ThoughtSpot

ThoughtSpot Deployment Guide for Google Cloud Platform

Release 6.0

December, 2019

910 Hermosa Court, Sunnyvale, California 94085

Table of Contents

GCP configuration options

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

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

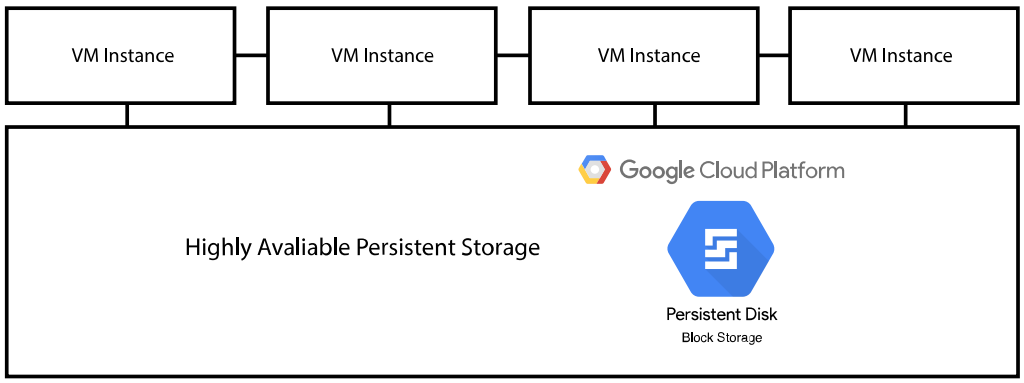
For more information about Persistant Storage, see Zonal Persistent SSD disks [See page 0] in Google's Cloud documentation.

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.

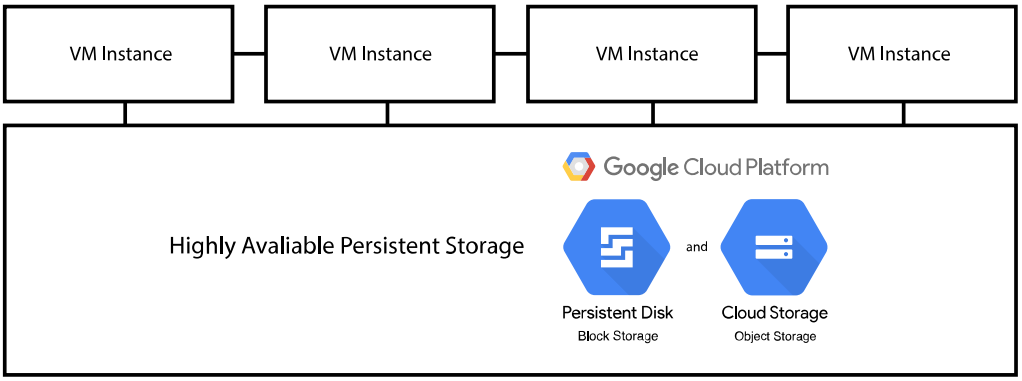
ThoughtSpot GCP instance types

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

Summary: Set up your GCP virtual machines.

After you determine your configuration options, set up your virtual machines (VMs). ThoughtSpot will share the ThoughtSpot base image for booting the VMs and some other aspects of system setup with you on the GCP platform [See page 0].

About the ThoughtSpot and Google Cloud Platform

ThoughtSpot uses a custom image to populate VMs in GCP. The base image is a Centos derived image. Find the ThoughtSpot custom image under Custom Images in **Boot Disk Options** within your Google Compute Engine project.

Ask your ThoughtSpot contact 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 to create ThoughtSpot VMs.

Overview

Before you can create a ThoughtSpot cluster, you must set up your VMs. Use the Google Compute Engine (GCP) platform to create and run VMs.

The following topics walk you through this process.

Prerequisites

1. Ensure that your **Network Service Tier** on the Google Cloud Console [See page 0] is set to **Premium** for the best performance of all your VMs.
2. A ThoughtSpot cluster requires 10 Gb/s bandwidth (or better) between any two nodes. You must ensure this *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. If you are not using GCS, skip this step and create an instance [See page 6].

1. Sign in to the Google Cloud Console [See page 0].
2. Go to the Storage dashboard from the navigation bar on the side of your screen.
3. Click **CREATE BUCKET**.
4. Enter a name for your bucket, and click **CONTINUE**.
5. For location type, select **Region**.
6. Use the Location drop-down menu to select the region where you are going to set up your instance.
7. Click **CONTINUE**.
8. For default storage class, make sure **Standard** is selected.
9. Click **CONTINUE**.
10. For access control model, make sure **Set permissions uniformly at bucket-level** is selected.
11. Click **CONTINUE**.
12. Do not edit the advanced settings.
    Leave Encryption set to **Google-managed key** and do not set a retention policy.
13. Click **CREATE**.

