CDP Journey HDP Migration Workshop

venky@cloudera.com

Version 0.2, July 27, 2021: Draft

Table of Contents

Colophon

Version: 0.2 : July 27, 2021 : Draft

Intro

CDP Journey , hands-on training and first look at HDP Migration.

Pre-Requisites (For Instructors)

1. Create a Centos 7.x VM
   a. On AWS , eu-west-2 use `ami-00846a67`.
2. Create a VM instance in the normal way and attach SSH Keys.
3. Login to the VM
   a. Change hostname of the VM

```
curl -sSL https://gist.github.com/abajwa-hw/9d7d06b8d0abf705ae311393d2ecdeec/raw |
sudo -E sh
```

   b. Install HDP 2.6.5 with Ambari 2.5.1.

```
curl -sSL https://gist.github.com/vsellappa/483a2ebf8ffd21735fe501616a24fc26/raw |
sudo -E sh
```

   c. Change to root. All steps below need to run as `root`.

```
sudo -i
```

   d. Install `postgresql10-server`

```
yum install https://download.postgresql.org/pub/repos/yum/reporpms/EL-7-
x86_64/pgdg-redhat-repo-latest.noarch.rpm
yum install postgresql10-server
```

   e. Install `openconnect`

```
yum install epel-release
yum install openconnect
```

   OpenConnect is required to login to the corporate network to get a version of the `cloudera-installer-bin`. This can be downloaded via the standard trial page as well.

   f. Login to the corporate network from the AWS VM. Open 2 separate bash windows and do *not* use FullTunnel-VPN.

```
openconnect -u <username> connect.cloudera.com
```

i. Download AM2CM.

```
git clone https://github.infra.cloudera.com/Starship/am2cm.git
```

ii. Download `cloudera-manager-installer.bin` .

```
wget http://cloudera-build-us-west-
1.vpc.cloudera.com/s3/build/3023178/cm7/7.1.1/cloudera-manager-installer.bin

chmod 755 cloudera-manager-installer.bin

export CLOUDERA_REPO_FIX_CMD="sed -e
's,https://archive.cloudera.com/,http://cloudera-build-us-west-
1.vpc.cloudera.com/s3/build/3023178/,g' -i /etc/yum.repos.d/cloudera-
manager.repo";

echo $CLOUDERA_REPO_FIX_CMD

./cloudera-manager-installer.bin
```

```
Resolving Dependencies
--> Running transaction check
---> Package cloudera-manager-server.x86_64 0:7.1.1-3023178.el7 will be installed
--> Processing Dependency: cloudera-manager-daemons = 7.1.1 for package: cloudera-manager-server-7.1.1-3023178.el7.x86_64
--> Running transaction check
---> Package cloudera-manager-daemons.x86_64 0:7.1.1-3023178.el7 will be installed
--> Finished Dependency Resolution

Dependencies Resolved

================================================================================
 Package                  Arch    Version              Repository         Size
================================================================================
Installing:
 cloudera-manager-server   x86_64  7.1.1-3023178.el7    cloudera-manager   12 k
Installing for dependencies:
 cloudera-manager-daemons  x86_64  7.1.1-3023178.el7    cloudera-manager  1.4 G

Transaction Summary
================================================================================
Install  1 Package (+1 Dependent package)

Total download size: 1.4 G
Installed size: 1.6 G
Downloading packages:
--------------------------------------------------------------------------------
Total                                           2.0 MB/s | 1.4 GB  12:00
Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
```

NOTE	This is a slow process and can take anywhere upto 45 minutes depending on network bandwidth.

iii. Install CM-agents

```
yum install cloudera-manager-daemons cloudera-manager-agent
```

iv. Change CM agent config

```
vi /etc/cloudera-scm-agent/config.ini

Change server_host=<CM_HOST>
```

g. On completion of install, go to <host>:7180 and activate the cloudera trial license from the UI.

```
uname: admin
pwd: admin
```

h. The use of CM here is to only validate the Ambari blueprint by using the Swagger API. The cluster itself is not expected to be used in any other way.

```
http://<hostname>:7180/static/apidocs/ui/index.html#!/ClouderaManagerResource/updat
eDeployment2
```

i. Regenerate the hosts file on reboot.

```
curl -sSL https://gist.github.com/abajwa-hw/4bf004d0fb065d404760eaeebc15e74e/raw |
sudo -E sh
```

j. AMI in AWS eu-west2 with complete install: ami-0b2ad1125cf85a075.

k. Create a VPN to allow access to all ports when participants log in. I use OpenConnect(Optional)

For Participants

IMPORTANT	Screenshots below might be for other versions of HDP / CDP. Please follow the instruction carefully and reach out to the instructor in case of questions.

1. Login to your cluster via ssh How To
2. If services are switched off re-start all of them via the Ambari UI.

Phase 1 : CDP-PvC-Base aka CDP-Dc With Ambari

CDP-PvC-Base aka CDP-Dc replaces Ambari with Cloudera Manager as the operational frontend,however for migration its easier to migrate all the services underneath HDP using Ambari and then migrate Ambari itself i.e. get a HDP with all the CDP-Dc services with Ambari and then move the services under Cloudera-Manager in Phase 2.

NOTE	This phase is an intermediate stage and this version of HDP is only for migration and will not be available externally.

