

cdh6 oozie调度shell中文乱码

解决方法

在文件第一行加入

```
export LANG=zh_CN.UTF-8
```

在spark中

老集群跑的原始代码：`val e = md5.digest(encryptStr.getBytes())` 这个是没有问题的

新集群上跑这个有问题，中文字符串的`.getBytes()`出来的结果是???, 编码问题，代码中要加上

```
val e = md5.digest(encryptStr.getBytes(Charset.forName("utf-8")))
```

在spark 配置里加上，解决根本问题

```
--conf 'spark.driver.extraJavaOptions=-Dfile.encoding=UTF-8'
```

```
--conf 'spark.executor.extraJavaOptions=-Dfile.encoding=UTF-8'
```

