大数据调度框架：Azkaban

实验要求

【作业要求】：

1、安装azkaban

2、打开azkaban网站门户

3、登录网站门户，并将之前的查询日志条数及用户数的程序转为azkaban任务

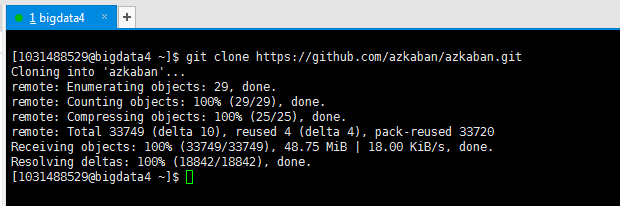
4、执行azkaban任务

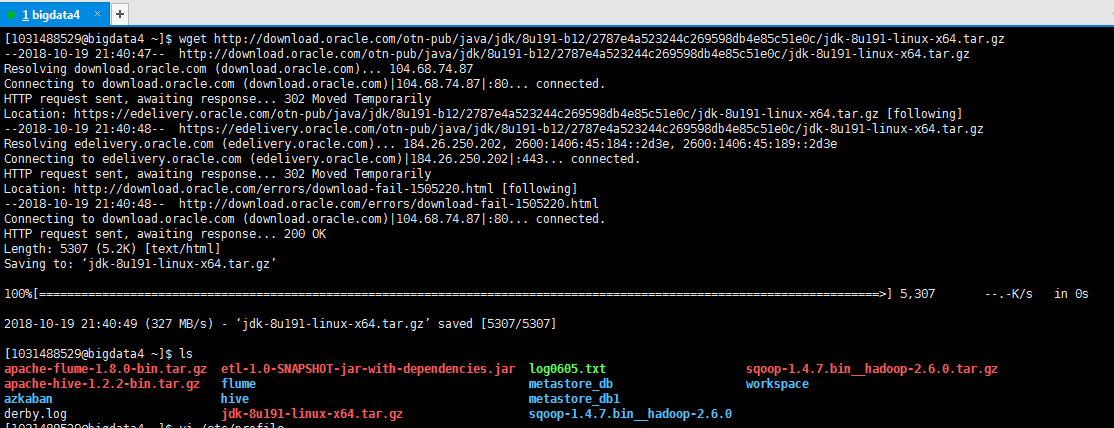
5、将过程截图写入实验报告，并提交附件

操作步骤

1，安装azkaban

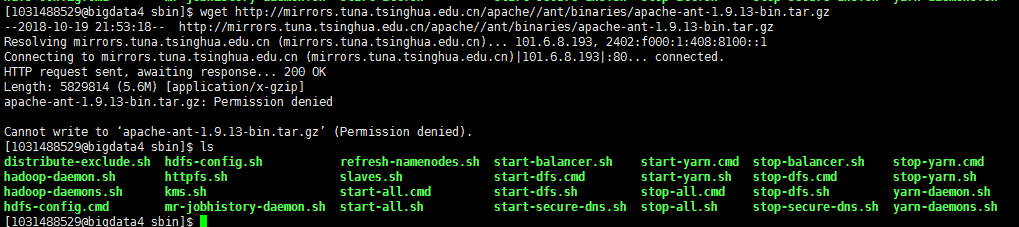
1)下载azkaban

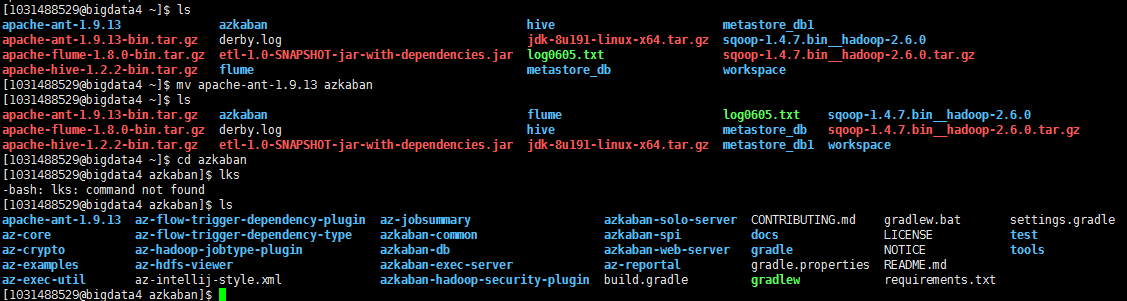




2)下载apache-ant，并解压、移动到azkaban目录下

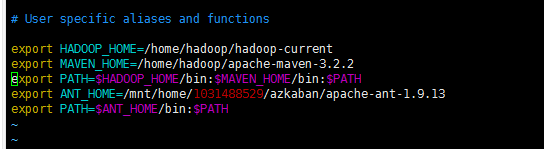
wget http://mirrors.tuna.tsinghua.edu.cn/apache//ant/binaries/apache-ant-1.9.13-bin.tar.gz





