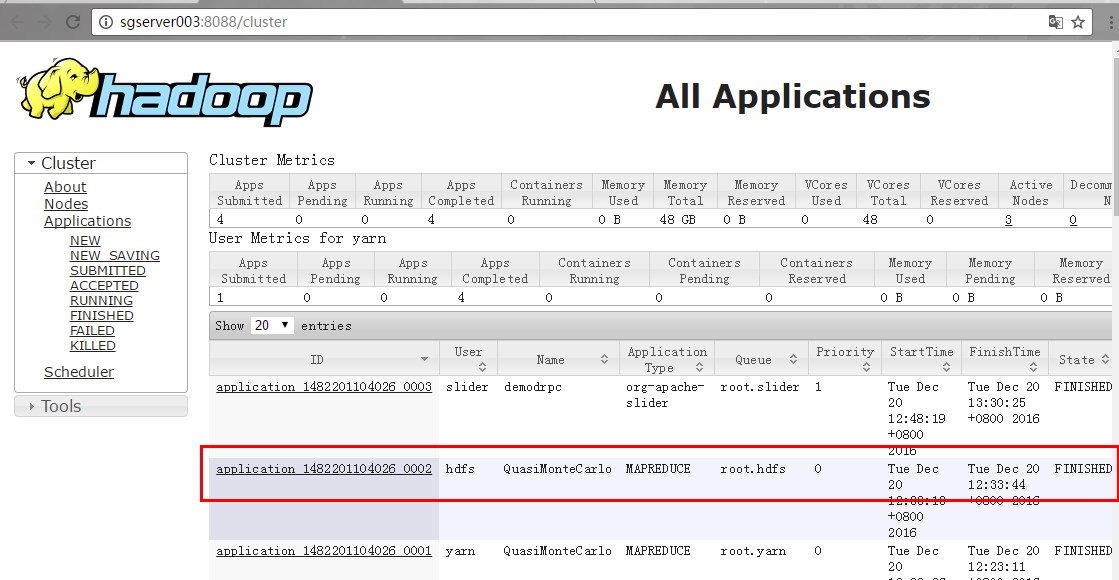
YARN日志收集与查看分析

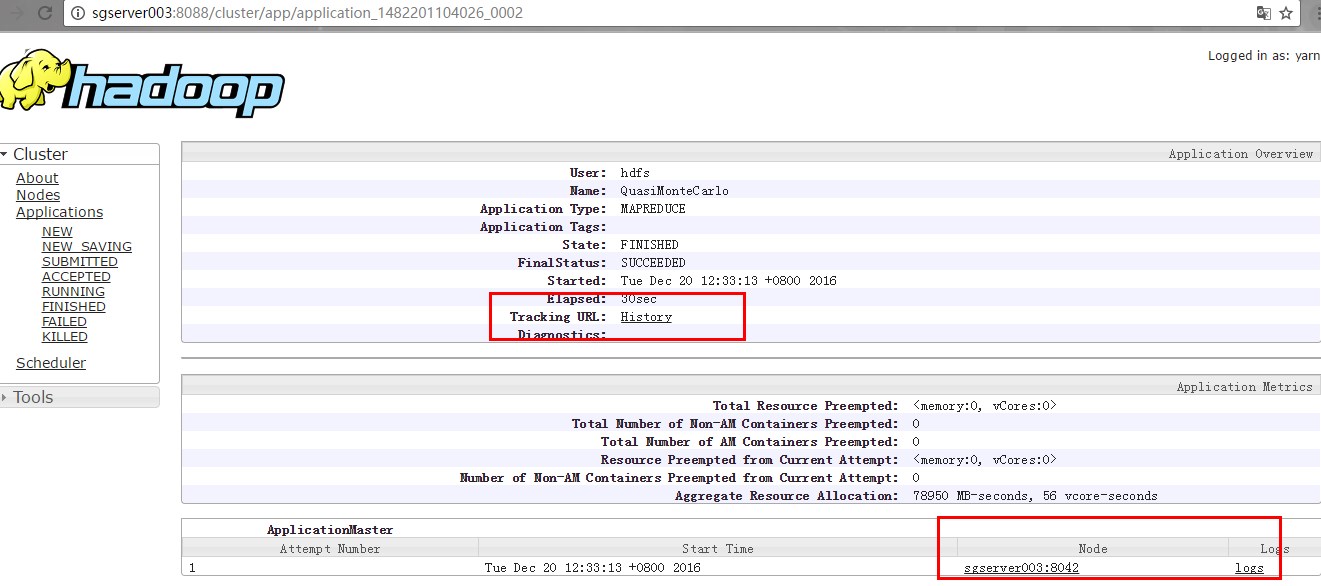
YARN在运行应用后，需要将日志收集，上传到hdfs上，以便以后查看。启用日志分为两类，一是MR job，另外是long-live job的日志。container在运行过程中，可以通过NM:8042 web页面查看，执行结束后，通过yarn的日志聚集功能将上传上传，MR job比较完善，但是long-live job（slider app）还没有对应机制，下面分析MR Job的日志收集与查看功能，并对slider app的日志功能进行优化。

