

Zewail City of Science and Technology

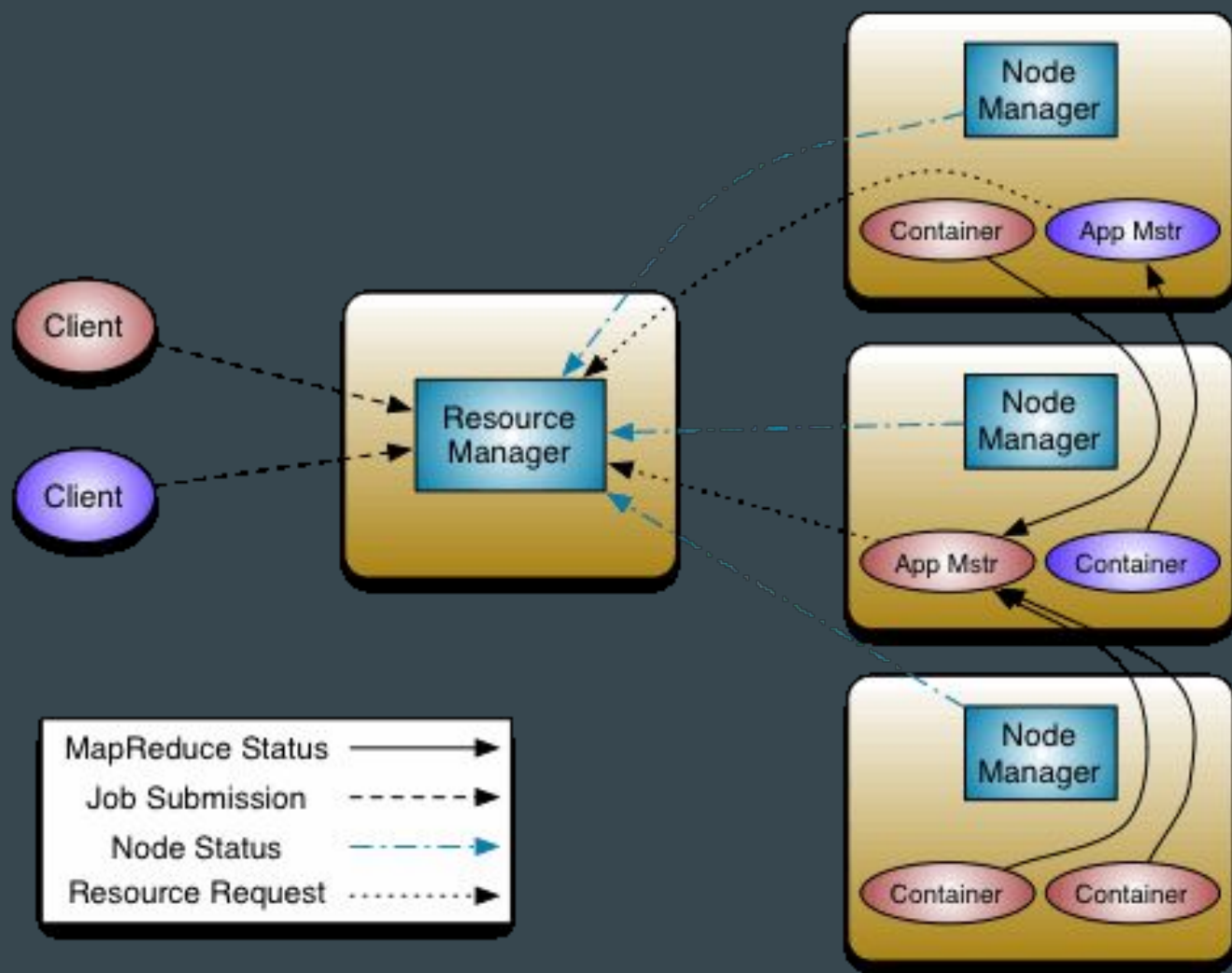
Communications and Information
Engineering (CIE) Program

Big Data Analytics (CIE 427) - Fall 2020



YARN





Hadoop Pseudo-Distributed Operation on YARN

Prerequisites: Shuffle Service

In the file `etc/hadoop/yarn-site.xml`, enable the shuffle service needed by MapReduce,

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

Prerequisites: MapReduce Framework

In the file `etc/hadoop/mapred-site.xml`, set YARN as the default framework for MapReduce jobs,

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

Run the Daemons

```
$ sbin/yarn-daemon.sh start resourcemanager  
$ sbin/yarn-daemon.sh start nodemanager
```

This should start local ResourceManager and NodeManager daemons!
You should be able to connect to the ResourceManager's web interface at
<http://localhost:8088/>!

Pseudo-Distributed Word Count Demo on YARN

Requirement: Pseudo-Distributed Pseudo-PageRank on YARN

Repeat the Pseudo-PageRank requirement from last time, but this time on YARN!

- Run the same MapReduce job after doing the previous YARN configurations and should *just work*.

Acknowledgement

The graph and the material introduced in YARN architecture were adapted from the official [Apache Hadoop documentation](#).

A lot of the material introduced in running Hadoop Pseudo-Distributed Operation on YARN were adapted from the official [Apache Hadoop documentation](#).

Thank you!