1. 启用nodelabel

在yarn-site.xml中添加配置：

*<property>*

*<name>yarn.node-labels.enabled</name> //启用nodelabel*

*<value>true</value>*

*</property>*

*<property>*

*<name>yarn.node-labels.fs-store.root-dir</name> //配置nodelabel持久化存储hdfs路径*

*<value>/hadoop/yarn/node-labels</value>*

*</property>*

在hdfs上会自动创建node-label目录：/hadoop/yarn/node-labels

设置node-label-manager的实现类，配置如下：

*<property>*

*<name>yarn.node-labels.manager-class</name>*

*<value>org.apache.hadoop.yarn.server.resourcemanager.nodelabels.RMNodeLabelsManager</value>*

*</property>*

1. 创建NodeLabel名称及为nm添加nodelabel

*$yarn rmadmin -addToClusterNodeLabels s1,s2*

指定nodemanager上的标签，例如指定fys1.cmss.com:45454的标签为s1

*$yarn rmadmin -replaceLabelsOnNode "fys1.cmss.com:45454,s1"*

*$yarn rmadmin -replaceLabelsOnNode "fys2.cmss.com:45454,s1"*

*$yarn rmadmin -replaceLabelsOnNode "fys3.cmss.com:45454,s2"*

1. 配置队列

在fair-scheduler.xml中为队列指定nodelabel，如下所示：

*<queue name="s1test">*

*<accessibleNodeLabel>s1</accessibleNodeLabel>*

*<maxResources>8096mb,8vcores</maxResources>*

*</queue>*

*<queue name="s2test">*

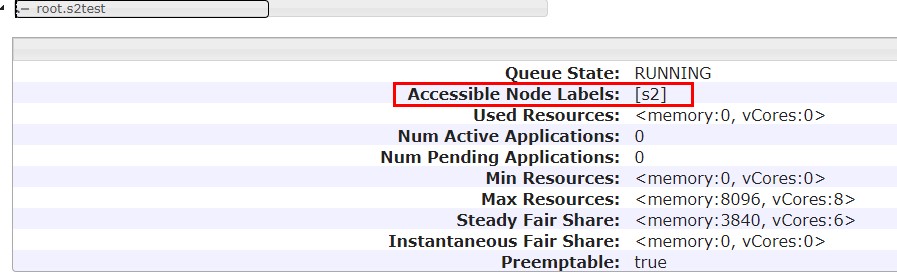
*<accessibleNodeLabel>s2</accessibleNodeLabel>*

*<maxResources>8096mb,8vcores</maxResources>*

*</queue>*

至此节点和队列都添加了nodelabel，查看rm的node页面，可以看到fys1.cmss.com:45454的标签：

查看fairscheduler页面,队列显示的accessibleNodeLable:



查看hdfs中的/hadoop/yarn/node-lable目录，有nodelabel持久化信息：

*$hdfs dfs -cat /hadoop/yarn/node-labels/nodelabel.editlog*

*s2*

*s1*

*fys1.cmss.comဒs1*

*fys2.cmss.comဒs1*

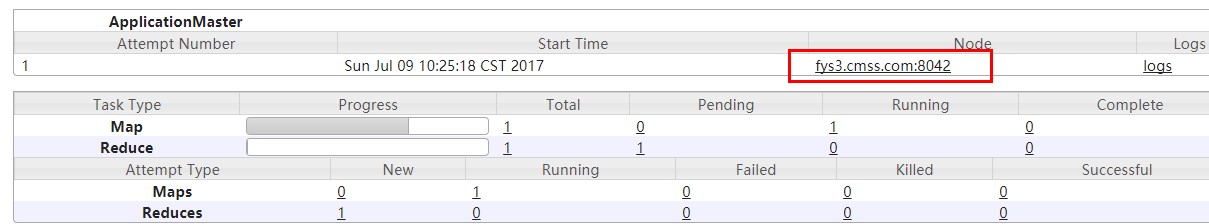
*fys3.cmss.comဒs2*

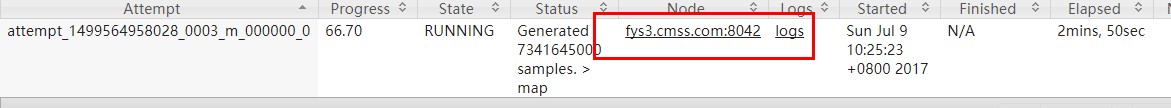
1. 提交程序

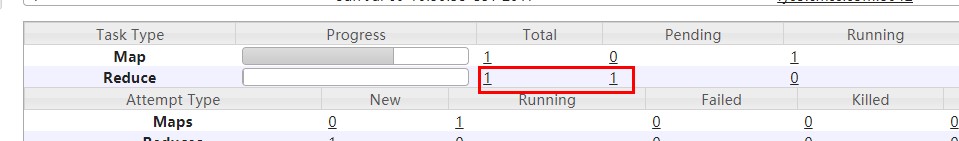
向s2中提交测试程序，对应的队列为root.s2test如下所示：

*$bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.0-bc1.3.5.jar pi -D mapreduce.job.queuename=root.s2test 1 100000*

MRAppMaster及MR Task均运行在fys3.cmss.com主机上（即nodelabel:s2）







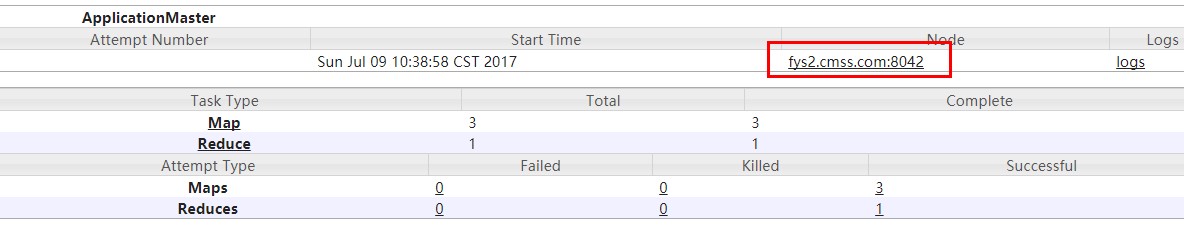
在其他节点资源充足的情况下，Reduce仍然是Pending状态。当Reduce启动时，执行的NM，也是fys3.cmss.com:

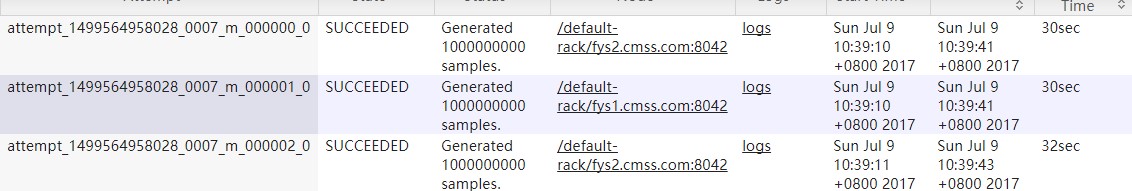


向s1中提交测试程序，对应的队列：

*$bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.0-bc1.3.5.jar pi -D mapreduce.job.queuename=root.s1test 1 100000*

执行的nm，





符合预期

注：

rmadmin nodelabel其他命令：

*yarn rmadmin -addToClusterNodeLabels "<label1>(exclusive=<true|false>),<label2>(exclusive=<true|false>)"*

NOTE: exclusive并不是必须的参数，默认值为true

false是指被nodelabel标签的节点资源不作调度分配。

可以用以下命令删除已创建的nodelabel

|  |
| --- |
| *yarn rmadmin -removeFromClusterNodeLabels "<label1>,<label2>"* |

也可以通过replaceLabelsOnNode命令来删除nodeLabel，只要将Label参数置为空即可。例如，可以通过下列命令来将node-1节点的Label置空。

|  |
| --- |
| *yarn rmadmin -replaceLabelsOnNode "node-1"* |

参考：

http://223.105.0.132:8090/display/BCH/YARN+NodeLabel