Installing hadoop

1. Install java 8 on your system
2. Add other systems host names to etc/hosts
3. Make same user in all systems
4. Ssh through all systems
5. Download hadoop from http://hadoop.apache.org/releases.html(use axel for downloading)
6. Once you download files unzip them using **tar xvzf filename**
7. Set java home in hadoop-env "**exprt JAVA_HOME=java home path"**
8. Set ssh port in hadoop-env "**export HADOOP_SSH_OPTS="-p PORTNUMBER -o ConnectionTimeout=1 -o SendEnv=HADOOP_CONF_DIR"**
9. you'll also need to add some environment variables to run Hadoop and its modules. They should be added to the bottom of the file so it looks like the following
    a. export HDFS_NAMENODE_USER="username"
    b. export HDFS_DATANODE_USER="username"
    c. export HDFS_SECONDARYNAMENODE_USER="username"
    d. export YARN_RESOURCEMANAGER_USER="username"
    e. export YARN_NODEMANAGER_USER="username"
10. With the hadoop-env.sh script updated and sourced, we need to create a data directory for the Hadoop Distributed File System (HDFS) to store all relevant HDFS files.
11. Make data dir ~/hadooop/hdfs/data
12. Set the permissions for this file with your respective user.
13. Update core-site.xml

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://server-ip:9000</value>
  </property>
</configuration>
```

14. Update hdfs-site.xml in master

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>2</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:///home/username/hadoop/hdfs/data</value>
  </property>
</configuration>
```

16. Upadte hdfs-site.xml in slaves

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>3</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:///home/username/hadoop/hdfs/data</value>
  </property>
</configuration>
```

15. Update maperd-site.xml

```
<configuration>
  <property>
    <name>mapreduce.jobtracker.address</name>
    <value>master-id:54311</value>
  </property>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```

16. Update yarn-site.xml

```
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>
  <property>
    <name>yarn.resourcemanager.hostname</name>
    <value>master-ip</value>
  </property>
</configuration>
```

17. Set master hostname in masters file
18. set slaves hostnames in slaves
19. Format the namenode

```
./bin/hdfs namenode -format
```

20. Start service using ./sbin/start-all.sh