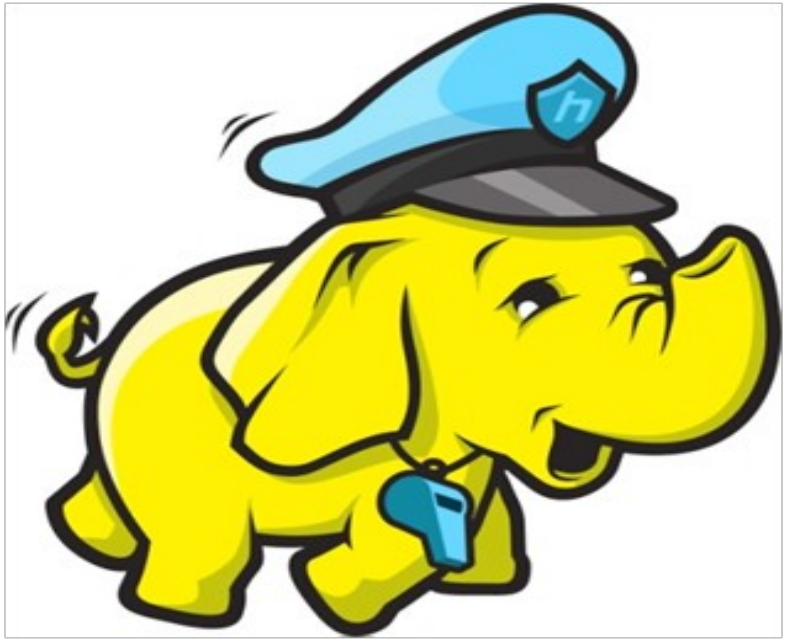
Commissioning
Decommissioning
ACL (Access Control List)



For online Hadoop training, send mail to neeraj.ymca.2k6@gmail.com

Agenda

What is Commissioning
Role of include & exclude file
Entry for include & exclude file
How to commission a new node
What is Decommissioning
How to decommission an old node
What is ACL
Why do we need ACL
Modify core-site.xml
Role of hadoop-policy.xml
Refresh service ACL

What is Commissioning

Hadoop is scalable.

We can increase/decrease number of nodes in a Hadoop cluster.

Adding a new machine (node) to Hadoop cluster is known as commissioning.

Commissioning a new node will increase the storage & processing capacity of Hadoop cluster.

Role of include & exclude file

Create 2 files in HADOOP_HOME/conf directory

include
exclude

include file contains the list of nodes which are allowed to connect to master machine.

exclude file contains the list of nodes which are supposed to be decommissioned.

If the name is available in both files, it can connect to master machine but can't participate in storage/processing.

Entry for include file and exclude file

Add the following entry to **hdfs-site.xml** file in HADOOP_HOME/ conf directory

```
<property>

    <name>dfs.hosts</name>
    <value>/home/neeraj/local_cluster_home/hadoop-1.2.1/conf/include</value>
    <description>Names a file that contains a list of hosts which are permitted to connect to
                the Namenode. The full pathname of the file must be specified. If
                 the value is empty, all hosts are permitted.
    </description>

</property>

<property>

    <name>dfs.hosts.exclude</name>
    <value>/home/neeraj/local_cluster_home/hadoop-1.2.1/conf/exclude</value>
    <description>Names a file that contains a list of hosts which are not permitted to
                connect to the Namenode. The full pathname of the file must be
                specified. If the value is empty, no hosts are excluded.
    </description>

</property>
```

How to commission a new node

1. Add the network addresses of the new nodes to the **include** file.

2. Update the **Namenode** with the new set of permitted nodes using below command:

   ./hadoop dfsadmin -refreshNodes

3. Update the **Jobtracker** with the new set of permitted nodes using below command:

   ./hadoop mradmin -refreshNodes

4. Update the **slaves** file with the new nodes, so that they are included in future operations performed by the Hadoop control scripts.

How to commission a new node

5. Start the new Datanodes.

```
./hadoop-daemon.sh start datanode
./hadoop-daemon.sh start tasktracker
```

6. Check that the new **Datanodes** and **Tasktrackers** appear in the web UI (http://master:50070).

7. You have successfully commissioned a new node.

What is Decommissioning

We can remove any node from Hadoop cluster, if required.

The data should be copied to alternate node, before this node leaves cluster.

Removing an old machine (node) from Hadoop cluster is known as decommissioning

Decommissioning a node from cluster will decrease the storage & processing capacity of Hadoop cluster.

How to decommission an old node

1. Add the network addresses of the nodes to be decommissioned to the **exclude** file.

2. Do not update the include file at this time.

3. Update the **Namenode** with the new set of permitted nodes using below command:

   ./hadoop dfsadmin -refreshNodes

4. Update the **Jobtracker** with the new set of permitted nodes using below command:

   ./hadoop mradmin -refreshNodes

5. Go to the web UI (http://master:50070) and check whether the admin state has changed to "**Decommission In Progress**" for the Datanodes being decommissioned.

How to decommission an old node

6. Hadoop will start copying the blocks to other Datanodes in the cluster.

7. When all the Datanodes report their state as "**Decommissioned**" then all the blocks have been replicated. Shut down the decommissioned nodes.

8. Remove the nodes from the include file, and run:

   ./hadoop dfsadmin -refreshNodes

9. Remove the nodes from the **slaves** file.

10. You have successfully decommissioned a node.

What is ACL

ACL stands for **A**ccess **C**ontrol **L**ist.

ACL contains list of authorized users & groups, who can perform specific activity on Hadoop cluster.

Hadoop doesn't allow the users to perform activity who are not part of ACL.

Hadoop Admin manage ACL.

Why do we need ACL

By default, Hadoop is not secure.