3)配置环境变量

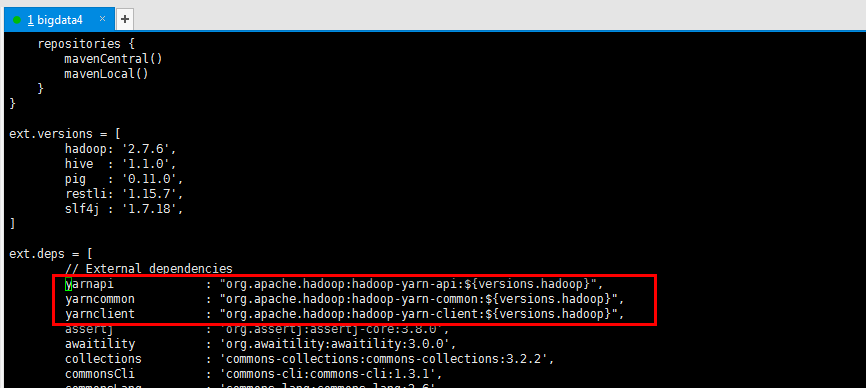




配置azkaban编译参数：

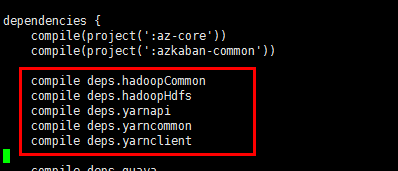


增加红框3行



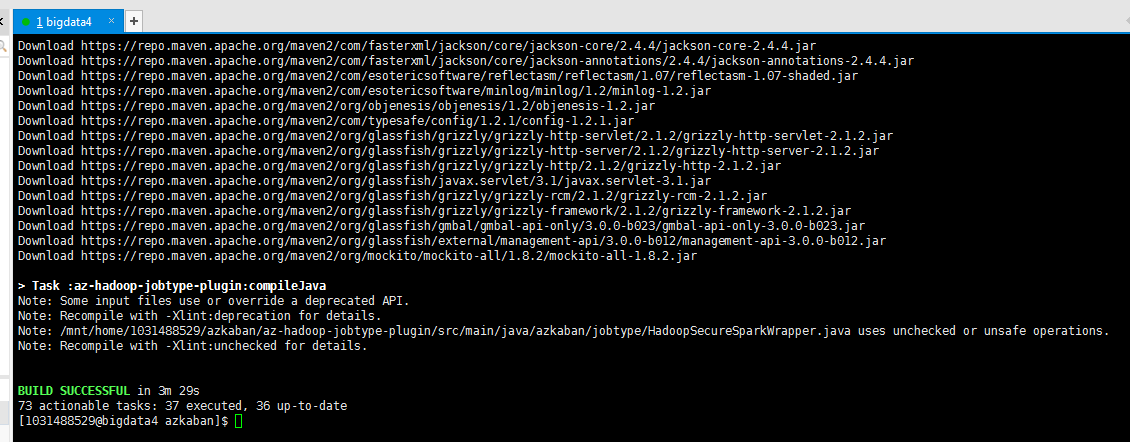


增加红框5行



4)编译azkaban

[1031488529@bigdata4 azkaban]$ ./gradlew build -x test



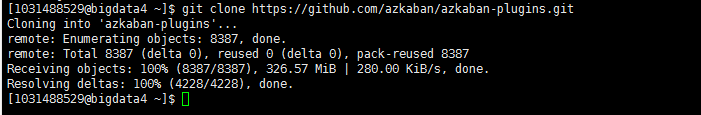
检查ant版本

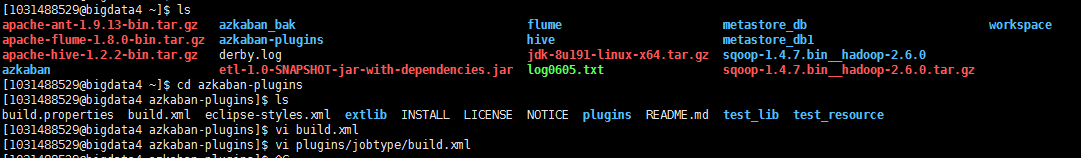


Azkaban编译完成

5)配置plugins

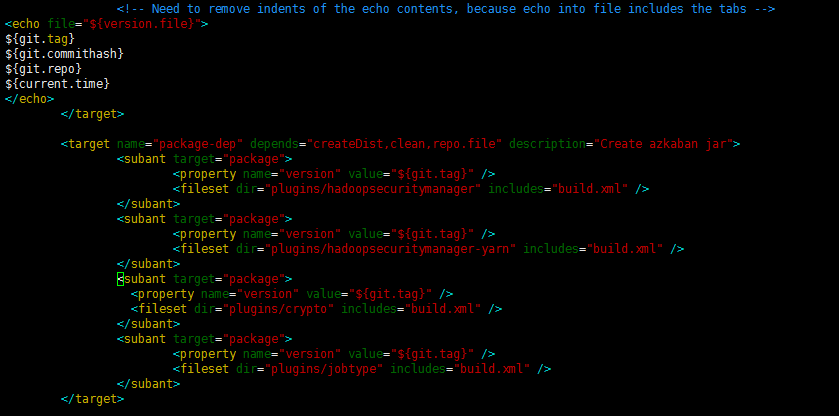
下载:git clone https://github.com/azkaban/azkaban-plugins.git