Upgrade Ambari to 7.1.x aka Ambari-Dc

- Upgrading to Ambari-Dc is much like any other upgrade in HDP. The big differences are :
  a. The underlying platform will be running Cloudera Runtime 7.x for all services.
  b. This is important because it provides an upgrade path to using the latest parts of the underlying stack. These include upgrades to HDFS, Hive, Spark, Ranger, Oozie, HBase, etc.
  c. A key aspect is Ambari and the HDP-7.1.x clusters being managed by Ambari can be upgraded independently , this separation of reduces the chances of failure and allows additional testing phases with minimal downtime.
- **Upgrade Process**
  1. Change to root.

```
sudo -i
```

  2. Stop all agents and ambari server.

```
ambari-agent stop
ambari-server stop
```

```
[root@ip-172-31-28-227 ~]# ambari-agent stop
Verifying Python version compatibility...
Using python  /usr/bin/python
Found ambari-agent PID: 10674
Stopping ambari-agent
Removing PID file at /run/ambari-agent/ambari-agent.pid
ambari-agent successfully stopped
[root@ip-172-31-28-227 ~]# ambari-server stop
Using python  /usr/bin/python
Stopping ambari-server
Waiting for server stop...
Ambari Server stopped
[root@ip-172-31-28-227 ~]#
```

3. Backup ambari properties.

```
cp /etc/ambari-server/conf/ambari.properties /etc/ambari-
server/conf/ambari_orig.properties
```

4. Backup ambari database.

```
mkdir -p /tmp/dbdump
cd /tmp/dbdump/

pg_dump -U [AMBARI_DB_USERNAME] -f ambari.sql
Password: [AMBARI_DB_PASSWORD]
```

Variable	Description	Default
AMBARI_DB_USERNAME	Database username	ambari
AMBARI_DB_PASSWORD	Database password	bigdata

TIP Open `ambari.sql` from above and make a note of the schema and content of the `blueprint` table.

5. Backup server metainfo.

```
ambari-server backup
```

NOTE In some cases this will give an error with regards to a runtime folder `/var/run/ambari-server/bootstrap/` , which is an artifact of the way the VM has been created, it can be safely ignored.

```
[root@ip-172-31-28-227 ~]# ambari-server backup
Using python  /usr/bin/python
Backing up Ambari File System state... *this will not backup the server database*
Backup requested.
No path specified. Will use /var/lib/ambari-server/Ambari_State_Backup.zip
Backup process initiated.
Creating zip file...
Zip file created at /var/lib/ambari-server/Ambari_State_Backup.zip
Backup complete.
Ambari Server 'backup' completed successfully.
[root@ip-172-31-28-227 ~]#
```

1. Check ambari server info.

```
yum info ambari-server
```

2. Make a copy of the ambari repo file.

```
cp /etc/yum.repos.d/ambari.repo /etc/yum.repos.d/ambari.repo_orig
```

3. Change the ambari repo to point to the latest version.

```
wget -nv https://cd43d367-6fcf-4a12-a807-
da22f1069c7d:13f2a6a1e854@archive.cloudera.com/p/ambaridc/7.x/7.1.6.0/centos7/ambaridc
.repo -O /etc/yum.repos.d/ambaridc.repo
```

4. Check if this url is reachable. Save this link.
5. Upgrade ambari server and agent.

```
yum upgrade ambari-server
yum upgrade ambari-agent
```

6. Check if the upgrades are successful.

```
yum info ambari-server
yum info ambari-agent
rpm -qa | grep ambari-agent
```

```
[root@ip-172-31-28-227 ~]# yum info ambari-server
Loaded plugins: fastestmirror
Loading mirror speeds from cached hostfile
 * base: d36uatko69830t.cloudfront.net
 * extras: d36uatko69830t.cloudfront.net
 * updates: d36uatko69830t.cloudfront.net
Installed Packages
Name        : ambari-server
Arch        : x86_64
Version     : 7.1.0.0
Release     : 70
Size        : 440 M
Repo        : installed
From repo   : ambari-7.1.0.0-70
Summary     : Ambari Server
URL         : https://www.apache.org
Licence     : 2012, Apache Software Foundation
Description : Maven Recipe: RPM Package.

[root@ip-172-31-28-227 ~]# yum info ambari-agent
Loaded plugins: fastestmirror
Loading mirror speeds from cached hostfile
 * base: d36uatko69830t.cloudfront.net
 * extras: d36uatko69830t.cloudfront.net
 * updates: d36uatko69830t.cloudfront.net
Installed Packages
Name        : ambari-agent
Arch        : x86_64
Version     : 7.1.0.0
Release     : 70
Size        : 70 M
Repo        : installed
From repo   : ambari-7.1.0.0-70
Summary     : Ambari Agent
URL         : https://www.apache.org
Licence     : 2012, Apache Software Foundation
Description : Maven Recipe: RPM Package.

[root@ip-172-31-28-227 ~]# rpm -qa | grep ambari-agent
ambari-agent-7.1.0.0-70.x86_64
[root@ip-172-31-28-227 ~]#
```

7. Upgrade the Ambari server DB.

```
ambari-server upgrade
```

8. Start ambari server and agent.

```
ambari-server start
ambari-agent start
```

   a. Check server log at: `/var/log/ambari-server/ambari-server.log`

   b. Check agent log at: `/var/log/ambari-agent/ambari-agent.log`

9. Restart all services from the Ambari UI menu. This should be all successful and green ready for migrating to the latest HDP.

CAUTION	The above is an abbreviated *golden path* version of the upgrade process intended as an exercise.

