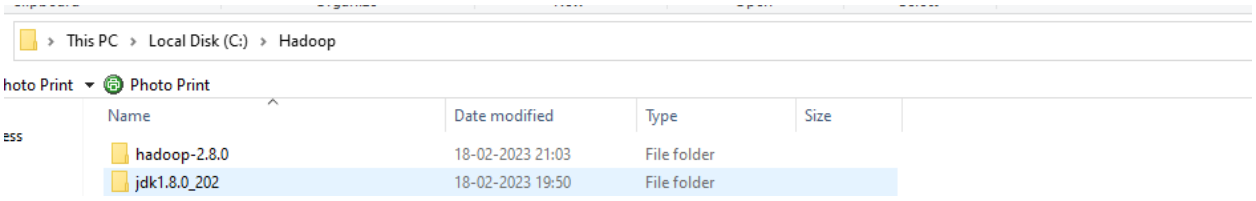
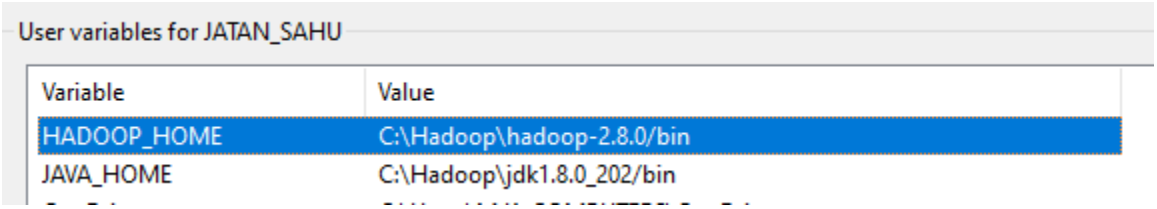
IT609: Big Data Processing
Assignment 02:Creating a Single Node Cluster using Hadoop

NAME -JATAN SAHU
ID - 202218061

The steps to install Java and Hadoop are as follows:
1. Visit https://www.oracle.com/in/java/technologies/javase/javase8-archive-downloads.HTML and download Java SE DK 8u202 for Windows x64.
2. Extract and install Java in the desired folder. (Here, C:\Program Files\Java).
3. Download Hadoop 2.8.0 tar.gz file from https://hadoop.apache.org/release/2.8.0.html
4. Open Command Prompt in the folder where Hadoop tar.gz file is saved (here, C:\Setups) and type the following command: tar -xvf .\hadoop-2.8.0.tar.gz
5. Copy the Java JDK folder from C:\Program Files\Java and paste in the same folder as Hadoop (here, C:\Setups).



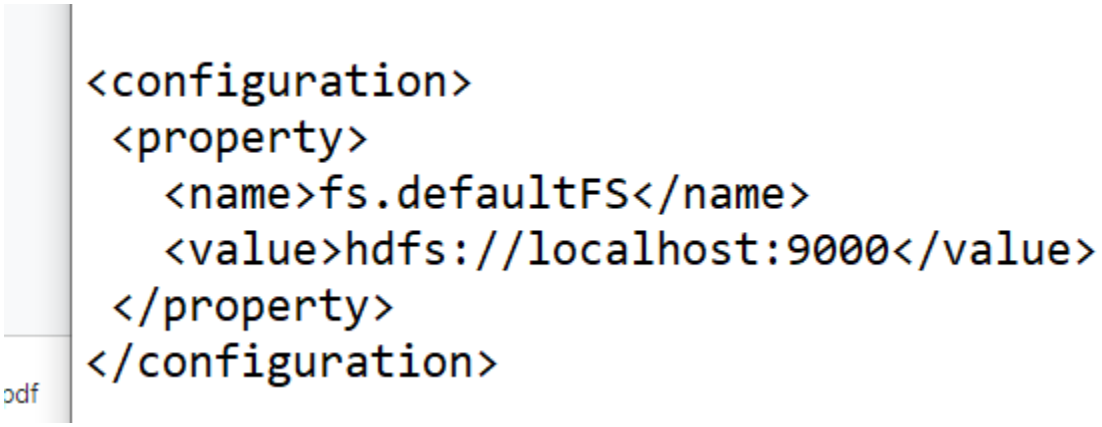
6. Go to Search and open 'Advanced System Settings'. Click on Environment Variables. Create a new variable in user variables category. Set Variable Name as 'HADOOP_HOME' and value as 'C:\Hadoop\hadoop-2.8.0/bin'. Similarly, create another variable with name 'JAVA_HOME' and set the value as 'C:\Hadoop\jdk1.8.0_202/bin'. Save the changes



7. Select Path Variable and edit it. Add new paths values used above here as well and move them to the top.

The procedure to update the configuration on file in etc/Hadoop folder is as follows:
1. Go to C:\Hadoop\hadoop-2.8.0\etc\hadoop
. 2. Open core-site.xml with Notepad.
3. Paste the following XML code within the tags and save the file.

