Kerberos Integration with Hadoop

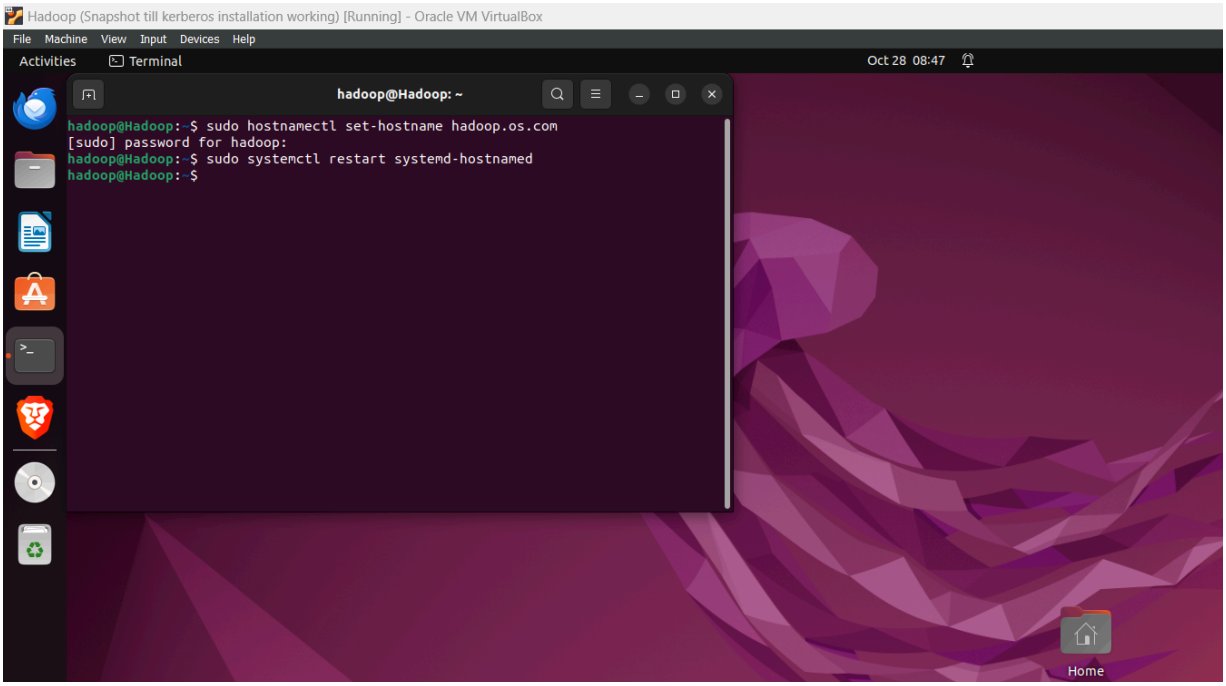
Change the hostname if not already changed

Run the command

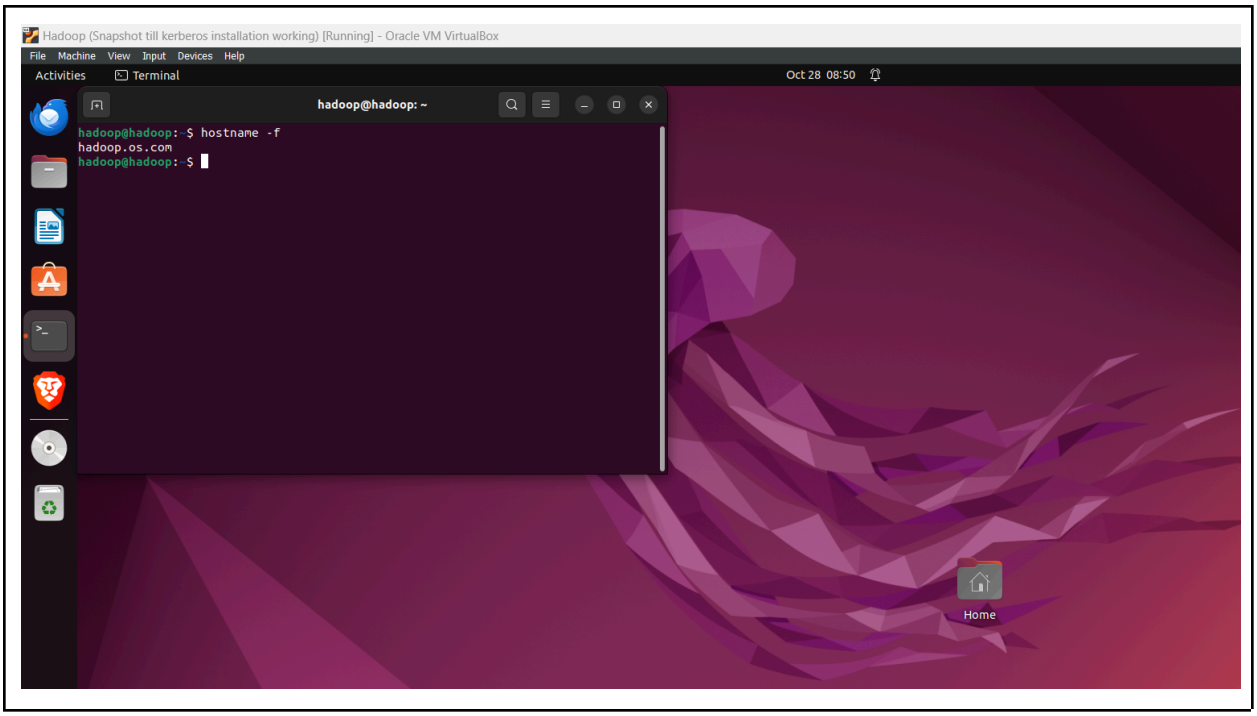
```
> sudo hostnamectl set-hostname hadoop.os.com # Here we are setting the hostname as os.com

> sudo systemctl restart systemd-hostnamed

> sudo reboot
```



