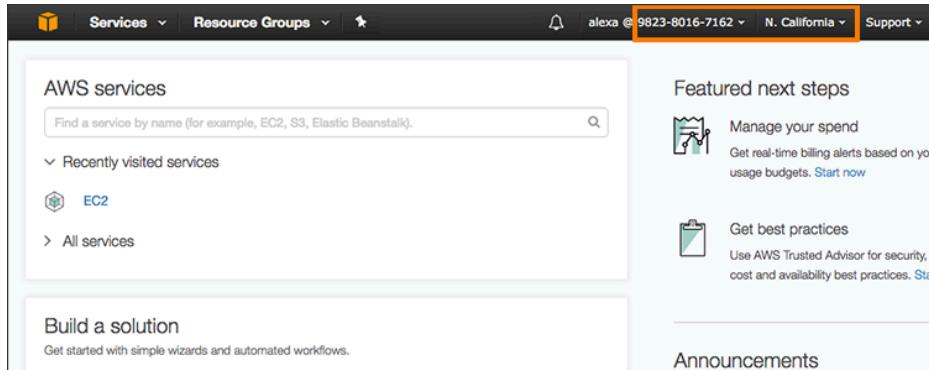
- external access may be required to bring data in/out of the VMs to your network.
- Number of EC2 instances needed: Based on the datasets, this number will vary. Please check [ThoughtSpot cloud instance types](#) for recommended nodes for a given data size.
 - Staging larger datasets (> 50 GB per VM), may require provisioning additional attached EBS volumes that are SSD (gp2).

Setting up your ThoughtSpot cluster in AWS

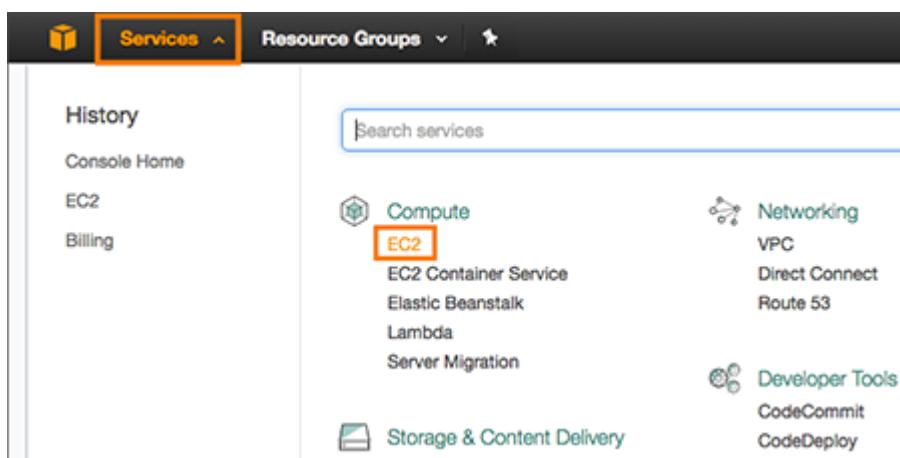
To set up a ThoughtSpot cluster in AWS, do the following:

- Log in to your AWS account from the [AWS Amazon sign in page](#).
- Provide ThoughtSpot support with your AWS account ID and the region where the VMs will be hosted. Support will grant you permissions and share the current ThoughtSpot base AMI with you.

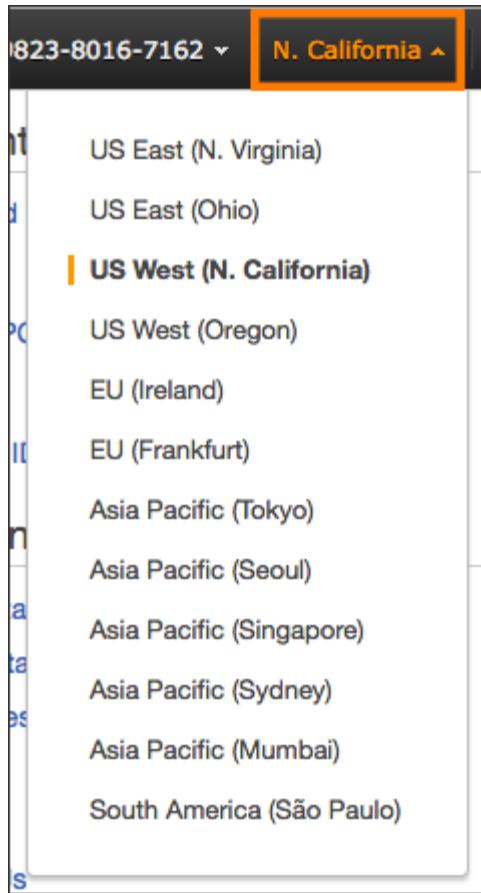
Note: You can find your account ID and region on the top-right corner of the AWS console.



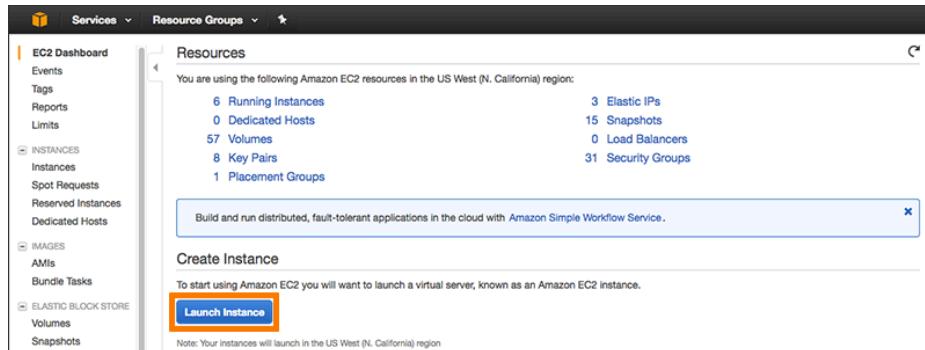
- Navigate to the EC2 service dashboard by clicking **Services**, then select **EC2**.



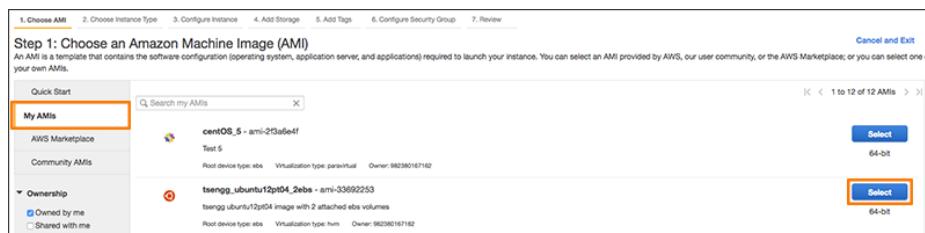
- Make sure your selected region is correct in the top right corner of the dashboard. If not, select a different region you would like to launch your instance in. Let ThoughtSpot support know if you change your region.



