spark异常处理

## 异常：

java.io.IOException: Could not locate executable null\bin\winutils.exe in the Hadoop binaries.

解决方案：

<https://blog.csdn.net/baidu_19473529/article/details/54693523>

## 异常：

java.lang.NoClassDefFoundError：org / apache / spark / Logging

解决方案：

因为spark-streaming-kafka-0-8\_2.11版本太低，需要换到2.1.0

## 异常：

java.lang.UnsatisfiedLinkError: org.apache.hadoop.util.NativeCrc32.nativeComputeChunkedSumsByteArray(II[BI[BIILjava/lang/String;JZ)V

解决方案：

由于hadoop.dll 版本问题出现的，这是由于hadoop.dll 版本问题，2.4之前的和之后的需要的不一样，需要选择正确的版本(包括操作系统的版本),并且在 Hadoop/bin上将其替换。

我的hadoop是2.7.5的，我之前用的是1.2的hadoop.dll,后来根据网上的说法换了hadoop.dll,找了2.6的还是不行，原来我找的2.6的是32位操作系统的，后来幸运找到了2.6的64位的hadoop.dll。

Hadoop的在windows系统中运行时需要添加win系统的支持,最开始找的版本比较老换成新的dell文件即可

## 异常：

Permission denied: user=administrator, access=WRITE, inode="/":root:supergroup:drwxr-xr-x

解决方案：

总结下来解决办法大概有三种：

1、在系统的环境变量或java JVM变量里面添加HADOOP\_USER\_NAME，这个值具体等于多少看自己的情况，以后会运行HADOOP上的Linux的用户名。（修改完重启eclipse，不然可能不生效）

2、将当前系统的帐号修改为hadoop

3、使用HDFS的命令行接口修改相应目录的权限，hadoop fs -chmod 777 /user,后面的/user是要上传文件的路径，不同的情况可能不一样，比如要上传的文件路径为hdfs://namenode/user/xxx.doc，则这样的修改可以，如果要上传的文件路径为hdfs://namenode/java/xxx.doc，则要修改的为hadoop fs -chmod 777 /java或者hadoop fs -chmod 777 /，java的那个需要先在HDFS里面建立Java目录，后面的这个是为根目录调整权限。

4、右击计算机-->管理--->用户和组-->admin。。。改为与Linux下集群主机名相同

## 异常

When running with master ‘yarn’ either HADOOP\_CONF\_DIR or YARN\_CONF\_DIR must be set in the environment

解决方案：

编辑$SPARK\_HOME/conf/spark-env.sh文件

hadoop@master:~$ vi spark-1.6.0-bin-hadoop2.4/conf/spark-env.sh

加入以下行：

HADOOP\_CONF\_DIR=/home/hadoop/hadoop-2.4.0/etc/hadoop/

然后，将集群上的这个文件都更新。

spark-submit的提交方式：

<https://www.cnblogs.com/weiweifeng/p/8073553.html>

实例：bin/spark-submit --class com.objecteye.spark.Kafka2Redis --master yarn --deploy-mode cluster --driver-memory 1g --executor-memory 1g --executor-cores 1 --queue thequeue /home/shiyu2018/face\_recognition/SparkStreamDemo-1.0-SNAPSHOT.jar

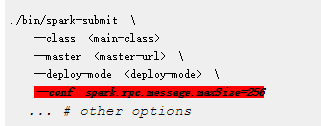
## 异常

Please try specifying another one using the -encoding option

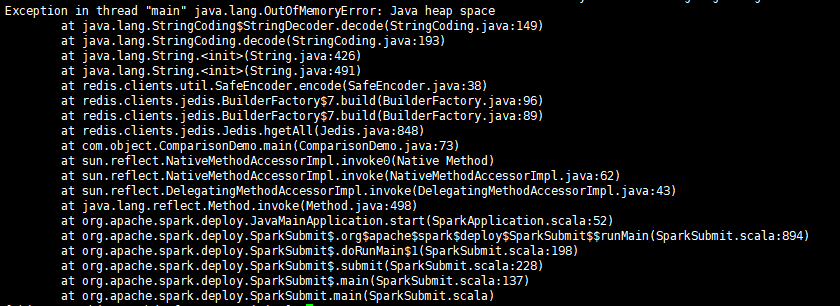
## 异常

Exception in thread "main" org.apache.spark.SparkException: Job aborted due to stage failure: Serialized task 0:0 was 166286923 bytes, which exceeds max allowed: spark.rpc.message.maxSize (134217728 bytes). Consider increasing spark.rpc.message.maxSize or using broadcast variables for large values.

解决方案：



## 异常



解决方案：

$SPARK\_HOME/conf/spark-env.sh 修改export SPARK\_WORKER\_MEMORY=2g的内存大小，同时参考异常7.