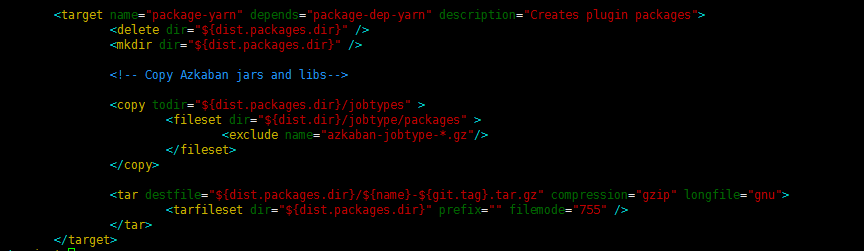




编辑build.xml，这个版本没有java viewer方面的配置，没有需要改的地方

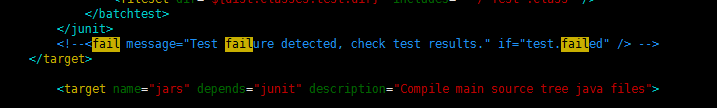




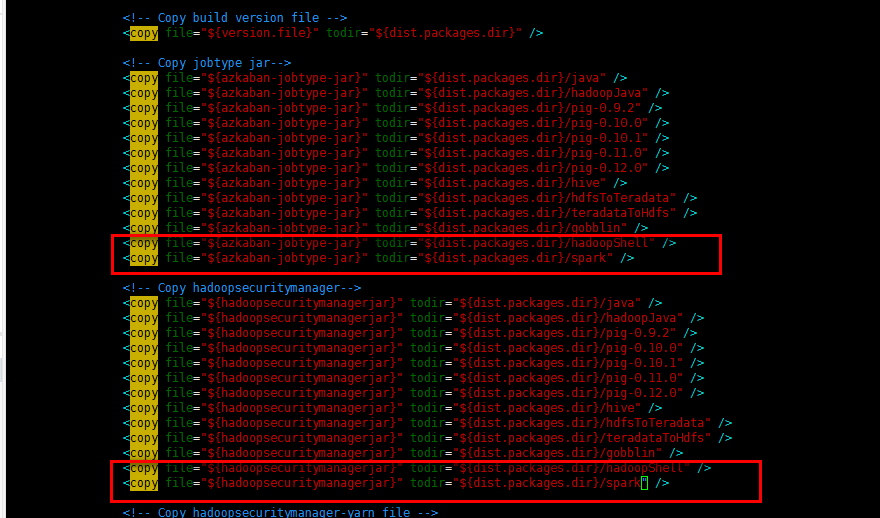


编辑azkaban-plugins/plugins/jobtype/build.xml

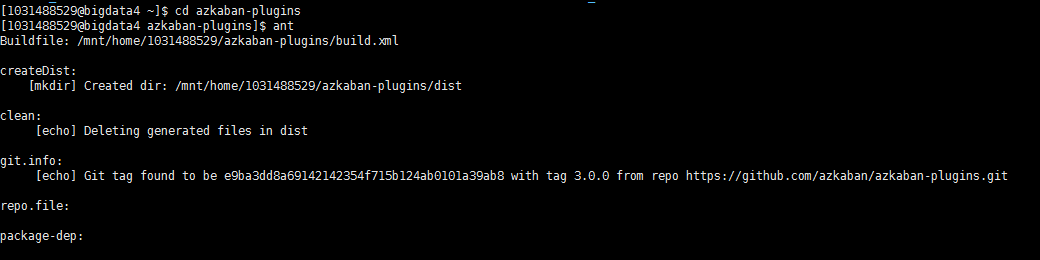
取消失败检查

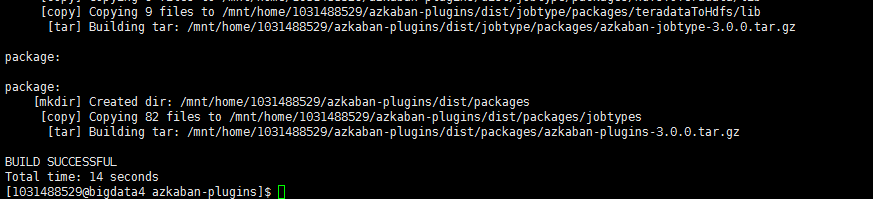


增加红框4行，给hadoopShell和spark增加jar包



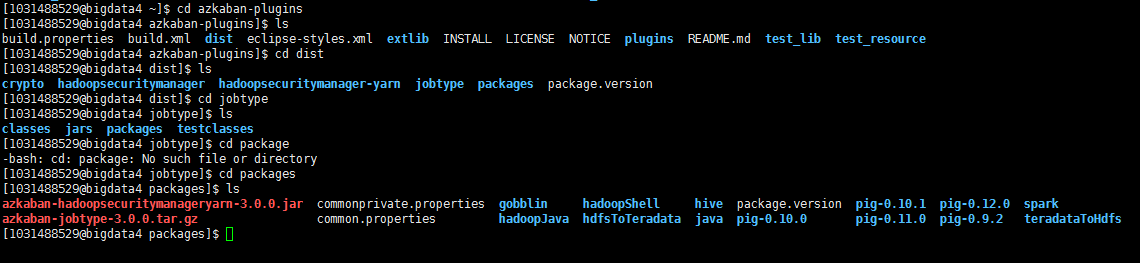
ant编译





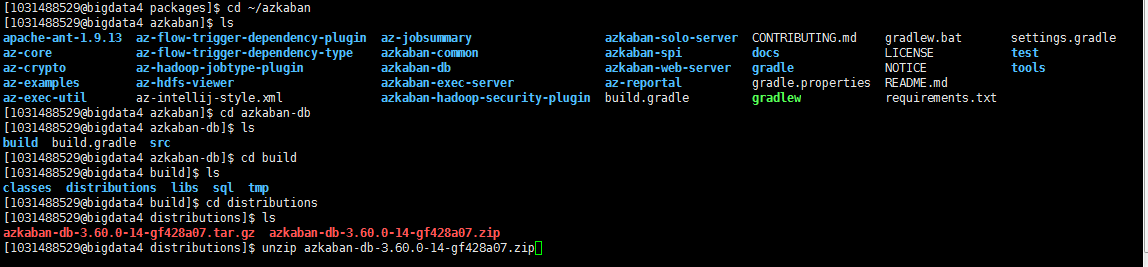
编译成功

6)查看打包结果

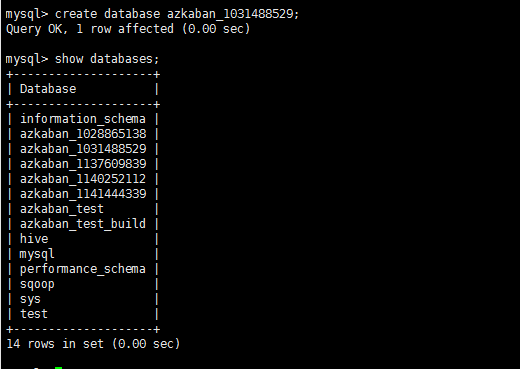