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

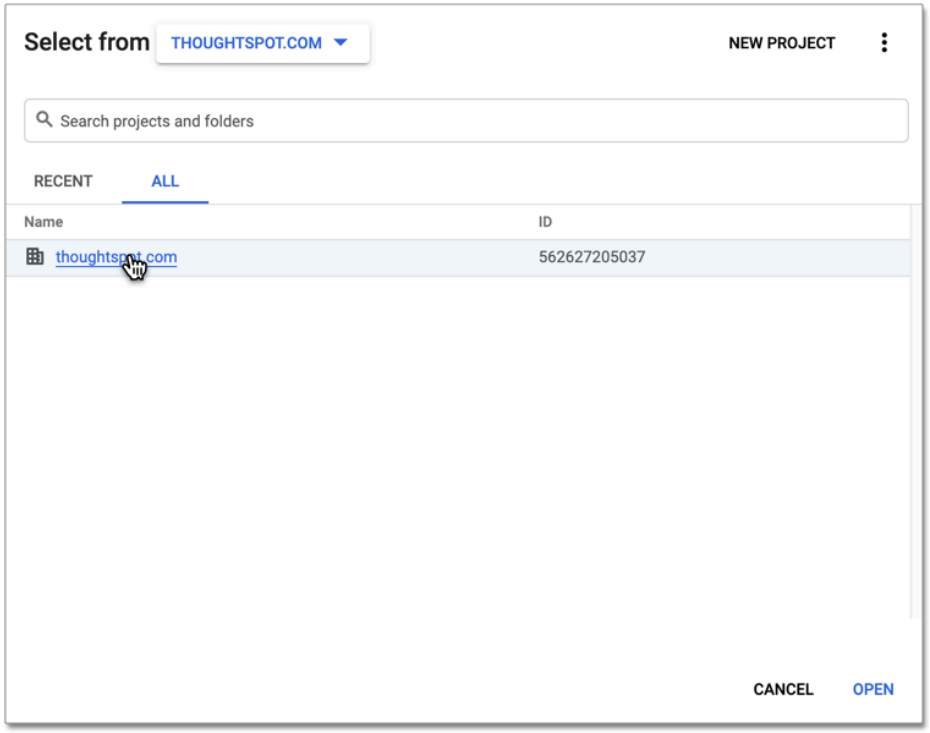
Create an instance

1. Sign in to the Google Cloud Console [See page 0].

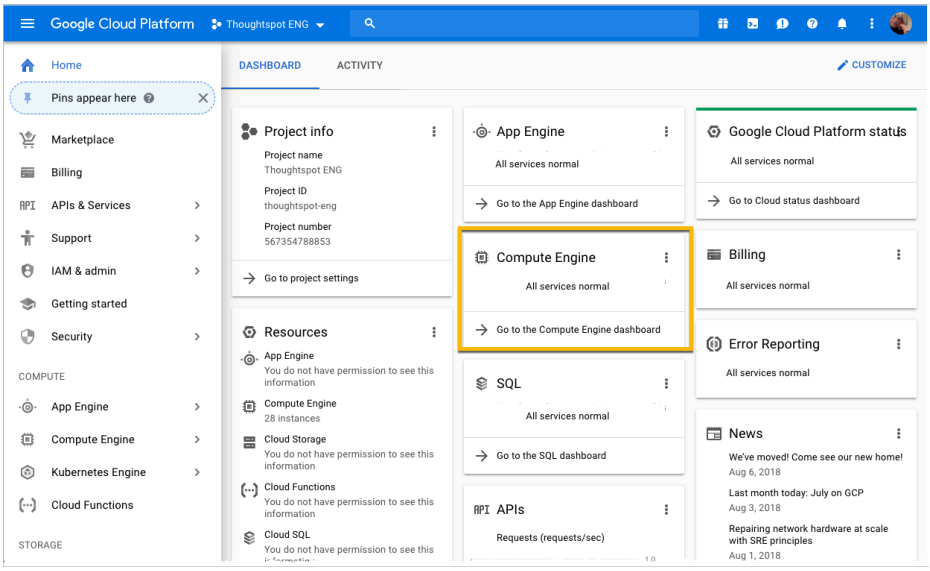
2. Click **Select a Project** from the top bar.

