ThoughtSpot

ThoughtSpot Deployment Guide for Amazon Web Services

Release 6.0

December, 2019

910 Hermosa Court, Sunnyvale, California 94085

Table of Contents

AWS configuration options

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

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

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

The cost of infrastructure for deploying ThoughtSpot is cheaper when using S3. However, there are differences in where data is loaded, as well as in the backup and restore procedure. For assistance in choosing the best mode for your organization, contact your ThoughtSpot representative. For more information on purchasing ThoughtSpot in AWS, see: ThoughtSpot Pricing [See page 0].

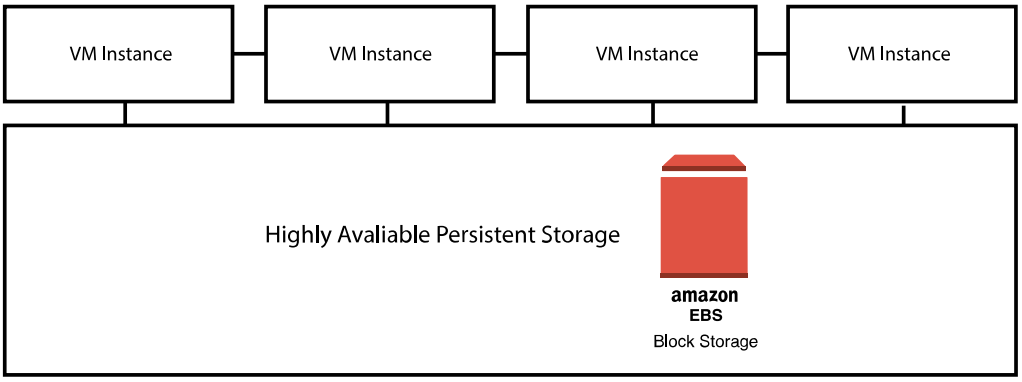
All AWS VMs in a ThoughtSpot cluster must be in the same availability zone (and therefore, also in the same region). ThoughtSpot does not support deploying VMs in the same cluster across availability zones. For more information, see Regions and Availability Zones [See page 0] in Amazon's AWS documentation.

ThoughtSpot AWS instance types

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

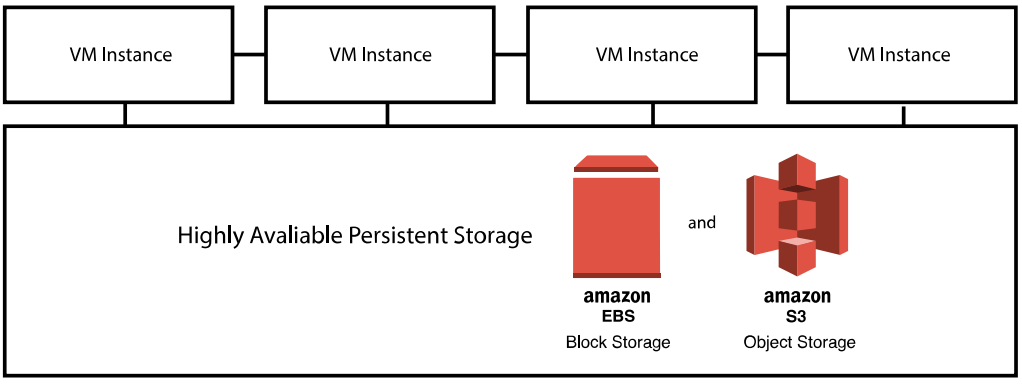
For example: If you were deploying a total cluster data size of 1 TB using the standard r5.16xlarge instance type, you would need 4 instances (VMs), because the per-VM user data capacity of that instance type is 250 GB. If you were deploying EBS-only data volumes, you would need 2x1 TB data volumes per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

VMs with EBS-only persistent storage



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

VMs with EBS and S3 persistent storage



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

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

Related information

- EC2 instance types [See page 0]
- EC2 pricing [See page 0]
- EBS pricing [See page 0]
- Placement groups [See page 0]

Set up AWS Resources for ThoughtSpot

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

Overview of ThoughtSpot setup in AWS

Follow these steps to set up your ThoughtSpot VMs in AWS.