Once the system is rebooted, check the hostname



Kerberos installation

/etc/krb5.conf

```
[libdefaults]
        default_realm = OS.COM

# The following krb5.conf variables are only for MIT Kerberos.
        kdc_timesync = 1
        ccache_type = 4
        forwardable = true
        proxiable = true

# The following encryption type specification will be used by MIT Kerberos
# if uncommented.  In general, the defaults in the MIT Kerberos code are
# correct and overriding these specifications only serves to disable new
# encryption types as they are added, creating interoperability problems.
#
# The only time when you might need to uncomment these lines and change
# the enctypes is if you have local software that will break on ticket
# caches containing ticket encryption types it doesn't know about (such as
# old versions of Sun Java).

#       default_tgs_enctypes = des3-hmac-sha1
#       default_tkt_enctypes = des3-hmac-sha1
#       permitted_enctypes = des3-hmac-sha1
```

```
# The following libdefaults parameters are only for Heimdal Kerberos.
        fcc-mit-ticketflags = true

[realms]
        OS.COM = {
                kdc = localhost
                admin_server = localhost
        }

[domain_realm]
       .os.com = OS.COM
        os.com=OS.COM
```

```
-rw-r--r--  1 root root    1072 Oct 28 08:58 krb5.conf
hadoop@hadoop:~$ ls -lrt /etc/ |grep krb
-rw-r--r--  1 root root    2883 Oct 10 22:10 krb5.conf_bkp
-rw-r--r--  1 root root    1113 Oct 10 22:15 krb5.conf_bkp_1
drwx------  2 root root    4096 Oct 10 22:21 krb5kdc
-rw-r--r--  1 root root    1072 Oct 28 08:58 krb5.conf
hadoop@hadoop:~$ cat /etc/krb5.conf
[libdefaults]
        default_realm = OS.COM

# The following krb5.conf variables are only for MIT Kerberos.
        kdc_timesync = 1
        ccache_type = 4
        forwardable = true
        proxiable = true

# The following encryption type specification will be used by MIT Kerberos
# if uncommented.  In general, the defaults in the MIT Kerberos code are
# correct and overriding these specifications only serves to disable new
# encryption types as they are added, creating interoperability problems.
#
# The only time when you might need to uncomment these lines and change
# the enctypes is if you have local software that will break on ticket
# caches containing ticket encryption types it doesn't know about (such as
# old versions of Sun Java).

#       default_tgs_enctypes = des3-hmac-sha1
#       default_tkt_enctypes = des3-hmac-sha1
#       permitted_enctypes = des3-hmac-sha1

# The following libdefaults parameters are only for Heimdal Kerberos.
        fcc-mit-ticketflags = true

[realms]
        OS.COM = {
                kdc = localhost
                admin_server = localhost
        }

[domain_realm]
        .os.com = OS.COM
         os.com=OS.COM
hadoop@hadoop:~$
```

Installing JSVC

> echo $JAVA_HOME (Make sure this command is pointing to right JDK)

