ThoughtSpot

ThoughtSpot Deployment Guide for GCP

Release 6.0

December, 2019

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

910 Hermosa Court, Sunnyvale, California 94085

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

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

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

Table of Contents

Installing ThoughtSpot in GCP 2

Configuration options 10

Set up ThoughtSpot in GCP 12

Installing Google Cloud Platform

Summary: Learn how to install a ThoughtSpot cluster on Google Cloud Platform (GCP).

Installation Prerequisites

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

1. **Review configuration options** Refer to GCP configuration options [See page 10] for detailed instance specs.
2. **Create the instance** Refer to Set up GCP for ThoughtSpot [See page 12] to create and launch your instance.
3. **Review required ports** Refer to Network Policies [See page 0] to view the required ports for successful operation of ThoughtSpot.

Configure Nodes

After creating the instance, you must configure the nodes.

Step 1: Log into your cluster

Use Terminal on a Mac or a terminal emulator on Windows to log into your cluster.

1. Run `ssh admin@clusterIP` or `ssh admin@hostname`, replacing ‘clusterIP’ or ‘hostname’ with your specific network information.
2. When prompted, enter the admin password. Ask your network administrator if you do not know the admin credentials.

```
$ ssh admin@clusterIP
```

Step 2: Get a list of nodes to configure

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

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

Step 3: Configure the network of nodes

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

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

Step 4: Configure the nodes

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

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

Set-config

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

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

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

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

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

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

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

The command output should no longer have a warning:

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

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

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

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

Confirm node configuration

```
$ tscli cluster get-config

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

Install Cluster

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

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

Step 1: Run the Installer

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

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

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

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

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

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

Step 2. Check Cluster Health

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

Cluster Status

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

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

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

Step 3. Finalize Installation

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

Follow these steps:

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

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

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

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

References

Use these references for successful installation and administration of ThoughtSpot.

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

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).

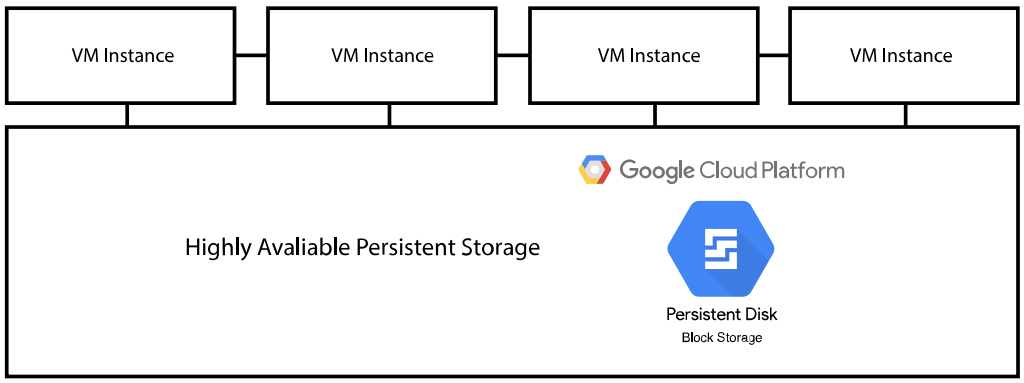
For more information about Persistant Storage, see Zonal Persistent SSD disks [See page 0] in Google's Cloud documentation.

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.

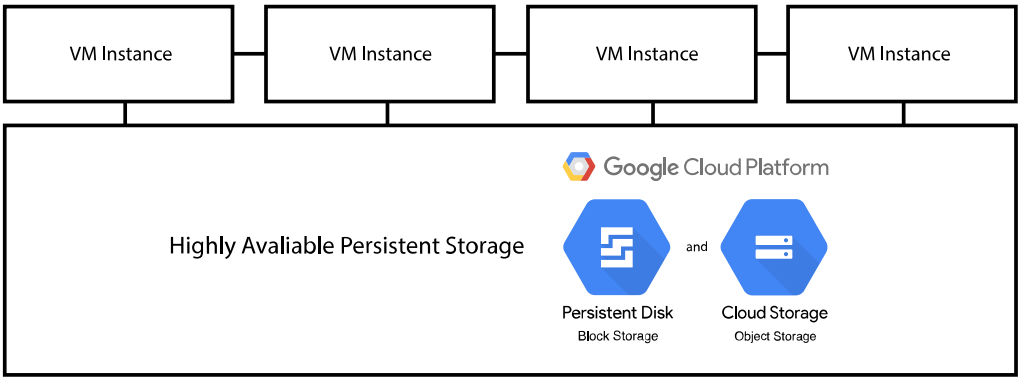
ThoughtSpot GCP instance types

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

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