Upgrade HDP to 7.1.x

- Upgrading HDP to 7.1.x is not a rolling upgrade i.e. it requires cluster downtime, potential

conflicting config changes and *cannot* be rolled back. Key considerations include:

a. For kerberised clusters KDC needs to be migrated independently outside of Ambari.

b. The upgrade process does not back up the Hive MetaStore, nor does it compact ACID tables. Before upgrading Hive, you must:

   i. Manually make a backup of your Hive metastore database after using the pre-upgrade tool and before upgrading.

   ii. If you have ACID tables in your Hive metastore, enable ACID operations using Ambari Web or set Hive configuration properties to enable ACID.

- **Upgrade Process**

a. The upgrade can be run and orchestrated from the Ambari UI. Understand the process here.

b. In short, the process is :

   i. Choose a HDP version to upgrade to.

   ii. Register the version using Ambari UI.

   iii. Install the version on the cluster. Not all packages may be available.

   iv. Sanity check the cluster.

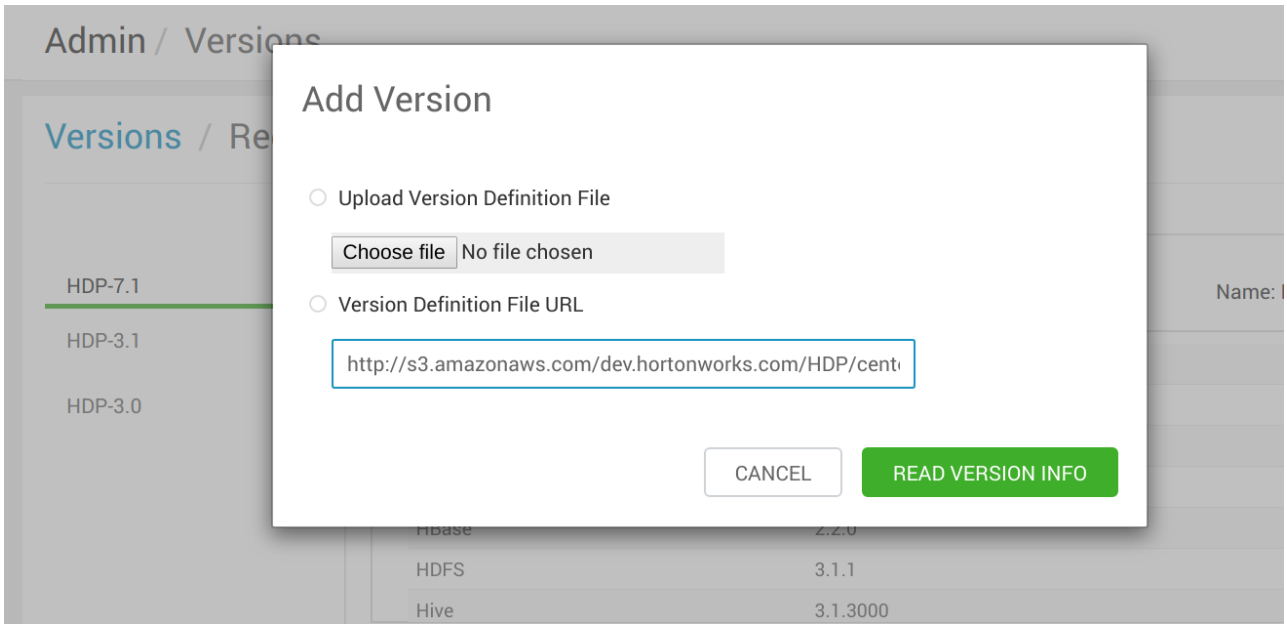
   v. Upgrade.

c. Check if the following VDF URL is accessible.

```
https://cd43d367-6fcf-4a12-a807-
da22f1069c7d:13f2a6a1e854@archive.cloudera.com/p/HDPDC/7.x/7.1.6.0/centos7/HDP-
7.1.6.0-297.xml
```