7）创建数据库和表

从azkaban-db/build/distributions目录下获取azkaban-db….zip压缩包，里面包括配置plugins所需要的表和数据所对应的mysql脚本，unzip解压出来供使用

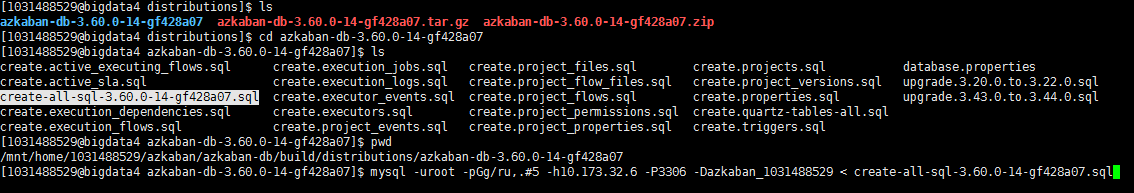


建立自己的数据库azkaban\_1031488529，后面配置web和exec的properties文件时需要指向这个库

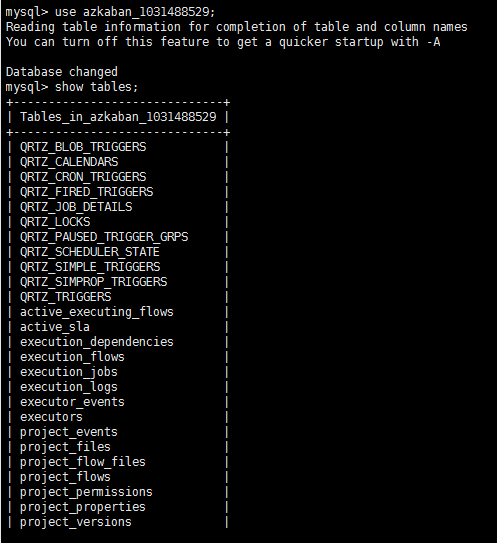


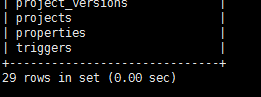
进入azkaban/azkaban-db/build/distributions/azkaban-db-3.60.0-14-gf428a07目录下，执行以下命令建表，会生成所需要的表和数据：

mysql -uroot -pGg/ru,.#5 -h10.173.32.6 -P3306 -Dazkaban\_1031488529 < create-all-sql-3.60.0-14-gf428a07.sql