# 1.MR job日志收集与查看

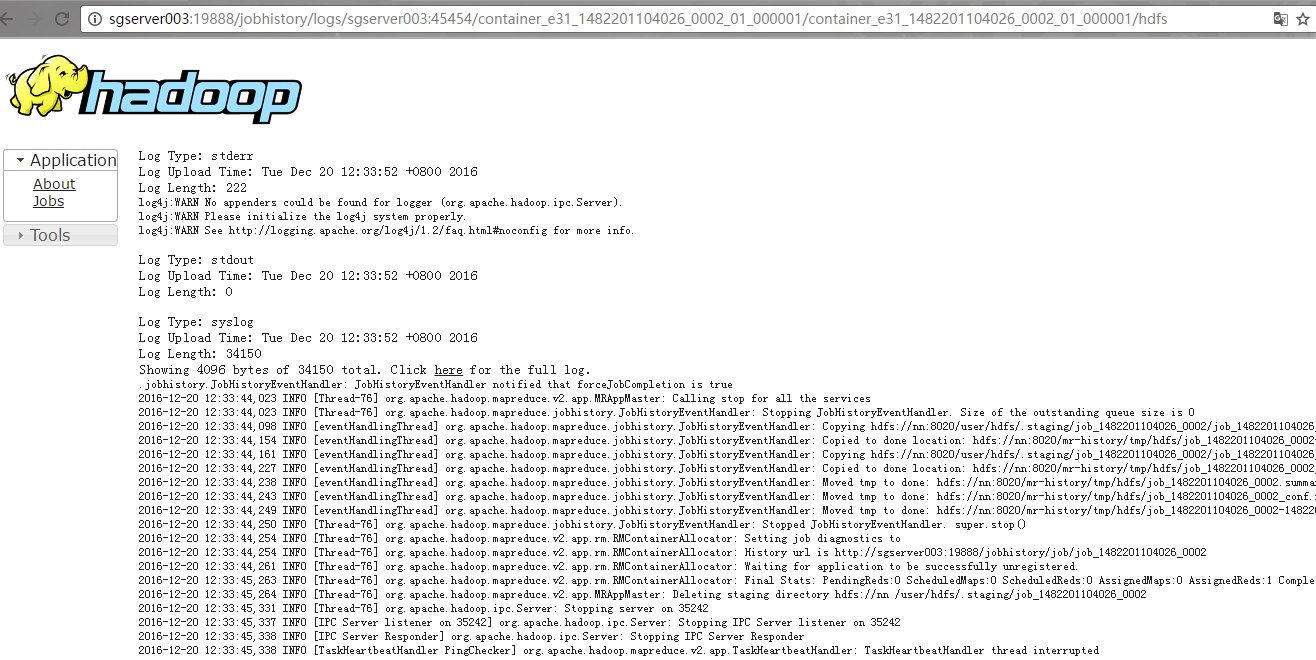
MR job，查看某个Job的日志，进入rm:8088后，如下所示：



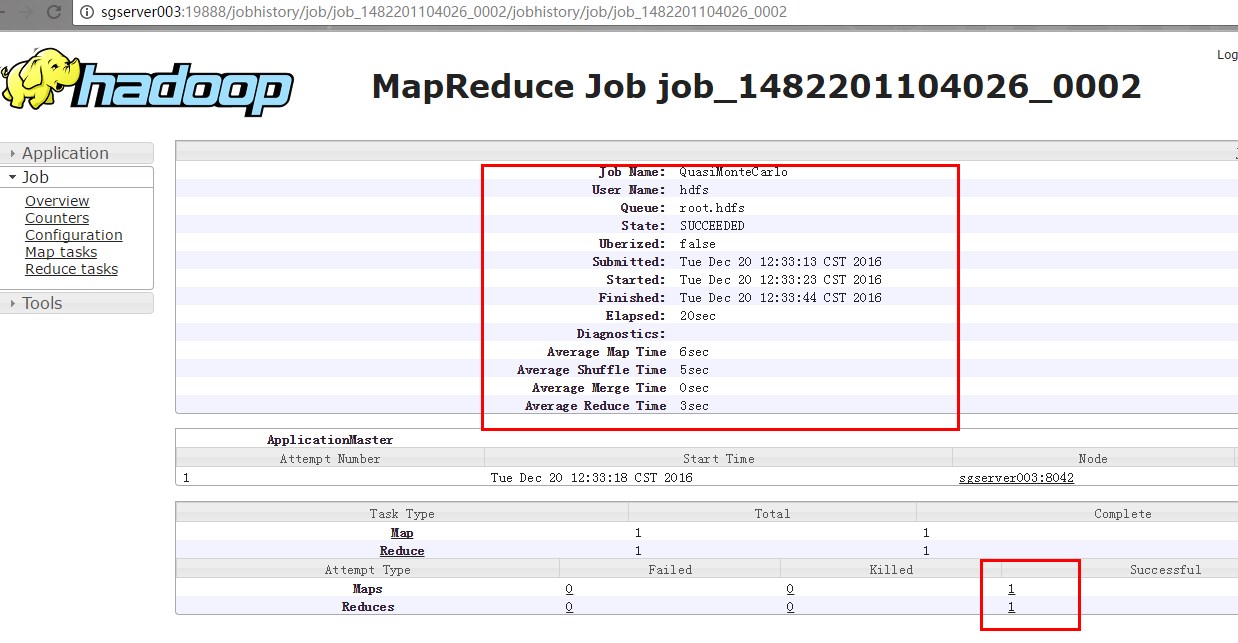