3. Under **Select From** pick THOUGHTSPOT.COM.

4. Select the **thoughtspot.com** project.



5. Go to the Compute Engine dashboard, and select the associated ThoughtSpot project.



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

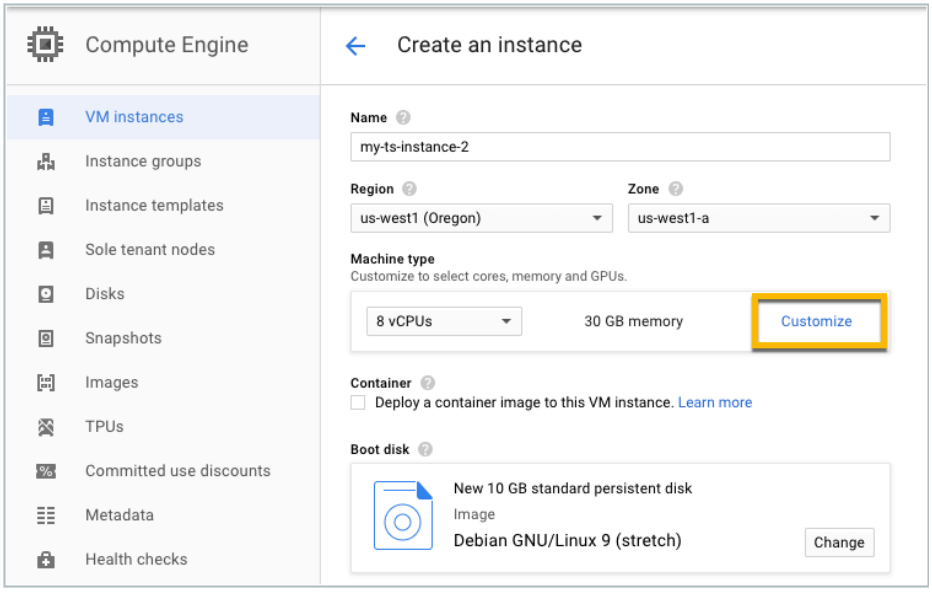
7. Provide a name for the instance.

8. Select the region you are creating the instance in.

9. Select the number of CPUs you need.

   Refer to ThoughtSpot GCP instance types [See page 2] to determine the number of CPUs your cluster needs.

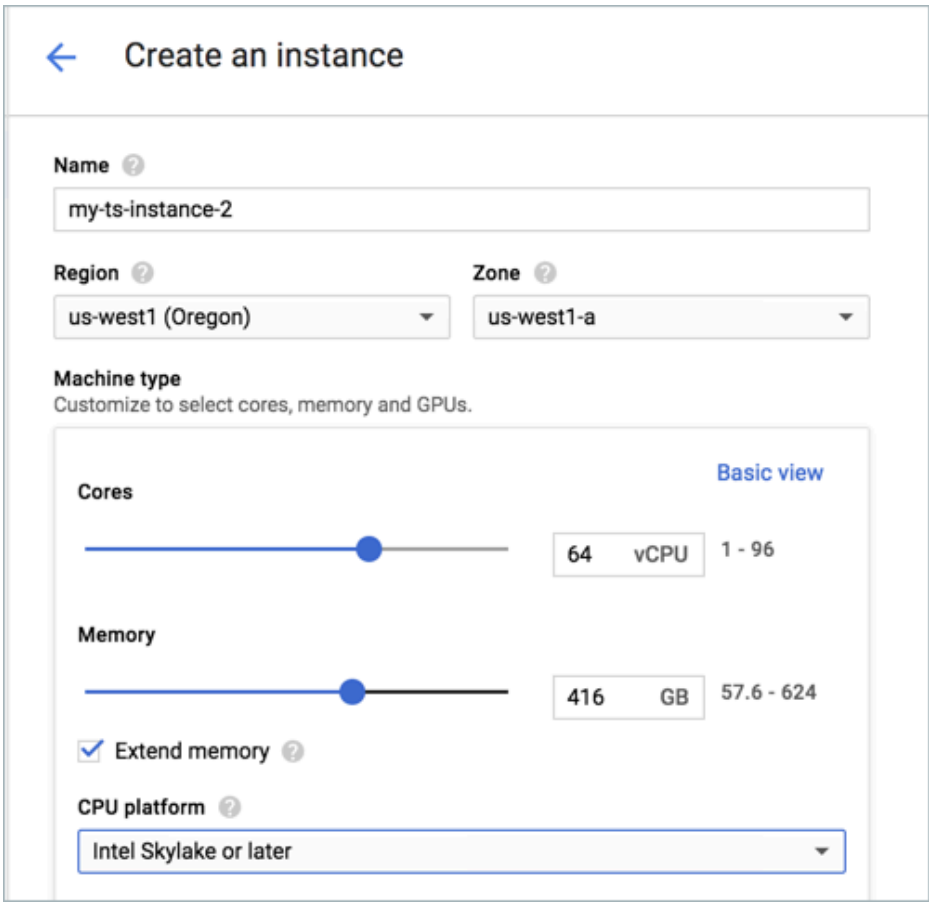
10. Click **Customize** to further configure CPUs and memory.