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



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



- On the Choose and Instance Type page, select a ThoughtSpot-supported instance type. (See

[ThoughtSpot cloud instance types \)](#)

8. Click **Next: Configure Instance Details**.
9. Configure the instances by choosing the number of EC2 instances you need. The instances need to be on the same VPC and subnetwork. ThoughtSpot will set up the instances to be in the same ThoughtSpot cluster.
10. Click **Next: Add Storage**. The default storage specified by the ThoughtSpot AMI should be populated. Optionally, you can add extra storage. Based on the dataset size requirement you might need to provision and prepare (formatting/file system placement) an extra storage of 400 GB per VM that is SSD gp2 provisioned.
11. Click **Next: Add Tags** when you are done modifying the storage size.
12. Set a name for tagging your instances and click **Next: Configure Security Group**.
13. Select an existing security group to attach new security groups to such 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 needs to be accessible from one another. For easier configuration, it is recommended that you enable full access between VMs in a cluster.
- Additionally, more ports need to be opened on the VM to provide data staging capabilities to your network. Check [Network policies](/appliance/firewall-ports.html) to determine the minimum required ports that must be opened for your ThoughtSpot appliance.

14. Click **Review and Launch**. After you have reviewed your instance launch details, click **Launch**.
15. 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.
16. Click **Launch Instances**. Wait a few minutes for it to fully start up. Once it has started up, it will show up on the EC2 console.
17. Contact ThoughtSpot support to complete your ThoughtSpot installation. They will set up the VM instances to be part of the cluster.
18. When the setup is complete, you can load data into ThoughtSpot for search analytics.

Open the required network ports

See [Network policies](#) for a complete list of network ports that must be open for traffic from end users as well as between ThoughtSpot nodes in a cluster.

Related information

[EC2 Best Practices Loading data from an AWS S3 bucket](#)

Deploy AWS using CloudFormation CLI

Deploying your AWS instance using the CloudFormation CLI

To deploy your AWS instance using CloudFormation CLI, do the following:

1. A ThoughtSpot cluster requires AWS nodes to be configured with EBS volumes.
2. Below is a Sample JSON file to deploy an AWS instance, please edit the values to the keys as per requirement.

```
{  
  "AWSTemplateFormatVersion" : "2010-09-09",  
  "Description" : "Thoughtspot Ec2 Instance",  
  "Resources" : {  
    "EC2Instance" : {  
      "Type" : "AWS::EC2::Instance",  
      "Properties" : {  
        "AvailabilityZone" : "us-west-2c",  
        "ImageId" : "ami-08aaf182e43599854",  
        "KeyName" : "<key pair>",  
        "InstanceType" : "r4.16xlarge",  
        "BlockDeviceMappings" : [  
          {  
            "DeviceName" : "/dev/xvda",  
            "Ebs" : {  
              "SnapshotId": "snap-01e554619c99184cf",  
              "VolumeType" : "gp2",  
              "DeleteOnTermination" : "true",  
              "VolumeSize" : "200"  
            }  
          },  
          {  
            "DeviceName" : "/dev/xvdb",  
            "Ebs" : {  
              "SnapshotId": "snap-012d2e12313e1cdb2",  
              "VolumeType" : "gp2",  
              "DeleteOnTermination" : "true",  
              "VolumeSize" : "200"  
            }  
          }  
        ]  
      }  
    }  
  }  
}
```

```
"VolumeType" : "gp2",
"DeleteOnTermination" : "true",
"VolumeSize" : "1024"
}

},
{

"DeviceName": "/dev/xvdc",
"Ebs": {
"SnapshotId": "snap-0ab3b43966023820a",
"DeleteOnTermination": true,
"VolumeType": "gp2",
"VolumeSize": "1024"
}
}

],
"EbsOptimized" : true,
"NetworkInterfaces": [
{
"DeviceIndex": 0,
"Description": "Primary network interface",
"SubnetId": "subnet-3aaf9e63",
"DeleteOnTermination": true,
"GroupSet": [
"sg-0e2725bca060835ea"
]
}
]
}
```

- aws cloudformation create-stack --stack-name testing --template-body file:/// --capabilities CAPABILITY_IAM
 - Verify the AWS instance is deployed from console.

Set up high availability for AWS

Setting up High Availability (HA) for your ThoughtSpot cluster using AWS Elastic File System (EFS)

To set up HA for your ThoughtSpot cluster, do the following:

1. Create EFS File System spanning across different availability zones and different subnets.

| Availability Zone | Subnet | IP address | Security groups |
|--|---|------------|--|
| <input checked="" type="checkbox"/> us-west-1a | subnet-02b10aa9d03698bc5 - public-subnet-1a | Automatic | sg-4cb75f2b - Outbound Internet Inbound HQ |
| <input checked="" type="checkbox"/> us-west-1c | subnet-d9274083 - 76-PROD | Automatic | sg-4cb75f2b - Outbound Internet Inbound HQ |

| Name | File system ID | Metered size | Number of mount targets | Creation date |
|------------------|----------------|--------------|-------------------------|--------------------------|
| fs-211edc38 | | 12.0 KB | 0 | 11/09/2018, 20:51:00 UTC |
| fs-3b65a122 | | 6.6 GB | 0 | 10/02/2018, 23:05:30 UTC |
| PM Backup Server | fs-6476979d | 6.0 KB | 1 | 06/18/2018, 23:59:34 UTC |
| | fs-c17ba4d8 | 6.0 KB | 2 | 01/16/2019, 23:59:19 UTC |
| TESTHA | fs-7756f1ee | 6.0 KB | 2 | 02/11/2019, 18:33:36 UTC |

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

The screenshots show the AWS EC2 Dashboard with two instances listed:

- Instance 1 (Top):** Instance ID: i-0337a06a762aaf8b2, Instance Type: r4.2xlarge, Availability Zone: us-west-1c, Status: running, 2/2 checks passed. Public DNS (IPv4): ex2-13-57-254-52.us-west-1.compute.amazonaws.com, IPv4 Public IP: 13.57.254.52. Private DNS: ip-10-83-77-23.us-west-1.compute.internal, Private IP: 10.83.77.23. Secondary private IP: vpc-100e7a95. Subnet ID: subnet-d1d274083. Network interfaces: eth0. Source/Dest. check: True. T2/T3 Unlimited.
- Instance 2 (Bottom):** Instance ID: i-0473aec09ddc299c0, Instance Type: r4.2xlarge, Availability Zone: us-west-1a, Status: running, 2/2 checks passed. Public DNS (IPv4): -, IPv4 Public IP: -. Private DNS: ip-10-83-99-209.us-west-1.compute.internal, Private IP: 10.83.98.209. Secondary private IP: vpc-100e7a95. Subnet ID: subnet-d2b10a8d02369bbcd. Network interfaces: eth0. Source/Dest. check: True. T2/T3 Unlimited.

