Lab 21.1: Setting up Knox gateway

|  |  |
| --- | --- |
| **Objective:** | Setting up “Knox gateway” on a new node (hadoop-master) and access Hadoop cluster (node1-node4) through Knox gateway. |
| **Successful Outcome:** | Once you run a mapReduce job without accessing Hadoop cluster using Knox gateway. |
| **Before You Begin:** | SSH into hadoop-master. |

1. Install Knox
   1. Login to **hadoop-master**
   2. Execute following command to install Knox gateway:

# yum –y install knox

1. Configure Knox
   1. Create a copy of the default topology file ‘/etc/knox/conf/topologies/sandbox.xml’:

# cd /etc/knox/conf/topologies

# cp sandbox.xml cluster-horton.xml

* 1. Look for ‘sandbox’ keyword and replace ‘**sandbox,sandbox.hortonworks.com’** with ‘**hadoop-master**’:

<provider>

<role>hostmap</role>

<name>static</name>

<enabled>true</enabled>

<param><name>localhost</name><value>**hadoop-master**</value></param>

</provider>

* 1. Look for ‘NAMENODE’, ’JOBTRACKER’, ‘WEBHDFS’, WEBHCAT’ & ‘OOZIE’ role and change their hostname from ‘localhost’ to following:

NAMENODE: localhost -> node1

JOBTRACKER: localhost -> node2

WEBHDFS: localhost -> node1

WEBHCAT: localhost -> node2

OOZIE: localhost -> node3

* 1. Save and close the file.
  2. Open /etc/knox/conf/users.ldif’ file and replace ‘guest’ keyword with ‘root’. Also change ‘userPAssword’ to ‘hadoop’:

# Entry for a sample end user

# Please replace with site specific values

dn: uid=**root**,ou=people,dc=hadoop,dc=apache,dc=org

objectclass:top

objectclass:person

objectclass:organizationalPerson

objectclass:inetOrgPerson

cn: Guest

sn: User

uid: **root**

userPassword:**hadoop**

* 1. Now run following setup command. When prompted for “master secret”. Leave it blank and ‘enter’:

# su -l knox -c "/usr/lib/knox/bin/gateway.sh setup"

* 1. Start knox gateway and ldap services:

su -l knox -c "/usr/lib/knox/bin/gateway.sh start"

su -l knox -c "/usr/lib/knox/bin/ldap.sh start"

* 1. Now check whether you can access HDFS from ‘hadoop-master’ by executing following:

# curl -iku root:hadoop -X GET 'https://node1:8443/gateway/cluster-horton/webhdfs/v1/user?op=LISTSTATUS'

It provides listing of /user directory in HDFS:



Now you can execute various Hadoop commands and MapReduce jobs using regular WebHDFS commands on Knox gateway without exposing your hadoop cluster.