Cloudera Data Science Workbench

Migrating a Deployment to a New Set of Hosts

Date published: 2020-02-28
Date modified: 2021-02-25

CLOUDERA

Legal Notice

Contents

Migrating a Deployment to a New Set of Hosts

This section describes how to migrate a Cloudera Data Science Workbench deployment to a new set of gateway hosts.

Migrating a CSD Deployment

This section describes how to migrate a CSD-based Cloudera Data Science Workbench service to a new set of gateway hosts.

Add and Set Up the New Hosts

Add and set up the new hosts for migrating a CSD deployment.

Procedure

1. Add new hosts to your cluster as needed.

   Make sure they are gateway hosts that have been assigned gateway roles for HDFS, YARN, and Spark 2. Do not run any other services on these hosts.

2. Set up the new hosts as per the Cloudera Data Science Workbench hardware requirements listed here.

   - Disable Untrusted SSH Access on the new hosts.
   - Configure Block Devices on the new hosts.

Copy the JDK to the new host

Provides information on how to copy the JDK to the new host.

Procedure

Copy the /usr/java directory to the new host.

Copy the DNS Nameserver to the new host

Provides information on how to copy the DNS nameserver to the new host.

Procedure

Copy the /etc/resolv.conf file to the new host.

Copy the Kerberos Configurations

Provides information on how to copy the Kerberos configurations.

Procedure

Copy the /etc/jkr5.conf file to the new host.

Stop the CDSW Service

Use Cloudera Manager to stop all roles of the CDSW service.

About this task

On Cloudera Data Science Workbench 1.4.0 (and lower), do not stop or restart Cloudera Data Science Workbench without using the cdsw_protect_stop_restart.sh script. This is to help avoid the data loss issue detailed in TSB-346.

Procedure

1. Log into the Cloudera Manager Admin Console.
2. On the Home Status tab, click ▼ to the right of the CDSW service and select Stop from the dropdown.
3. Confirm your choice on the next screen. When you see a Finished status, the action is complete.

Backup Application Data

In Cloudera Data Science Workbench all stateful data is stored on the master host at /var/lib/cdsw. Backup the contents of this directory before you begin the migration process.

Procedure

1. After stopping CDSW, and before running the following tar command, wait 2-5 minutes (depending on your disk speed) to ensure that all data from CDSW is successfully written to the disks. Otherwise the tar command may not capture all recent changes.
2. To create the backup, run the following command on the master host:

```
tar cvzf cdsw.tar.gz /var/lib/cdsw/*
```

Delete CDSW Roles from Existing Hosts

This provides the steps for deleting CDSW roles from an existing hose.

Procedure

1. Log into the Cloudera Manager Admin Console.
2. Go to the CDSW service.
3. Click the Instances tab.
4. Select all the role instances.
5. Select Actions for Selected > Delete. Click Delete to confirm the deletion.

Move Backup to the New Master

This provides steps to move the backup to the new master.

Procedure

1. Copy the backup taken previously to the host that will be the new Cloudera Data Science Workbench master.
2. Unpack the contents of the backup into /var/lib/cdsw.

```
tar xvzf cdsw.tar.gz -C /var/lib/cdsw
```

Update DNS Records for the New Master

Update your DNS records with the IP address for the new master host.

Add Role Instances for the New Hosts

This provides steps to add role instances for the new host.

Procedure

1. Log into the Cloudera Manager Admin Console.
2. Go to the CDSW service.
3. Click the Instances tab.

4. Click Add Role Instances. Assign the Cloudera Data Science Workbench Master, Application, and Docker Daemon roles to the new master host. If you want to configure worker hosts, assign the Cloudera Data Science Workbench Worker and Docker Daemon roles to the new workers.

> ⚠️ **Important:** Do not assign the Master and Worker roles to the same host. Even if you only have one host for Cloudera Data Science Workbench, the Master can automatically perform the functions of a Worker host as needed.

5. Click Continue. On the Review Changes page, review the configuration changes to be applied. The wizard finishes by performing any actions necessary to add the new role instances.

   Do not start the new roles at this point. You must run the Prepare Node command as described in the next step before the roles are started.

Run the Prepare Node command on the New Hosts

This provides information on how to run the prepare command on the new hosts.