3. Change the IP addresses of the cluster (if needed).
4. Create efs directory in /home/admin path and issue below command to mount the above created file system, editing the below sample link: `sudo mount -t nfs -o nsize=4.1,rsize=1048576,wszie=1048576,hard,timeo=600,retrans=2,noresvport fs-f756f1ee.efs.us-west-1.amazonaws.com:/ /home/admin/efs/` Make sure read and write permissions are provided on all cluster EFS mount points. use: `chmod 777 /home/admin/efs`
5. Go to First cluster. Create a snapshot and backup from that snapshot, on to the mount point where EFS is mounted. Below are sample commands: `tscli snapshot create EfsTest HA 2 tscli backup create --mode full --type full --storage_type local EfsTest /home/admin/efs/Efs-backup`
6. Make sure the backup is created and accessible from all the clusters where EFS is mounted. In our cases both the clusters.
7. Now bring down the first cluster instances.
8. Go to Second cluster, delete existing cluster and create new one by restoring from the first cluster backup which is accessible from efs mount point. Example: `tscli cluster restore /home/admin/EFS/Efs-backup`

Cluster should now be successfully restored on the second cluster from the backup provided my EFS, achieving HA for ThoughtSpot clusters.

Azure node overview

Azure can provide lots of memory and CPU for your ThoughtSpot instance, and it can be easily updated from development instances.

Your database capacity will determine the number of instances you'll need and the memory and storage requirements. Depending on the amount of data you want to load, you can use multiple nodes together in a single instance.

The security group setting of your ThoughtSpot instance on Azure is up to you. You can find more information about which ports to open in the [network ports reference](#).

A little about Azure

Azure is a secure cloud services platform offered by Microsoft. Using ThoughtSpot on Azure allows you to easily add instances as your dataset grows. You can do everything you'd normally want to do in a traditional data center with Azure. It features an on-demand delivery of IT resources and applications via the Internet. When you are connected to an instance, you can use it just like you use any other server.

Using Azure lets you develop and deploy applications faster since there is no need to manage hardware. Therefore, it is easy to scale and manage computing capacity.

Azure uses [role based access control \(RBAC\)](#) to manage permissions on resources. Permissions are assigned to a [scope](#) in the following hierarchy: Subscription, management group, resource group, and individual resource. See [Organize your resources](#) in the Azure documentation for information on how to use this capability.

Azure configuration options

ThoughtSpot engineering has performed extensive testing of the ThoughtSpot appliance on various Azure configurations for best performance, load balancing, scalability, and reliability.

You can find information here on which configuration of memory, CPU, storage, and networking capacity you should be running for your instances. There are also details on how to configure your scope and permissions.

Hardware configurations

The following Azure resources are required for provisioning a Thoughtspot image in Azure:

- Virtual Machine Size: Standard E32s and E64s v3 (64 vcpus, 432 GB memory)
- Storage: Two 1TB SSD disks
- Network Ports: See PORTS SETTINGS

Data capacity per node

ThoughtSpot sizes nodes in multiples of 200 GB of data in Azure. This size refers to the amount of data in the CSV files you will be loading into ThoughtSpot. The 200GB number takes into account all replication of data done automatically by ThoughtSpot to provide redundancy and fast performance.

You can start with 200 GB, and add increments of 200 GB each time your data capacity needs increase. You can also choose to start off with more than 200 GB of data, as long as you know the best fit configuration for your data volume.

Set up Azure for ThoughtSpot

After you've determined your configuration options, you must setup your virtual machines (VMs) using a ThoughtSpot image for Azure.

About the ThoughtSpot image

To provision a ThoughtSpot image 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.

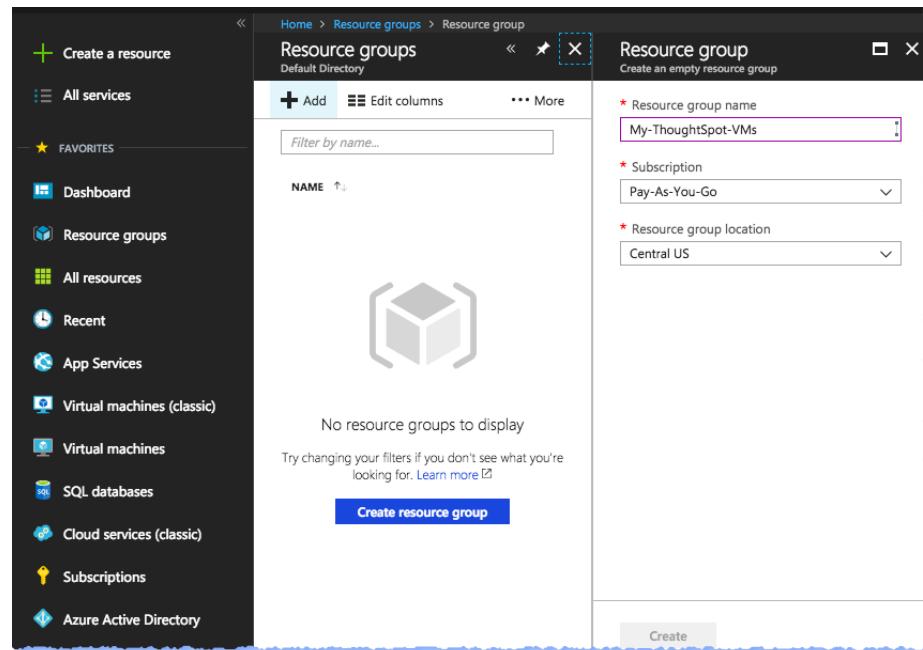
Create an instance

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.

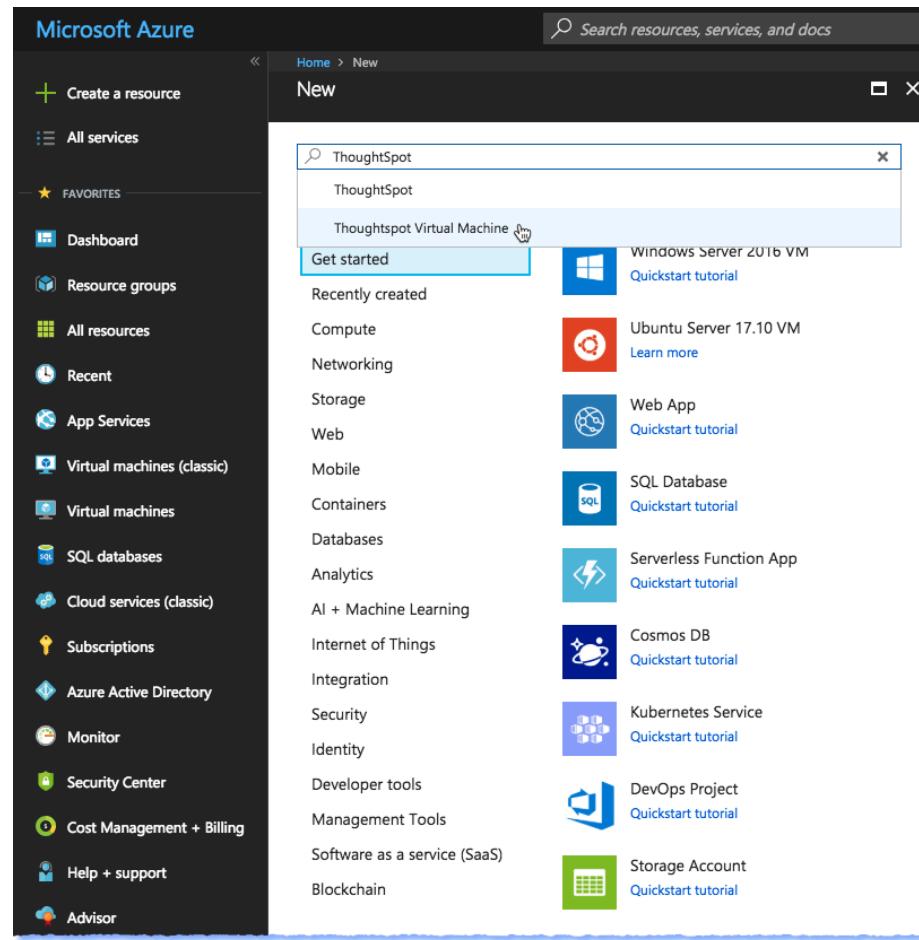
1. Log in to the Azure portal.

