Big Data Frameworks
## CSE6701

Prof. Ramesh Ragala

July 9, 2019

- **ssh**
    - **Check whether ssh is working or not**
    - **$ ssh localhost**
    - if the result like connection refused on port 22 then start **sshd** service
    - command to start sshd service in fedora
        - $ sudo systemctl start sshd → it asks password
        - check ssh
        - **$ ssh localhost** → if it shows any error, then install ssh
        - command to install ssh
        - **$ sudo dnf install openssh-server***
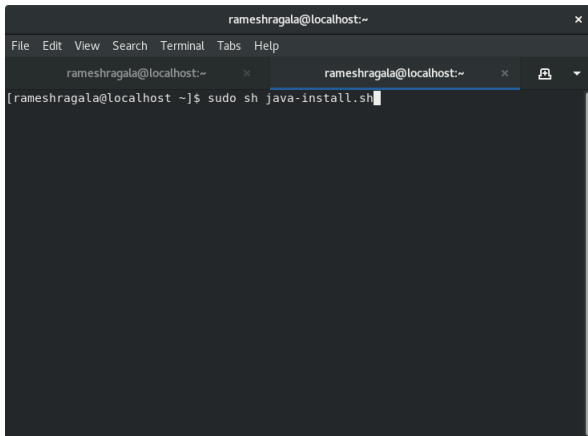   - check the ssh in machine

- **java**
  - **Check java version on your local machine**
  - command is $ **java -version**
  - if the java version is **openjdk**, then install oracle JDK
  - **procedure to install oracle java**
  - Download JDK from Oracle jdk website
    http://www.oracle.com/technetwork/java/javase/downloads/index.html
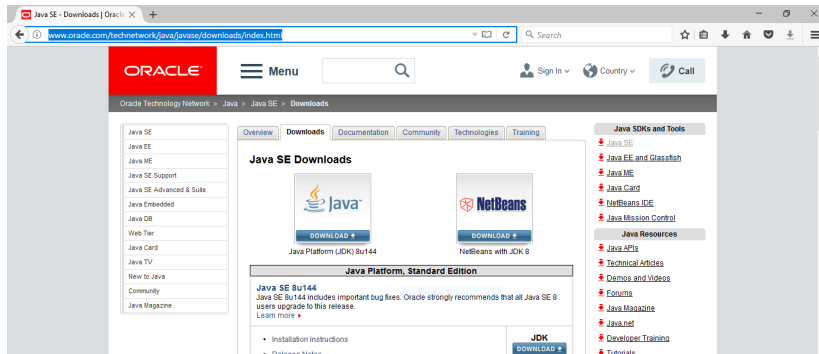
- create a directory **in** /opt directory
  - command is $ **sudo mkdir -p /opt/java8**
- change the permission of the directory
  - command is $ **sudo chmod 777 -R java**
- copy the jdk-x.x.x.gz into /opt/java8 directory
  - command is $ **cp /home/vitchennai/Downloads/jdk-x.x.x.gz /opt/java8**
- untar the jdk file
  - command is $ **tar -xvzf jdk-x.x.x.gz**
- change the permission of java directory
  - command is $ **sudo chmod 777 -R /opt/java8/jdk-x.x.x**

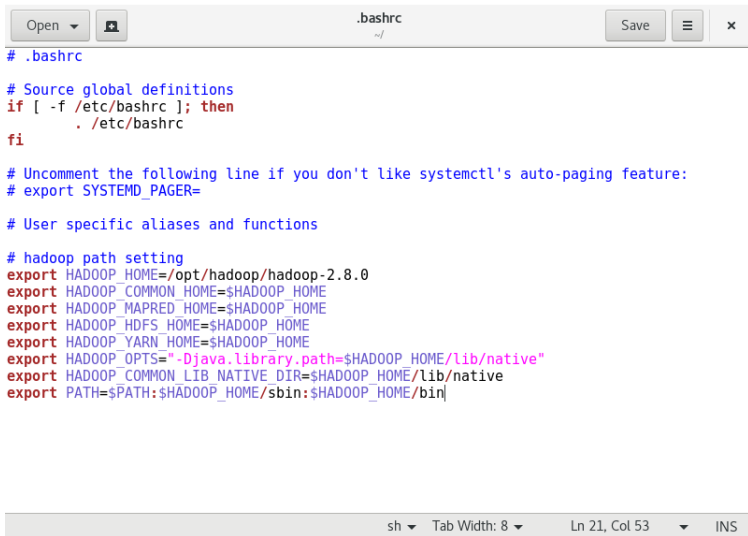- path setting details are set in java-install.sh(personal file)
- Execute java-install.sh file in terminal