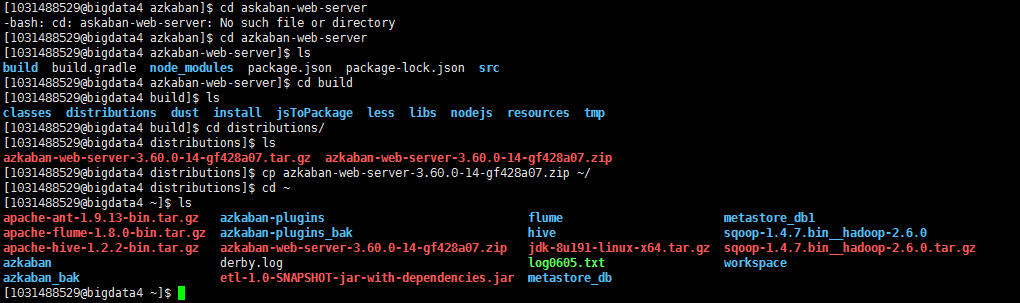


Mysql端查看结果，29个表已生成，没问题：

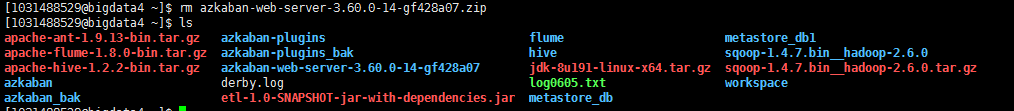




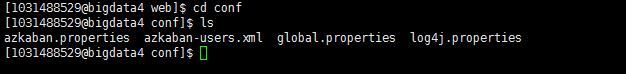
8）进入azkaban/azkaban-web-server/build/distributions，拷贝azkaban-web-server-3.60.0-14-gf428a07.zip到azkaban同级目录，这个包是用来启动azkaban的web服务







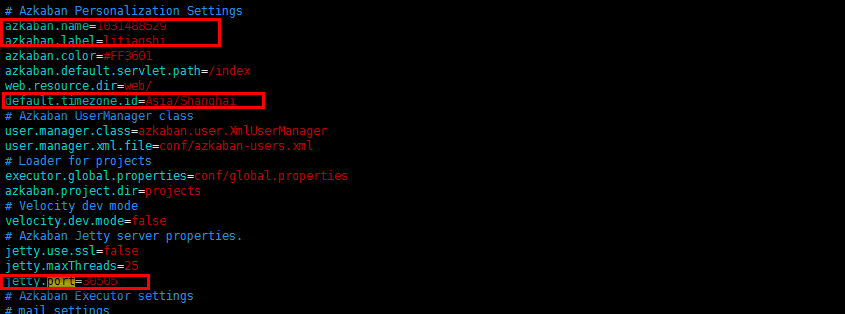
该版本已经直接把配置文件放到conf目录下，不需要再复制conf，直接编辑properties即可：



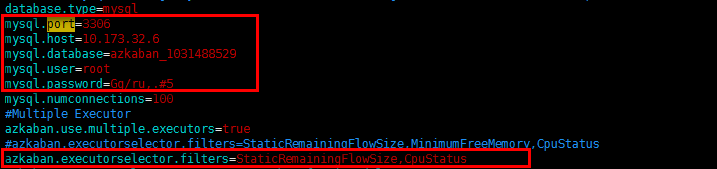
修改配置文件：



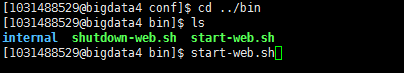
修改azkaban的名称和标签以验证确保是自己的配置；时区为上海；jetty.port=30505，这个端口将作为前端URL访问的端口：



设置mysql连接的参数，指定自己的数据库azkaban\_1031488529



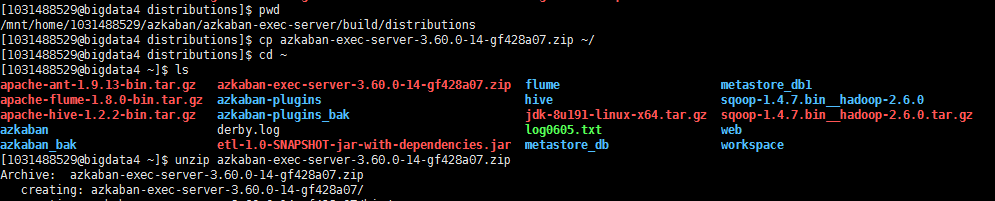
9）启动web服务

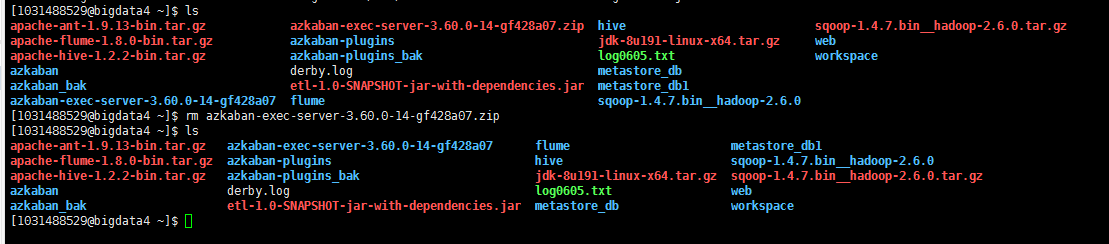


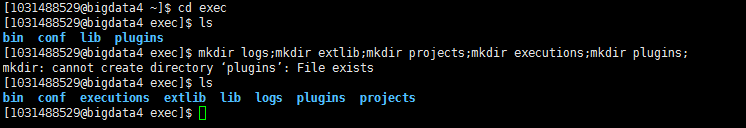
10）部署execution

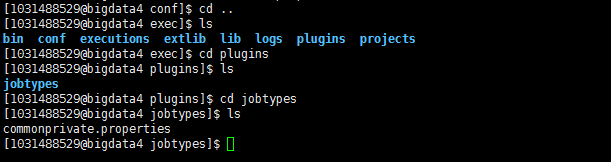
进入azkaban/azkaban-exec-server/build/distributions 目录，拷贝azkaban-exec-server-3.60.0-14-gf428a07.zip 到azkaban同级目录；