> wget https://dlcdn.apache.org//commons/daemon/source/commons-daemon-1.4.0-src.tar.gz

> tar -zxvf commons-daemon-1.4.0-src.tar.gz

> cd commons-daemon-1.4.0-src/src/native/unix

> ./configure

> make

>sudo cp jsvc /usr/bin

```
hadoop@hadoop:~$ echo $JAVA_HOMe

hadoop@hadoop:~$ echo $JAVA_HOME
/usr/lib/jvm/java-11-openjdk-amd64
hadoop@hadoop:~$ which java
/usr/bin/java
```

```
hadoop@hadoop:~$ ^C
hadoop@hadoop:~$  wget https://dlcdn.apache.org//commons/daemon/source/commons-daemon-1.4.0-src.tar.gz
--2024-10-28 09:02:03--  https://dlcdn.apache.org//commons/daemon/source/commons-daemon-1.4.0-src.tar.gz
Resolving dlcdn.apache.org (dlcdn.apache.org)... 151.101.2.132, 2a04:4e42::644
Connecting to dlcdn.apache.org (dlcdn.apache.org)|151.101.2.132|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 291610 (285K) [application/x-gzip]
Saving to: 'commons-daemon-1.4.0-src.tar.gz'

commons-daemon-1.4.0-src.tar.gz                   100%[==========================================================================================================>] 284.78K  --.-KB/s    in 0.1s

2024-10-28 09:02:03 (2.71 MB/s) - 'commons-daemon-1.4.0-src.tar.gz' saved [291610/291610]
```

```
hadoop@hadoop:~$ tar -zxvf commons-daemon-1.4.0-src.tar.gz
commons-daemon-1.4.0-src/CONTRIBUTING.md
commons-daemon-1.4.0-src/HOWTO-RELEASE.txt
commons-daemon-1.4.0-src/LICENSE.txt
commons-daemon-1.4.0-src/NOTICE.txt
commons-daemon-1.4.0-src/PROPOSAL.html
commons-daemon-1.4.0-src/README.md
commons-daemon-1.4.0-src/RELEASE-NOTES.txt
commons-daemon-1.4.0-src/pom.xml
commons-daemon-1.4.0-src/src/
commons-daemon-1.4.0-src/src/assembly/
commons-daemon-1.4.0-src/src/changes/
commons-daemon-1.4.0-src/src/docs/
commons-daemon-1.4.0-src/src/main/
commons-daemon-1.4.0-src/src/main/java/
commons-daemon-1.4.0-src/src/main/java/org/
commons-daemon-1.4.0-src/src/main/java/org/apache/
commons-daemon-1.4.0-src/src/main/java/org/apache/commons/
commons-daemon-1.4.0-src/src/main/java/org/apache/commons/daemon/
```

```
hadoop@hadoop:~$
hadoop@hadoop:~$
hadoop@hadoop:~$ cd commons-daemon-1.4.0-src/src/native/unix
hadoop@hadoop:~/commons-daemon-1.4.0-src/src/native/unix$ ./configure
*** Current host ***
checking build system type... x86_64-pc-linux-gnu
checking host system type... x86_64-pc-linux-gnu
checking cached host system type... ok
*** C-Language compilation tools ***
checking for gcc... gcc
checking whether the C compiler works... yes
checking for C compiler default output file name... a.out
checking for suffix of executables...
checking whether we are cross compiling... no
checking for suffix of object files... o
checking whether the compiler supports GNU C... yes
checking whether gcc accepts -g... yes
checking for gcc option to enable C11 features... none needed
checking for ranlib... ranlib
```

```
*** All done ***
Now you can issue "make"
hadoop@hadoop:~/commons-daemon-1.4.0-src/src/native/unix$ make
(cd native; make  all)
make[1]: Entering directory '/home/hadoop/commons-daemon-1.4.0-src/src/native/unix/native'
gcc -g -O2 -DOS_LINUX -DDSO_DLFCN -DCPU=\"amd64\" -Wall -Wstrict-prototypes   -I/usr/lib/jvm/java-11-openjdk-amd64/include -I/usr/lib/jvm/java-11-openjdk
jsvc-unix.c: In function 'get_pidf':
jsvc-unix.c:666:5: warning: ignoring return value of 'lockf' declared with attribute 'warn_unused_result' [-Wunused-result]
  666 |     lockf(fd, F_LOCK, 0);
      |     ^~~~~~~~~~~~~~~~~~~~
jsvc-unix.c:668:5: warning: ignoring return value of 'lockf' declared with attribute 'warn_unused_result' [-Wunused-result]
  668 |     lockf(fd, F_ULOCK, 0);
      |     ^~~~~~~~~~~~~~~~~~~~
jsvc-unix.c: In function 'wait_child':
jsvc-unix.c:762:9: warning: ignoring return value of 'lockf' declared with attribute 'warn_unused_result' [-Wunused-result]
```

```
gcc -g -O2 -DOS_LINUX -DDSO_DLFCN -DCPU=\"amd64\" -Wall -Wstrict-prototypes   -I/usr/lib/jvm/java-11-openjdk-amd64/include -I/usr/lib/jvm/
ar cr libservice.a arguments.o debug.o dso-dlfcn.o dso-dyld.o help.o home.o java.o location.o replace.o locks.o signals.o
ranlib libservice.a
gcc   jsvc-unix.o libservice.a -ldl -lpthread -o ../jsvc
make[1]: Leaving directory '/home/hadoop/commons-daemon-1.4.0-src/src/native/unix/native'
hadoop@hadoop:~/commons-daemon-1.4.0-src/src/native/unix$ sudo cp jsvc /usr/bin
hadoop@hadoop:~/commons-daemon-1.4.0-src/src/native/unix$
```