HADOOP INSTALLATION STEPS
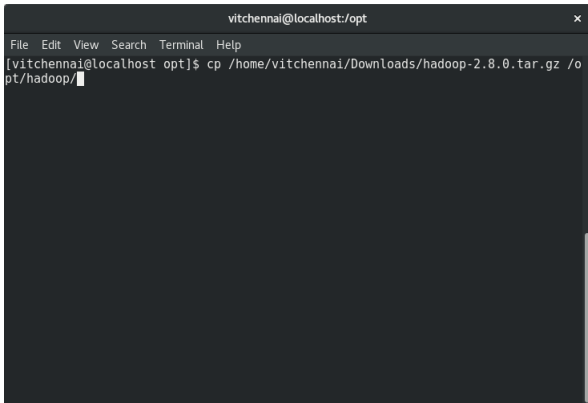
- Download latest hadoop binary file from
  http://hadoop.apache.org/releases.html
- Let us assume it have downloaded in Download directory

- create hadoop directory in /opt directory

- change the permission of this directory
  - command is $ **sudo chmod 777 -R /opt/hadoop**
- copy the hadoop-2.8.0.tar.gz into /opt/hadoop directory
  - command is $ **cp /home/vitchennai/Downloads/hadoop-2.8.0.tar.gz /opt/hadoop/**

Hadoop Installation steps

- untar the file
  - command is $ **tar -xvzf hadoop-2.8.0.tar.gz**
- change the permission of **hadoop** directory
  - command is $ **sudo chmod 777 -R /opt/hadoop**

- open bashrc file for hadoop path setting
  - command is $ **gedit ∼/.bashrc**

- Append the Hadoop and java paths
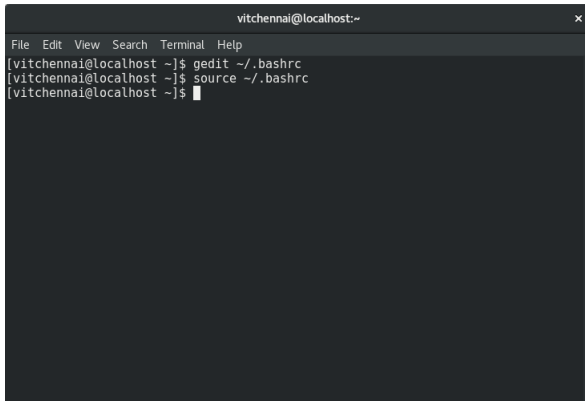
- Restart the terminal
    - command to restart is $**source ∼/.bashrc**

- Hadoop requires SSH access to manage its nodes, i.e. remote machines plus your local machine if you want to use Hadoop on it
- Commands is to create an RSA key pair with an empty password.

- we have to enable SSH access to local machine with this newly created key.
- Commands is

- Go to hadoop directory for Configuration purpose
- add java path in **hadoop_env.sh** file
- command to open hadoop_env.sh is $ **gedit hadoop_env.sh**