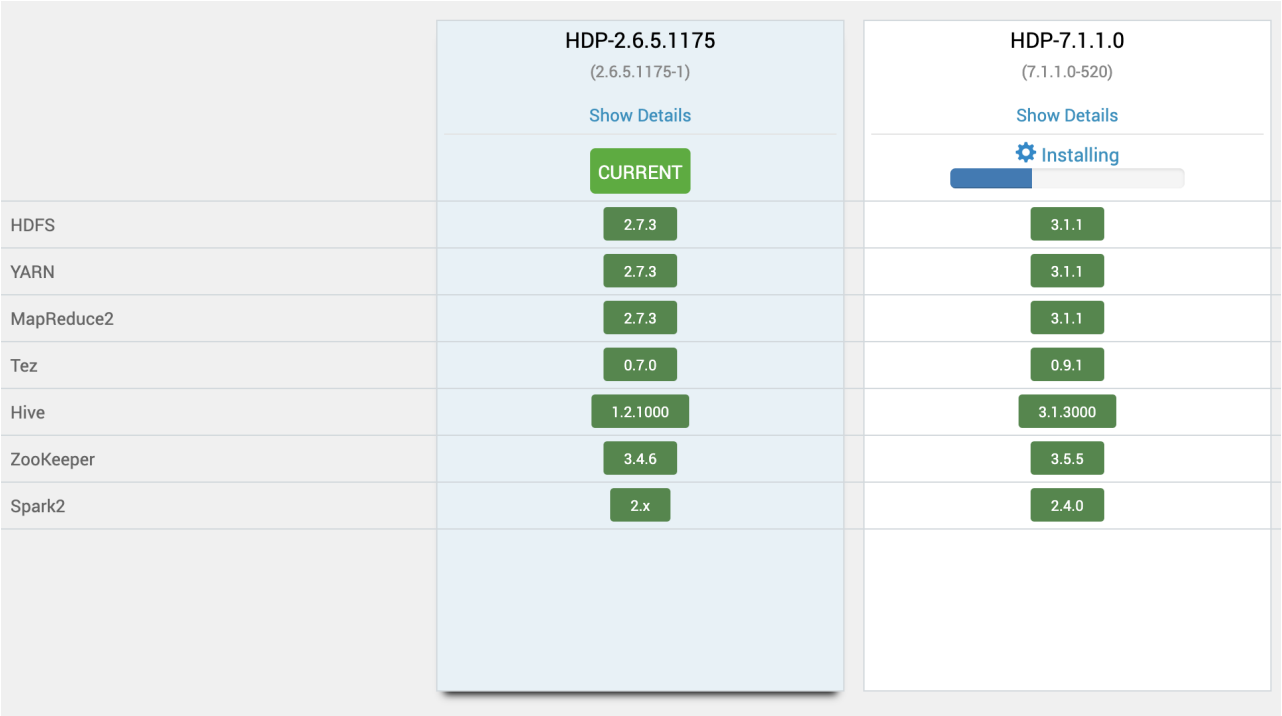
> **TIP** Do not proceed if you cannot reach the URL , ping the instructor.

d. Add the VDF from above via Ambari UI.

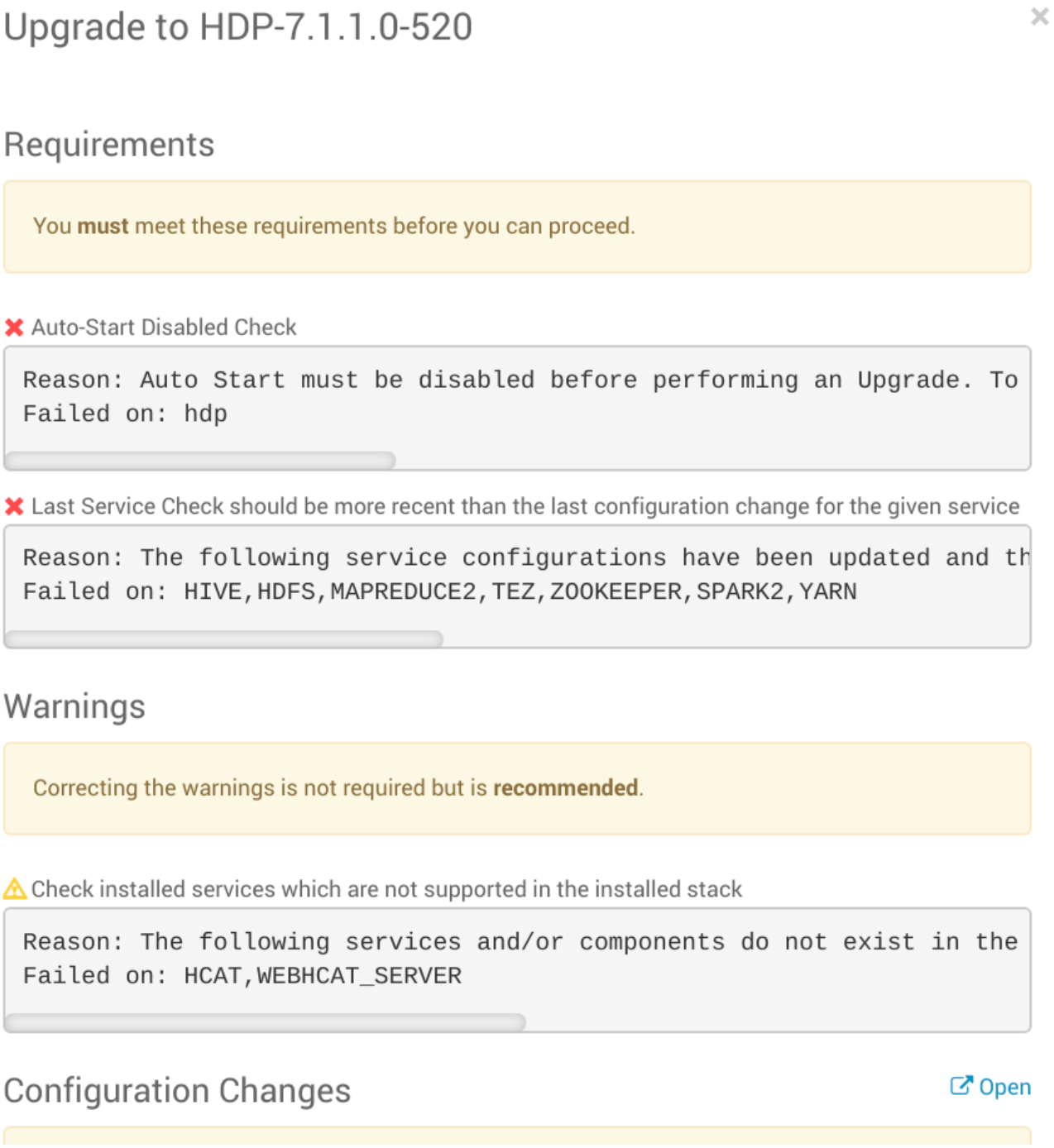


e. Save and register the version HDP-7.1.x.

f. Upgrade the HDP version.



g. Choose *Express Upgrade* only. This should fail with the below. These are service checks and auto-start failures.