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



4. Rename “mapred-site.xml.template” to “mapred-site.xml” and edit this file as done earlier. Paste the following code and save the file:

```
<configuration>
  <property>
      <name>mapreduce.framework.name</name>
      <value>yarn</value>
  </property>
</configuration>
```

```
<configuration>
    <property>
        <name>mapreduce.framework.name</name>
        <value>yarn</value>
    </property>
</configuration>
```

5. Create a folder with data under 'D:\Setups\hadoop-2.8.0'

6. Create folder 'datanode' under C:\Setups\hadoop-2.8.0

7. Create folder 'namenode' under C:\Setups\hadoop-2.8.0

8. Edit file C:\Hadoop-2.8.0/etc/hadoop/hdfs-site.xml,

paste xml code and save this file.

```
<configuration>
  <property>
        <name>dfs.replication</name>
        <value>1</value>
  </property>
  <property>
        <name>dfs.namenode.name.dir</name>
        <value>/hadoop-2.8.0/data/namenode</value>
  </property>
  <property>
        <name>dfs.datanode.data.dir</name>
```

<value>/hadoop-2.8.0/data/datanode</value>

</property>

</configuration>

```
<configuration>
   <property>

       <name>dfs.replication</name>

       <value>1</value>

   </property>

   <property>

       <name>dfs.namenode.name.dir</name>

       <value>/hadoop-3.3.0/data/namenode</value>

   </property>

   <property>

       <name>dfs.datanode.data.dir</name>

       <value>/hadoop-3.3.0/data/datanode</value>
```

9.Edit file C:/Hadoop-2.8.0/etc/hadoop/yarn-site.xml,

paste xml code and save this file.

<configuration>

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

```
  <property>
                <name>yarn.nodemanager.aux-
services</name>
                <value>mapreduce_shuffle</value>
   </property>
   <property>

<name>yarn.nodemanager.auxservices.mapreduce.shuffle.cla
ss</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>
   </property>
</configuration>
```

10.Edit file C:/Hadoop-2.8.0/etc/hadoop/hadoop-env.cmd

by closing the command line

“JAVA_HOME=%JAVA_HOME%” instead of set “JAVA_HOME=C:\Java”

-

```
@rem The java implementation to use.  Required.
@rem set JAVA_HOME=%JAVA_HOME%
set JAVA_HOME=C:\Hadoop\jdk1.8.0_202

@rem The jsvc implementation to use. Jsvc is required to
```

10.Hadoop Configurations

Download

https://github.com/brainmentorspvtltd/BigData_RDE/blob/master/Hadoop%20Configuration.zip

or (for hadoop 3)

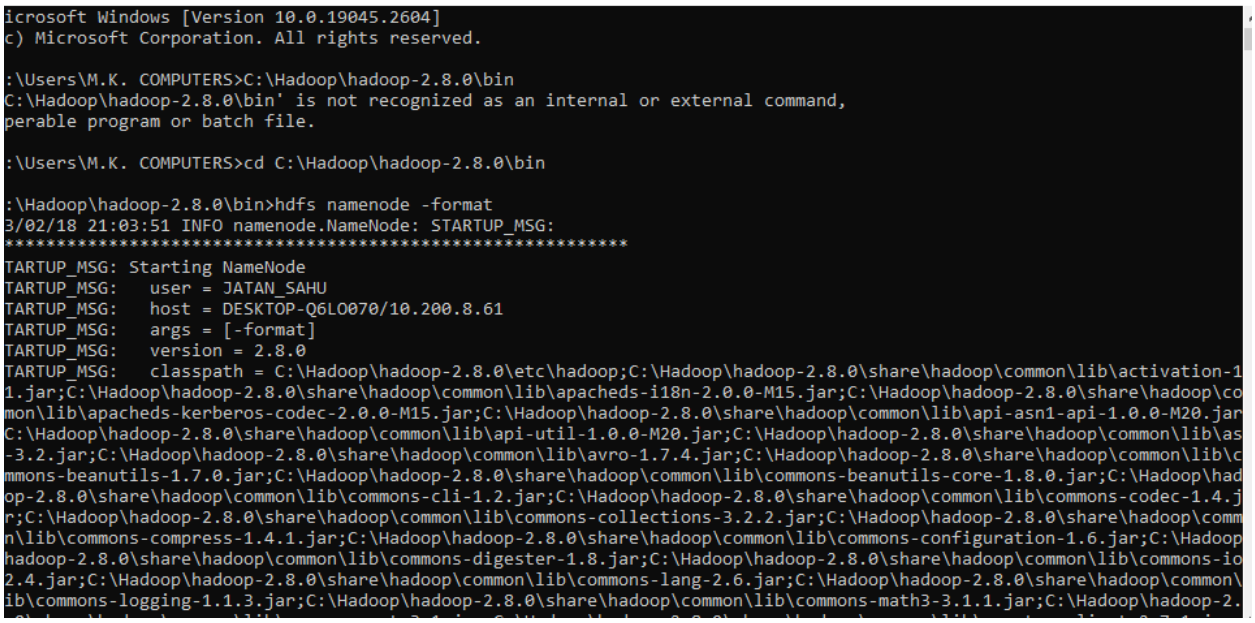
https://github.com/s911415/apache-hadoop-3.1.0-winutils