点击app id链接，进入job页面：



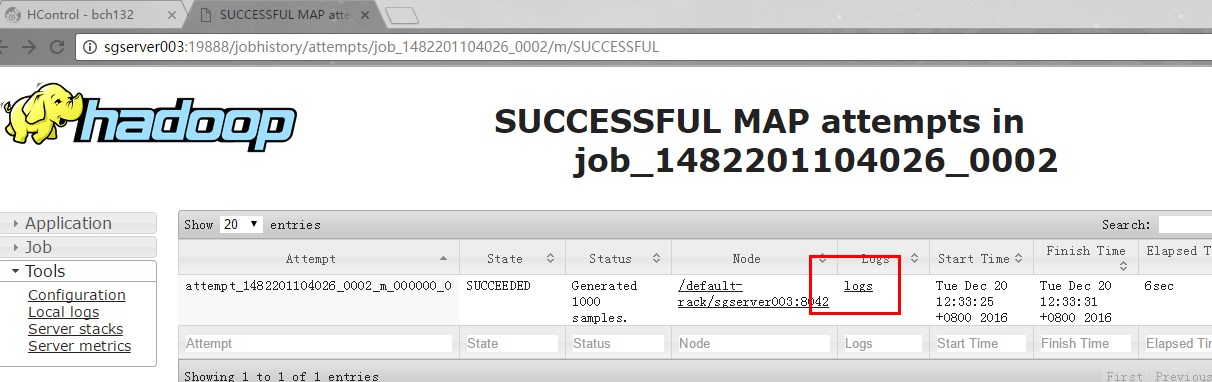
其中sgserver003:8042:logs，是AM的日志，如下所示：



History Tracking URL，为MR的执行统计结果及日志：



进入Map或者Reduce:

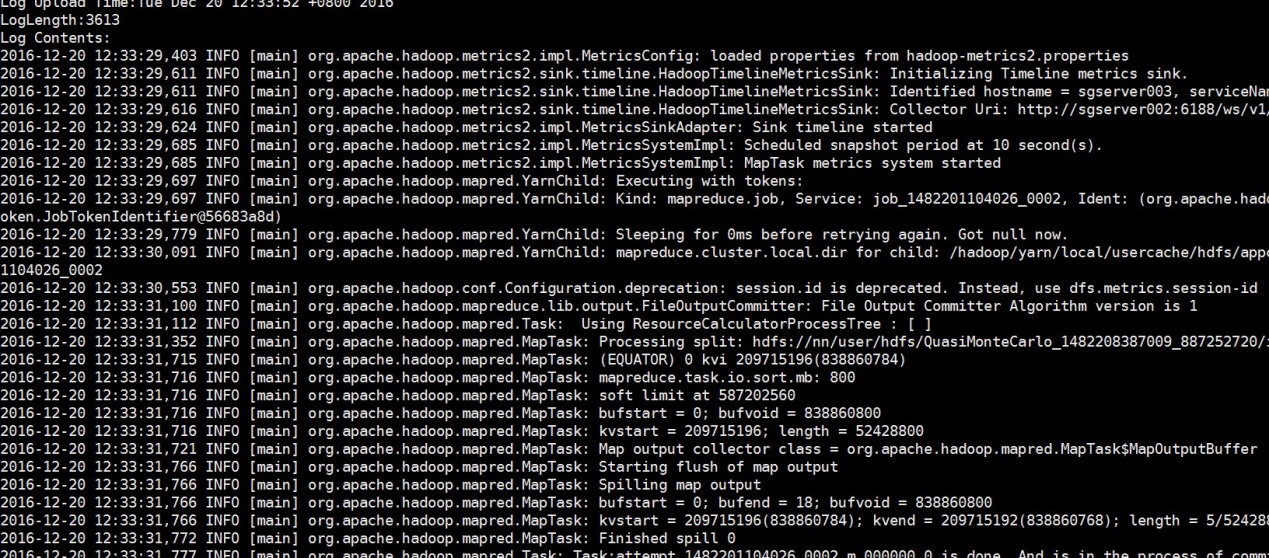


点击即可查看Map/Reduce日志

开启日志聚集后，也可以手动查看，如下所示：

*$bin/yarn logs -applicationId application\_1482201104026\_0002*

查看到的日志如下所示：



## 1.2 Long-live app日志

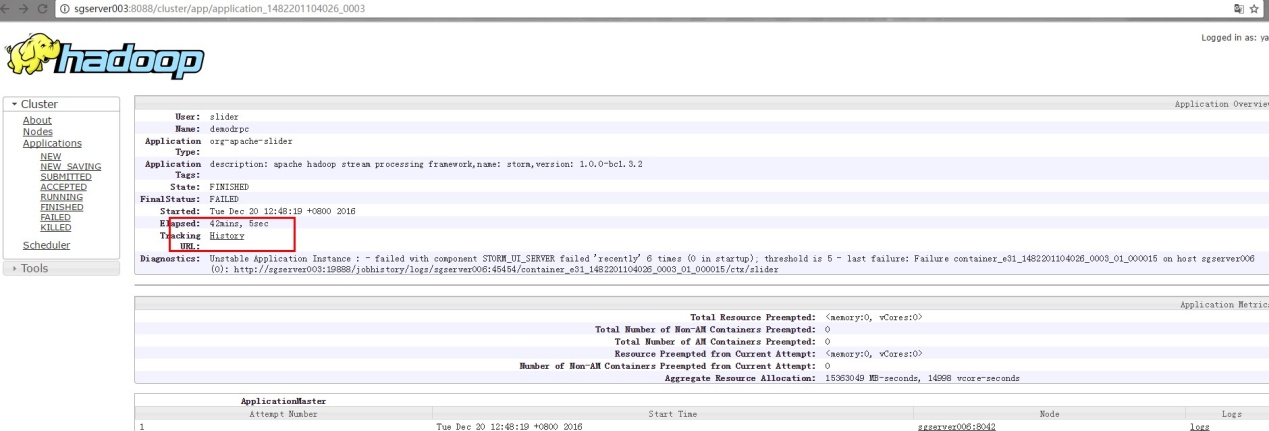
通过yarn logs命令查看日志如下所示：

*$ bin/yarn logs -applicationId application\_1482201104026\_0002*

可以显示所有的日志，但是不友好，通过以下URL也可以看到日志

*http://sgserver003:19888/jobhistory/logs/sgserver006:45454/container\_e31\_1482201104026\_0004\_01\_000003/container\_e31\_1482201104026\_0004\_01\_000003/slider*

其中的container为组件的启动ContainerID，但是没有入口，需要在yarn:8088 页面上提供链接，如下图所示：



红框中的Tracking URL目前是无效的链接，需要进行开发。

## 1.3 YARN日志聚合

配置项：yarn.nodemanager.remote-app-log-dir <= /app-logs 默认配置目录