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

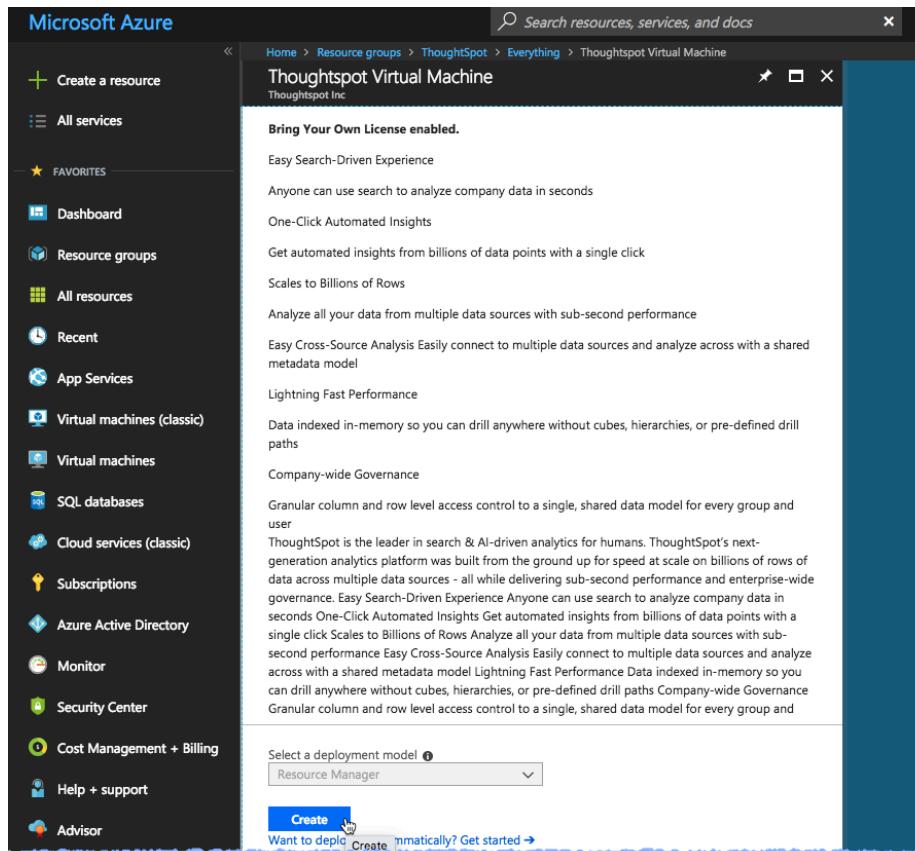
2. Create a Resource Group.



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**.



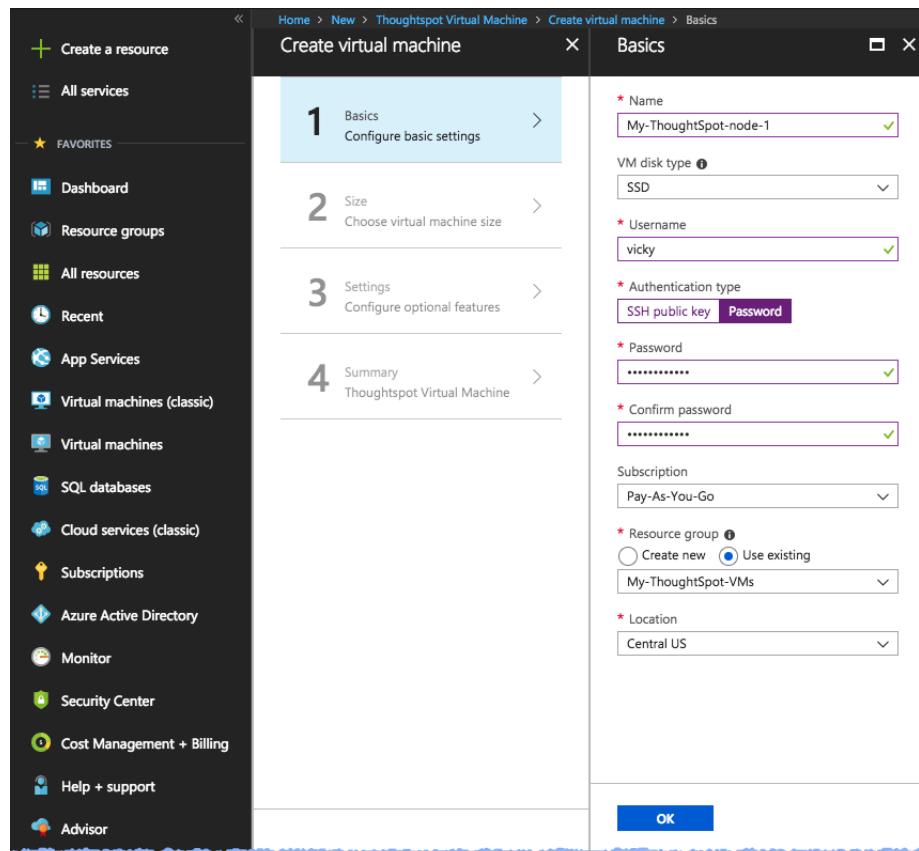
Configure basic settings

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 example below.



Choose a machine size

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

| RECOMMENDED | SKU | TYPE | COMPUTATION | VCPUS | GB RAM | DATA DISKS | MAX IOPS | LOCAL SSD | Premium | Additional | USD/MONTH |
|-------------|---------|----------|---------------|-------|--------|------------|----------|-----------|---------|------------|------------|
| | E64S_v3 | Standard | Memory optim. | 64 | 432 | 32 | 128000 | 864 GB | SSD | | \$3,282.53 |

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.

