Module 4–Backup, Recovery and Maintenance

Hands On Guide

edureka!

edureka!

Version 2.1

Module 4 – Backup, Recovery and Maintenance

Securing Your Hadoop Cluster

Table of Contents

1. Kerberos

a. core-site.xml

```
<property>
        <name>hadoop.security.authentication</name>
        <value>kerberos</value>
        <!-- Giving value as "simple" disables security.-->
```

```
        </property>

        <property>
                <name>hadoop.security.authorization</name>
                <value>true</value>
        </property>
```

b. hdfs-site.xml

```
        <property>
                <name>dfs.block.access.token.enable</name>
                <value>true</value>
        </property>

        <!-- NameNode security config -->

        <property>
                <name>dfs.https.address</name>
                <value><fully qualified domain name of NN>:50470</value>
        </property>

        <property>
                <name>dfs.https.port</name>
                <value>50470</value>
        </property>

        <property>
                <name>dfs.namenode.keytab.file</name>
                <value>/etc/hadoop/conf/hdfs.keytab</value> <!-- path to the HDFS keytab -->
        </property>

        <property>
                <name>dfs.namenode.kerberos.principal</name>
                <value>admin/_HOST@queryiorealm</value>
        </property>

        <property>
                <name>dfs.namenode.kerberos.https.principal</name>
                <value>admin/_HOST@queryiorealm</value>
        </property>

        <!-- Secondary NameNode security config -->

        <property>
                <name>dfs.secondary.https.address</name>
                <value><fully qualified domain name of Standby NN>:50495</value>
        </property>

        <property>
                <name>dfs.secondary.https.port</name>
```

```xml
            <value>50495</value>
    </property>

    <property>
            <name>dfs.secondary.namenode.keytab.file</name>
            <value>/etc/hadoop/conf/hdfs.keytab</value> <!-- path to the HDFS keytab -->
    </property>

     <property>
            <name>dfs.secondary.namenode.kerberos.principal</name>
            <value>admin/_HOST@queryiorealm</value>
    </property>

    <property>
            <name>dfs.secondary.namenode.kerberos.https.principal</name>
            <value>admin/_HOST@queryiorealm</value>
    </property>

    <!-- DataNode security config -->

    <property>
            <name>dfs.datanode.data.dir.perm</name>
            <value>700</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.datanode.keytab.file</name>
            <value>/etc/hadoop/conf/hdfs.keytab</value> <!-- path to the HDFS keytab -->
    </property>

    <property>
            <name>dfs.datanode.kerberos.principal</name>
            <value>admin/_HOST@queryiorealm</value>
    </property>

    <property>
            <name>dfs.datanode.kerberos.https.principal</name>
            <value>admin/_HOST@queryiorealm</value>
    </property>
    <property>
      <name>dfs.web.authentication.kerberos.principal</name>
```

```
<value>admin/_HOST@queryiorealm</value>
</property>
```

a. Update for Jobtracker

Add Hosts:

1. dfs.hosts in the hdfs-site.xml, mapred.hosts

b. Cluster Balancing

hadoop balancer -threshold 40

c. Add Disk Space to a datanode

```
How do you add storage to cluster
<property>
        <name>dfs.hosts</name>
        <value>/home/hadoop/include</value>
        <final>true</final>
</property>
```