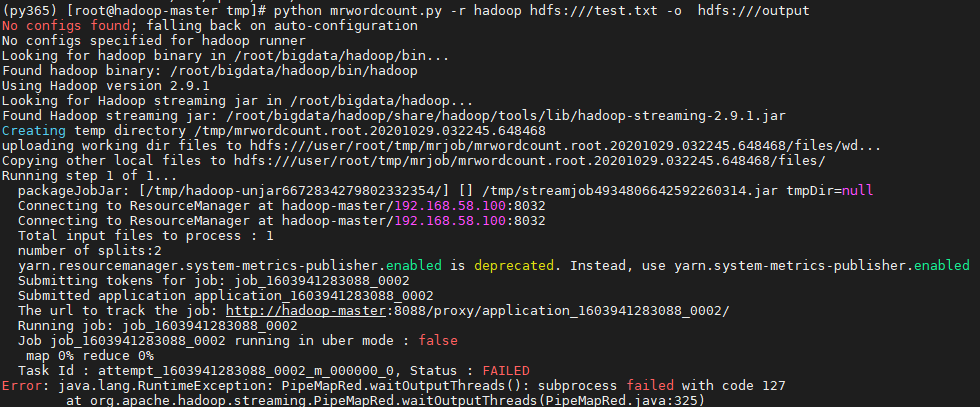
**！！！注意：有很多问题可以通过百度查询得到解决，这些问题不列在本文件。本文件的问题均是难以通过网络查询直接得到的问题以及花费很长时间才自行解决的问题。**

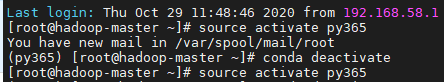
1.hadoop提交任务，要推出canda python环境



退出py365，再运行就可：

source activate py365

conda deactivate



1. Hadoop的重启
2. 删除主节点hadoop/hdfs/目录下面的所有内容，删除hadoop/目录下面的logs目录和tmp目录
3. 删除从节点的hadoop目录
4. 主节点Hadoop/下运行bin/hdfs namenode -format格式化，然后打开hdfs和yarn即可。
5. 如何启动/关闭docker，如何在docker中进入mysql命令行？

启动 systemctl start docker

service docker start

守护进程重启 systemctl daemon-reload

重启docker服务 systemctl restart docker

重启docker服务 service docker restart

关闭docker service docker stop

关闭docker systemctl stop docker

启动mysql：docker start mysql

进入mysql命令行 docker exec -ti mysql bash

使用docker MySQL中可输入中文 docker exec -ti mysql env LANG=C.UTF-8 bash ,每次想要可输入中文的mysql都需要这条命令

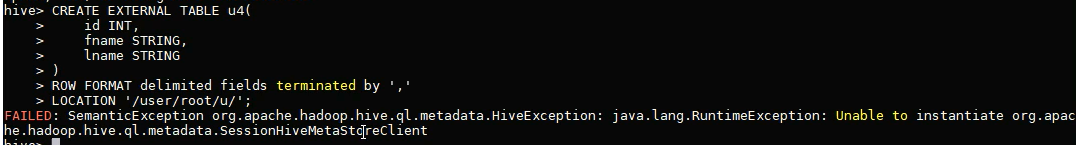
Username:root 密码:password

[root@localhost ~]# docker exec -ti mysql bash

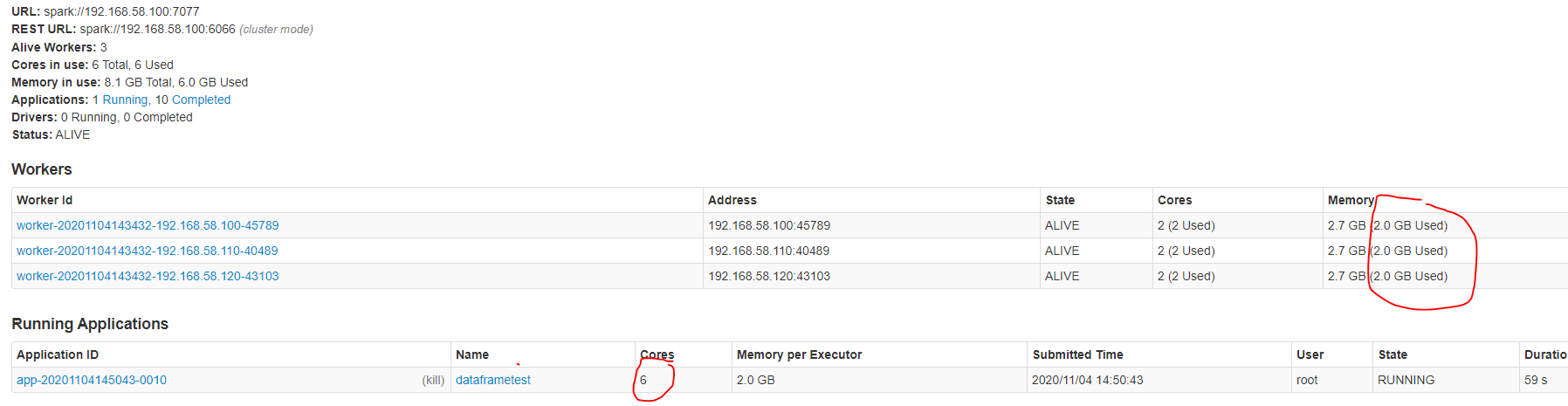
root@eb3dbfb0958f:/# mysql -uroot -p

Enter password: password

4. hive出现下图failed，要输入 hive --service metastore& 重连



1. spark集群启动问题：标红位置是0



查看设置发现标红位置设置了6核，但是虚拟机每台只有2核，因此注释掉此行选用默认值或者设置为2



1. 使用sqoop将sql导入hive，一直running。

关闭spark on yarn，再导入

1. Hive连接不上mysql

sqoop list-databases --connect jdbc:mysql://192.168.58.100:3306/ --username root -P失败