- ❏ 1. Gain access to ThoughtSpot's AMI. [See page 0]
- ❏ 2. Choose a VM instance configuration recommended by ThoughtSpot. [See page 0]
- ❏ 3. Set up your Amazon S3 bucket (optional). [See page 0]
- ❏ 4. Set up your ThoughtSpot cluster in AWS. [See page 0]
- ❏ 5. Configure security groups. [See page 0]
- ❏ 6. Open the required network ports for communication for the nodes in your cluster and end users. [See page 0]
- ❏ 7. Configure your nodes and install the cluster. [See page 0]

About the ThoughtSpot AMI

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

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

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

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

Prerequisites

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

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

For more information about IAM, see: What Is IAM? [See page 0] in Amazon's AWS documentation.

Required AWS components

- ❏ An AWS Virtual Private Cloud (VPC). An AWS VPC is a virtual network specifically for your AWS account. It exists in all availability zones in your region, but you can specify a local zone for even lower latency. For more details, see VPCs and Subnets [See page 0] in Amazon's AWS documentation.
- ❏ A ThoughtSpot AMI. For details, see Setting up your EC2 instances [See page 0].
- ❏ AWS security groups. For required open ports, see Network Policies [See page 0].
- ❏ AWS VM instances. For instance type recommendations, see ThoughtSpot AWS instance types [See page 2].
- ❏ EBS volumes for data storage.
- ❏ (Optional) If deploying with S3 persistent storage, you need one S3 bucket for each ThoughtSpot cluster.

Setting up your EC2 instances

1. Sign into your AWS account [See page 0].
2. Copy the following ThoughtSpot public AMI to your AWS region:

AMI Name: thoughtspot-image-20191031-8ae15008336-prod

AMI ID: ami-06276ece42ed96994

Region: N. California

> **Note:** The AMI is based in the N. California region. You may have to temporarily switch to the N. California region on the AWS website to access it. Then you can return to your own region.

> **Note:** The AMI is backward-compatible with ThoughtSpot releases 5.1.x - 6.0.x.

3. Choose the appropriate EC2 instance type: See ThoughtSpot AWS instance types [See page 2] for help choosing the correct instance type for your cluster.
4. Networking requirements: 10 GbE network bandwidth is needed between the VMs. Ensure that you have this bandwidth.
5. Ensure that all your VMs are on the same Amazon Virtual Private Cloud (VPC) and subnetwork. This is necessary because VMs that are part of a cluster need to be accessible by each other. Additional external access may be required to bring data in/out of the VMs to your network.
6. Determine the number of EC2 instances you need: Based on the datasets, this number will vary. Refer to ThoughtSpot AWS instance types [See page 0] for recommended nodes for a given data size.

> **Note:** Staging larger datasets (> 50 GB per VM), may require provisioning additional attached EBS volumes that are SSD (gp2).

Setting up your Amazon S3 bucket (recommended)

If you are going to deploy your cluster using the S3-storage option, you must set up that bucket before you set up your cluster. Contact ThoughtSpot Support [See page 0] to find out if your specific cluster size can benefit from the S3 storage option.

Follow these steps to set up an S3 bucket in AWS.

1. On the AWS website, navigate to the S3 service dashboard by clicking **Services**, then **S3**.

2. Make sure the selected region in the top-right corner of the dashboard is the same region in which you plan to set up your cluster.
3. Click **Create bucket**.
4. In the **Name and region** page, enter a name for your bucket.
5. Select your region.
6. Click **Next**.
7. On the **Properties** page, click **Next**.
8. On the Configure options page, ensure that **Block *all* public access** is selected.
9. Click **Next**.
10. On the Set permissions page, click **Create bucket**.

Encrypting your data at rest

ThoughtSpot makes use of EBS for the data volumes to store persistent data (in the EBS deployment model) and the boot volume (in the EBS and S3 deployment models). ThoughtSpot recommends that you encrypt your data volumes prior to setting up your ThoughtSpot cluster. If you are using the S3 persistent storage model, you can encrypt the S3 buckets using SSE-S3 or AWS KMS.

