开启pyspark

C:\Users\sherrytima\AppData\Roaming\Tencent\Users\540310916\QQ\WinTemp\RichOle\$%@{AZ$O)]MG@UCJ0XGSA}T.png2.0.0

Locate查找文件位置

Vim修改文件

Cat 查看文件

Cp 复制 cp ~/.bashrc /root/mt

Rm 移除

Mv 移动档案

**配置spark环境：**

export JAVA\_HOME=/opt/Beaver/jdk

export SPARK\_HOME=/opt/Beaver/spark

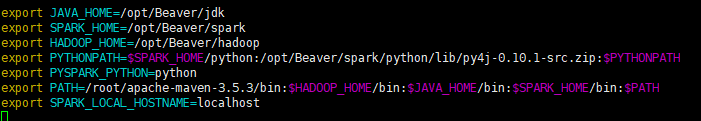
export HADOOP\_HOME=/opt/Beaver/hadoop

export PYTHONPATH=$SPARK\_HOME/python:/opt/Beaver/spark/python/lib/py4j-0.10.1-src.zip:$PYTHONPATH

export PYSPARK\_PYTHON=python

export PATH=/root/apache-maven-3.5.3/bin:$HADOOP\_HOME/bin:$JAVA\_HOME/bin:$SPARK\_HOME/bin:$PATH

export SPARK\_LOCAL\_HOSTNAME=localhost



**添加本地文件到hdfs目录**

hadoop fs -put /root/mt hdfs://172.17.16.5:9000/user/root/

加内存打开方式：

spark-submit --driver-memory 4G /root/mt-code/LR.py

C:\Users\sherrytima\AppData\Roaming\Tencent\Users\540310916\QQ\WinTemp\RichOle\0185A3_Y2E8JGBON@090{QS.png

C:\Users\sherrytima\AppData\Roaming\Tencent\Users\540310916\QQ\WinTemp\RichOle\@_M77LR$(J4UYRL]KA34]17.png

C:\Users\sherrytima\AppData\Roaming\Tencent\Users\540310916\QQ\WinTemp\RichOle\W(M{(QSYXJ6T7}U_$DILPCK.png

key: 'sha1:1fee8f3e3641:4bd9f782de5b8e81cf4f1eb2b7244955094ec5ca'

C:\Users\SHERRY~1\AppData\Local\Temp\1532325910(1).png

c.NotebookApp.ip='\*'  
c.NotebookApp.password = u'sha:ce...刚才复制的那个密文'  
c.NotebookApp.open\_browser = False   
c.NotebookApp.port =8888 #随便指定一个端口

开启jupyter：

C:\Users\SHERRY~1\AppData\Local\Temp\1532325977(1).png

<http://118.25.178.40:8888>

登陆mt用户

su – mt

12345678

Sudo

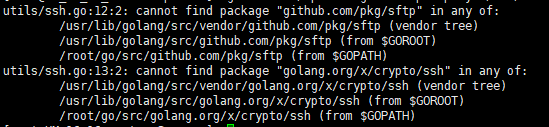
C:\Users\SHERRY~1\AppData\Local\Temp\1532590863(1).png

C:\Users\SHERRY~1\AppData\Local\Temp\1532590948(1).png

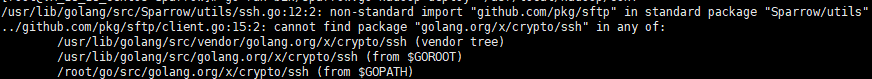
Issue:

C:\Users\SHERRY~1\AppData\Local\Temp\1532604014(1).png

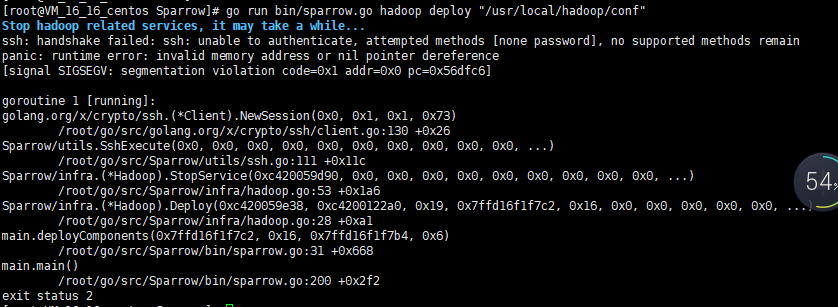
解决：Sparrow要放在GOPATH的src文件夹下



解决：缺两个包



解决：Sparrow要放在GOPATH的src文件夹下，不能放在GOROOT目录下的src文件下



解决：

更改slaves.custom文件 改成自己主机ip的master，及其他slaves

/usr/local/hadoop/conf，指定的这个目录下要有env和slaves.custom文件

另外hadoop部署时要初始化文件，要把utils.InitEnv(slaves, sparrowEnv) 加到infra/hadoop.go里面

C:\Users\SHERRY~1\AppData\Local\Temp\1533020512(1).png

Stop Hadoop时卡住：

**TPCx-BB:**

ERROR

Yarn command not found:

C:\Users\SHERRY~1\AppData\Local\Temp\1533109840(1).png

ERROR

Hdfs doesn’t exit：

重装了一下hadoop

之前安装的hive没有把conf文件移入hive文件目录下

C:\Users\SHERRY~1\AppData\Local\Temp\1533281735.png

ERROR

:line 15: /root/spark-git/bin/spark-sql: No such file or directory

修改spark\_sql下的conf BINARY变量

ERROR

C:\Users\SHERRY~1\AppData\Local\Temp\1533296108(1).png

加了PATH

ERROR（Q5）

C:\Users\SHERRY~1\AppData\Local\Temp\1533537366(1).png

ERROR（Q2）

重复建bigbench：