Lab 2:
Configuring HDFS and YARN for Hadoop

Set NameNode location

- Update hdfs-site.conf from the directory: /home/student/hadoop/etc/hadoop

```
<configuration>
    <property>
            <name>dfs.namenode.name.dir</name>
            <value>/home/student/hadoop/data/nameNode</value>
    </property>

    <property>
            <name>dfs.datanode.data.dir</name>
            <value>/home/student/hadoop/data/dataNode</value>
    </property>

    <property>
            <name>dfs.replication</name>
            <value>1</value>
    </property>
</configuration>
```

Configuring HDFS

- Update /home/student/hadoop/etc/hadoop/core-site.xml:

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
    <configuration>
        <property>
            <name>fs.default.name</name>
            <value>hdfs://localhost:9000</value>
        </property>
    </configuration>
```

- You can change ‘localhost’ to your PC’s IP address.
- Alternatively, you can put ‘file:///’ instead to indicate local file system

Configuring YARN

- Update mapred-site.xml from the same directory:

```
<configuration>
    <property>
            <name>mapreduce.framework.name</name>
            <value>yarn</value>
    </property>
    <property>
            <name>yarn.app.mapreduce.am.resource.mb</name>
            <value>512</value>
    </property>
    <property>
            <name>mapreduce.map.memory.mb</name>
            <value>256</value>
    </property>
    <property>
            <name>mapreduce.reduce.memory.mb</name>
            <value>256</value>
    </property>
</configuration>
```

Configuring YARN (Part 2)

- Update yarn-site.xml from the same directory:

```
<configuration>
    <property>
            <name>yarn.acl.enable</name>
            <value>0</value>
    </property>

    <property>
            <name>yarn.resourcemanager.hostname</name>
            <value>localhost</value>
    </property>

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

Configuring YARN (Part 2)

- Add these lines to yarn-site.xml:

```
<property>
        <name>yarn.nodemanager.resource.memory-mb</name>
        <value>1536</value>
</property>

<property>
        <name>yarn.scheduler.maximum-allocation-mb</name>
        <value>1536</value>
</property>

<property>
        <name>yarn.scheduler.minimum-allocation-mb</name>
        <value>128</value>
</property>

<property>
        <name>yarn.nodemanager.vmem-check-enabled</name>
        <value>false</value>
</property>
```

Formatting HDFS

- Try `hdfs namenode -format`
  - Is it working? If not, how to make it work?
- If it is working, try `start-all.sh`
  - or run `start-dfs.sh` and `start-yarn.sh` separately
- You can go to http://localhost:50070 to check your HDFS status (namenode)
  - If port 50070 is not working, please try port 9870.
  - 50090 → secondary namenode

Try again the mapreduce sample

- It is working?
- If not, how to make it work?
- Hint:
- After the Hadoop settings, the default path for hadoop is now point to hdfs file system:

Try to run a set of commands and inspect their outcome!

- `hadoop fs -mkdir /user/hdfs`
- `hadoop fs -ls /user`
- `touch sample.txt`

`hdfs dfs -put sample.txt /user/hdfs/sample.txt`

- `hdfs dfs -ls /user/hdfs/`
- `echo "This is line 1." >> sample1.txt`
- `echo "This is line 2." >> sample1.txt`
- `echo "This is line 3." >> sample1.txt`
- `cat sample.txt`
- `hdfs dfs -appendToFile sample1.txt /user/hdfs/sample.txt`
- `hdfs dfs -get /user/hdfs/sample.txt`
- `hdfs dfs -rm /user/hdfs/sample.txt`
- `exit`