About the ThoughtSpot and Google Cloud Platform

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

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

Overview

Before you can create a ThoughtSpot cluster, you must provision VMs. You use the Google Compute Engine (GCP) platform for creating and running VMs [See page 0].

The following topics walk you through this process.

Prerequisites

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

Setting up your Google Cloud Storage (GCS) bucket

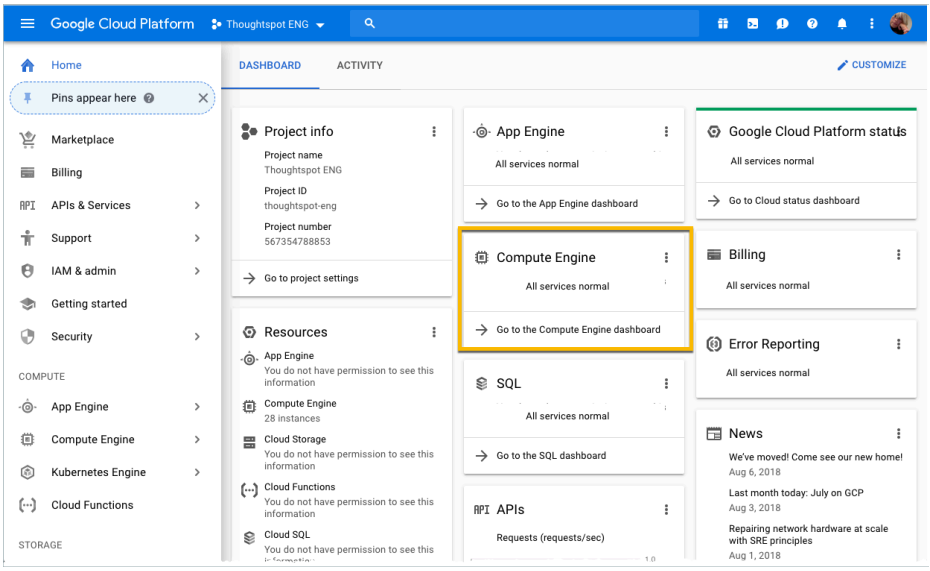
If you are going to deploy your cluster using the GCS-storage option, you must set up that bucket before you set up your cluster. Contact ThoughtSpot Support [See page 0] to find out if your specific cluster size will benefit from the GCS storage option.

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

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

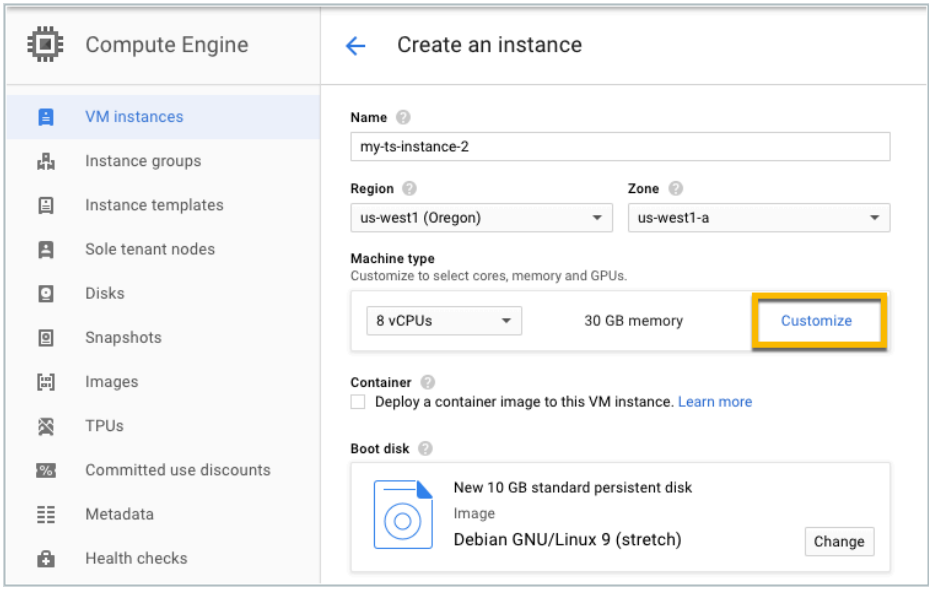
Create an instance