- open **core-site.xml** for configuration purpose
- command to open core-site.xml is
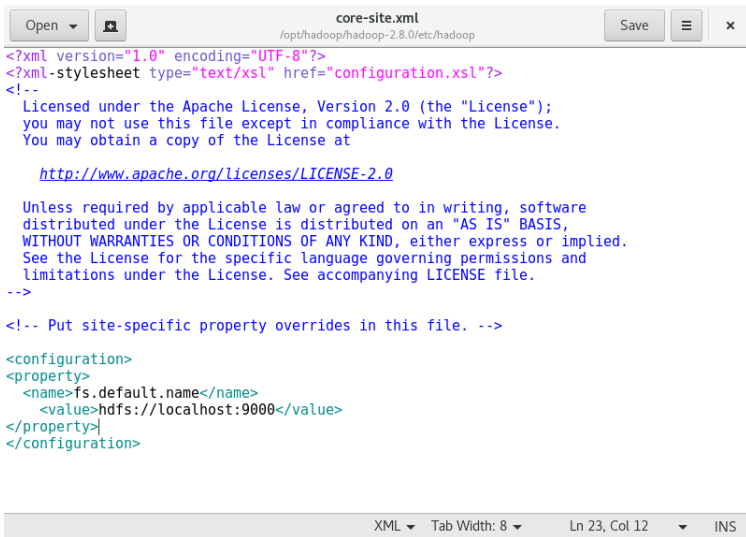
- do the modification in core-site.xml as shown below, save and exit


```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>fs.default.name</name>
    <value>hdfs://localhost:9000</value>
</property>
</configuration>
```
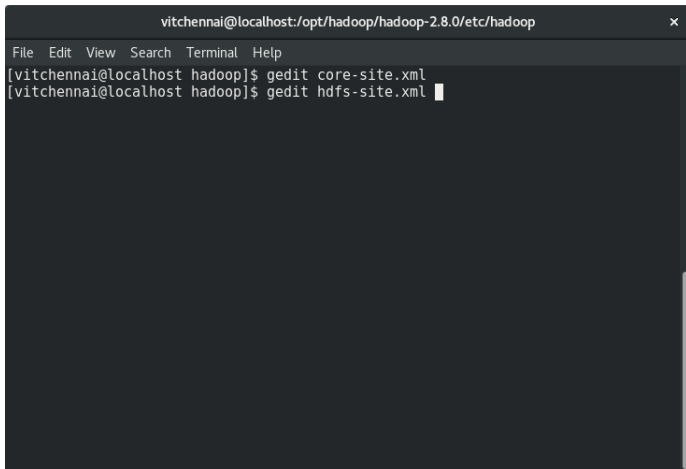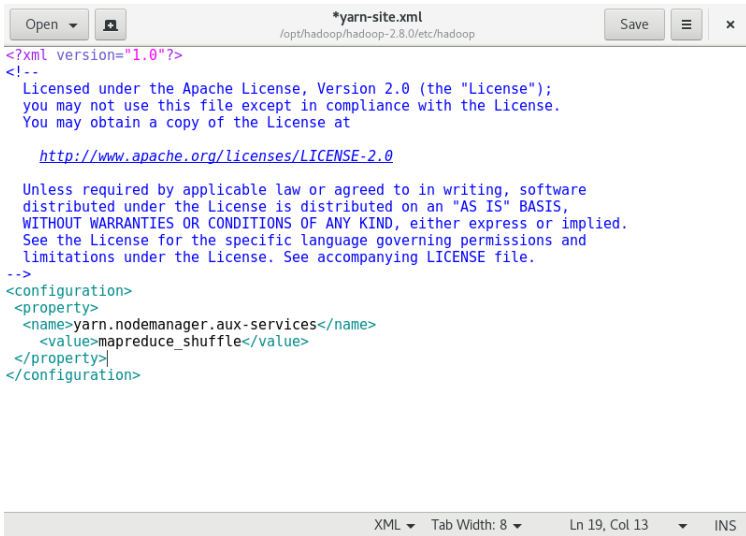
- open **hdfs-site.xml** for configuration purpose
- command to open hdfs-site.xml is

- do the modification in hdfs-site.xml as shown below, save and exit

- create **mapred-site.xml** from **mapred-site.xml.template**
  configuration purpose
- command to open mapred-site.xml is

- do the modification in mapred-site.xml as shown below, save and exit

- open **yarn-site.xml** for configuration purpose
- command to open yarn-site.xml is

- do the modification in yarn-site.xml as shown below, save and exit


```xml
<?xml version="1.0"?>
<!--
  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.
-->
<configuration>
 <property>
  <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
 </property>
</configuration>
```

