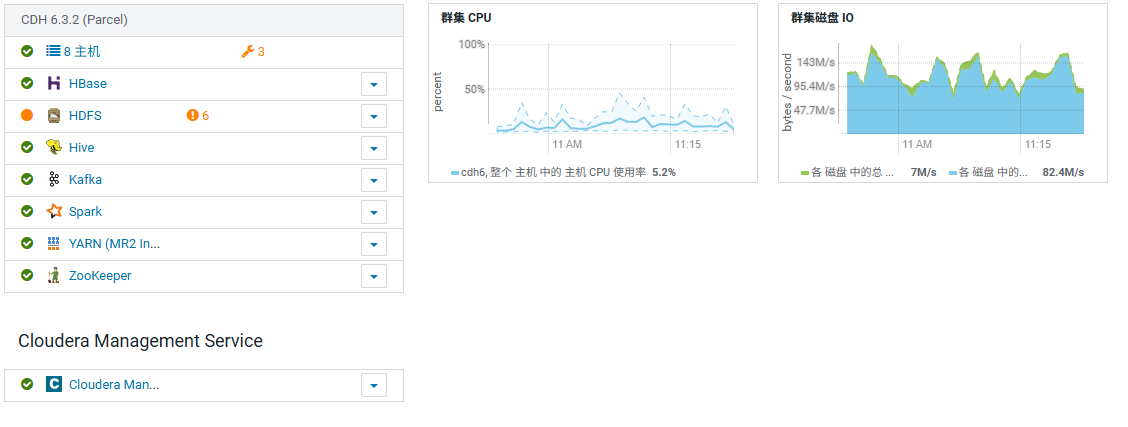
# ClouderaManager平台

[http:// 114.132.169.71:7180/cmf/login](http://gzcloud20:7180/cmf/login)

admin

admin!@#$%^

正常情况应该是这样：



**HDFS：最核心模块，不能出问题，不能随便重启，如果单个出问题，启动起来就行，如果修改参数后需要重启，在实例里面一台台启动**

HBASE：暂时没有业务，可以不用管

HIVE：[Hive Metastore Server](http://gzcloud20:7180/cmf/services/12/instances/159/status) 实例为trino读取数据的关键，不能出问题

KAFKA：保证数据实时入库的，出问题会影响实时数据

SPARK：spark历史日志相关，不用管，出问题也不影响真正的spark入数据

YARN：实时入库和批量入库的资源管理，一般来说不会出问题，需要保证/data目录的占用小于90%(95%)

