

— Setting up a single Node Cluster — # 3.4.C YARN on a single Node

The following instructions assume that you have configured and formatted HDFS, have created the user/ directories, and that the hdfs is not currently running.

1. Configure `mapred-site.xml` as shown below

```
> <configuration>
  <property>                <name>mapreduce.framework.name</name>
  <value>yarn</value>        </property>    <property>                <name>mapreduce.application.classpath</name>
  <value>${HADOOP_HOME}/share/hadoop/mapreduce/*:${HADOOP_HOME}/share/hadoop/mapreduce/lib/*</value>
  </property> </configuration>
```

The first property specifies you will be using YARN for mapReduce jobs. The second property specifies the location of Hadoop Libraries (HADOOP_HOME is already configured).
2. Configure `yarn-site.xml` as shown below

```
> <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>
```
3. Run the yarn daemon and ensure resource manager is running as a java process

```
> start-dfs.sh && start-yarn.sh jps
```
4. Run a MapReduce job, as shown in 3.4B

```
> NOTE: You can see the node status, application status, monitor its execution, check logs and much more using the ResourceManager configured at 'localhost:8088' by default. Open up your web browser and type in 'localhost:8088' > NOTE: If you do not have enough space (mostly occurs in a VM instance - min 10GB required), MR job hangs and YARN throws the following error ERROR message:  ACCEPTED: waiting for AM container to be allocated, launched and register with MR. > NOTE: MapReduce jobs may also hang because of incorrect DataNode configuration, or the DataNode not running. The only way to find out for sure is by looking at the logs in the /hadoop/logs/ directory.
```
5. Check output contents, ensure they are correct and then finally shutdown daemons using

```
> stop-dfs.sh && stop-yarn.sh
```

Sources

1. Dev - Hadoop installation tutorial (Main)
2. Apache Hadoop Documentation (Main)
3. MapReduce classpath application not setup in `mapred-site.xml`
4. MapReduce job hangs due to incorrect Yarn configuration (memory issues)