

16-mapreduce和yarn的配置

-----成都尚学堂-mr-zeng-----

ResourceManager

NodeManager

-----【在hdfs集群的基础上】-----

e) 配置etc/hadoop/mapred-site.xml : 这里是mapreduce的配置文件

把mapred-site.xml.template 改为 mapred-site.xml

mv mapred-site.xml.template mapred-site.xml

修改内容

vi mapred-site.xml

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

注 : mapreduce.framework.name用于配置mapreduce运算采用的方式-默认是local表示在当前单节点机器运算(默认的方式), yarn表示交给yarn-通过集群运算

f) etc/hadoop/yarn-site.xml

```
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>resourcemanager的地址</value>
</property>
```

注 :

yarn.nodemanager.aux-services : 配置nodemanager的服务列表-默认 (mapreduce.shuffle) 需要改为mapreduce_shuffle-否则会报错

yarn.resourcemanager.hostname : 配置resourcemanager的地址-resourcemanager用于给mapreduce任务分配资源

ps—》nodemanager地址直接读取slave的地址, 与datanode一样!!

-----启动yarn-----

a) 先启动 hdfs (任务需要使用hdfs) -在namenode的节点操作-
start-dfs.sh

b) 再启动yarn (在resourceManager的节点操作)
start-yarn.sh

c)检查各节点的进程-jps
ResourceManager
NodeManager

-----可选配置项-----

配置mapred-site.xml

```
<property>
  <name>mapreduce.job.reduces</name>
```

```
<value>2</value>
</property>
```

mapreduce.job.reduces : reduce的 任务数-默认为1,可提高性能。

Yarn-siste.xml

```
<property>
<name>yarn.log-aggregation-enable</name>
<value>true</value>
</property>
```

yarn.log-aggregation-enable : 启动日志,交给historyserver管理, mapreduce执行的任务日志就会从各节点上移动到hdfs中的logs里,当然需要提前开启任务日志服务器historyserver

- a) 启动hdfs
- b) 启动yarn
- c) 再运行日志服务器

mr-jobhistory-daemon.sh start historyserver

查看任务