*[hdfs@sgserver002 ~]$ hdfs dfs -ls /app-logs*

*ambari-qa hadoop /app-logs/ambari-qa*

*hdfs hadoop /app-logs/hdfs*

*kylin hadoop /app-logs/kylin*

*slider hadoop /app-logs/slider*

*yarn hadoop /app-logs/yarn*

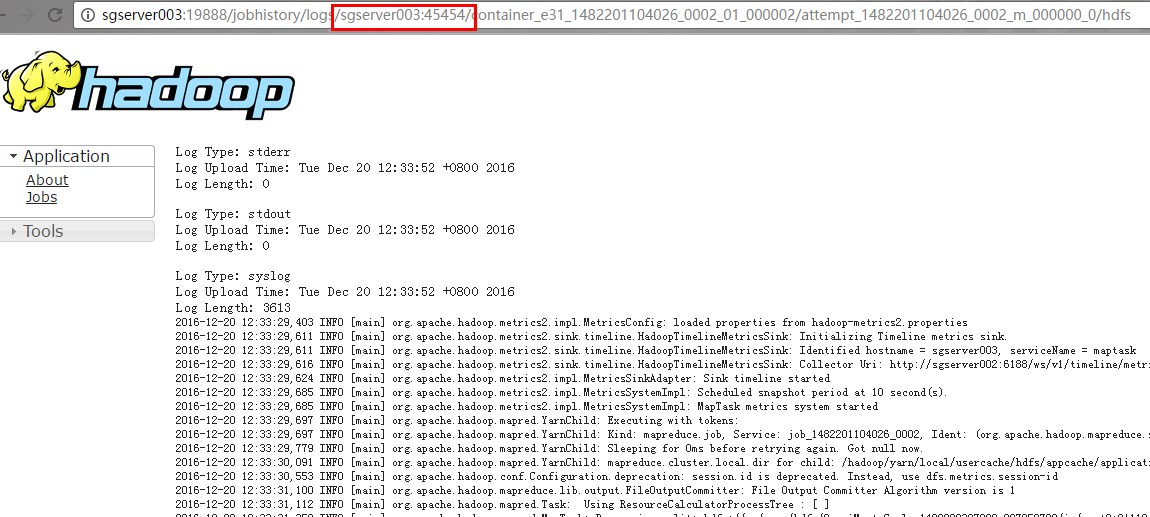
/app-logs下层目录是每个执行用户的执行job

*[hdfs@sgserver002 ~]$ hdfs dfs -ls /app-logs/hdfs/logs/application\_1482201104026\_0002*

*hdfs hadoop /app-logs/hdfs/logs/application\_1482201104026\_0002/sgserver002\_45454*

*hdfs hadoop /app-logs/hdfs/logs/application\_1482201104026\_0002/sgserver003\_45454*

对应的显示页面如下：



显示的是MapReduce的执行日志，在本例中sgserver003:45454也保存了AM的日志，每个节点只会上传一份聚合的文件。

通过yarn logs命令查看日志如下所示：

*$ bin/yarn logs -applicationId application\_1482201104026\_0002*

可以显示所有的日志