

WEEK-1 BDLAB Submission

- Name : P K Navin Shrinivas
- SRN : PES2UG20CS237
- Section : D

Screenshots :

ssh without password :

```
pes2ug20cs237@pes2ug20cs237:~$ ssh localhost
Welcome to Ubuntu 22.04 LTS (GNU/Linux 5.15.0-25-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

289 updates can be applied immediately.
114 of these updates are standard security updates.
To see these additional updates run: apt list --upgradable

Last login: Wed Aug 24 11:19:23 2022 from 127.0.0.1
pes2ug20cs237@pes2ug20cs237:~$ █
```

Bashrc :

```
#Hadoop Path Configs
export HADOOP_HOME=/home/pes2ug20cs237/Downloads/hadoop-3.2.4
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```

Hadoop-env.sh :

```
pes2ug20cs237@pes2ug20cs237: ~/Downloads/hadoop-3.2.4/etc/hadoop
export JAVA_HOME=/usr/lib/jvm/java-1.11.0-openjdk-amd64

#
# Licensed to the Apache Software Foundation (ASF) under one
# or more contributor license agreements. See the NOTICE file
# distributed with this work for additional information
# regarding copyright ownership. The ASF licenses this file
# to you 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.

# Set Hadoop-specific environment variables here.

##
## THIS FILE ACTS AS THE MASTER FILE FOR ALL HADOOP PROJECTS.
hadoop-env.sh 1,1 Top
```

core-site.xml :

```
pes2ug20cs237@pes2ug20cs237: ~/Downloads/hadoop-3.2.4/etc/hadoop
<?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>hadoop.tmp.dir</name>
    <value>/home/pes2ug20cs237/Downloads/hadoop-3.2.4/tmpdata</value>
  </property>
  <property>
    <name>fs.default.name</name>
  </property>
</configuration>
core-site.xml 1,1 Top
```

hdfs-site.xml :

```
pes2ug20cs237@pes2ug20cs237: ~/Downloads/hadoop-3.2.4/etc/hadoop
<?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>dfs.data.dir</name>
  <value>/home/pes2ug20cs237/Downloads/hadoop-3.2.4/namenode</value>
</property>
hdfs-site.xml 1,1 Top
```

mapred-site.xml :

```
pes2ug20cs237@pes2ug20cs237: ~/Downloads/hadoop-3.2.4/etc/hadoop
<?xml version="1.0"?>
<?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>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
~
mapred-site.xml 1,1 All
```

yarn-site.xml :

```
pes2ug20cs237@pes2ug20cs237: ~/Downloads/hadoop-3.2.4/etc/hadoop
<?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>
  <property>
    <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>
</configuration>

yarn-site.xml 1,1 Top
```

Formatting namenode :