Edit hadoop-env.sh

```
export JSVC_HOME=/usr/bin
export HADOOP_OPTS="-Djava.security.krb5.conf=/etc/krb5.conf"
export HDFS_DATANODE_SECURE_USER=root
```

```
hadoop@hadoop:~/hadoop-3.3.6/etc/hadoop$ pwd
/home/hadoop/hadoop-3.3.6/etc/hadoop
hadoop@hadoop:~/hadoop-3.3.6/etc/hadoop$ ls -l
total 180
-rw-r--r-- 1 hadoop hadoop  9213 Jun 18  2023 capacity-scheduler.xml
-rw-r--r-- 1 hadoop hadoop  1335 Jun 18  2023 configuration.xsl
-rw-r--r-- 1 hadoop hadoop  2567 Jun 18  2023 container-executor.cfg
-rw-r--r-- 1 hadoop hadoop  1158 Oct 17 18:44 core-site.xml
-rw-r--r-- 1 hadoop hadoop  3999 Jun 18  2023 hadoop-env.cmd
-rw-r--r-- 1 hadoop hadoop 16693 Jan 21  2024 hadoop-env.sh
-rw-r--r-- 1 hadoop hadoop  3321 Jun 18  2023 hadoop-metrics2.properties
-rw-r--r-- 1 hadoop hadoop 11765 Jun 18  2023 hadoop-policy.xml
-rw-r--r-- 1 hadoop hadoop  3414 Jun 18  2023 hadoop-user-functions.sh.example
-rw-r--r-- 1 hadoop hadoop   683 Jun 18  2023 hdfs-rbf-site.xml
-rw-r--r-- 1 hadoop hadoop  2050 Oct 17 18:45 hdfs-site.xml
-rw-r--r-- 1 hadoop hadoop  1484 Jun 18  2023 httpfs-env.sh
-rw-r--r-- 1 hadoop hadoop  1657 Jun 18  2023 httpfs-log4j.properties
-rw-r--r-- 1 hadoop hadoop   620 Jun 18  2023 httpfs-site.xml
-rw-r--r-- 1 hadoop hadoop  3518 Jun 18  2023 kms-acls.xml
-rw-r--r-- 1 hadoop hadoop  1351 Jun 18  2023 kms-env.sh
-rw-r--r-- 1 hadoop hadoop  1860 Jun 18  2023 kms-log4j.properties
-rw-r--r-- 1 hadoop hadoop   682 Jun 18  2023 kms-site.xml
-rw-r--r-- 1 hadoop hadoop 13700 Jun 18  2023 log4j.properties
-rw-r--r-- 1 hadoop hadoop   951 Jun 18  2023 mapred-env.cmd
-rw-r--r-- 1 hadoop hadoop  1764 Jun 18  2023 mapred-env.sh
-rw-r--r-- 1 hadoop hadoop  4113 Jun 18  2023 mapred-queues.xml.template
-rw-r--r-- 1 hadoop hadoop   844 Jan 21  2024 mapred-site.xml
drwxr-xr-x 2 hadoop hadoop  4096 Jun 18  2023 shellprofile.d
-rw-r--r-- 1 hadoop hadoop  2316 Jun 18  2023 ssl-client.xml.example
-rw-r--r-- 1 hadoop hadoop  2697 Jun 18  2023 ssl-server.xml.example
-rw-r--r-- 1 hadoop hadoop  2681 Jun 18  2023 user_ec_policies.xml.template
-rw-r--r-- 1 hadoop hadoop    10 Jun 18  2023 workers
-rw-r--r-- 1 hadoop hadoop  2250 Jun 18  2023 yarn-env.cmd
-rw-r--r-- 1 hadoop hadoop  6329 Jun 18  2023 yarn-env.sh
-rw-r--r-- 1 hadoop hadoop  2591 Jun 18  2023 yarnservice-log4j.properties
-rw-r--r-- 1 hadoop hadoop  1340 Jan 21  2024 yarn-site.xml
hadoop@hadoop:~/hadoop-3.3.6/etc/hadoop$
```

```
# Supplemental options for privileged registry DNS
# By default, Hadoop uses jsvc which needs to know to launch a
# server jvm.
# export HADOOP_REGISTRYDNS_SECURE_EXTRA_OPTS="-jvm server"
export JAVA_HOME=/usr/lib/jvm/java-1.11.0-openjdk-amd64
export JSVC_HOME=/usr/bin
export HADOOP_OPTS="-Djava.security.krb5.conf=/etc/krb5.conf"
export HDFS_DATANODE_SECURE_USER=root

hadoop@hadoop:~/hadoop-3.3.6/etc/hadoop$ cat hadoop-env.sh
```