Without using ACL, any user can do any activity on Hadoop cluster.

A small command can delete everything from HDFS

```
./hadoop fs -rmr /
```

ACL helps us to control the access of different user to Hadoop.

Using ACL, we can make our Hadoop secure.

Modify core-site.xml

Add the following property to core-site.xml to enable security in Hadoop.

```
<property>

    <name>hadoop.security.authorization</name>
    <value>true</value>
    <description>To enable authorization(ACL) in Hadoop </description>

</property>
```

Hadoop-policy.xml

hadoop-policy.xml file in HADOOP_HOME/conf directory contains all security related setting.

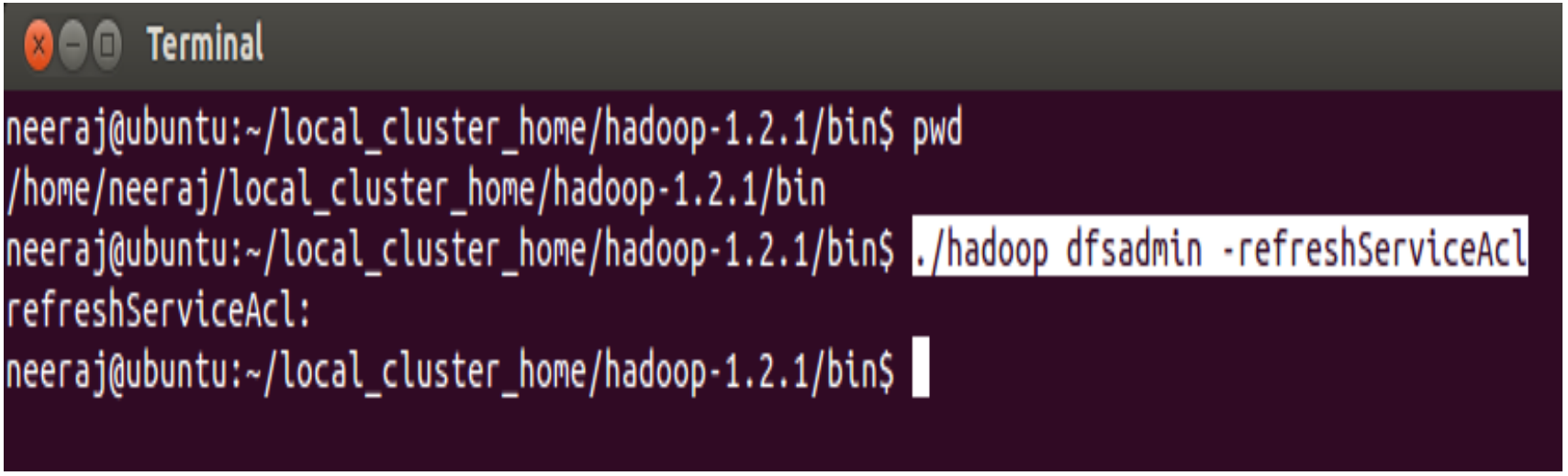
```
<property>

        <name>security.job.submission.protocol.acl</name>
        <value>*</value>
        <description>ACL for JobSubmissionProtocol, used by job clients
                     to communicate with the Jobtracker for job
                     submission, kill etc. The ACL is a comma separated
                     list of user and group names. The user and group
                     list is separated by a blank. For e.g.
                     "ravi,swathi developers,testers". A special value of "*"
                     means all users are allowed.
        </description>

</property>
```

Refreshing Service ACL

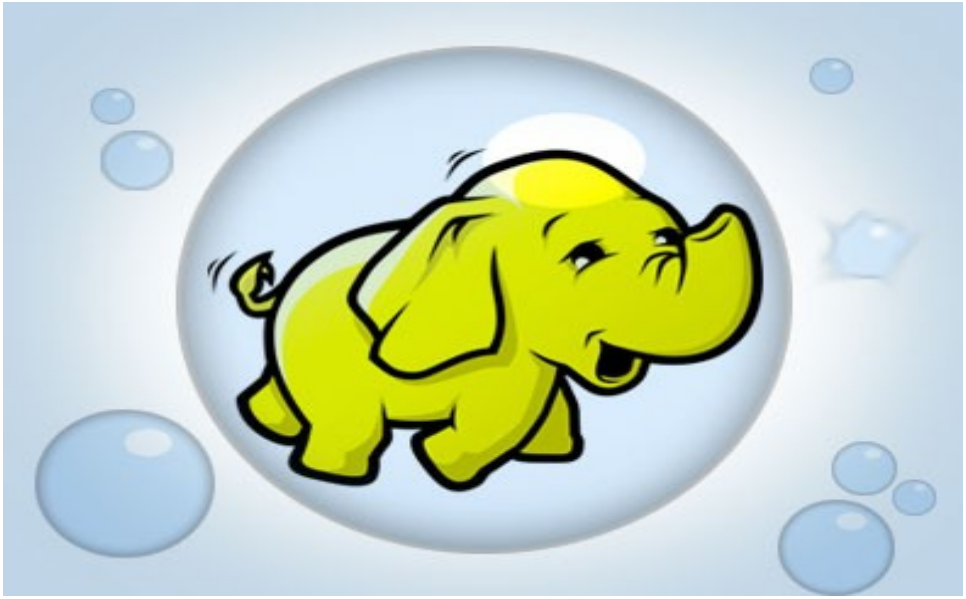
ACL is Access Control List

ACL contains the names of authorized user & groups
Who are allowed to submit/kill MR job.

Whenever we make any change to ACL, we need to refresh ACL.



...Thanks...



For online Hadoop training, send mail to neeraj.ymca.2k6@gmail.com