1. Click on the parcels icon in CM and open configuration. Remove the generic CDH parcels link and update it with this " https://archive.cloudera.com/cm6/6.2.0/redhat7/yum/cloudera-manager.repo"
2. Add the below mentioned services in the cluster using this link as reference: http://www.cloudera.com/documentation/enterprise/latest/topics/cluster\_setup\_wizard.html

Flume

Hbase

HDFS

Hive

Hue

Impala

Oozie

Solr

Spark

Spark2

Sqoop1 client

YARN

Zookeeper

1. Create symlinks for all /var/lib/ and /var/log folders related to the cluster to /apps/lib and /apps/log

mkdir /apps/var/lib

chmod 771 /apps/var/lib

mv /var/lib/zookeeper /apps/var/lib/.

ln -s /apps/var/lib/zookeeper /var/lib/zookeeper

mkdir /apps/log

chmod 771 /apps/log

mv /var/log/zookeeper /apps/log/.

ln -s /apps/log/zookeeper /var/log/zookeeper

-----------------------------------------------------------------------------------------------

mv /var/lib/cloudera-scm-headlamp /apps/var/lib/.

ln -s /apps/var/lib/cloudera-scm-headlamp /var/lib/cloudera-scm-headlamp

--------------------------------------------------------------------------------------------------

on Hive Server 2 Hosts 773 and 796:

cd /var/log/hive

mv /var/log/hive /apps/log/.

ln -s /apps/log/hive /var/log/hive

---------------------------------------------------------------------

mv /var/log/hcatalog /apps/log/.

ln -s /apps/log/hcatalog /var/log/hcatalog

--------------------------------------------------------------------------

on 796 host:

mv /var/log/hue /apps/log/.

ln -s /apps/log/hue /var/log/hue

--------------------------------------------------------------------------

for Impala on 777 Host:

mv /var/log/statestore /apps/log/.

ln -s /apps/log/statestore /var/log/statestore

mv /var/log/catalogd /apps/log/.

ln -s /apps/log/catalogd /var/log/catalogd

------------------------------------------------------------------------------

fro Impala daemon on all datanodes: 778, 793, 792, 776:

mkdir /apps/log

chmod 771 /apps/log

mv /var/log/impalad /apps/log/.

ln -s /apps/log/impalad /var/log/impalad

------------------------------------------------------------------------------------

oozie on 796 Host:

mv /var/log/oozie /apps/log/.

ln -s /apps/log/oozie /var/log/oozie

---------------------------------------------------------------------------------------

SOLR on all datanodes: 778, 793, 792, 776:

mv /var/log/solr /apps/log/.

ln -s /apps/log/solr /var/log/solr

----------------------------------------------------------------------------

for Spark on 796 Host:

mv /var/log/spark /apps/log/.

ln -s /apps/log/spark /var/log/spark

---------------------------------------------------------------------

for YARN on 777 host

mv /var/log/hadoop-mapreduce /apps/log/.

ln -s /apps/log/hadoop-mapreduce /var/log/hadoop-mapreduce

-----------------------------------------------------------------------------------

on all Node Managers on all datanodes: 778, 793, 792, 776:

mv /var/log/hadoop-yarn /apps/log/.

ln -s /apps/log/hadoop-yarn /var/log/hadoop-yarn

mkdir -p /apps/var/lib

chmod 771 /apps/var

chmod 771 /apps/var/lib

mv /var/lib/hadoop-yarn/yarn-nm-recovery /apps/var/lib/.

ln -s /apps/var/lib/hadoop-yarn/yarn-nm-recovery /var/lib/hadoop-yarn/yarn-nm-recovery

----------------------------------------------------------------------------------

on RM 796 host:

mv /var/log/hadoop-yarn /apps/log/.

ln -s /apps/log/hadoop-yarn /var/log/hadoop-yarn

----------------------------------------------------------------------------------

for HDFS hosts 778, 792, 793, 776, 795, 796, 773

mv /var/log/hadoop-hdfs /apps/log/.

ln -s /apps/log/hadoop-hdfs /var/log/hadoop-hdfs

on 795:

mkdir /apps/log

chmod 771 /apps/log

on 796 host:

mv /var/log/hadoop-httpfs /apps/log/.

ln -s /apps/log/hadoop-httpfs /var/log/hadoop-httpfs

-----------------------------------------------------------------------------------

for HBase on 773, 796, 776, 778, 792, 793 host

mv /var/log/hbase /apps/log/.

ln -s /apps/log/hbase /var/log/hbase

1. Change the configuration of all the services in CM to point to /apps/heapdumpdir for the heap dump directory and restart the services

**Heap Dump Directory - /apps/heapdumpdir**

1. Enable HA in HDFS
2. Enable kerberos in the cluster using the CM wizard using this link as reference: https://www.cloudera.com/documentation/enterprise/5-11-x/topics/cm\_sg\_intro\_kerb.html

copy krb5 from DEV to all servers in RND to /etc/krb5.conf and change permissions to

chmod a+r /etc/krb5.conf

1. Add sentry service in the cluster and enable sentry in solr, hive and hdfs services and restart the cluster
2. Enable TLS in all the services of the cluster except Hive and restart the cluster
3. Tune YARN for the right memory and container size
4. Retune the heap size used for all the services in the cluster
5. Add hdfs in yarn whitelist. Remove hdfs from yarn banned list and add parallel bdr configs in HDFS configuration

HDFS Client Advanced Configuration Snippet (Safety Valve) for hdfs-site.xml (Add this both in the source and destination cluster)

distcp.enable.listing.parallel

true

For parallel BDR

distcp.fetch.threads

80

For parallel BDR

dfs.client.block.write.locateFollowingBlock.retries

10

for fixing the issue -does not have enough number of replicas.