Reference:
https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/SecureMode.html

Add the following properties in core-site.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed 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 writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
        <property>
                <name>hadoop.tmp.dir</name>
                <value>/home/hadoop/tmpdata</value>
```

```
        </property>
        <property>
                <name>fs.default.name</name>
                <value>hdfs://127.0.0.1:9000</value>
        </property>
        <property>
                <name>hadoop.security.authentication</name>
                <value>kerberos</value>
        </property>
        <property>
                <name>hadoop.security.authorization</name>
                <value>true</value>
        </property>
</configuration>
```

Add the following properties in hdfs-site.xml (Please check the hostname and realm)

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed 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 writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
        <property>
                <name>dfs.namenode.name.dir</name>
                <value>/home/hadoop/dfsdata/namenode</value>
        </property>
        <property>
                <name>dfs.datanode.data.dir</name>
                <value>/home/hadoop/dfsdata/datanode</value>
        </property>
        <property>
                <name>dfs.replication</name>
                <value>1</value>
        </property>
        <property>
                <name>dfs.namenode.kerberos.principal</name>
```

```
                <value>hdfs/hadoop.os.com@OS.COM</value>
        </property>
        <property>
                <name>dfs.namenode.keytab.file</name>
                <value>/etc/nn.keytab</value>
        </property>
        <property>
                <name>dfs.datanode.kerberos.principal</name>
                <value>hdfs/hadoop.os.com@OS.COM</value>
        </property>
<property>
  <name>dfs.datanode.keytab.file</name>
   <value>/etc/nn.keytab</value>
</property>
<property>
  <name>dfs.block.access.token.enable</name>
  <value>true</value>
</property>
 <property>
  <name>dfs.block.access.token.enable</name>
  <value>true</value>
</property>
 <property>
   <name>dfs.datanode.address</name>
   <value>0.0.0.0:1004</value>
```

```
  </property>

  <property>

    <name>dfs.datanode.http.address</name>

    <value>0.0.0.0:1006</value>

  </property>

<property>

    <name>dfs.permissions.enabled</name>

    <value>false</value>

  </property>


</configuration>
```

Create new realm with the new hostname

```
sudo krb5_newrealm
```

NOTE: if any error comes then delete the files present in /var/lib/krb5kdc folder

```
sudo rm -rf /var/lib/krb5kdc/*
```

Now again add the newrealm

Generating the keytab

```
sudo kadmin.local -q "addprinc -randkey hdfs/hadoop.os.com@OS.COM"
sudo kadmin.local -q "ktadd -k /etc/nn.keytab hdfs/hadoop.os.com@OS.COM"
```

→ Change the permission of hadoop user directory to 755

Run the below command to become root

```
sudo su -
```

Run name node and datanode from root user:

Now run the below command to start the namenode and datanode

```
hadoop namenode
```

Open new terminal and run the command to start datanode

```
hadoop datanode
```

Similarly SNN, RM, NM needs to be kerberized

TBD: Steps needs to be added