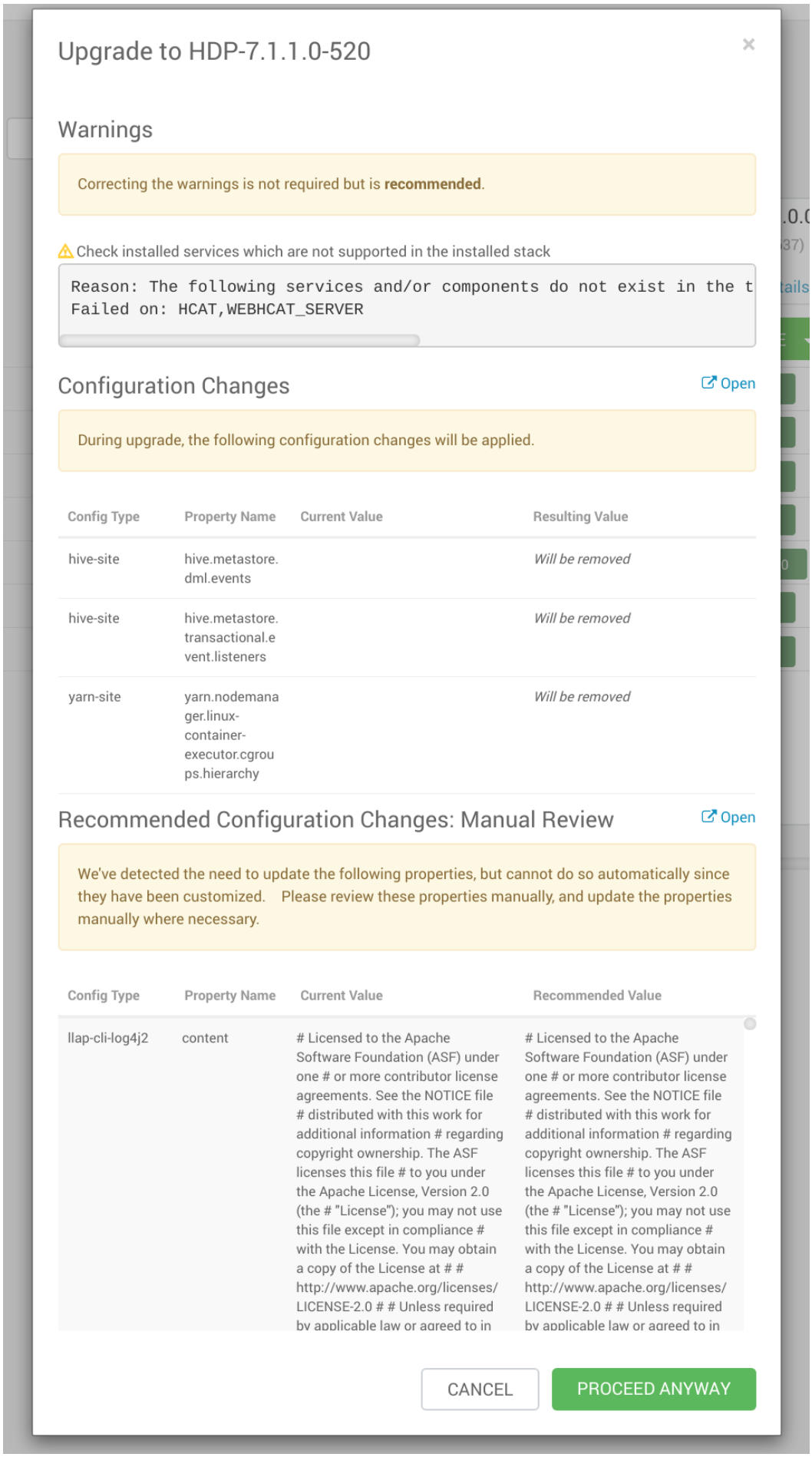


h. For removing service check failures , there are 2 ways:

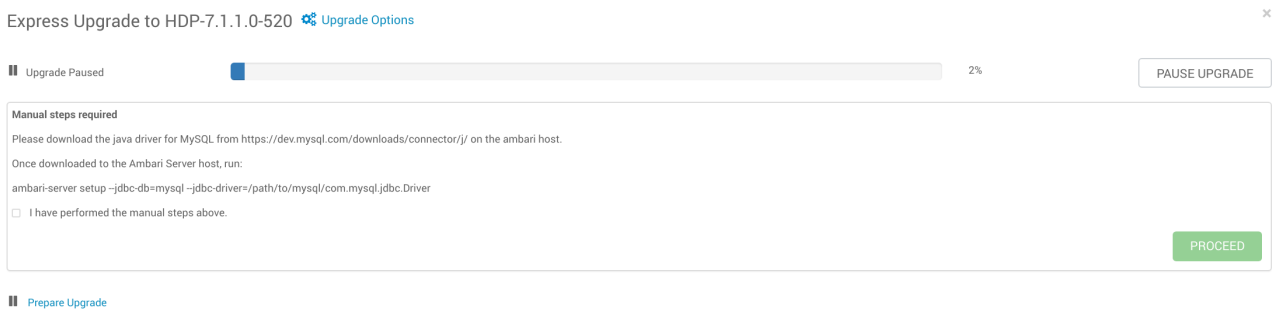
   i. Do it manually via the UI.

   ii. Use this repo.

i. If all service checks pass and auto-start services is shut down you should now be able to proceed with the upgrade.