解决方法：Docker启动mysql成功之后，再连；默认导入default数据库中。

1. Spark rdd toDF失败

(1)指定Java路径

import os

JAVA\_HOME = '/root/bigdata/jdk'

PYSPARK\_PYTHON = '/miniconda2/envs/py365/bin/python'

os.environ['JAVA\_HOME'] = JAVA\_HOME

os.environ['PYSPARK\_PYTHON'] = PYSPARK\_PYTHON

os.environ['PYSPARK\_DRIVER\_PYTHON'] = PYSPARK\_PYTHON

(2)另一个案例如下：

cate\_count\_df.write.csv("cate\_count.csv", header=True)

schema = StructType([

StructField("userId", IntegerType()),

StructField("cateId", IntegerType()),

StructField("pv", IntegerType()),

StructField("fav", IntegerType()),

StructField("cart", IntegerType()),

StructField("buy", IntegerType())

])

# 从hdfs加载CSV文件,这种加载有一个好处：加载出的新df可以指定新的schema

# 我尝试过如果指定，pv-buy的数据类型是bigint(即longtype)，这种类型在下面toDF那步会报错，转换成int就不会

cate\_count\_df = spark.read.csv("/cate\_count.csv", header=True, schema=schema)

1. show显示 不缩减 truncate=False
2. spark在hadoop集群上运行结巴分词时，要保证每台电脑都有词典，否则就会报错，找不到：no such files or directory。
3. 如何在docker-mysql中存入.sql文件，并运行？

本地文件例如area.sql文件，复制到docker中

docker cp 文件路径/area.sql mysql:/tmp/area.sql

进入mysql容器中确保文件已经复制过来了

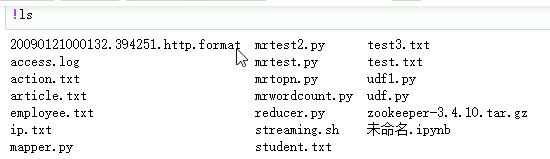
docker exec -it mysql bash

ls -l

运行文件

source /tmp/area.sql

1. Centos linux打开windows系统的jupyter notebook之后，想要在notebook上运行linux命令，需要在命令前加！，如下图所示：



1. Centos linux 以及linux打开的jupyter notebook不显示中文

vim /etc/profile

添加下面三行：



保存之后source /etc/profile

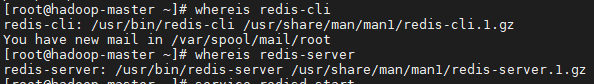
1. 如何查看linux是否安装了redis

(1)rpm -qa|grep redis



(2)查看安装的位置 rpm -ql redis-3.2.12-2.el7.x86\_64

(3) whereis redis-cli 和 whereis redis-server 出现下图，说明安装好了：



(4)检测后台进程是否存在 ps -ef |grep redis

(5)检测6379端口是否在监听 netstat -lntp | grep 6379

(6)使用redis-cli客户端检测连接是否正常

redis-cli

127.0.0.1:6379> keys \*

(empty list or set)

127.0.0.1:6379> set key "hello world"

OK

127.0.0.1:6379> get key

"hello world"

(7)停止redis:

使用客户端redis-cli

shutdown

因为Redis可以妥善处理SIGTERM信号，所以直接kill -9也是可以的

kill -9 PID

1. redis问题 报错：

-DENIED Redis is running in protected modebecause protected mode is enabled, no bind address was specified, noauthentication password is requested to clients.

In this mode connections areonly accepted from the loopback interface. If you want to connect from externalcomputers to Redis you may adopt one of the following

solutions: 1) Justdisable protected mode sending the command 'CONFIG SET protected-mode no' fromthe loopback interface by connecting to Redis from the same host

the server isrunning, however MAKE SURE Redis is not publicly accessible from internet ifyou do so. Use CONFIG REWRITE to make this change permanent.

2) Alternativelyyou can just disable the protected mode by editing the Redis configurationfile, and setting the protected mode option to 'no', and then restarting theserver.

3) If you started the server manually just for testing, restart it withthe '--protected-mode no' option.

4) Setup a bind address or an authenticationpassword. NOTE: You only need to do one of the above things in order for theserver to start accepting connections from the outside.

解决方案：

(1)修改redis服务器的配置文件 vi redis.conf

注释以下绑定的主机地址

# bind 127.0.0.1

(2)修改redis服务器的参数配置

修改redis的守护进程为no ，不启用

127.0.0.1:6379> config set daemonize "no"

OK

修改redis的保护模式为no，不启用

127.0.0.1:6379> config set protected-mode "no"

OK

1. 清除redis中所有的数据

打开Redis客户端redis-cli 输入命令 ：

flushall 可以清除所有数据库的值

flushdb 清除当前数据库的值

1. redis的数据存在哪里？

先进入客户端 redis-cli

再运行 config get dir

本虚拟机在 /var/lib/redis下的dump.rdb –>持久化

1. spark运行在yarn，spark on yarn，如何优雅的kill基于yarn的Spark [Streaming]程序

yarn application --list

yarn application -kill application\_1518954379926\_0063