11. Under **Machine type**, specify your configuration information. Refer to ThoughtSpot GCP instance types [See page 2].

    Your configuration may look something like the following, but with your specific information.

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)



12. Configure the Boot disk.

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

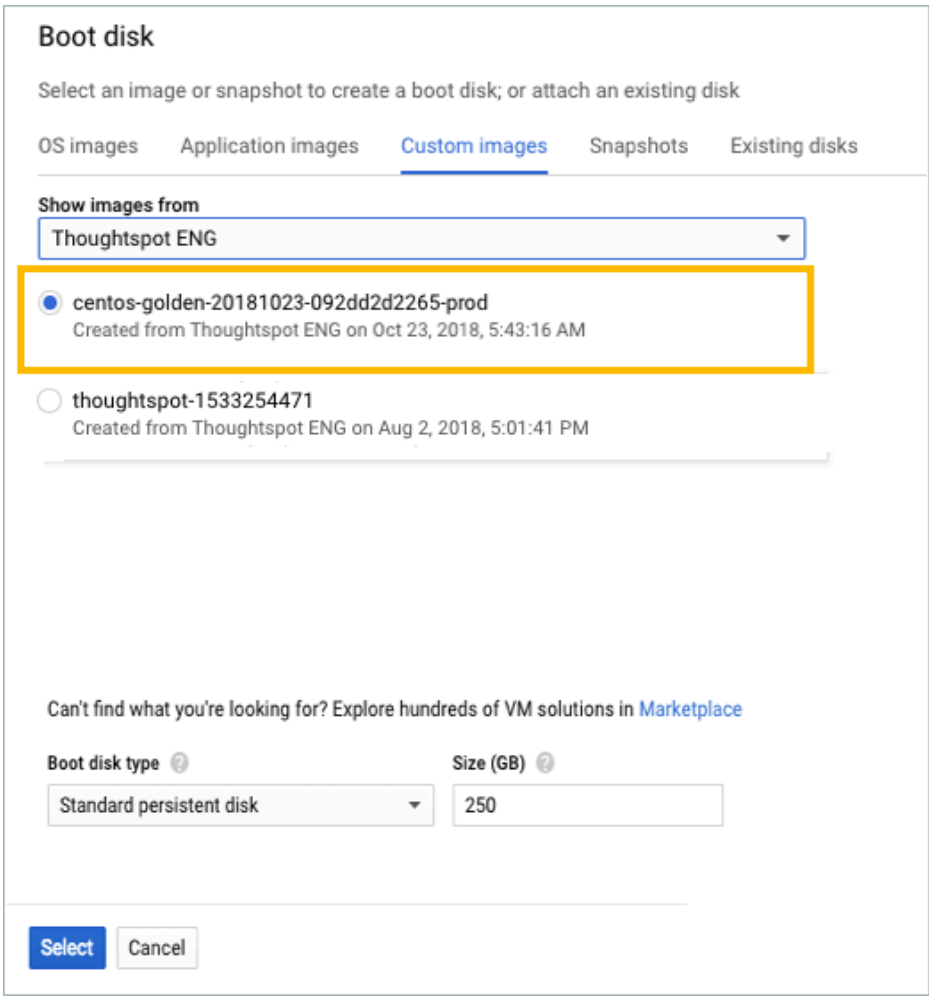
b. Click **Custom Images** from the options under **Boot disk**.

c. Select your ThoughtSpot project under **Show images from**.

d. Select one of the ThoughtSpot base images. The image at the top of the list is the latest one.

e. Configure the boot disk as follows:

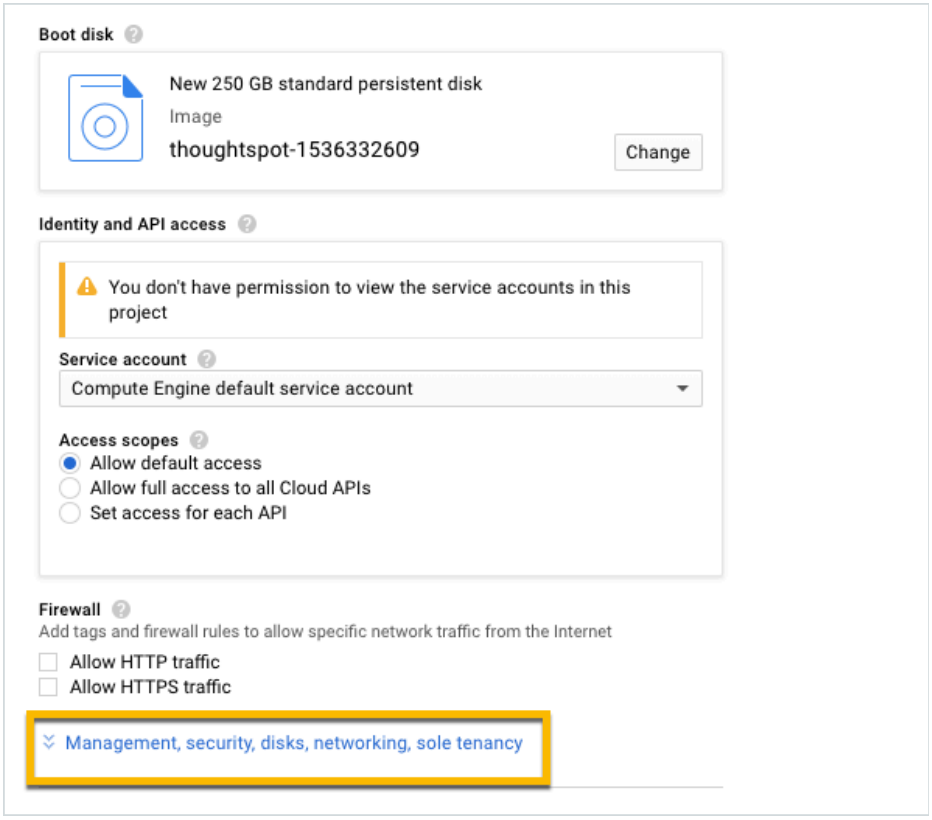
Setting	Value
Image	`ThoughtSpot`
Boot disk type	`Standard persistent disk`
Size (GB)	`250`



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.

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

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



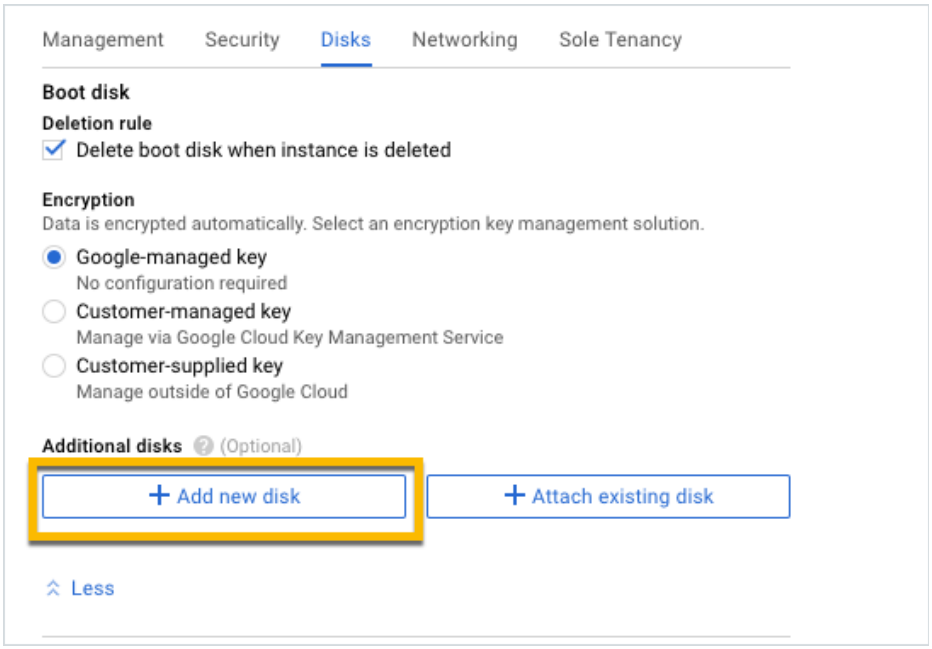
14. Under **Networking**, open required ports.

    These are the minimum ports you must open.

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

15. Attach two 1 TB SSD drives for data storage. If you are using GCS, attach only 1 SSD drive, with 500 GB instead of 1 TB.