Upgrade to HDP-7.1.1.0-520
Warnings
Correcting the warnings is not required but is recommended.
Check installed services which are not supported in the installed stack
Reason: The following services and/or components do not exist in the t
Failed on: HCAT,WEBHCAT_SERVER
Configuration Changes
Open
During upgrade, the following configuration changes will be applied.
Config Type
Property Name
Current Value
Resulting Value
hive-site
hive.metastore.dml.events
Will be removed
hive-site
hive.metastore.transactional.event.listeners
Will be removed
yarn-site
yarn.nodemanager.linux-container-executor.cgroups.hierarchy
Will be removed
Recommended Configuration Changes: Manual Review
Open
We've detected the need to update the following properties, but cannot do so automatically since they have been customized. Please review these properties manually, and update the properties manually where necessary.
Config Type
Property Name
Current Value
Recommended Value
llap-cli-log4j2
content
Licensed to the Apache Software Foundation (ASF) under one # or more contributor license agreements. See the NOTICE file # distributed with this work for additional information # regarding copyright ownership. The ASF licenses this file # to you under the Apache License, Version 2.0 (the # "License"); you may not use this file except in compliance # with the License. You may obtain a copy of the License at # # http://www.apache.org/licenses/LICENSE-2.0 # # Unless required by applicable law or agreed to in
Licensed to the Apache Software Foundation (ASF) under one # or more contributor license agreements. See the NOTICE file # distributed with this work for additional information # regarding copyright ownership. The ASF licenses this file # to you under the Apache License, Version 2.0 (the # "License"); you may not use this file except in compliance # with the License. You may obtain a copy of the License at # # http://www.apache.org/licenses/LICENSE-2.0 # # Unless required by applicable law or agreed to in
CANCEL
PROCEED ANYWAY

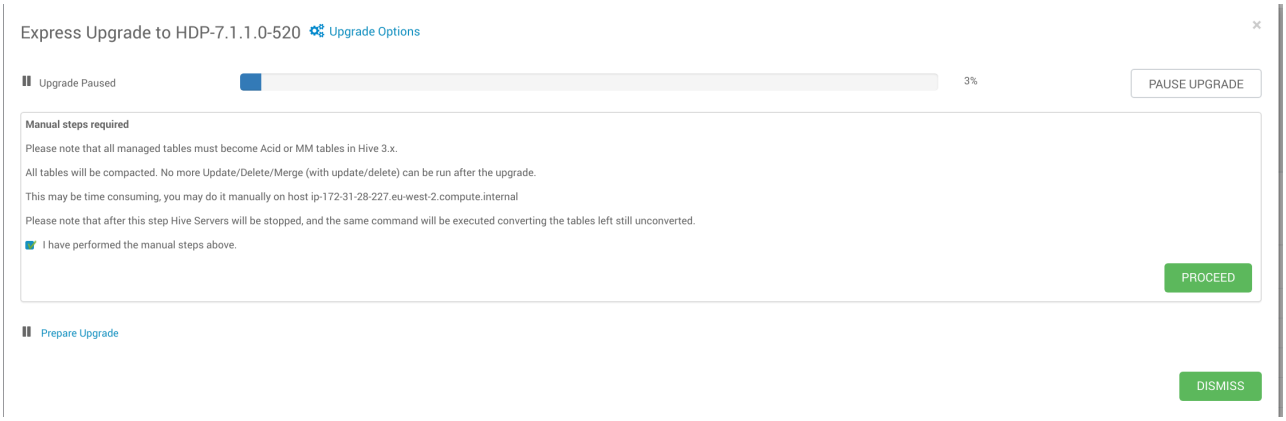
j. Hive metastore upgrade requirement



k. Upgrading MySQL for Hive

```
curl -SL  https://dev.mysql.com/get/Downloads/Connector-J/mysql-connector-java-8.0.20-1.el7.noarch.rpm -o mysql-connector-java-8.0.20-1.el7.noarch.rpm

rpm2cpio mysql-connector-java-8.0.20-1.el7.noarch.rpm | cpio -idmv

cp /usr/share/java/mysql-connector-java.jar /usr/share/java/mysql-connector-java_orig.java

chmod 644 /usr/share/java/mysql-connector-java.jar

ambari-server setup --jdbc-db=mysql --jdbc-driver=/usr/share/java/mysql-connector-java.jar
```

l. Proceed after the above steps are successfully completed. If everything goes well , the below will appear.



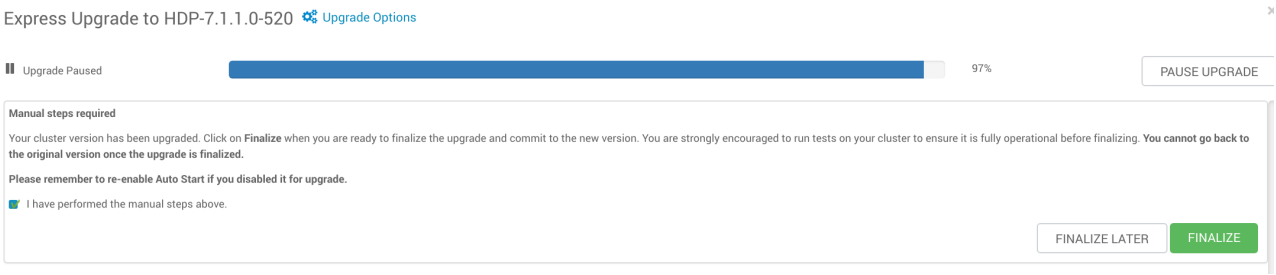
NOTE	As mentioned above, Hive table upgrades are manual and a one-time operation. In a real scenario this will have to be carefully considered before proceeding.

m. The upgrade can take upto 30 minutes , the progress can be tailed on the UI or on the command line.