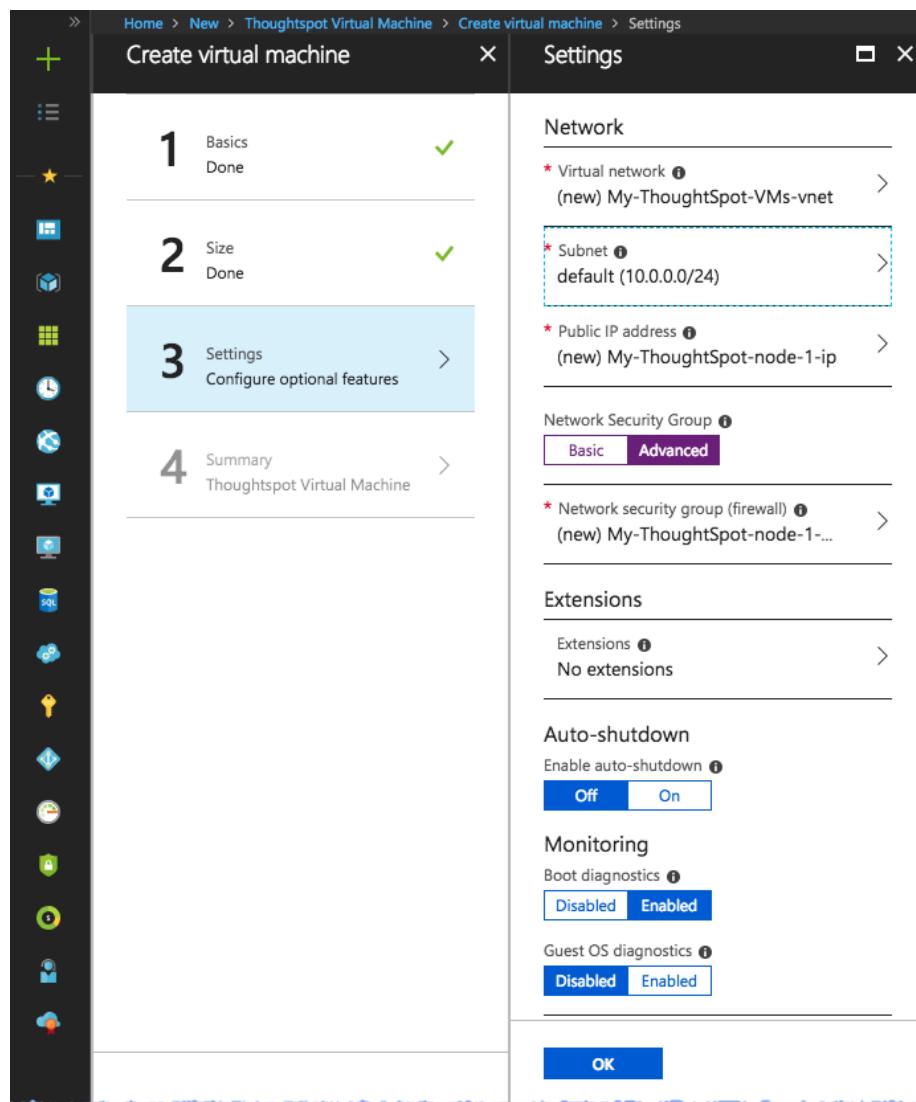
1. For storage, select **Yes to use managed disks**.
2. Under **Network**, select **Virtual network**, then **Subnet**, then **Public IP addresses**, and set those names, addresses, and ranges appropriately for your network.
3. Open the necessary Inbound and Outbound ports to ensure that the ThoughtSpot 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 |

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 will 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 (below SETTINGS on the left side of the screen).
2. Connect to the VM via SSH, using the private key provided for the admin user.
 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 node overview

Google Cloud Platform (GCP) can provide memory and CPU for your ThoughtSpot instance.

Your database capacity will determine the number of instances you'll need and the instance network/storage requirements. In addition, you can go with multiple virtual machines (VMs) based on your dataset size.

You will need to setup the appropriate Firewall Rules in your GCP environment for your ThoughtSpot deployment. See the [GCP Firewall Rules](#) article for configuration details.

You can find more information about appropriate network policies for your ThoughtSpot deployment in the [network ports reference](#).

A little about GCP

GCP is a secure cloud services platform offered by Google. Using ThoughtSpot on GCP allows you to easily add instances as your dataset grows. You can do everything you'd normally want to do in a traditional data center with GCP. It features an on-demand delivery of IT resources and applications via the Internet. When you are connected to an instance, you can use it just like you use any other server.

Using GCP lets you develop and deploy applications faster since there is no need to manage hardware. Therefore, it is easy to scale and manage computing capacity.

The ThoughtSpot image deploys on [Google Compute Engine](#), the GCP platform for creating and running VMs (*instances*) on Google Cloud infrastructure. GCP and Google Compute Engine use [Identity and Access Management \(IAM\)](#) roles to manage permissions to resources. Google Compute engine instances and persistent disks are [zonal resources](#), and therefore must be in the same zone to work together.

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 will need the Google account/email ID of the individual who will be logged 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 need to provision VMs. We'll do this on Google Compute Engine, the GCP platform for [creating and running VMs](#).

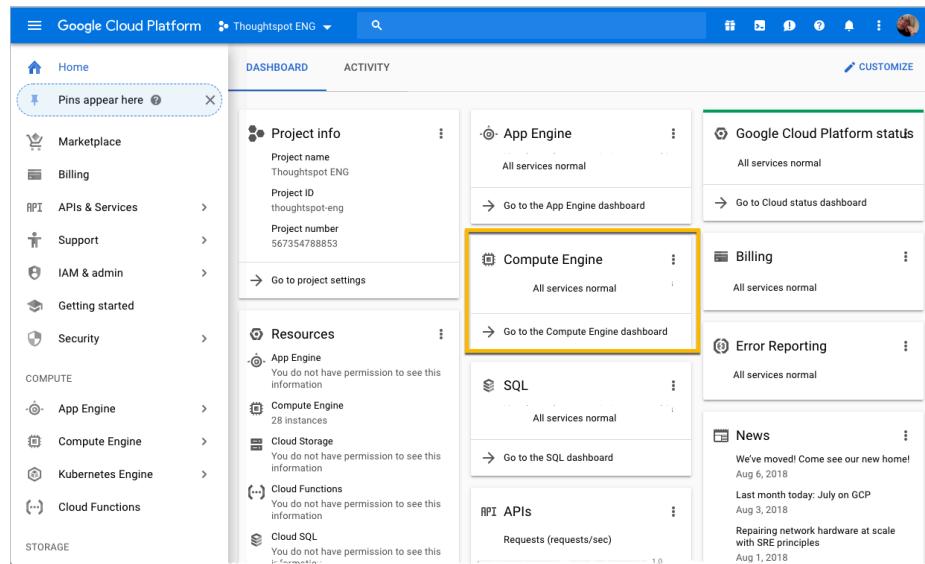
In a nutshell, the required configuration ThoughtSpot is:

- 64 vCPU
- 416 GB RAM
- 250GB SSD for the boot disk, provisioned with a ThoughtSpot base image
- 2 1TB SSD for data