ZOOKEEPER：一般不会出问题，出问题记得不要整体重启

Cloudera Manager：右侧报表的东西，可以随便重启，不影响数据

**HIVE的METADATE数据保存在一个自建mysql中，这个一定不能出问题，不然会影响hive中的数据**，现在每天会对mysql进行全量备份，注意定时检查磁盘，删除老数据防止磁盘满了。

备份脚本地址：

gz19\_211.159.188.153

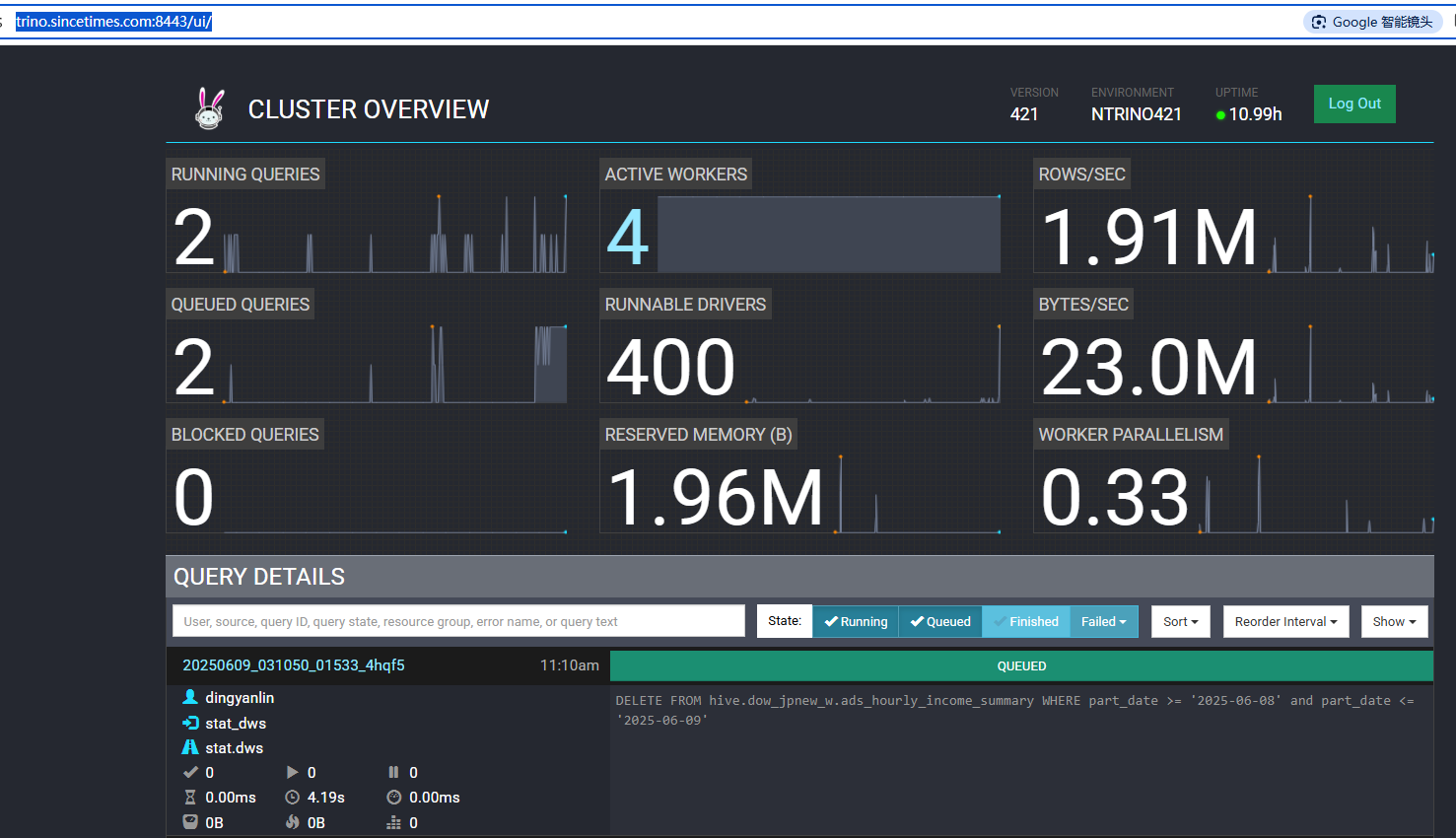
**/data/bigdata/dbbackup/hivedbback.sh**

