Hadoop Single-Node cluster Installation

- step 1. Prepare machine with JDK and ssh installation.

```
sudo apt install openjdk-8-jdk ssh
```

- step 2. In /etc/hosts ensure entry of standalone hostname.

```
127.0.0.1   localhost
```

- step 3. Enable password-less login for SSH:

```
ssh-keygen -t rsa -P ""
ssh-copy-id $USER@localhost
ssh localhost
```

- step 4. Download & Extract Hadoop into $HOME.
  - Download from https://archive.apache.org/dist/hadoop/common/hadoop-3.3.2/hadoop-3.3.2.tar.gz

```
cd ~
tar xvf ~/Downloads/hadoop-3.3.2.tar.gz
```

- step 5. In $HOME/.bashrc

```
export PDSH_RCMD_TYPE=ssh

export HADOOP_HOME=$HOME/hadoop-3.3.2
export PATH=$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$PATH
```

- step 6. In $HADOOP_HOME/etc/hadoop/hadoop-env.sh

```
export JAVA_HOME="/usr/lib/jvm/java-11-openjdk-amd64"
```

- step 7: In $HADOOP_HOME/etc/hadoop/core-site.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
```

```
<configuration>
    <property>
        <name>fs.defaultFS</name>
        <value>hdfs://localhost:9000</value>
    </property>
</configuration>
```

- step 8: In $HADOOP_HOME/etc/hadoop/hdfs-site.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<configuration>
    <property>
        <name>dfs.name.dir</name>
        <value>${user.home}/bigdata/hd-data/nn</value>
    </property>
    <property>
        <name>dfs.data.dir</name>
        <value>${user.home}/bigdata/hd-data/dn</value>
    </property>
    <property>
        <name>dfs.replication</name>
        <value>1</value>
    </property>
</configuration>
```

- step 9: In $HADOOP_HOME/etc/hadoop/mapred-site.xml

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<configuration>
    <property>
        <name>mapreduce.framework.name</name>
        <value>yarn</value>
    </property>
    <property>
        <name>mapreduce.application.classpath</name>

<value>$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/*:$HADOOP_MAPRED_HOM
E/share/hadoop/mapreduce/lib/*</value>
    </property>
</configuration>
```

- step 10: In $HADOOP_HOME/etc/hadoop/yarn-site.xml

```
<?xml version="1.0"?>
<configuration>
```

```
    <property>
        <name>yarn.resourcemanager.hostname</name>
        <value>localhost</value>
    </property>
    <property>
        <name>yarn.nodemanager.aux-services</name>
        <value>mapreduce_shuffle</value>
    </property>
    <property>
        <name>yarn.nodemanager.local-dirs</name>
        <value>${user.home}/bigdata/hd-data/yarn/data</value>
    </property>
    <property>
        <name>yarn.nodemanager.logs-dirs</name>
        <value>${user.home}/bigdata/hd-data/yarn/logs</value>
    </property>
    <property>
        <name>yarn.nodemanager.disk-health-checker.max-disk-
utilization-perdisk-percentage</name>
        <value>99.9</value>
    </property>
    <property>
        <name>yarn.nodemanager.vmem-check-enabled</name>
        <value>false</value>
    </property>
    <property>
        <name>yarn.nodemanager.env-whitelist</name>

<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,C
LASSPATH_PREPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
    </property>
</configuration>
```

- step 11: In $HADOOP_HOME/etc/hadoop/workers

```
localhost
```

- step 12: Format namenode

```
hdfs namenode -format
```

- step 13: Start HDFS & YARN. Then verify using jps command.

```
start-dfs.sh
start-yarn.sh
jps
```

- step 14: Check Hadoop web interface in browser.

```
http://localhost:9870/
```

- step 15: HDFS commands

```
hadoop fs -ls /
hadoop fs -mkdir /user/nilesh
hadoop fs -put localfilepath /user/nilesh
hadoop fs -get /user/nilesh/filepath localfilepath
```

- step 16: Stop HDFS & YARN. Then verify using jps command

```
stop-yarn.sh
stop-dfs.sh
jps
```