Data Analytics Studio Installation

Date of Publish: 2018-08-14

Contents

Installation Overview

You must install the Data Plane Service, Data Analytics Studio Engine, and the Data Analytics Studio (DAS) application in the same order.

To install Data Analytics Studio, you must install the following components:

1. Data Plane Service
2. Data Analytics Studio Engine
3. Data Analytics Studio app

You are strongly encouraged to read completely through this entire document before starting the installation process, so that you understand the interdependencies and order of the steps.

Installing Data Analytics Studio Engine on Clusters

Install the DAS Engine on clusters to begin the installation of Data Analytics Studio service.

Procedure

1. Make sure all the prerequisites are met.
2. Configure an external database.
3. Install the Data Analytics Studio Engine.

Prerequisites for Data Analytics Studio Engine

Perform these tasks before you try to install the Data Analytics Studio Engine on the cluster.

Procedure

1. Ensure that the clusters are running the latest version of HDP.
2. Ensure that the following HDP components are installed and configured:
   - Hive
   - Knox

Configure Postgres database

If you want to use and manage your own database instead of the default database, you must configure the Postgres database and create the required roles in the database.

Procedure

1. Install the supported version of Postgres using the following commands:

```
yum install https://download.postgresql.org/pub/repos/yum/9.6/redhat/
rhel-7-x86_64/pgdg-centos96-9.6-3.noarch.rpm
```

```
yum install postgresql96-contrib postgresql96-server
```

For more information about the supported version of Postgres, see the DAS Support Matrix.

2. Initialize the Postgres database by running the following command:

```
service postgresql-9.6 initdb
```

3. Open the Postgres configuration file pg_hba.conf for editing by entering the following command:

```
vi /var/lib/pgsql/9.6/data/pg_hba.conf
```

4. Configure Postgres to be accessible from Event Processor and DAS hosts. Add similar lines as follows to the configuration file:

```
local   all             <dbuser>                                md5
host    all             <dbuser>      0.0.0.0/0                 md5
host    all             <dbuser>      ::/0                      md5
local   all             postgres         ident
```

5. Edit the Postgres configuration file postgresql.conf by opening and adding the following information:

```
vi /var/lib/pgsql/9.6/data/postgresql.conf
```

```
listen_addresses = '*'
```

6. Start the Postgres database by running the following command:

```
service postgresql-9.6 start
```

7. Run the following commands to create roles in Postgres:

```
psql -tc "SELECT 1 FROM pg_database WHERE datname = <dbuser>" | grep 1 ||
 (
psql -c "CREATE ROLE <dbuser> WITH LOGIN PASSWORD <dbpass>;" &&
psql -c "ALTER ROLE <dbuser> SUPERUSER;" &&
psql -c "ALTER ROLE <dbuser> CREATEDB;" &&
psql -c "CREATE DATABASE <dbuser>;" &&
psql -c "GRANT ALL PRIVILEGES ON DATABASE <dbuser> TO <dbuser>;")
```

Replace <dbuser> with the database username/database name and <dbpass> with database password.

> **Note:** The database user and database name must be the same. It should be the one that is used as database username in the DAS configuration on Ambari.

Install the Data Analytics Studio Engine

Data Analytics Studio requires that the DAS Engine be installed on all custers. The engine is installed on the Ambari host, using an Ambari management pack (MPack). An MPack bundles service definitions, stack definitions, and stack add-on service definitions.

About this task

This task must be completed on all the clusters to be used with DAS.

Before you begin

You must have root access to the Ambari Server host node to perform this task.

> **Important:** Download the required repository tarballs from the Hortonworks customer portal by following the instructions provided as part of the product procurement process. The repository tarballs for the DAS Engine are different from the DAS app repository tarballs.

Procedure

1. Log in as root to an Ambari host on a cluster.
2. Install the Data Analytics Studio MPack by running the following command, replacing <mpack-file-name> with the name of the MPack.

```
ambari-server install-mpack --mpack=<mpack-file-name> --verbose
```

3. Restart the Ambari server.

```
ambari-server restart
```

4. Launch Ambari in a browser and log in.
   http://<ambari-server-host>:8080

   Default credentials are:

   Username: admin
   Password: admin
5. In the Ambari Services navigation pane, click **Actions** > **Add Service**.
   The **Add Service Wizard** displays.
6. On the **Choose Services** page of the Wizard, select the Data Analytics Studio service to install in Ambari, and then follow the on-screen instructions.

   Other required services are automatically selected.
7. When prompted to confirm addition of dependent services, give a positive confirmation to all.