- we need to create directories for namenode and datanode, which are specified in hdfs-site.xml
- The following commands are for creating namenode, Datanode and permission settings

- Now we need to format the namenode
- The following commands is used to format the namenode

- **Hadoop-2.8.0 installation has completed**
- **Check the Hadoop status**

- Now start the Hadoop Distributed File System
- The following commands (**start-dfs.sh**) is used for this

- If Datanode is not started, then change the permissions of datanode
- The following command is used

sudo chown vitchennai:root
/opt/Hadoop/hadoop_tmp/hdfs/datanode

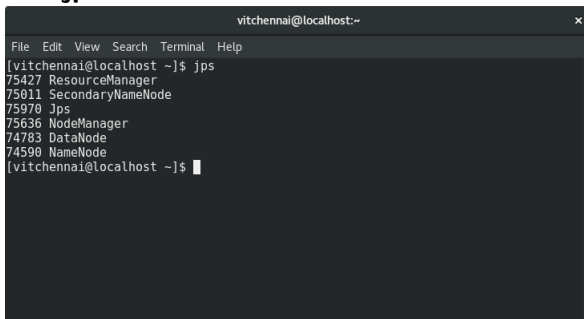- Now check the background process for hadoop distributed file system

- Now start yarn resources for hadoop
- The following commands (**start-yarn.sh**) is used for this
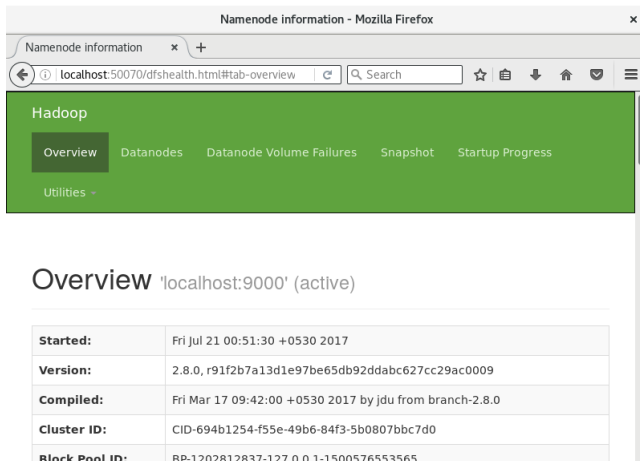
- The total number of daemons to execute hadoop-2.8 on local machine are
- **NameNode**
- **DataNode**
- **SecondaryNameNode**
- **NodeManager**
- **ResourceManager**
- we have to use **jps** command to check

- **UI view of Hadoop**
- Open **http://localhost:50070** in browser → Namenode

- Click on **Utilities** menu bar and then click on **Browse File System**
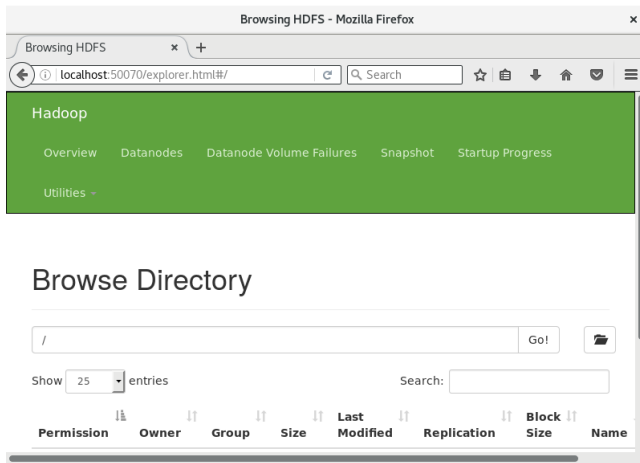
- command to stop hadoop are $**stop-dfs.sh**$ and $**stop-yarn.sh**$

- command to passwordless promt
- $ **sudo chmod 700 ∽ ╱.ssh╱**
- $ **sudo chmod 600 ∽ ╱.ssh╱authorized_keys**