# Trino管理

Trino状态页面：<https://trino.sincetimes.com:8443/ui/>

taskdeploy

G5dAToqEzv7QK1UM



Trino所有的节点都配置了进程挂掉自动重启脚本

**手动重启需要使用用户trino执行**

Worker节点重启不影响业务，一个一个重启，不要多个同时重启

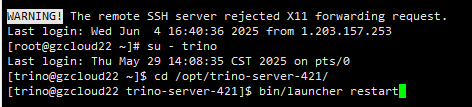
Master节点重启会影响业务使用

重启命令：

[root@gzcloud22 ~]# su - trino

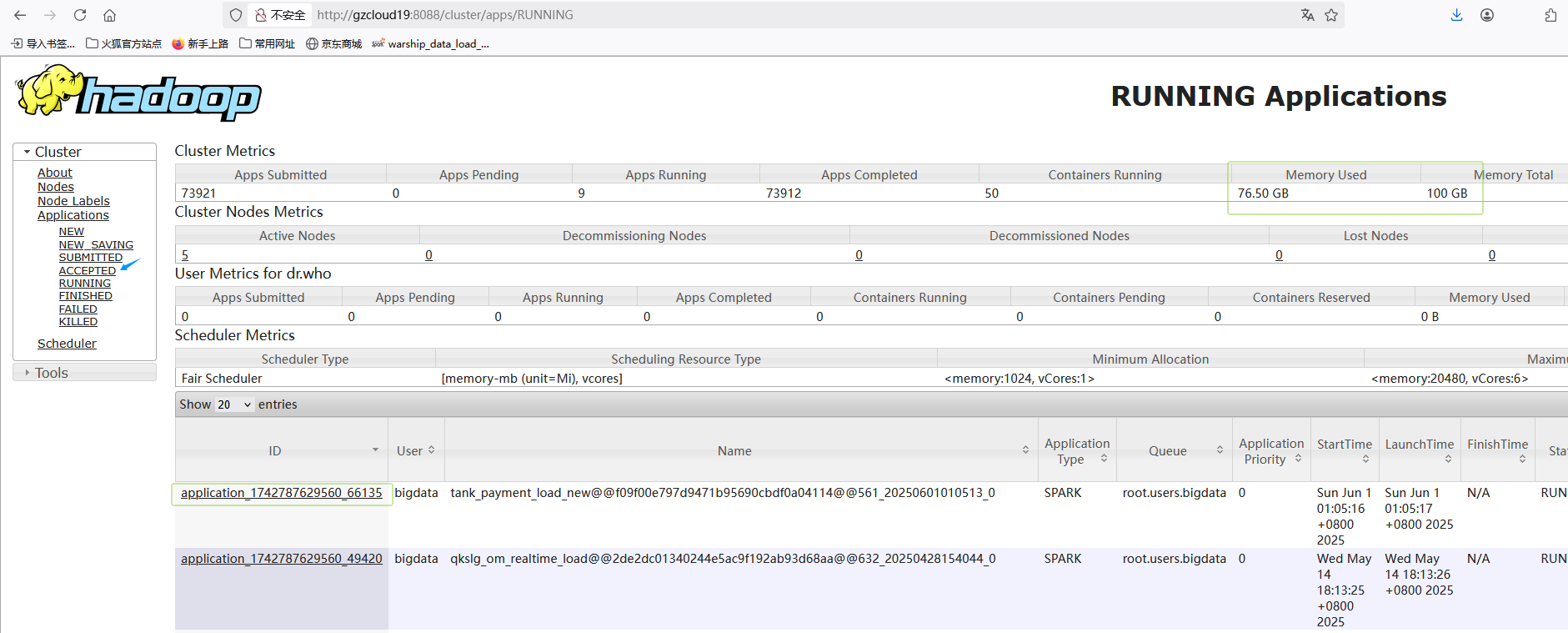
[trino@gzcloud22 ~]$ cd /opt/trino-server-421/

[trino@gzcloud22 trino-server-421]$ bin/launcher restart



# YARN资源状态监控

[http:// 211.159.188.153:8088/cluster/apps/RUNNING](http://gzcloud19:8088/cluster/apps/RUNNING)



1.需要保证/data目录的占用小于90%(95%)

2.需要注意不要有太多的排队任务，长期的大量任务排队（超过10个任务排队半小时以上）可能会导致YARN崩溃

3.手动kill yarn任务命令,bigdata用户下运行：

yarn application –kill application\_1742787629560\_xxxxxx

4.注意：spark streaming任务必须先通过yarn的命令kill，等待2分钟左右，查看linux内进程，如果还存在，再使用linux的kill，直接linux kill会导致kafka的offset写入失败，在后续启动后会出现数据重复录入

# 服务器相关的Grafana监控

<http://123.207.39.39:3100/dashboards>

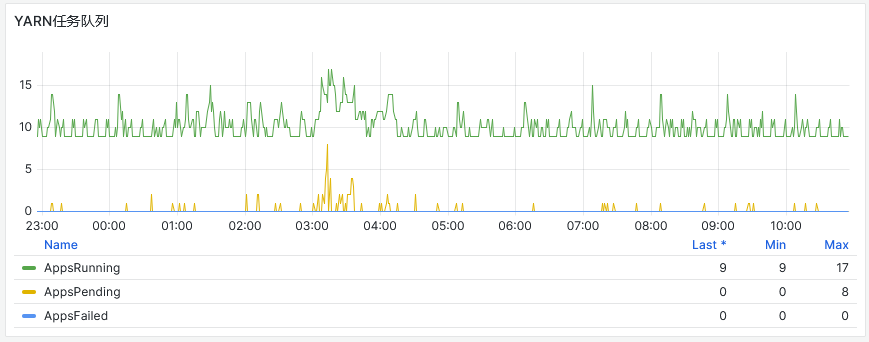
ranhualin

JFhVZbJeYOagCxLKP4IajN7QPKrzwcki



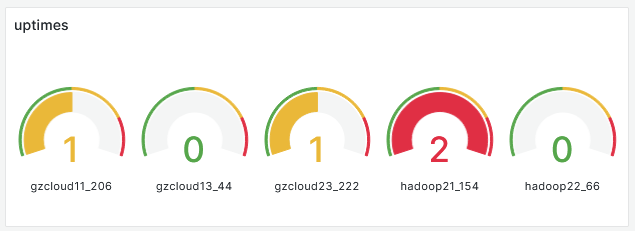
主要关注点：

1.CDH监控



黄线（AppsPending）不应该长时间很多，如果一直在5以上或者10以上(超过半小时)，需要调整yarn的任务运行时间

2.Trino



图形表示这个trino节点在24小时内的重启次数，如果重启太频繁需要进一步跟进调优

3.服务器监控

主要关注磁盘占用