```
pes2ug20cs237@pes2ug20cs237: ~/Downloads/hadoop-3.2.4
2022-08-24 11:50:24,121 INFO util.GSet: Computing capacity for map cachedBlocks
2022-08-24 11:50:24,121 INFO util.GSet: VM type = 64-bit
2022-08-24 11:50:24,122 INFO util.GSet: 0.25% max memory 3.8 GB = 9.6 MB
2022-08-24 11:50:24,122 INFO util.GSet: capacity = 2^20 = 1048576 entries
2022-08-24 11:50:24,131 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
2022-08-24 11:50:24,131 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
2022-08-24 11:50:24,132 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
2022-08-24 11:50:24,136 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
2022-08-24 11:50:24,136 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
2022-08-24 11:50:24,138 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2022-08-24 11:50:24,138 INFO util.GSet: VM type = 64-bit
2022-08-24 11:50:24,138 INFO util.GSet: 0.029999999329447746% max memory 3.8 GB = 1.2 MB
2022-08-24 11:50:24,138 INFO util.GSet: capacity = 2^17 = 131072 entries
Re-format filesystem in Storage Directory root= /home/pes2ug20cs237/Downloads/hadoop-3.2.4/tmpdata/dfs/name; location= null ? (Y or N) Y
2022-08-24 11:50:26,135 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1841145699-127.0.1.1-1661322026125
2022-08-24 11:50:26,136 INFO common.Storage: Will remove files: [/home/pes2ug20cs237/Downloads/hadoop-3.2.4/tmpdata/dfs/name/current/seen_txid, /home/pes2ug20cs237/Downloads/hadoop-3.2.4/tmpdata/dfs/name/current/fsimage_00000000000000000000, /home/pes2ug20cs237/Downloads/hadoop-3.2.4/tmpdata/dfs/name/current/edits_00000000000000000000, /home/pes2ug20cs237/Downloads/hadoop-3.2.4/tmpdata/dfs/name/current/edits_inprogress_00000000000000000000, /home/pes2ug20cs237/Downloads/hadoop-3.2.4/tmpdata/dfs/name/current/VERSION, /home/pes2ug20cs237/Downloads/hadoop-3.2.4/tmpdata/dfs/name/current/edits_00000000000000000001, /home/pes2ug20cs237/Downloads/hadoop-3.2.4/tmpdata/dfs/name/current/fsimage_00000000000000000000, /home/pes2ug20cs237/Downloads/hadoop-3.2.4/tmpdata/dfs/name/current/edits_00000000000000000000, /home/pes2ug20cs237/Downloads/hadoop-3.2.4/tmpdata/dfs/name/current/fsimage_00000000000000000000]
2022-08-24 11:50:26,166 INFO common.Storage: Storage directory /home/pes2ug20cs237/Downloads/hadoop-3.2.4/tmpdata/dfs/name has been successfully formatted.
2022-08-24 11:50:26,217 INFO namenode.FSImageFormatProtobuf: Saving image file /home/pes2ug20cs237/Downloads/hadoop-3.2.4/tmpdata/dfs/name/current/fsimage.ckpt_00000000000000000000 using no compression
2022-08-24 11:50:26,335 INFO namenode.FSImageFormatProtobuf: Image file /home/pes2ug20cs237/Downloads/hadoop-3.2.4/tmpdata/dfs/name/current/fsimage.ckpt_00000000000000000000 of size 408 bytes saved in 0 seconds.
2022-08-24 11:50:26,357 INFO namenode.MNStorageRetentionManager: Going to retain 1 images with txid >= 0
2022-08-24 11:50:26,403 INFO namenode.FSNamesystem: Stopping services started for active state
2022-08-24 11:50:26,403 INFO namenode.FSNamesystem: Stopping services started for standby state
2022-08-24 11:50:26,409 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2022-08-24 11:50:26,410 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at pes2ug20cs237/127.0.1.1
*****/
pes2ug20cs237@pes2ug20cs237: ~/Downloads/hadoop-3.2.4$
```

JPS output

```

pes2ug20cs237@pes2ug20cs237:~/Downloads/hadoop-3.2.4$ ./sbin/start-yarn.sh
Starting resourcemanager
Starting nodemanagers
pes2ug20cs237@pes2ug20cs237:~/Downloads/hadoop-3.2.4$ jps
24323 ResourceManager
23828 DataNode
24473 NodeManager
24860 Jps
23692 NameNode
24078 SecondaryNameNode
pes2ug20cs237@pes2ug20cs237:~/Downloads/hadoop-3.2.4$ █

```

MapReduce job output 1 :

```

pes2ug20cs237@pes2ug20cs237:~/github/UE20CS30X-Submissions/BDLAB/WEEK-1$ hdfs dfs -cat /example/output/part-r-00000 | head
2022-08-24 12:22:35,432 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicabl
e
"      4
"Morshipful  1
"a      1
"*Do      1
"*Hin      1
"*Stint 1
"*Tell      1
"*Woe      1
"*take      1
"."      1
cat: Unable to write to output stream.

```

MapReduce job output 2 :

```

pes2ug20cs237@pes2ug20cs237:~/github/UE20CS30X-Submissions/BDLAB/WEEK-1$ hdfs dfs -cat /handson/output-tech/part-00000
2022-08-24 12:25:39,315 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicabl
e
54
pes2ug20cs237@pes2ug20cs237:~/github/UE20CS30X-Submissions/BDLAB/WEEK-1$ █

```