– Copy folder bin and replace existing bin folder in

C:\Hadoop-2.8.0\bin

– Format the NameNode

– Open cmd and type command “hdfs namenode –format”



7. Testing

– Open cmd and change directory to C:\Hadoop-2.8.0\sbin

– type start-all.cmd

```
23/02/18 21:03:54 INFO util.GSet: 1.0% max memory 889 MB = 8.9 MB
23/02/18 21:03:54 INFO util.GSet: capacity      = 2^20 = 1048576 entries
23/02/18 21:03:54 INFO namenode.FSDirectory: ACLs enabled? false
23/02/18 21:03:54 INFO namenode.FSDirectory: XAttrs enabled? true
23/02/18 21:03:54 INFO namenode.NameNode: Caching file names occurring more than 10 times
23/02/18 21:03:54 INFO util.GSet: Computing capacity for map cachedBlocks
23/02/18 21:03:54 INFO util.GSet: VM type       = 64-bit
23/02/18 21:03:54 INFO util.GSet: 0.25% max memory 889 MB = 2.2 MB
23/02/18 21:03:54 INFO util.GSet: capacity      = 2^18 = 262144 entries
23/02/18 21:03:54 INFO namenode.FSNamesystem: dfs.namenode.safemode.threshold-pct = 0.9990000128746033
23/02/18 21:03:54 INFO namenode.FSNamesystem: dfs.namenode.safemode.min.datanodes = 0
23/02/18 21:03:54 INFO namenode.FSNamesystem: dfs.namenode.safemode.extension     = 30000
23/02/18 21:03:54 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
23/02/18 21:03:54 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
23/02/18 21:03:54 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
23/02/18 21:03:54 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
23/02/18 21:03:54 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time
is 600000 millis
23/02/18 21:03:54 INFO util.GSet: Computing capacity for map NameNodeRetryCache
23/02/18 21:03:54 INFO util.GSet: VM type       = 64-bit
23/02/18 21:03:54 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
23/02/18 21:03:54 INFO util.GSet: capacity      = 2^15 = 32768 entries
23/02/18 21:03:54 INFO namenode.FSImage: Allocated new BlockPoolId: BP-623877158-10.200.8.61-1676734434799
23/02/18 21:03:54 INFO common.Storage: Storage directory C:\hadoop-3.3.0\data\namenode has been successfully formatted.
23/02/18 21:03:54 INFO namenode.FSImageFormatProtobuf: Saving image file C:\hadoop-3.3.0\data\namenode\current\fsimage.c
kpt_0000000000000000000 using no compression
23/02/18 21:03:55 INFO namenode.FSImageFormatProtobuf: Image file C:\hadoop-3.3.0\data\namenode\current\fsimage.ckpt_000
0000000000000000 of size 327 bytes saved in 0 seconds.
23/02/18 21:03:55 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
23/02/18 21:03:55 INFO util.ExitUtil: Exiting with status 0
```

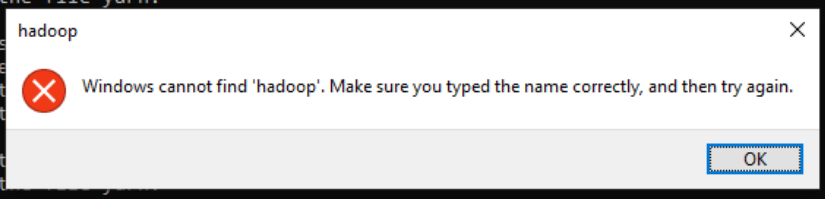
```
arting yarn daemons
e system cannot find the file yarn.
e system cannot find the file yarn.

\Hadoop\hadoop-2.8.0\s
is script is Deprecate
e system cannot find t
e system cannot find t
arting yarn daemons
e system cannot find t
e system cannot find t

\Hadoop\hadoop-2.8.0\sbin>start-all.cmd
is script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
e system cannot find the file hadoop.
e system cannot find the file hadoop.
arting yarn daemons
e system cannot find the file yarn.
e system cannot find the file yarn.

\Hadoop\hadoop-2.8.0\sbin>

\Hadoop\hadoop-2.8.0\sbin>start-all.cmd
is script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
```



Tried in system variables trying moving the jdk, jre and hadoop path to top still it is not working