解压、删除压缩包、重命名、创建需要的文件夹

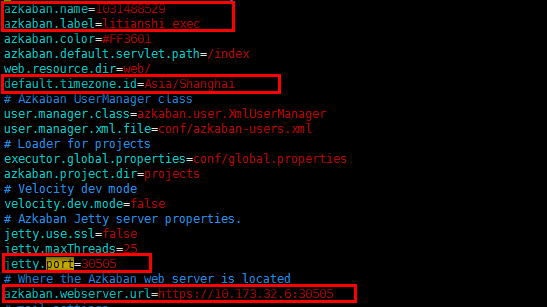




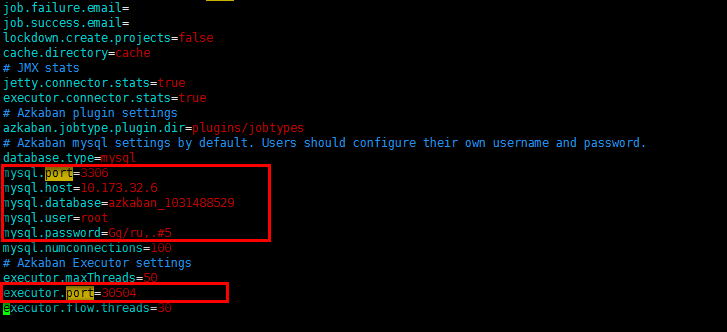


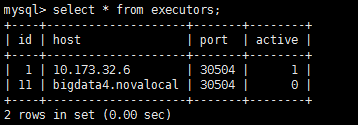


修改exec下的配置：前半部分和web基本一致

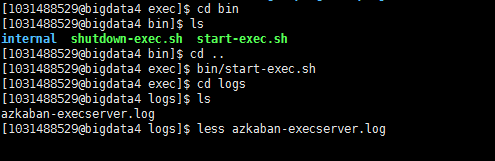


设置executor端口为30504，这个端口和mysql的表executor里一致：





启动executor服务：



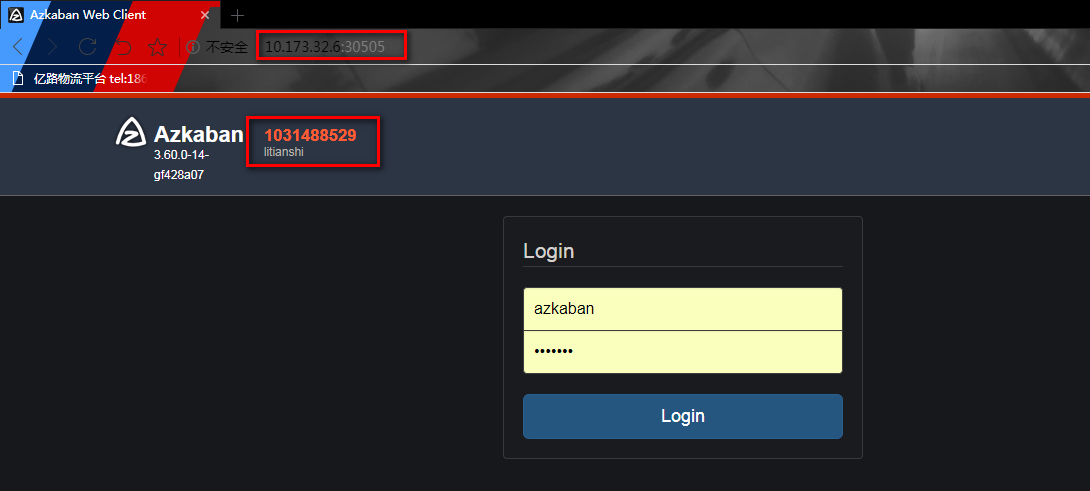
应该是exec自动在executions增加一笔novalocal的纪录，暂时不清楚其中原理