1. Add informatica impersonation configs
2. "Temp hue access : Hive / Hive

[desktop]

http\_500\_debug\_mode=1

django\_debug\_mode=1

secure\_hsts\_seconds=0

[[auth]]

backend=desktop.auth.backend.LdapBackend,desktop.auth.backend.AllowFirstUserDjangoBackend

[[ldap]]

ignore\_username\_case=true

force\_username\_lowercase=true

[hbase]

hbase\_conf\_dir={{HBASE\_CONF\_DIR}}

[impala]

close\_queries=true

server\_host=itsusralsp07811.jnj.com

query\_timeout\_s=600

session\_timeout\_s=1800

[beeswax]

close\_queries=true

auth\_username=saedlldd@JNJ.COM

auth\_password=LdapOnHive100

[hadoop]

[[yarn\_clusters]]

[[[default]]]

spark\_history\_server\_url=https://itsusraedld01.jnj.com:18488"

1. Test Hive on LDAP connection string. Add the truststore file in HiveServer2 TLS/SSL Certificate Trust Store File to make sure hive with ldap works
2. Make sure policy file sentry is enabled in oozie.
3. keytool -import -alias tomcat -file /opt/cloudera/security/CAcerts/combinedtrust.pem -keystore /usr/java/latest/jre/lib/security/cacerts
4. Check if postfix is running and use mailx command to test mails from all the servers

keytool -importcert -keystore /apps/java/jdk1.8.0\_74/jre/lib/security/cacerts -alias ldapcaa3 -storepass changeit -keypass Cloudera@5 -trustcacerts -file INT-PROD-CA-A3.cer

keytool -importcert -keystore /apps/java/jdk1.8.0\_74/jre/lib/security/cacerts -alias ldapcab3 -storepass changeit -keypass Cloudera@5 -trustcacerts -file INT\_PROD-CA-B3.cer

keytool -importcert -keystore /apps/java/jdk1.8.0\_74/jre/lib/security/cacerts -alias ldapcac3 -storepass changeit -keypass Cloudera@5 -trustcacerts -file INT-PROD-CA-C3.cer

keytool -importcert -keystore /apps/java/jdk1.8.0\_74/jre/lib/security/cacerts -alias ldaproot2 -storepass changeit -keypass Cloudera@5 -trustcacerts -file INT-PROD-Root-2.cer

keytool -importcert -keystore /apps/java/jdk1.8.0\_74/jre/lib/security/cacerts -alias ldapt01 -storepass changeit -keypass Cloudera@5 -trustcacerts -file jnjinternalrootca.crt

keytool -importcert -keystore /apps/java/jdk1.8.0\_74/jre/lib/security/cacerts -alias ldapt02 -storepass changeit -keypass Cloudera@5 -trustcacerts -file jnjinternalonlinecaa2.crt

keytool -importcert -keystore /apps/java/jdk1.8.0\_74/jre/lib/security/cacerts -alias ldap -storepass changeit -keypass Cloudera@5 -trustcacerts -file ldapt01.cer

For CM LDAP to work, the LDAP certificate needs to be imported into cacerts - java default cacerts. Refer the certs in itsusralsp07935

1. Add root and issuing to cacerts - default file

hdfs dfs -chown -R yarn:supergroup /tmp/hadoop-yarn

1. Oozie Server Advanced Configuration Snippet (Safety Valve) for oozie-site.xml - sync up with the old cluster

LDAP in navigator

Create INFA\_Pushdown\_Hive

INFA\_Pushdown\_Spark folder in all datanodes

#-> hdfs dfs -setfacl -R -m default:group:its-app-edl-dev-proj12-etl-usr:rwx /app/INFA\_Pushdown\_Hive

[root@ITSUSRALSP07814:/home/sbiswa25]#

#-> hdfs dfs -setfacl -R -m group:its-app-edl-dev-proj12-etl-usr:rwx /app/INFA\_Pushdown\_Hive "

"Impala Daemon Command Line Argument Advanced Configuration Snippet (Safety Valve)

-force\_lowercase\_usernames=1

-fe\_service\_threads=128

-idle\_session\_timeout=1800

-convert\_legacy\_hive\_parquet\_utc\_timestamps=true

-idle\_query\_timeout=1800"

Install anaconda parcels

Install teradata connector parcels

1. Point hive, oozie to impala load balancer
2. YARN: Enable Fair Scheduler Continuous Scheduling - set it to false
3. YARN: Fair Scheduler Assign Multiple Tasks - set it to true - enable check box
4. Correct: "Oozie Server Advanced Configuration Snippet (Safety Valve) for oozie-site.xml" to "log4j.rootLogger=INFO,console log4j.appender.console=org.apache.log4j.ConsoleAppender log4j.appender.console.target=System.err log4j.appender.console.layout=org.apache.log4j.PatternLayout log4j.appender.console.layout.ConversionPattern=%d{yy/MM/dd HH:mm:ss} %p %c{2}: %m%n"
5. Set "Auto Logout Timeout" for Hue to 20 minutes

Proof: A cloudera manager web ui snapshot will be taken with the timestamp showing all the services which have been deployed.