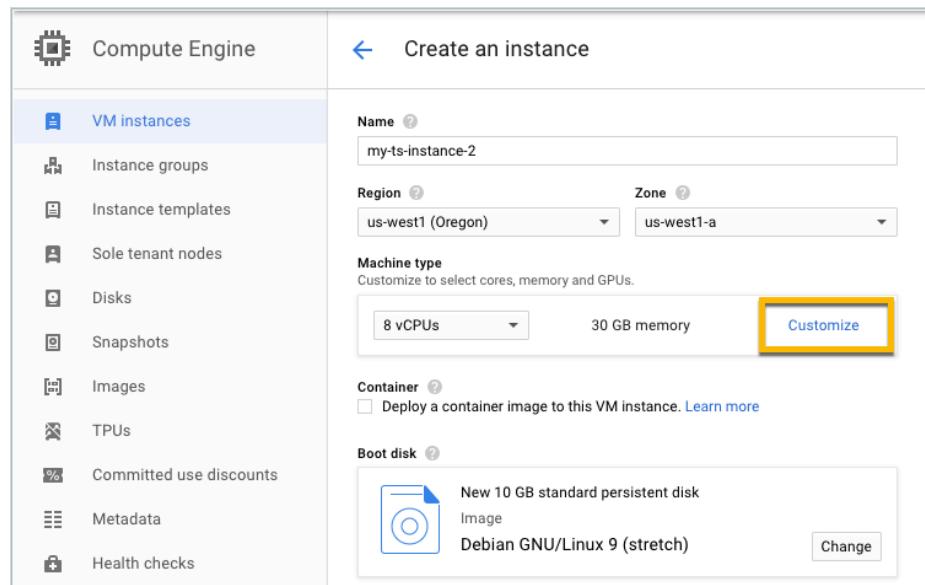
The following topics walk you through this process.

Create an instance

1. Log 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 image, 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) |

| Setting | Value |
|--------------|--|
| 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 ? Zone ?
us-west1 (Oregon) 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

Thoughtspot ENG

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](#)

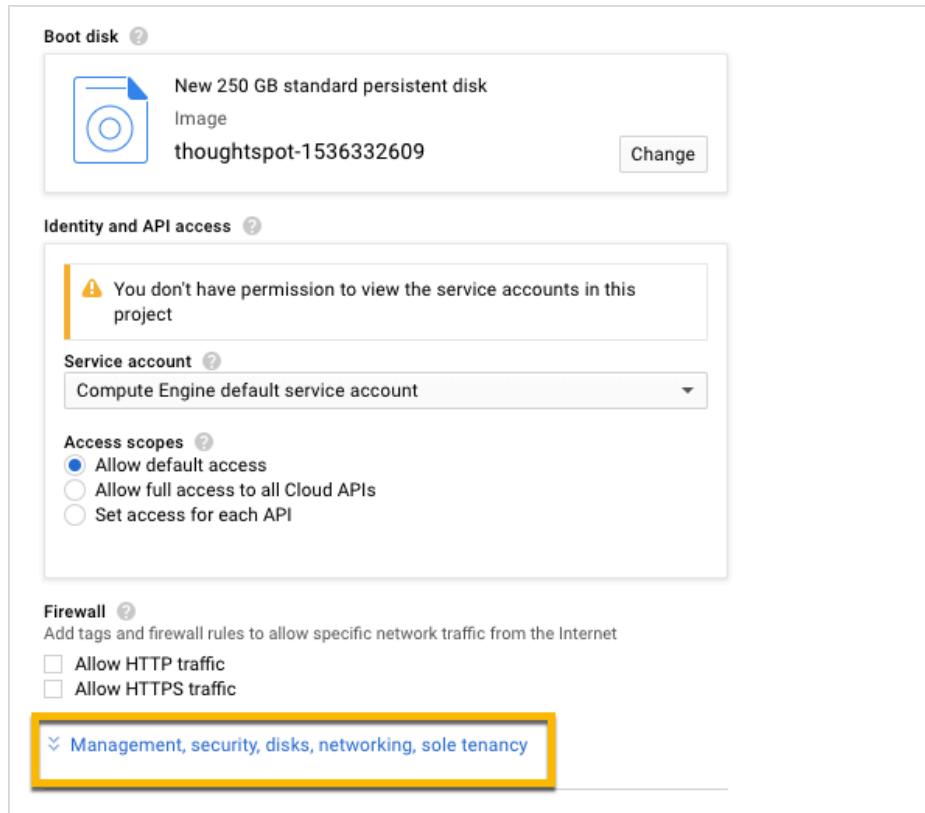
Boot disk type
Standard persistent disk

Size (GB)
250

Select Cancel

Note: ThoughtSpot updates these base images with patches and enhancements. If more than one image is available, the newest one is always at the top of the list. Both will work, but we recommend taking the newer image as it likely includes security or other types of 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 2 1TB 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 Image Blank disk

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 |

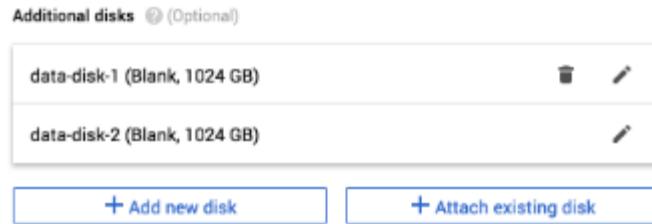
Encryption
Data is encrypted automatically. Select an encryption key management solution.

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. Customize the network settings as needed, preferably use your default VPC settings.
10. 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 in to each VM via 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

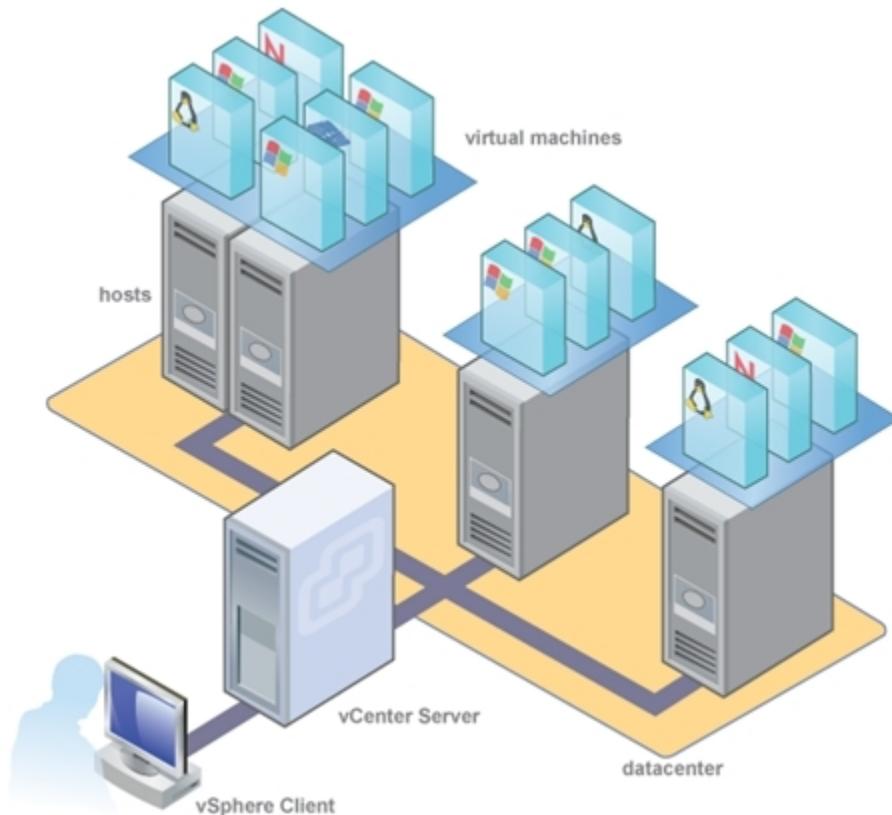
Upload the TS tarball to one of the machines and proceed with the normal cluster creation process, using `tscli cluster create`.