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

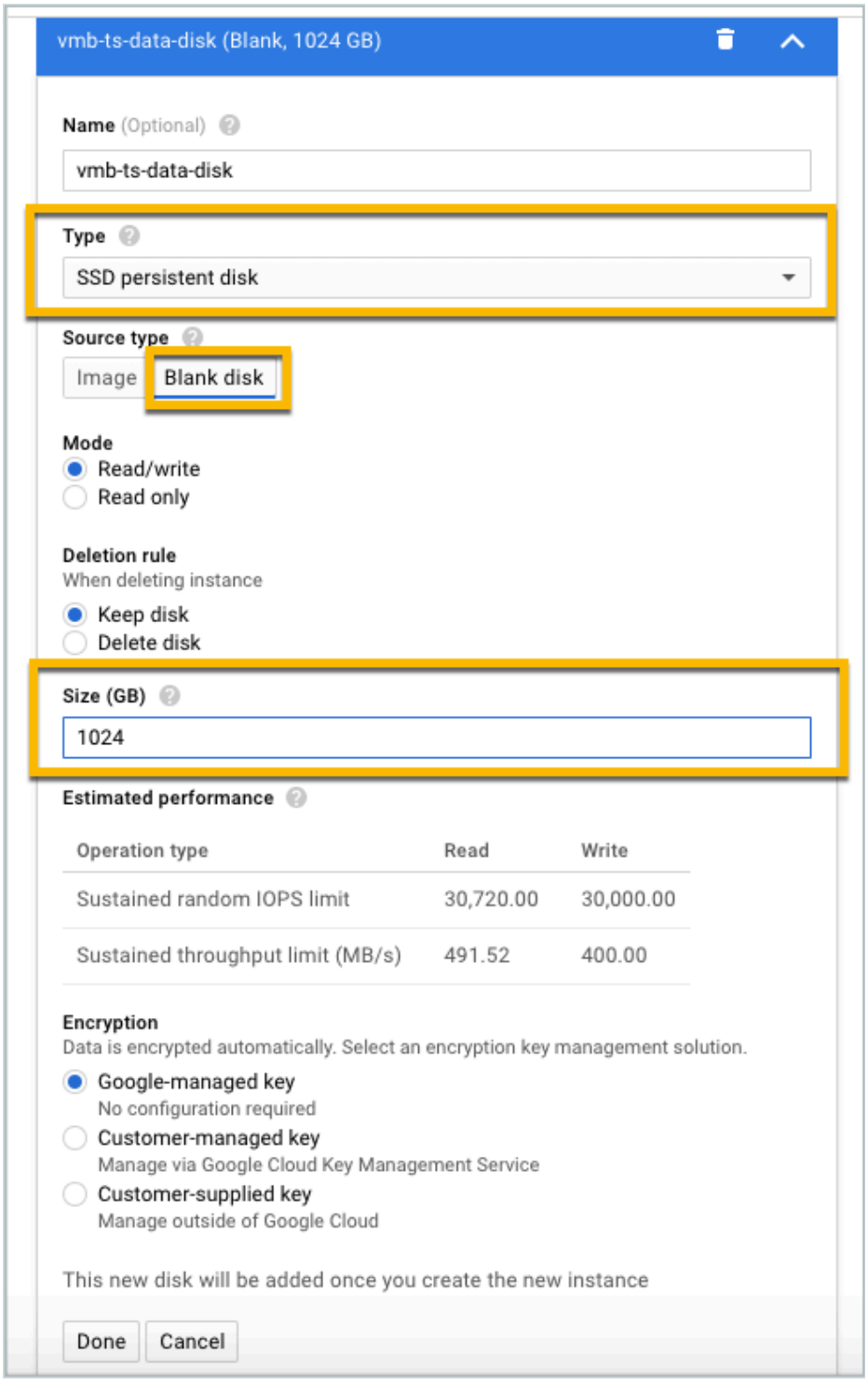


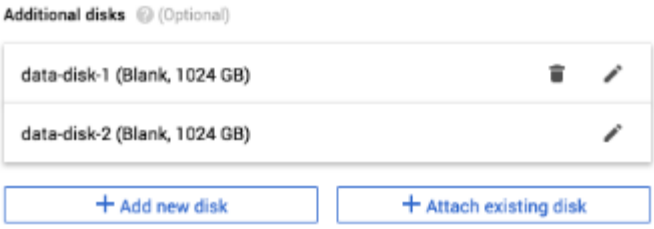
    You can select or unselect the **Deletion rule**, depending on your preferences.

    b. Configure the following settings for each disk. Refer to ThoughtSpot GCP instance types [See page 2] to determine the size in GB when you have GCS.

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

    Under **Deletion rule**, select either **keep disk** or **delete disk**, depending on your preference.


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



16. (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**.
17. (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**.
18. Customize the network settings as needed. Use your default VPC settings, if you know them. Ask your network administrator if you do not know your default VPC settings.
19. Repeat these steps to create the necessary number of VMs for your cluster.