   This adds other required services.
8. On the **Assign Masters** page, you can choose the default settings.
9. On the **Customize Services** page, expand **Advance_data_analytics_studio-database** and fill in the database details and other required fields that are highlighted.

   a. If you installed Postgres on your own:

      1. Uncheck **Create Data Analytics Studio database**.
      2. Set the database host in the **Data Analytics Studio database hostname**.
      3. Set the database username in **Data Analytics Studio database username**.

         Note: The hostname is ignored if the **Create Data Analytics Studio database** option is checked, the database will be installed on the same host as webapp.

   b. Database Password - Enter the password.

   You can set credentials to whatever you want.
10. If Hive SSL is enabled, set the **Hive session params** in DAS configuration as follows:

```
sslTrustStore=/etc/security/serverKeys/
hivetruststore.jks;trustStorePassword=your_password
```

11. If KNOX SSO is enabled, update **admin_users** under **Advanced data_analytics_studio-security-site**, with the list of users who need admin access to DAS.

    Note: Only admin users have access to all the queries. Non-admin users can access only their queries.
12. Complete the remaining installation wizard steps and exit the wizard.
13. Ensure that all components required for your DAS service have started successfully.
14. Make sure to restart all the affected services in Ambari.

Configure SSL/TLS

If your HDP cluster is SSL enabled, then you can configure SSL. You can use one of the two options to set up SSL certificates.

- Setup trusted CA certificates
- Setup self-signed certificates

Set up trusted CA certificate

You can enable SSL for the DAS Engine using a certificate from a trusted Certificate Authority (CA). Certificates from a trusted CA are primarily used in production environments. For a test environment, you can use a self-signed certificate.

Before you begin

- You must have root user access to the clusters on which DAS Engine is installed.
- You must have obtained a certificate from your CA, following their instructions.

Procedure

1. Log in as root user on the cluster with DAS Engine installed.
2. Import the Certificate Chain Certificate and the certificate you obtained from your CA.

```
keytool -import -alias root -keystore <path_to_keystore_file> -
trustcacerts -file <certificate_chain_certificate>
```

```
keytool -import -alias jetty -keystore <path_to_keystore_file> -file
 <certificate_from_CA>
```

Set up self-signed certificates

You can enable SSL for the DAS Engine using a self-signed certificate. Self-signed certificates are primarily used in test environments. For a production environment, you should use a certificate from a trusted CA.

Before you begin

You must have root user access to the clusters on which DAS Engine is installed.

Procedure

1. Log in as root user on the cluster with DAS Engine installed.
2. Generate a key pair and keystore for use with DAS Engine.

```
keytool -genkey -alias jetty -keystore <certificate_file_path>
-storepass <keystore_password> -dname 'CN=das.host.com, OU=Eng, O=ABC
 Corp,
L=Santa Clara, ST=CA, C=US' -keypass <key_password>
```

Follow the prompts and enter the required information.

- CN must be the FQDN of the DAS Engine host.
- Default value for the key password is *password*.

  If you change the password, then you have to update the DAS configuration.

Following is a sample command output:

```
keytool -genkey -alias jetty -keystore ~/tmp/ks -storepass password
```

```
What is your first and last name?
  [Unknown]:  das.host.com
What is the name of your organizational unit?
  [Unknown]:  Eng
What is the name of your organization?
  [Unknown]:  ABC Corp
What is the name of your City or Locality?
  [Unknown]:  Santa Clara
What is the name of your State or Province?
  [Unknown]:  CA
What is the two-letter country code for this unit?
  [Unknown]:  US
Is CN=das.host.com, OU=Eng, O=ABC Corp, L=Santa Clara, ST=CA, C=US correct?
  [no]:  yes

Enter key password for <jetty>
  (RETURN if same as keystore password):
```

Note: You will have to use this keystore file while configuring the DAS Engine for TLS in Ambari.

3. Export the certificate.

```
keytool -exportcert -alias jetty -keystore /my/file.keystore -file
 <certificate file path> -storepass <keystore_password> -rfc
```

Configure SSL/TLS in Ambari

In the Ambari UI, you enable TLS for DAS Engine and update the DAS Engine configuration if settings change.

Procedure

1. Copy the keystore files generated in the earlier procedures to webapp and event processor hosts. Make sure they are owned by configured user for DAS. The default user is hive.

   For example:

   ```
   /etc/security/certs/das-cert.jks
   ```

2. Navigate to **Data Analytics Studio** > **Configs**.
3. Set the following properties in **Advanced data_analytics_studio-security-site** section.