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.

Configuration (minimum)

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:

- 512GB Memory
- 200GB SSD
- 3X 2TB Hard disks (6TB in total in hard disk space)
- 72 Hyper-threaded Cores (Additional spare cores can also be added. Oversubscription is not supported).
- Intel Xeon 2600 series operating at clock frequencies 2.1GHz (Faster is better).

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 (~4s 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 phone, mail, email, or by filing a support ticket.

Set up VMware for ThoughtSpot

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 one or three node cluster. A one node cluster is suitable for a sandbox environment but is insufficient for a production environment.

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

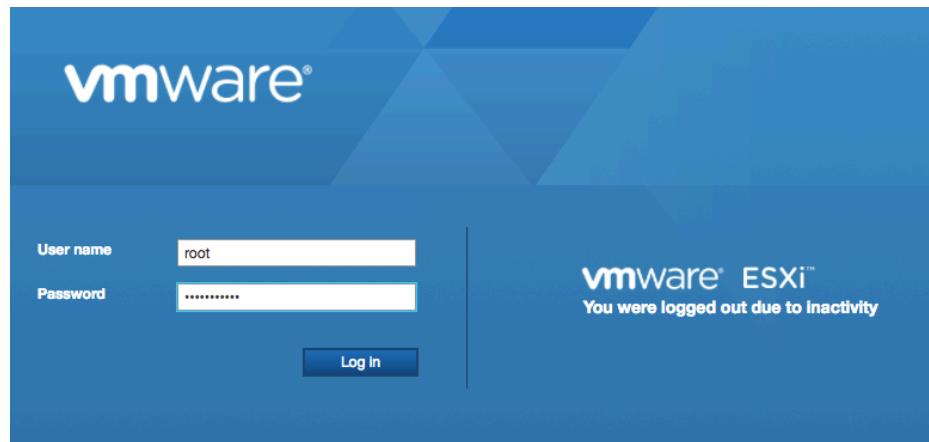
The VM template, by default, captures a 72-core configuration. If your physical host has more than 72 cores, you may want to edit VM to have (**n-2**) cores (for a physical host with n cores) to fully take advantage of computing power of the physical host. Extra cores help performance.

You should aim to allocate 490 GB or more RAM.

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

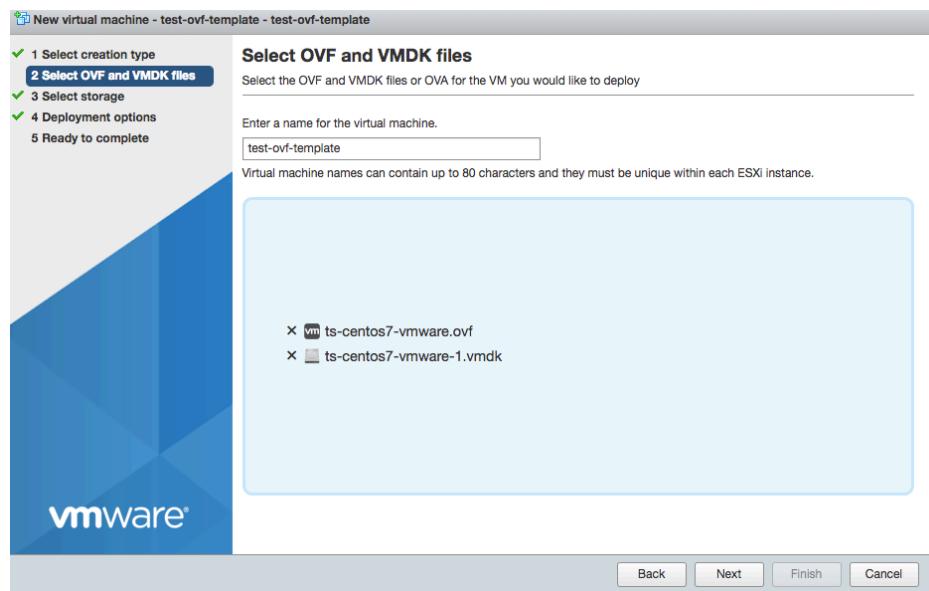
Use the OVF to Create a VM

1. Download the [ThoughtSpot OVF](#) from the [Downloads page here](#) 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 press **Next**.

The system prompts you to select a storage.

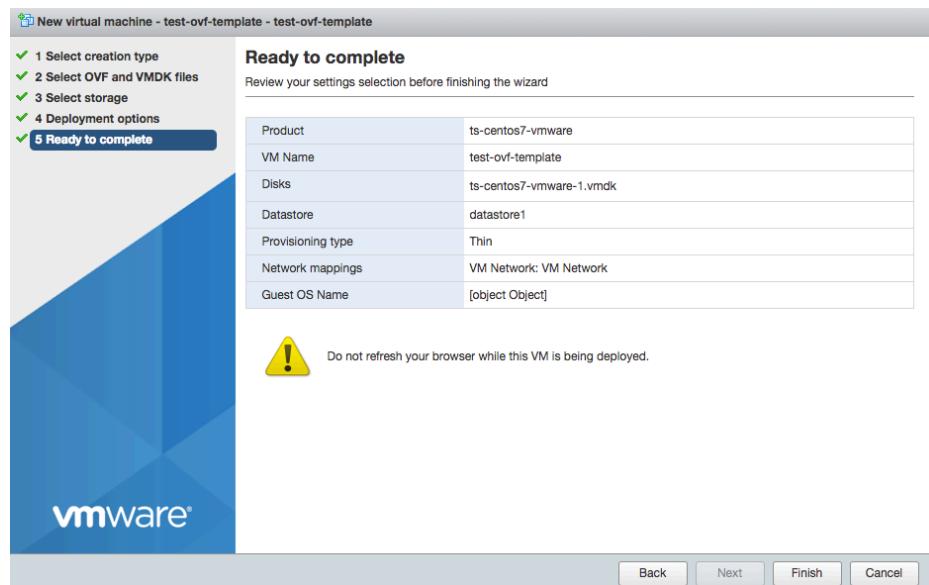
5. Choose the SSD as the destination and press **Next**.