Prepare the VMs

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

1. Open a terminal application on your machine and ssh into one of your VMs.

```
ssh admin@<VM-IP>
```

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

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

3. Configure the VM based on the site-survey.
4. Repeat this process for each of your VMs.

Install cluster

To install your ThoughtSpot cluster, complete the installation process outlined in Installing ThoughtSpot in GCP [See page 17].

Related information

Connecting to Google Cloud Storage buckets [See page 0]

Loading data from a GCP GCS bucket [See page 0]

Configure ThoughtSpot Nodes in GCP

Summary: Prepare to install your ThoughtSpot cluster by configuring nodes.

Before you can install a ThoughtSpot cluster in GCP, you must configure your nodes.

Installation Prerequisites

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

1. **Review configuration options** Refer to GCP configuration options [See page 2] for detailed instance specs.
2. **Create the instance** Refer to Set up GCP for ThoughtSpot [See page 4] to create and launch your instance.
3. **Review required ports** Refer to Network Policies [See page 0] to view the required ports for successful operation of ThoughtSpot.

Configure Nodes

After creating the instance, you must configure the nodes. Follow the steps in this checklist.

- ❑ Step 1: Log into your cluster [See page 0]
- ❑ Step 2: Get a template for network configuration [See page 0]
- ❑ Step 3: Prepare node configuration [See page 0]
- ❑ Step 4: Configure the nodes [See page 0]
- ❑ Step 5: Confirm node configuration [See page 0]

Step 1: Log into your cluster

Log into your cluster with admin credentials from Terminal on a Mac or a terminal emulator on Windows. Ask your network administrator if you do not know the admin credentials.

1. Run `ssh admin@<clusterIP>` or `ssh admin@<hostname>`.

   Replace `clusterIP` or `hostname` with your specific network information.

   ```
   $ ssh admin@<clusterIP>
   ```

2. Enter your admin password at the prompt.

   Ask your network administrator if you don't know the password.

Note: The password does not appear on the screen as you type it.

Step 2: Get a template for network configuration

Run the `tscli cluster get-config` command to get a template for network configuration 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: Prepare node configuration

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.

Do not edit any part of the `nodes.config` file except the sections described in Parameters of the nodes.config file [See page 0]. If you delete quotation marks, commas, or other parts of the code, it may cause setup to fail.

Step 4: Configure the nodes

Configure the nodes in the `nodes.config` file using the `set-config` command.

1. Disable the `firewalld` service by running `sudo systemctl stop firewalld` in your terminal. The `firewalld` service is a Linux firewall that must be off for ThoughtSpot installation. After the cluster installer reboots the nodes, `firewalld` automatically turns back on.

```
$ sudo systemctl stop firewalld
```

2. To make sure you temporarily disabled `firewalld`, run `sudo systemctl status firewalld`. Your output should specify that `firewalld` is inactive. It may look something like the following:

```
$ sudo systemctl status firewalld
  ● firewalld.service - firewalld - dynamic firewall da
emon
     Loaded: loaded (/usr/lib/systemd/system/firewalld.s
ervice; disabled; vendor preset: enabled)
     Active: inactive (dead)
```

3. Run the configuration command: `$ cat nodes.config | tscli cluster set-config`.
   If the command returns an error, refer to set-config error recovery [See page 20].
   After you run the node configuration command, your output appears similar to the following:

```
$ 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

Use the `get-config` command to confirm node configuration.

Your output may look similar to the following:

```
$ 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 ThoughtSpot software

Next, install your ThoughtSpot clusters [See page 23].

Error recovery

`Set-config` error recovery

If you get a warning about node detection when you run the `set-config` command, restart the node-scout service.

Your error may look something like the following:

```
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 Configuration failed.
Please retry or contact support.
```

Restart the node-scout service with the following command.

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

Ensure that you restarted the node-scout by running `sudo systemctl status node-scout`. Your output should specify that the node-scout service is active. It may look something like the following:

```
$ sudo systemctl status node-scout
  ● node-scout.service - Setup Node Scout service
    Loaded: loaded (/etc/systemd/system/node-scout.service; ena
bled; vendor preset: disabled)
    Active: active (running) since Fri 2019-12-06 13:56:29 PS
T; 4s ago
```

Next, retry the `set-config` command.

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

The command output should no longer have a warning.

Related information

Use these references for successful installation and administration of ThoughtSpot.

- the nodes.config file [See page 0]
- Parameters of the nodes.config file [See page 0]
- Using the cluster create command [See page 0]

- Parameters of the `cluster create` command [See page 0]
- ThoughtSpot Documentation [See page 0]
- Contact Support [See page 0]

Install ThoughtSpot Clusters in GCP

Summary: Learn how to install ThoughtSpot clusters in GCP.