Procedure

1. The new hosts must have the following packages installed on it.

```
nfs-utils
libseccomp
lvm2
bridge-utils
libtool-ltdl
iptables
rsync
policycoreutils-python
selinux-policy-base
selinux-policy-targeted
ntp
ebtables
bind-utils
nmap-ncat
openssl
e2fsprogs
redhat-lsb-core
conntrack-tools
socat
```

   You can either manually install these packages now, or , allow Cloudera Manager to install them as part of the Prepare Node command later in this step.

   If you choose the latter, make sure that Cloudera Manager has the permissions needed to install the required packages. To do so, go to the CDSW service and click Configuration. Search for the Install Required Packages property and make sure it is enabled.

2. Run the Prepare Node command on the new hosts.
   a) Go to the CDSW service.
   b) Click Instances.
   c) Select all the role instances.
   d) Select Actions for Selected > Prepare Node. This will install the required set of packages on all the new hosts.

Start the CDSW Service

This provides the steps to start the CDSW service.

Procedure

1. Log into the Cloudera Manager Admin Console.

2.

On the Home Status tab, click ⏷ to the right of the CDSW service and select Start from the dropdown.

3. Confirm your choice on the next screen. When you see a Finished status, the action is complete.

Migrating an RPM Deployment

This section describes how to migrate an RPM-based Cloudera Data Science Workbench service to a new set of gateway hosts.

Add and Set Up the New Hosts

Provides the steps to add and set up the new hosts.

Procedure

1. Add new hosts to your cluster as needed. Make sure they are gateway hosts that have been assigned gateway roles for HDFS, YARN, and Spark 2. Do not run any other services on these hosts.

2. Set up the new hosts as per the Cloudera Data Science Workbench hardware requirements listed here.

- Disable Untrusted SSH Access on the new hosts.
- Configure Block Devices on the new hosts.

Copy the JDK to the new host

Provides the steps to copy the JDK to the new host.

Procedure

Copy the /usr/java directory to the new host.

Copy the DNS Nameserver to the New Host

Provides the steps to copy the DNS nameserver to the new host.

Procedure

Copy the /etc/resolv.conf file to the new host.

Copy the Kerberos Configurations

Povides the steps to copy the Kerberos configurations.

Procedure

Copy the /etc/jkr5.conf file to the new host.

Stop Cloudera Data Science Workbench

Provides the steps to stop the Cloudera Data Science Workbench service.

About this task

On Cloudera Data Science Workbench 1.4.0 (and lower), do not stop or restart Cloudera Data Science Workbench without using the cdsw_protect_stop_restart.sh script. This is to help avoid the data loss issue detailed in TSB-346.

Procedure

Run the following command on the master host to stop Cloudera Data Science Workbench.

```
cdsw stop
```

Backup Application Data

Provides the steps to back up the application data.

About this task

In Cloudera Data Science Workbench all stateful data is stored on the master host at /var/lib/cdsw. Backup the contents of this directory before you begin the migration process.

Procedure

1. After stopping CDSW, and before running the following tar command, wait 2-5 minutes (depending on your disk speed) to ensure that all data from CDSW is successfully written to the disks. Otherwise the tar command may not capture all recent changes.

2. To create the backup, run the following command on the master host:

```
tar cvzf cdsw.tar.gz /var/lib/cdsw/*
```

Remove Cloudera Data Science Workbench from Existing Hosts

Provides the steps to remove Cloudera ata Science Workbench from the existing hosts.

Procedure

Run the following commands on the existing master and any worker hosts you want to migrate:

```
cdsw stop
yum remove cloudera-data-science-workbench
```

Move Backup to New Master

Provides the steps to move the backup to a new master.

Procedure

Copy the backup taken in the previous step to the host that will be the new Cloudera Data Science Workbench master. Unpack the contents of the backup into /var/lib/cdsw.

```
tar xvzf cdsw.tar.gz -C /var/lib/cdsw
```

Update DNS Records for the New Master

Update your DNS records with the IP address for the new master host.

Install Cloudera Data Science Workbench on New Master Host

Provides information on how to install Cloudera Data Science Workbench on the new master host.

Procedure

For instructions, see Installing Cloudera Data Science Workbench 1.8.1 using Packages.