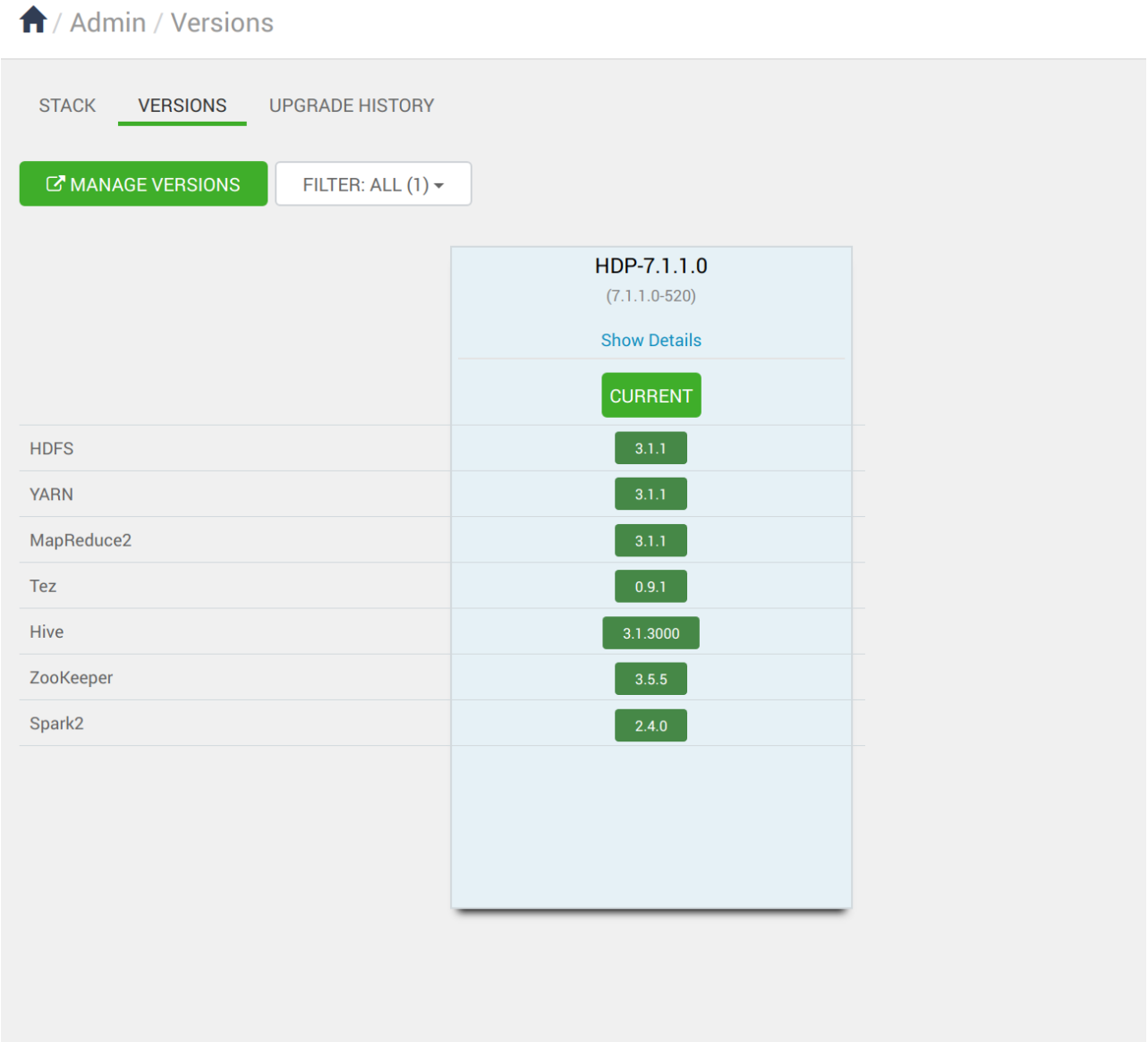
```
tail -100f /var/lib/ambari-agent/data/output-143.txt
```

TIP Get ready for Phase 2 by reading the AM2CMDesignDoc.pdf.