Prerequisites

Before you can install your ThoughtSpot clusters in GCP, complete these prerequisites.

1. **Review configuration options** Refer to GCP configuration options [See page 2] for detailed instance specs.
2. **Create the instance** Refer to Set up ThoughtSpot in GCP [See page 4] to create and launch your instance.
3. **Review required ports** Refer to Network Policies [See page 0] to view the required ports for successful operation of ThoughtSpot.
4. **Configure nodes** Refer to Configure ThoughtSpot Nodes in GCP [See page 17] to configure your nodes.

Install ThoughtSpot Software

Install the cluster using the release tarball. The estimated installation time is one hour. Follow the steps in this checklist.

- ❏ Step 1: Run the installer [See page 0]
- ❏ Step 2: Check cluster health [See page 0]
- ❏ Step 3: Finalize installation [See page 0]

Refer to your welcome letter from ThoughtSpot to find the link to download the release tarball. If you do not have a link, open a support ticket at ThoughtSpot Support [See page 0] to request access to the release tarball.

Step 1: Run the installer

1. Copy the downloaded release tarball to `/home/admin` using the following command:

```
 $ scp <release-number>.tar.gz admin@<hostname>:/home/ad
min/<file-name>
```

Note the following parameters:

- `release-number` is the release number of your ThoughtSpot instance, such as 5.3, 6.0, and so on.
- `hostname` is your specific hostname.
- `file-name` is the name of the tarball file on your local computer.

2. Create the cluster.

   Run `tscli cluster create` to create the cluster.

   If you are using a gcs bucket for object storage, include the flag `--enable_cloud_storage=gcs`.

```
 $ tscli cluster create <release-number>.tar.gz --enabl
e_cloud_storage=gcs
```

3. Edit the output with your specific cluster information.

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

The cluster installer automatically reboots all the nodes after a successful install. The `firewalld` service automatically turns on. At this time, the system is rebooting, which may take approximately 15 minutes.

Log into any node to check the current cluster status:

```
    $ tscli cluster status
```

Step 2: Check cluster health

After the cluster installs, check its status using the `tscli cluster status` command.

Your output may look similar to the following:

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

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

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

Step 3: Finalize installation

After the cluster status changes to `READY`, sign into ThoughtSpot on your browser. Follow these steps:

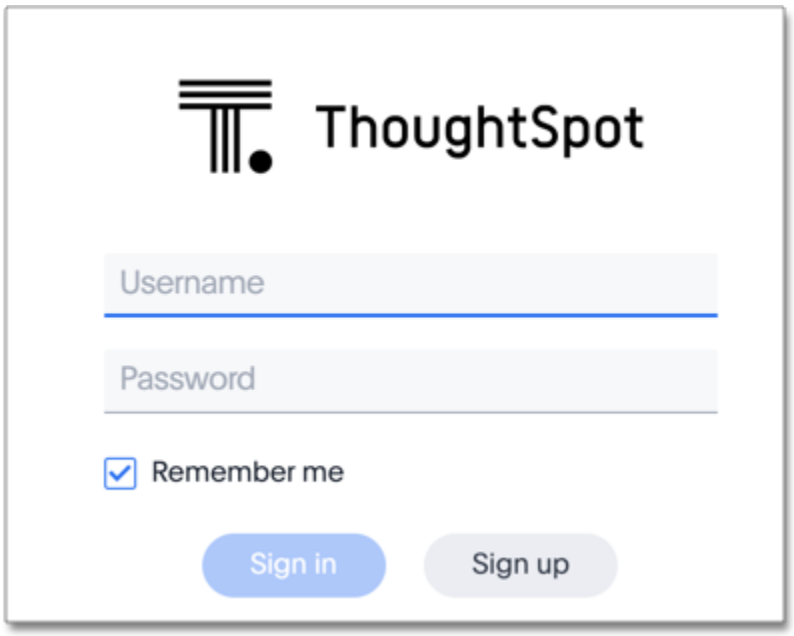
1. Start a browser from your computer.
2. Enter your secure IP information on the address line.

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

3. If you don't have a security certificate for ThoughtSpot, you must bypass the security warning:
   - Click **Advanced**
   - Click **Proceed**
4. The ThoughtSpot sign-in page appears.

5. In the ThoughtSpot sign-in window [See page 26], enter admin credentials, and click **Sign in**.

   ThoughtSpot recommends changing the default admin password.



ThoughtSpot's sign-in window

Related information

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]
- Deployment Overview [See page 0]
- Contact Support [See page 0]