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



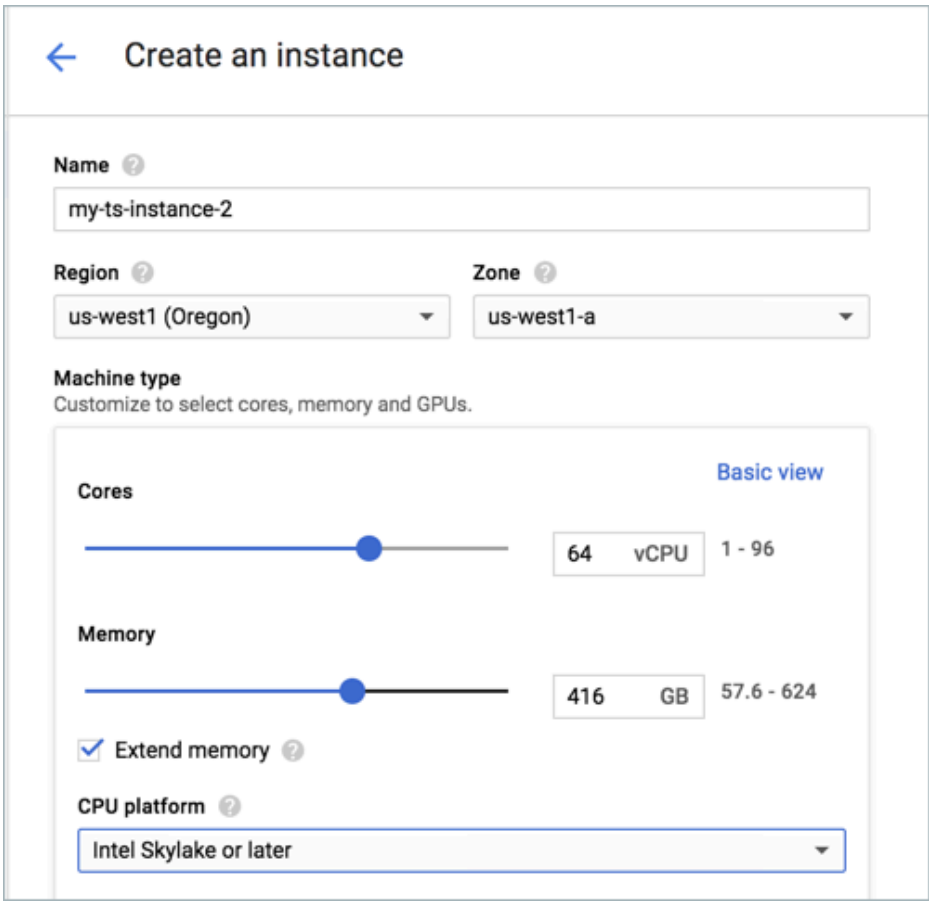
3. Select **VM instances** on the left panel and click **CREATE INSTANCE**.

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



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

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

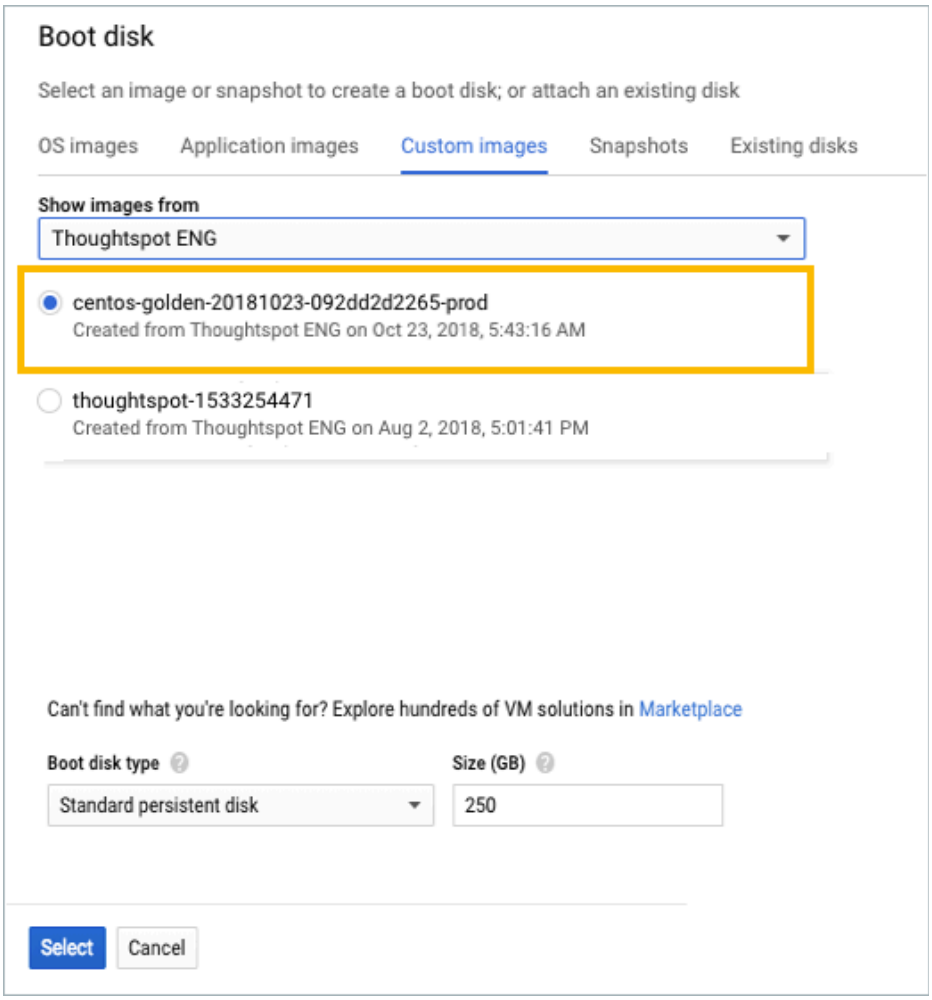


6. Configure the Boot disk.

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

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

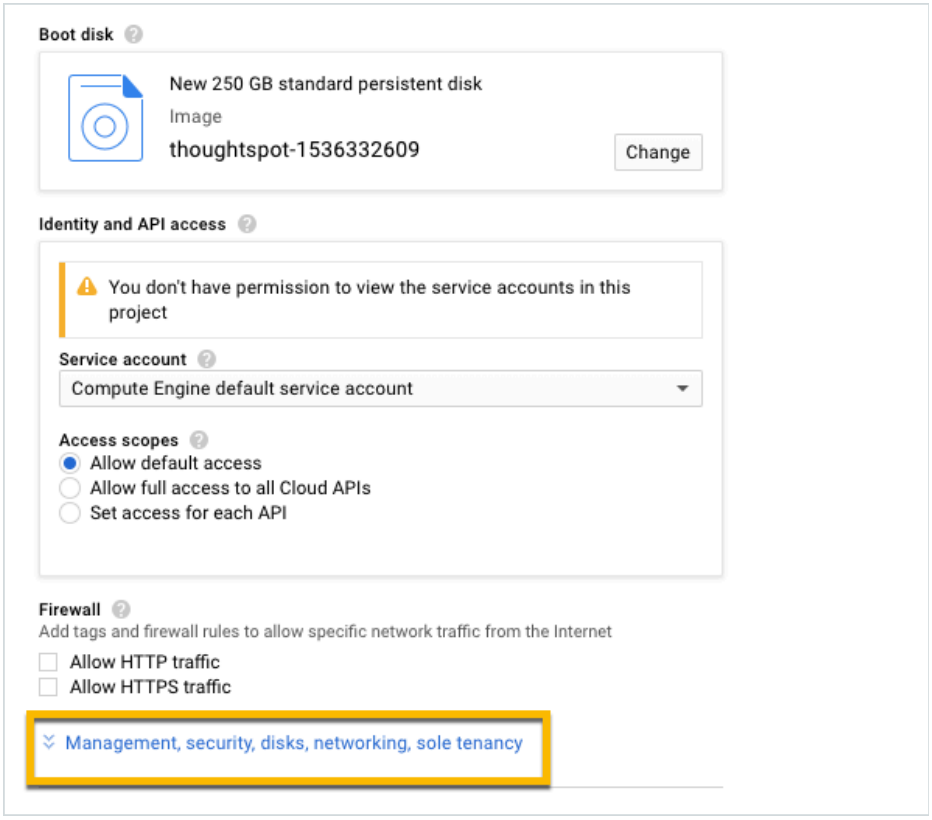
Setting	Value
Image	`ThoughtSpot`
Boot disk type	`Standard persistent disk`
Size (GB)	`250`



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

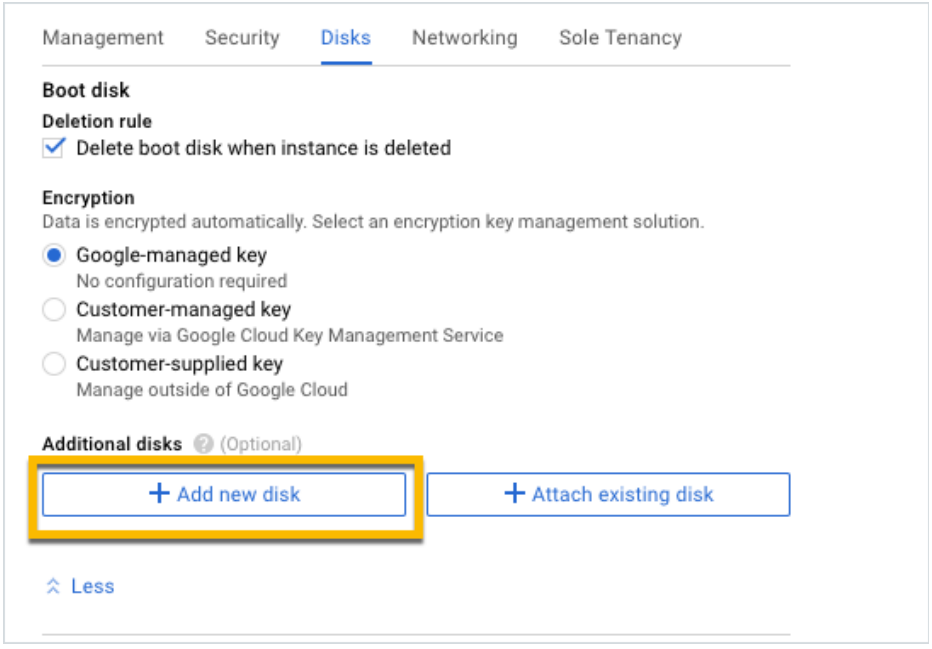
c. Click **Select** to save the boot disk configuration.

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



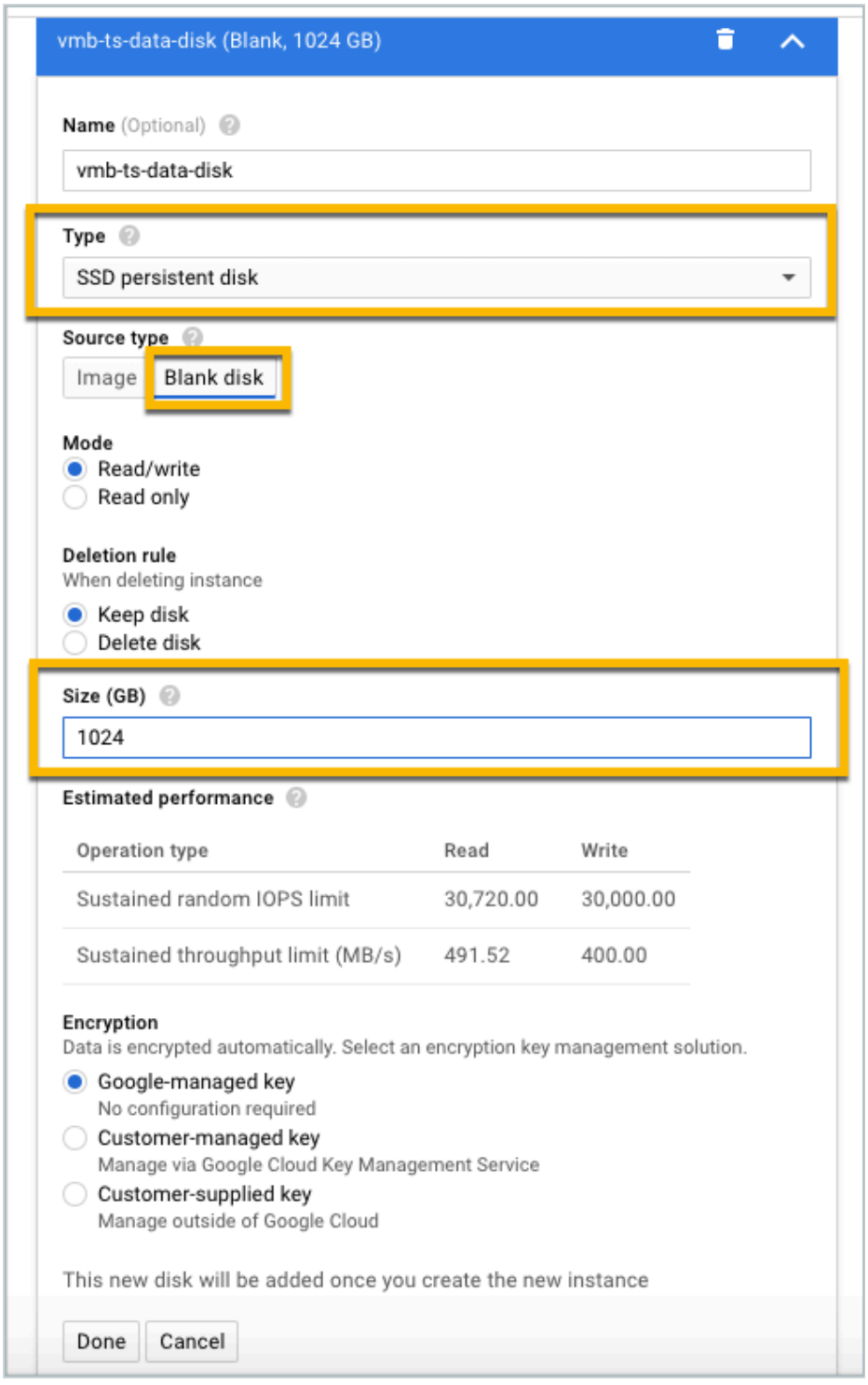
8. Attach two 1 TB SSD drives. These drives will be used for the data storage.

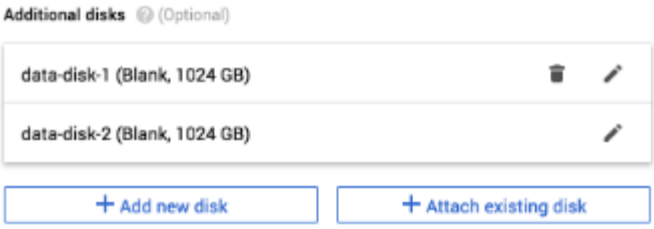
   a. Click the **Disks** tab, and click **Add new disk**.



b. Configure the following settings for each disk.

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


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



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

Prepare the VMs (ThoughtSpot Systems Reliability Team)

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

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

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

Launch the cluster

Upload the TS tarball to one of the VMs and proceed with the normal cluster creation process, using tscli cluster create [See page 0].

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

Open the required network ports

To determine which network ports to open for a functional ThoughtSpot cluster, see Network policies [See page 0].

Related information

Connecting to Google Cloud Storage buckets [See page 0]

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