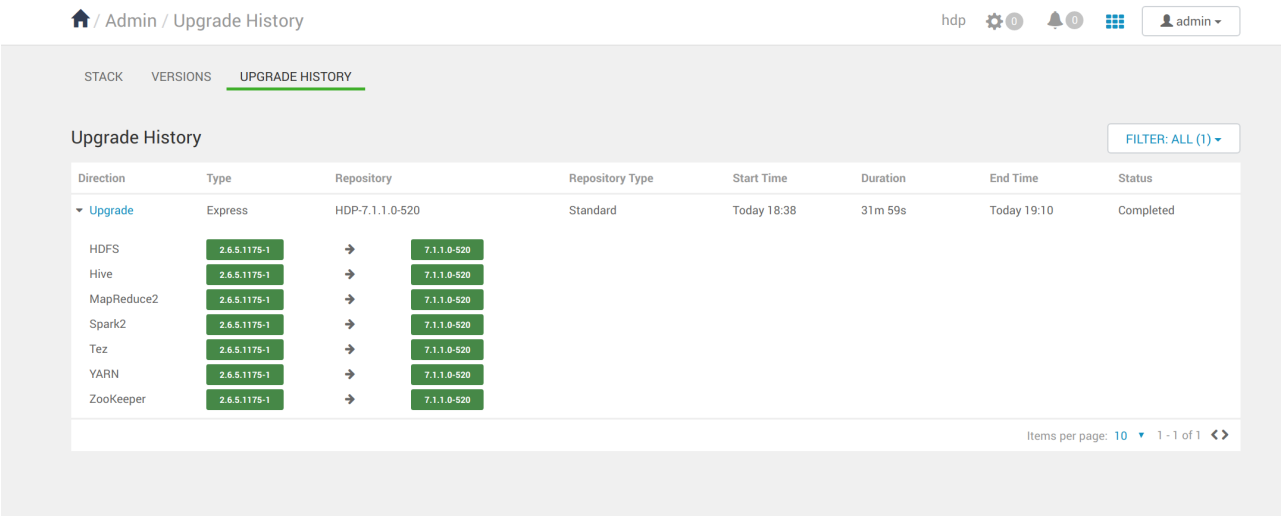
n. If all goes well you should reach a stage where the following manual notification is required.



o. On completion the version can be checked from Ambari UI and you should get the below.



This can also be checked via Stack and Versions → Upgrade History



p. Congrats , you now have a brand new shiny HDP , a CDP-Dc with an Ambari instance.

Phase 2 : Migrating An Ambari Cluster to Cloudera Manager Using AM2CM

AM2CM tool enables you to transition from Ambari Managed cluster to Cloudera Manager cluster. The purpose of this tool is to convert the Ambari blueprint to Cloudera Manager Deployment template. Do *not* proceed if you have not read the Design Doc.

1. Install maven

```
yum clean all
yum-config-manager --disable cloudera-manager
yum install maven
```

> **NOTE** Disabling the cloudera-manager repo only because it points to an internal CLDR URL that is not accessible outside the VPN. There might be other repos that need to be disabled as well.

2. Check if there is `am2cm` directory within `/root`
3. Prepare `am2cm` executable

```
cd am2cm
mvn clean install
```

4. Download the Ambari blueprint

```
curl -u admin:BadPass#1 -X GET <hostname>:8080/api/v1/clusters/hdp?format
=blueprint_with_hosts -o my_cluster_blueprint.json
```

5. Copy the blueprint to `target/am2cm/conf` directory.
6. Convert the ambari blueprint to Cloudera Manager deployment template.

```
cd target/am2cm

sh -x ./am2cm.sh -bp conf/my_cluster_blueprint.json  -dt
my_cm_deployment_template.json
```

```
[root@ip-172-31-28-227 am2cm]# sh -x ./am2cm.sh -bp conf/my_cluster_blueprint.json  -dt my_cm_deployment_template.json
+ export PATH=/bin:/usr/local/sbin:/sbin:/bin:/usr/sbin:/usr/bin:/root/bin
+ PATH=/bin:/usr/local/sbin:/sbin:/bin:/usr/sbin:/usr/bin:/root/bin
+++ dirname ./am2cm.sh
++ cd ./
++ pwd
+ export AM2CM_HOME_DIR=/root/am2cm/target/am2cm
+ AM2CM_HOME_DIR=/root/am2cm/target/am2cm
+ CLASSPATH='/root/am2cm/target/am2cm/conf:/root/am2cm/target/am2cm/lib/*:/root/am2cm/target/am2cm/*:.:'
+ java -Xms2G -cp '/root/am2cm/target/am2cm/conf:/root/am2cm/target/am2cm/lib/*:/root/am2cm/target/am2cm/*:.:' com.cloudera.migration.CMMigration -bp conf/my_cluster_blueprint
.json -dt my_cm_deployment_template.json

INPUT Ambari Blueprint : conf/my_cluster_blueprint.json
OUTPUT CM Template     : my_cm_deployment_template.json

Starting blueprint to CM Template migration
Total number of hosts in blueprint: 1
Processing: LIVY
Processing: HDFS
Processing: TEZ
Processing: OOZIE
Processing: SQOOP_CLIENT
Processing: ZOOKEEPER
Processing: HBASE
Processing: YARN
Processing: HIVE_ON_TEZ
Processing: HIVE
Processing: KAFKA
Processing: SPARK_ON_YARN
CM Template is generated at : /root/am2cm/target/am2cm/my_cm_deployment_template.json
Successfully completed
[root@ip-172-31-28-227 am2cm]#
```

7. Validate the generated cluster template
   a. Goto the following url on your browser

```
http://<CM_HOST>:7180/static/apidocs/ui/index.html#!/ClouderaManagerResource/update
Deployment2
```

8. Paste the contents of the previously generated `my_cm_deployment_template.json` file into the “body” section of the tool, and “try it out”! .

9. If you get the above, you now have a Ambari based cluster that can be taken over by CM. So what have we accomplished:
    a. Migrated a HDP 2.6.5 cluster to HDP 7.x
    b. Converted an HDP Ambari blueprint ready to be used with CDP-Dc.

Congratulations