Field	Value
ssl_enabled	Make sure it is checked.
webapp_keystore_file	Enter the keystore path on the webapp host.
webapp_keystore_password	Enter the password used in the previous procedure.
event_processor_keystore_file	Enter the keystore path on the event processor.
event_processor_keystore_password	Enter the password used in the previous procedure.

4. In the **Advanced data_analytics_studio-webapp-properties** section, set **Data Analytics Studio Webapp server protocol** property to **https**.
5. In the **Advanced data_analytics_studio-event_processor-properties** section, set **Data Analytics Studio Event Processor server protocol** property to **https**.

Configure Knox SSO for Data Analytics Studio

Update the Knox SSO settings in the DAS Configuration in Ambari.

Procedure

1. Export the Knox certificate:
   a) From the Knox Gateway machine, run the following command:

```
$JAVA_HOME/bin/keytool -export -alias gateway-identity -rfc -file
 <cert.pem> -keystore /usr/hdp/current/knox-server/data/security/
keystores/gateway.jks
```

   b) When prompted, enter the Knox master password.
   c) Note the location where you save the cert.pem file.
2. Enable the Knox SSO topology settings:
   a) From **Ambari** > **Data Analytics Studio** > **Configs** > **Advanced data_analytics_studio-security-site**, check to select **knox_sso_enabled**.
   b) Set **knox_sso_url** value as **https://<knox-host>:8443/gateway/knoxsso/api/v1/websso**.
   c) Copy the contents of the PEM file exported in Step 1 to knox_publickey. Make sure the certificate headers are not copied.
   d) Click **Save** and click through the confirmation pop-ups.
   e) Restart Data Analytics Studio webapp.
   f) Select **Actions** > **Restart All Required** to restart all other services that require a restart.

Installing the Data Analytics Studio App

After installing the DP Platform, you must install the Data Analytics Studio application.

About this task

You must install the app on the same host as DP Platform.

Procedure

1. Set up a local repository.
2. Create the repository configuration file.
3. Install the Data Analytics Studio app.
4. Enable the clusters for DAS in the Data Plane Service Platform.
5. Add users and assign roles for the DAS app.

Set Up a Local Repository

Setting up a local repository involves moving the tarball to the selected mirror server and extracting the tarball to create the repository.

Before you begin

Ensure that you have downloaded the required tarballs from the Hortonworks customer portal by following the instructions provided as part of the product procurement process.

You must have completed the preparatory tasks before setting up a repository.

Procedure

1. Copy the repository tarballs to the web server directory and expand (uncompress) the archive file:
   a) Navigate to the web server directory you previously created.
   cd /var/www/html/

All content in this directory is served by the web server.

b) Move the tarballs to the current directory and expand each of the repository tarballs that you downloaded.

Replace <file-name> with the actual name of the RPM tarball that you are expanding.

tar zxvf <file-name>.tar.gz

When you expand the tarball, subdirectories are created in /var/www/html/, such as DAS/centos7. These directories contain the repositories.

Expanding the DAS app tarball can take several seconds.

2. Confirm that you can browse to the newly created local repositories by using the base URLs: http://<webserver-host-name>/<repo-name>/<OS>/<service-version-X>

- <webserver-host-name>

  This is the FQDN of the web server host.
- <repo-name>

  This is composed of the abbreviated name of the repository, such as DAS.
- <OS>

  This is the operating system version.
- <service-version-X>

  This is the version number of the downloaded repository, appended with a unique version number.

Base URL Examples

DAS Base URL:

```
http://webserver.com:port/DAS/centos7/1.0.0.0-X
```

Note the base URLs because you need them to install the DAS app on the host and to install the associated agent on the clusters.

3. If you have configured multiple repositories in your environment, then install the following plugin on all the nodes in your cluster:

```
yum install yum-plugin-priorities
```

4. Edit the /etc/yum/pluginconf.d/priorities.conf file to add the following values:

```
[main]
enabled=1
gpgcheck=0
```

Results

The repositories for DAS are now prepared for installation.

What to do next

Create the configuration file for the DAS repository.

Create the Repository Configuration File

A repository configuration file must be created for the DAS Service on the DPS host. The file is required to identify the path to the repository data, and establish whether a GPG signature check should be performed on the repository packages. You need only one repository configuration file.

Procedure

1. Navigate to the repository directory.

```
cd /etc/yum.repos.d/
```

2. Create a repository file.

```
vi das.repo
```

Alternatively, you can copy an existing repository file to edit.

3. Add the following content in the repository file:
#VERSION_NUMBER=<downloaded-version#> [<service-name-abbreviation>]

This is composed of the service name abbreviation and version number (includes the build number). Example: DAS-APP-1.0.0.0-59

```
name=<service-name-abbreviation> Version - <service-name-abbreviation>
```

```
baseurl=http://<webserver-host-name>/<directory-containing-repo>
```

<webserver-host-name> is the FQDN of the web server host that contains the repository. This is the same base URL that you used in the task to prepare the repositories.