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

6. Enter the options and press **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 press **Finish**.



- Wait for the template to be loaded.

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

| Recent tasks | | | | | | |
|---|-------------------|-----|------|------|---|------|
| Task | Target | I | Q | S | Result | C |
| Upload disk - ts-centos7-vmware-template-1.vmdk ... | test-ovf-template | ... | 1... | 1... | <div style="width: 50%;"> </div> | R... |
| Destroy | test-ovf-template | ... | 1... | 1... | <div style="width: 100%; background-color: green;">Completed successfully</div> | 1... |
| Power Off VM | test-ovf-template | ... | 1... | 1... | <div style="width: 100%; background-color: green;">Completed successfully</div> | 1... |
| Import VApp | Resources | ... | 1... | 1... | <div style="width: 50%;"> </div> | R... |

- Make sure that VM is powered off.

Add hard disks to the VM

Use Case HDFS Disk Requirements

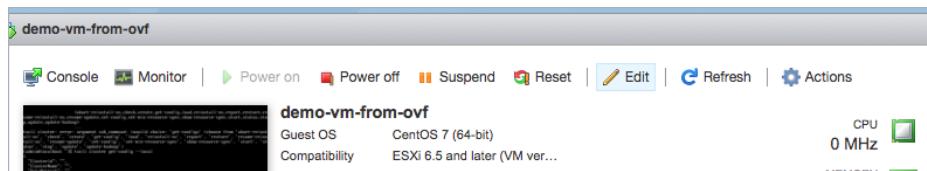
POC 2 x 1 TB on HDD

Production 3 x 2 TB on HDD

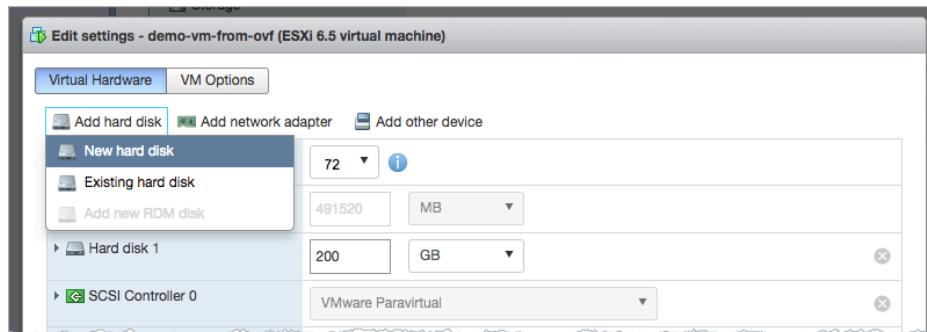
For a proof of concept (POC), follow these steps to create two 1TB HDFS disks on HDD storage, as shown here (2 x 1TB).

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.

- Edit the VM you just created.



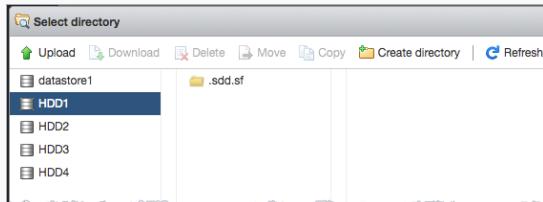
2. Select **Add hard disk > New hard disk**.



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

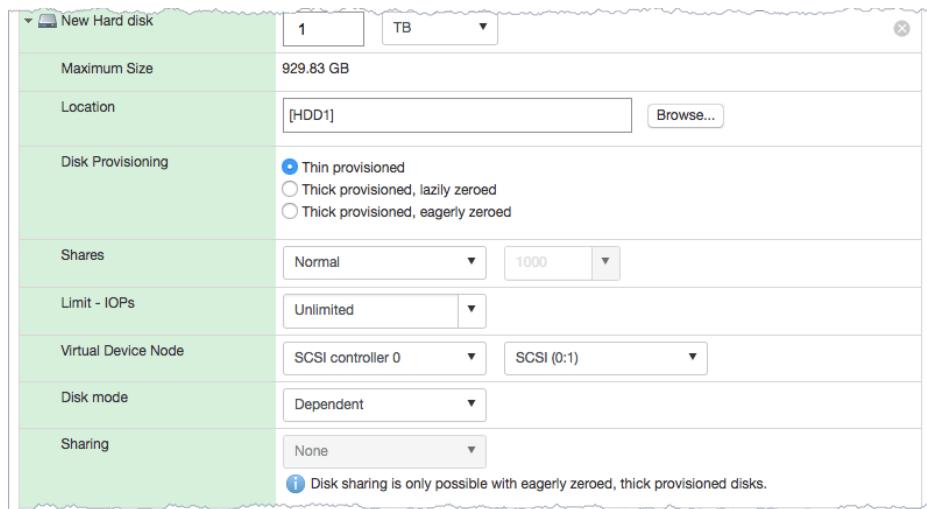
3. Click the **New Hard disk** to expand the detailed configuration options.
4. For a proof of concept, set the options as follows. (For production deployments, set the size to 2TB.)

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



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. Once 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. To make the VM node accessible from any terminal within local network, contact support@thoughtspot.com.

Network policies

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

For regular operations and for debugging, there are some ports you will need to 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. |
| 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 is 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 below. 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 |
| 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 |

| Port | Protocol | Service Name | Direction | Source | Dest. | Description |
|------|----------|--|---------------|----------------------------------|-----------|--|
| 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 |
| 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 |

| Port | Protocol | Service Name | Direction | Source | Dest. | Description |
|------|----------|---|---------------|----------------------------------|-----------|--|
| 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 |
| 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 |

| Port | Protocol | Service Name | Direction | Source | Dest. | Description |
|------|----------|--|---------------|----------------------------------|-----------|--|
| 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 |
| 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 |

| Port | Protocol | Service Name | Direction | Source | Dest. | Description |
|-------------|-----------------|--|------------------|----------------------------------|--------------|---|
| 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 |
| 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 |

| Port | Protocol | Service Name | Direction | Source | Dest. | Description |
|-------------|-----------------|--|------------------|---------------|--------------|---|
| 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 |
| 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 meta-data 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 |

| Port | Protocol | Service Name | Direction | Source | Dest. | Description |
|-------|-----------|---|---------------|-----------|-----------|--|
| 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 |
| 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 |

| Port | Protocol | Service Name | Direction | Source | Dest. | Description |
|---------------------|---------------|--|---------------|-----------|-----------|-------------|
| 20123 - 32768 | TCP (dynamic) | Dynamic port in this range used for various services and ancillary services like atlas, caffeine, call-home, callosum, falcon, monitoring, munshi server, nlp, object_search, postgres, sage UBR, spo-tiq 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 |
| 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. |

| Port | Protocol | Service Name | Direction | Source | Dest. | Description |
|-------|----------|---|---------------|---------------------|-----------|--|
| 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. |
| 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. |
| 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.

1 Palo Alto Square, Building 1, Suite 200

Palo Alto, CA 94306

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 |