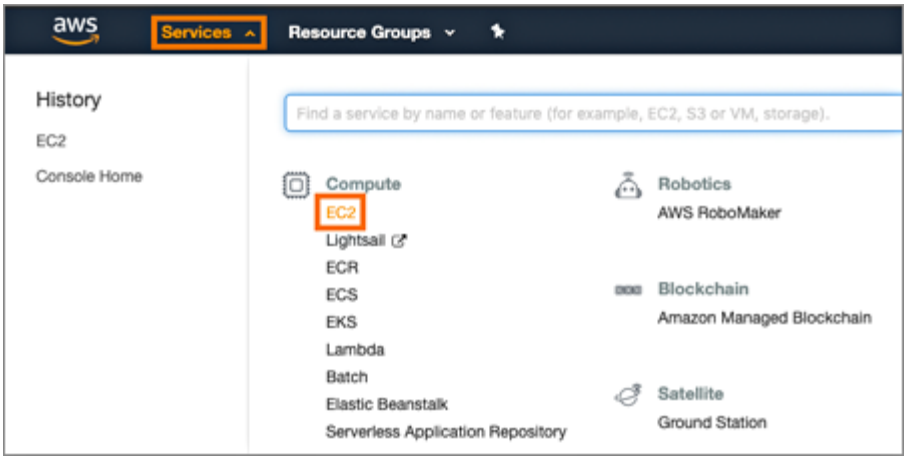
For more information on encryption supported with AWS:

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

Setting up your ThoughtSpot cluster

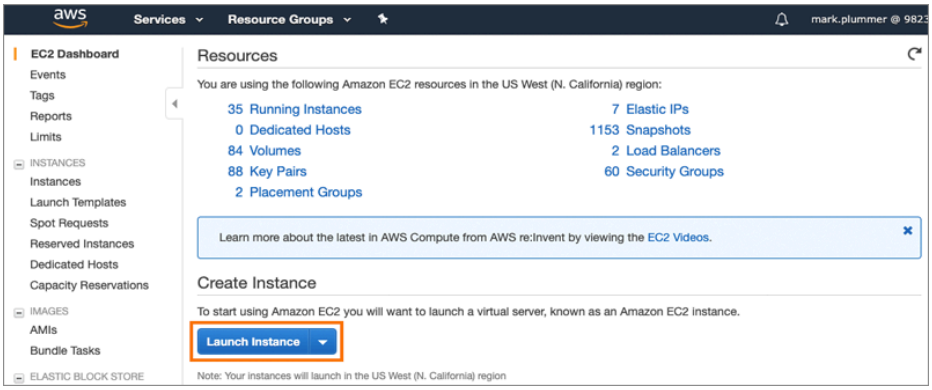
To set up a ThoughtSpot cluster in AWS, follow these steps:

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



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

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



4. Click the **My AMIs** tab and search **ThoughtSpot** to find the ThoughtSpot AMI.

5. Click **Select**.

6. On the **Choose an Instance Type** page, select a ThoughtSpot-supported instance type. (See ThoughtSpot AWS instance types [See page 0].)

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

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

S3 storage setting: If you are going to use the S3 storage option, ThoughtSpot recommends that you restrict access to a specific S3 bucket. Create a new IAM role that provides access to the specific bucket, and select it. For details on that, click **Create new IAM role**.

9. Click **Next: Add Storage**. Add the required storage based on your instance type (either EBS volumes or S3), and the amount of data you are deploying. For specific storage requirements, refer to ThoughtSpot AWS instance types [See page 0].

10. When you are done modifying the storage size, click **Next: Add Tags**.

11. Set a name for tagging your instances. This tag allows you to identify your instance more easily.

Configure security groups

1. Click **Next: Configure Security Group**.

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

   ☑ **Tip: Security setting for ThoughtSpot**

   - The VMs need intragroup security, i.e. every VM in a cluster must be accessible from one another. For easier configuration, ThoughtSpot recommends that you enable full access between VMs in a cluster.
   - Additionally, more ports must be opened on the VM to provide data staging capabilities to your network. Check Network policies [See page 0] to determine the minimum required ports you must open for your ThoughtSpot appliance.

3. Click **Review and Launch**.

4. After you have reviewed your instance launch details, click **Launch**.

5. 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. Without a key pair, you cannot SSH into the AWS instance later on.

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

Prepare the VMs

Before installing a ThoughtSpot cluster, an administrator must prepare the VMs. [See page 12]

Prepare AWS VMs for ThoughtSpot

Summary: Prepare the VMs before installing your ThoughtSpot cluster(s).

1. SSH into a VM.

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

2. Run `sudo /usr/local/scaligent/bin/prepare_disks.sh` to configure the VMs.

   Warning: Make sure you migrate any data off the disks. This command wipes the disks clean if they are not empty already.