<directory-containing-repo> is the path expanded from the tarball.

```
gpgcheck=1
gpgkey=http://<webserver-host-name>/<directory-containing-repo>/RPM-GPG-
KEY/RPM-GPG-KEY-Jenkins
enabled=1
priority=1
```

Example Repository File

```
#VERSION_NUMBER=1.0.0.0-1
[DAS-APP-1.0.0.0-59]
name=DAS-APP Version - DAS-APP-1.0.0.0-1
baseurl=http://<your_webserver>:port/DAS-APP/centos7/1.0.0.0
gpgcheck=1
gpgkey=http://<your_webserver>:port/DAS-APP/centos7/1.0.0.0/RPM-GPG-KEY/
RPM-GPG-KEY-Jenkins
enabled=1
priority=1
```

Install the Data Analytics Studio Service App

Follow the instructions to install the Data Analytics Studio Service app.

Before you begin

You must have successfully installed DPS Platform and DPS is running.

Procedure

1. Log in to the host on which you have set up the DPS repositories as a root user.
2. Install the RPMs for the DAS service application.

```
yum install das-app
```

A folder is created that contains the Docker image tarball files and a configuration script.

If the yum command fails, then the local repository was not set up correctly. Check the repository file /etc/yum.repos.d/das.repo on the host.

3. Navigate to the directory containing the installation scripts for the DAS service, for example:

```
cd /usr/das-app/das-xxx/apps/das/bin
```

where das-xxx refers to the version number of the DAS app.

4. Load the DAS Docker images and initialize the environment.

```
./dasdeploy.sh load
```

```
./dasdeploy.sh init
```

It prompts for the master password that was used for initializing the Data Plane platform. Make sure you enter the same master password.

Loading the images might take a while.

Note:

If you run into errors while deploying the DAS application, then destroy the deployment using the ./dasdeploy.sh destroy command and re-install the app. To check the logs of the das-app container, you can use the ./dasdeploy.sh logs command.

5. Verify that the container you installed is running.

```
./dasdeploy.sh ps
```

Make sure that the container with the name das-app is running.

Enable the clusters for DAS in the Data Plane Service

After installing the DAS app, you must enable clusters for it in the Data Plane Service Platform.

Procedure

1. Log in to the DP Platform as a DataPlane Admin user.
2. Select the clusters from the list of clusters.

   The Services page is displayed.
3. Move the cursor over the service and click the **Enable** button on the clusters on which you want to enable the DAS service.

   A verification page is displayed.
4. Click **Next**.

Results

The cluster is enabled for DAS service.

Add Users and Assign Roles for the DAS App

After you set up the LDAP configuration for DP Platform, you need to add users for the DAS app. During the LDAP configuration, you add users and groups that can log in as a DP admin. You must now assign roles to users and groups that allow the users to access the services that plug into DataPlane.

About this task

You must select the Data Analytics Studio User role for accessing the DAS Service. Users and groups should be assigned this role to access Data Analytics Studio service. To enable Data Analytics Studio User role, see Role Management section of the *Data Plane Service Administration Guide.*

Before you begin

User accounts must already exist within your corporate LDAP prior to adding the user to DPS Platform.

The DataPlane Admin role is required to perform this task.

Procedure

1. Log in to the DP Platform.
2. Click the (Users) icon in the DP Platform navigation pane.
3. On the Users and Groups page, click **Add User**.
4. Enter the name of the user.

   With your own LDAP server, the user must already exist within your corporate LDAP. If you are using the packaged LDAP, enter one of the predefined users (guest, sam, tom). The name auto-populates as you type.

   Tip:

   You must click the name of the user when it displays and ensure it appears in the Username field on a dark background.

   If the name appears on a white background, it means the name is not recognized and the action fails.
5. Select the Data Analytics Studio User role to assign to the user.

   With the Data Analytics Studio User role, you can perform all actions in the Data Steward Studio service UI and also manage DAS-enabled clusters in the DP Platform.
6. Click **Save**.

   You can log in and see service inside the DP Platform. If Data Anlaytics Studio User role is the only role assigned, you will be directed to the Data Analytics Studio Service. If you have more roles assigned, you can select the Data Analytics Studio Service in the navigation menu in the top left corner.

   Note: If you assign the Data Analytics Studio User role to yourself or to the group that you belong to, you must log out and log in again to verify that Data Analytics Studio Service is available.

   The new user displays in the list on the Users page.