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

3. Configure the VM based on your specific network information. Refer to your site-survey or ask your network administrator for that information.

4. Run `df -h` to confirm configuration.

   ```
   $ df -h
   ```

5. Repeat these steps for each of your VMs.

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

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

Open the required network ports

If you have not already opened the required network ports, see Network policies [See page 0] to determine which ports to open.

Install Cluster

Next, you must configure your nodes and install your cluster. Follow the steps in Installing AWS [See page 14].

Related information

EC2 Best Practices [See page 0]

Loading data from an AWS S3 bucket [See page 0]

Configure ThoughtSpot Nodes in AWS

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

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

Installation Prerequisites

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

1. **Review configuration options** Refer to AWS configuration options [See page 2] for detailed instance specs.
2. **Create the instance** Refer to Set up AWS for ThoughtSpot [See page 5] 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 17].
   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 20].

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

Install ThoughtSpot clusters in AWS

Summary: Learn how to install ThoughtSpot clusters in AWS.

Prerequisites

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

1. **Review configuration options** Refer to AWS configuration options [See page 2] for detailed instance specs.
2. **Create the instance** Refer to Set up AWS for ThoughtSpot [See page 5] 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 AWS [See page 14] 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 an s3 bucket for object storage, include the flag `--enable_cloud_storage=s3a`.

```
 $ tscli cluster create <release-number>.tar.gz --enabl
e_cloud_storage=s3a
```

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:

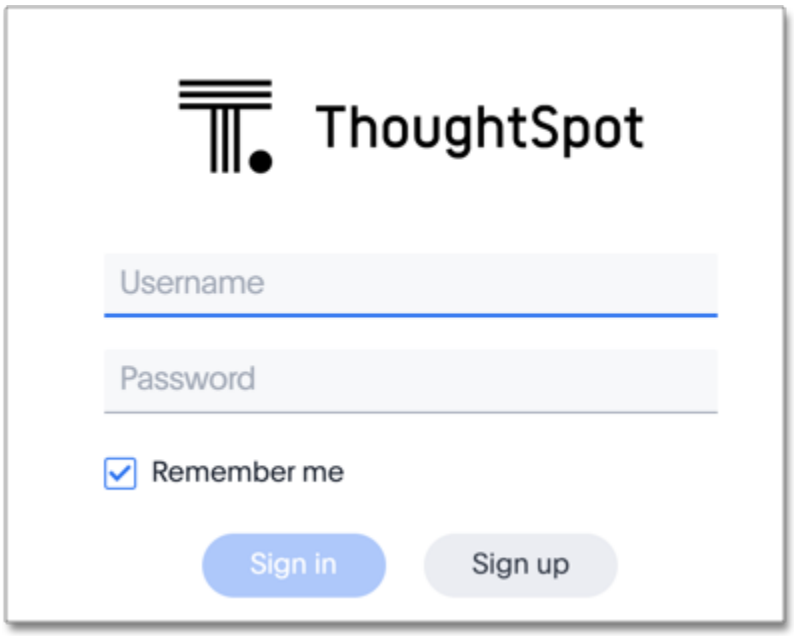
1. Start a browser from your computer.
2. Enter your secure IP information on the address line.

   ```
   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 23], 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]

Set up high availability for AWS

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

Configure high availabiity

Follow these steps to set up High Availability (HA) for your ThoughtSpot cluster using the AWS Elastic File System (EFS).

1. Create an EFS File System that spans across different availability zones, and across different subnets. Refer to AWS documentation on creating an EFS File System [See page 0].

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

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

4. Create an EFS directory in the `/home/admin` path.

5. Issue the following command to mount the new file system.

   Modify the fields as necessary for your installation.

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

6. Modify permissions to ensure that all clusters with EFS mount points have read and write permissions.

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

7. On the first cluster, create a snapshot on to the EFS mount point, and back it up.

```
tscli snapshot create EfsTest HA 2
```

8. Create a backup on the cluster.

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

9. Ensure that the backup is successful, and that it can be accessed from all clusters where EFS is mounted, by listing all backups and looking for the new one.

```
  tscli backup ls
```

10. Take down the first cluster instances.

11. On the second cluster, delete the existing cluster.

12. Create a new cluster by restoring from the first cluster backup. This is accessible from the EFS mount point.

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

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

Replace a cluster

For information on how to recover from infrastructure failure scenarios, see: Cluster replacement [See page 0].

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