至此，azkaban安装配置完成。

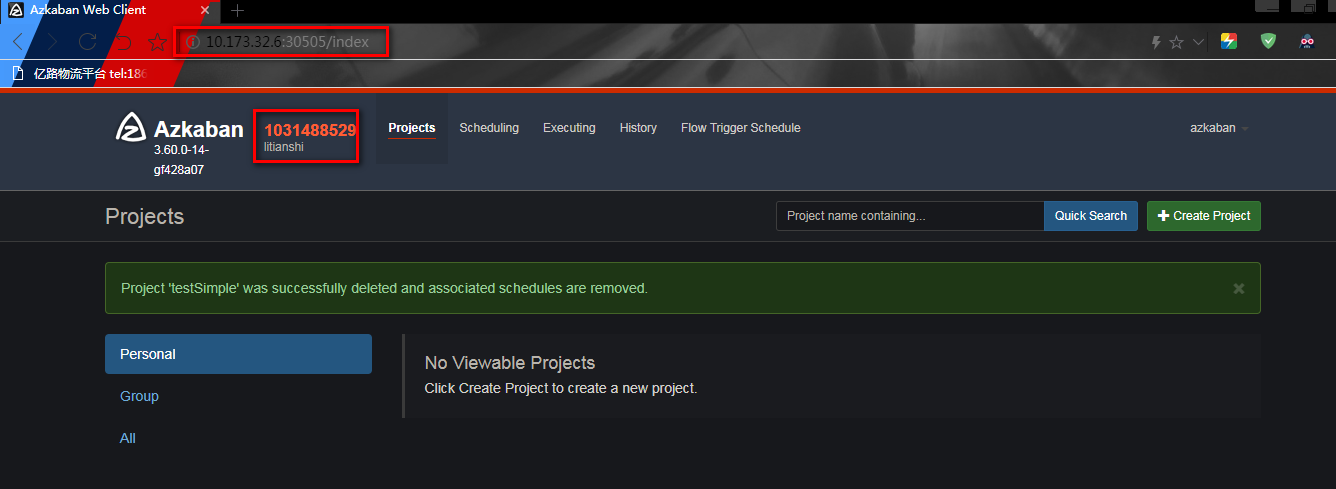
2，登陆azkaban

使用地址[http://10.173.32.6:3050](http://10.173.32.6:30504)5登录:

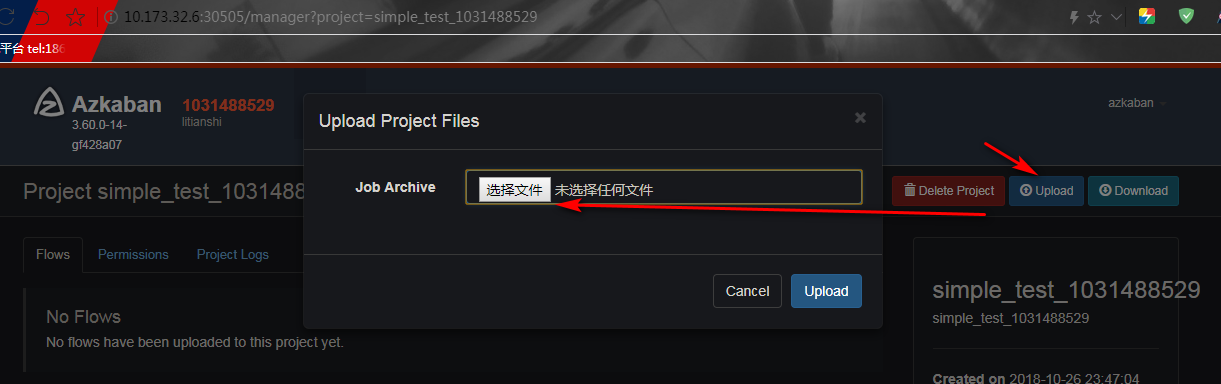
名称、标签和配置文件一致，证明是自己的配置



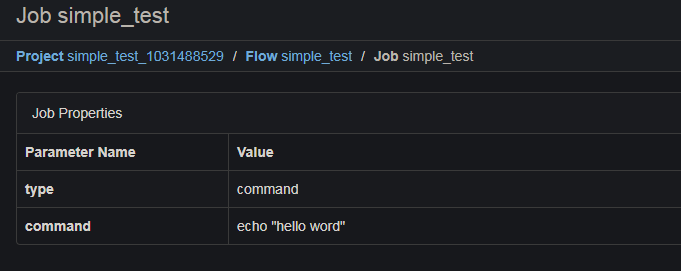
输入账号密码：azkaban/azkaban,登录：

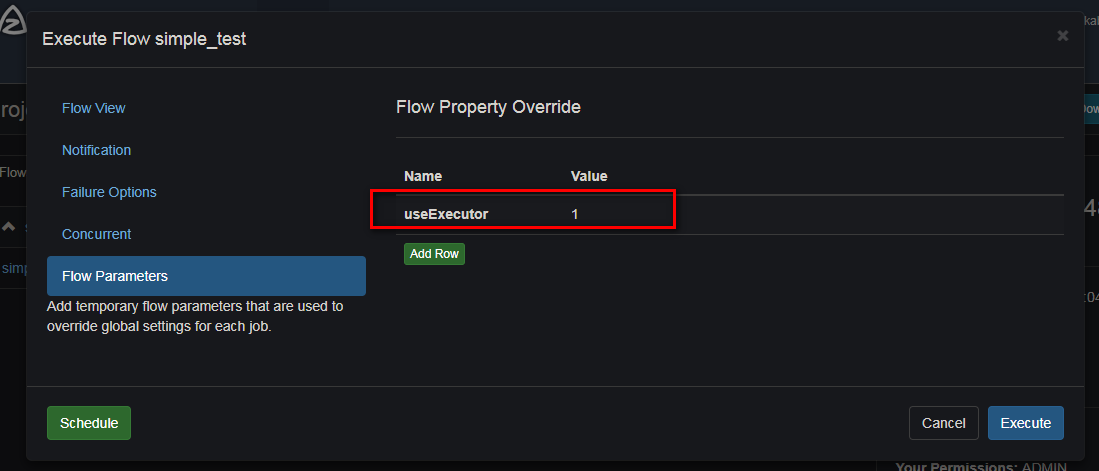


先做个简单的hello world任务验证是否正常：

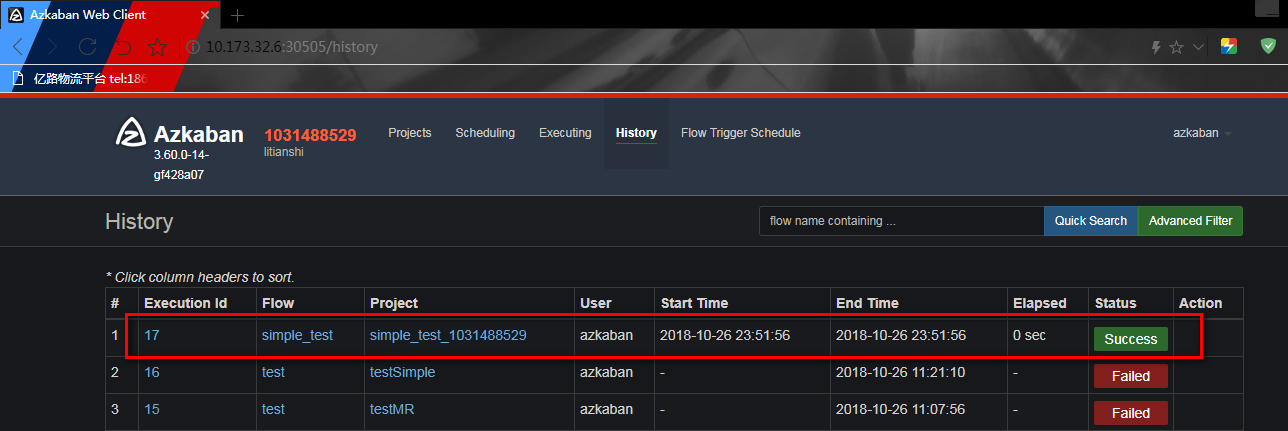


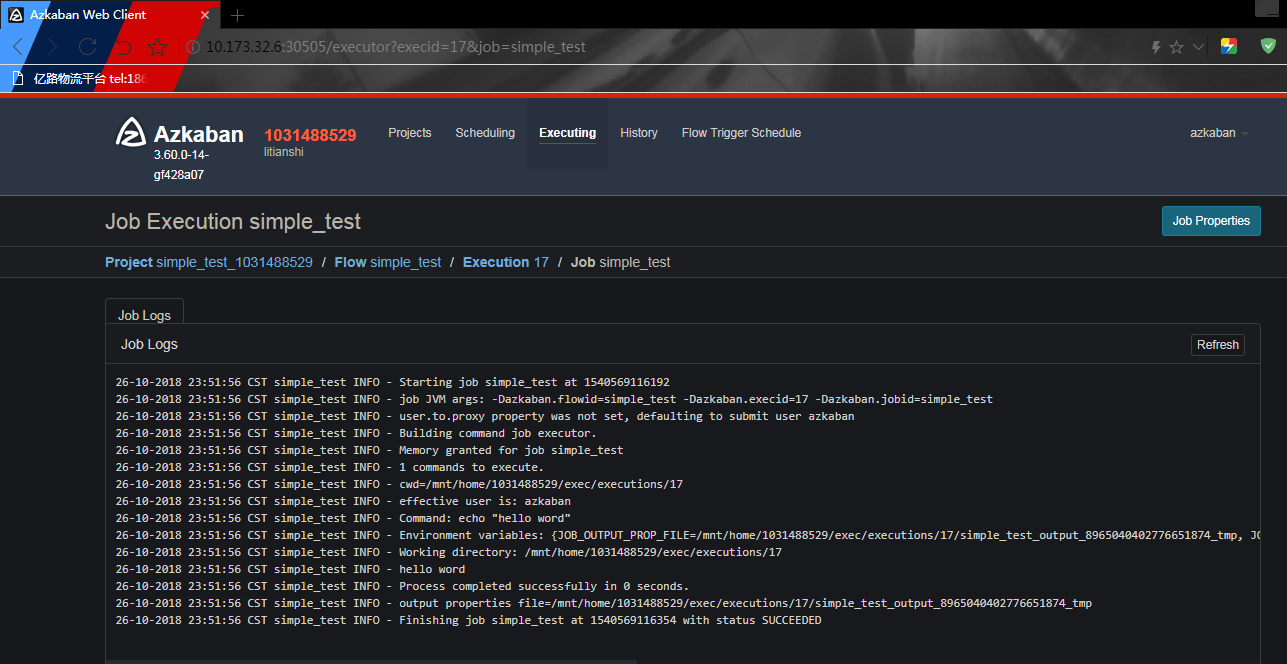
上传一个简单的hello world任务





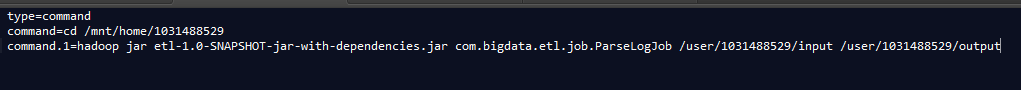
点击Executor，执行成功



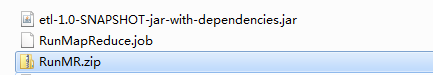


3，配置任务调用MR程序

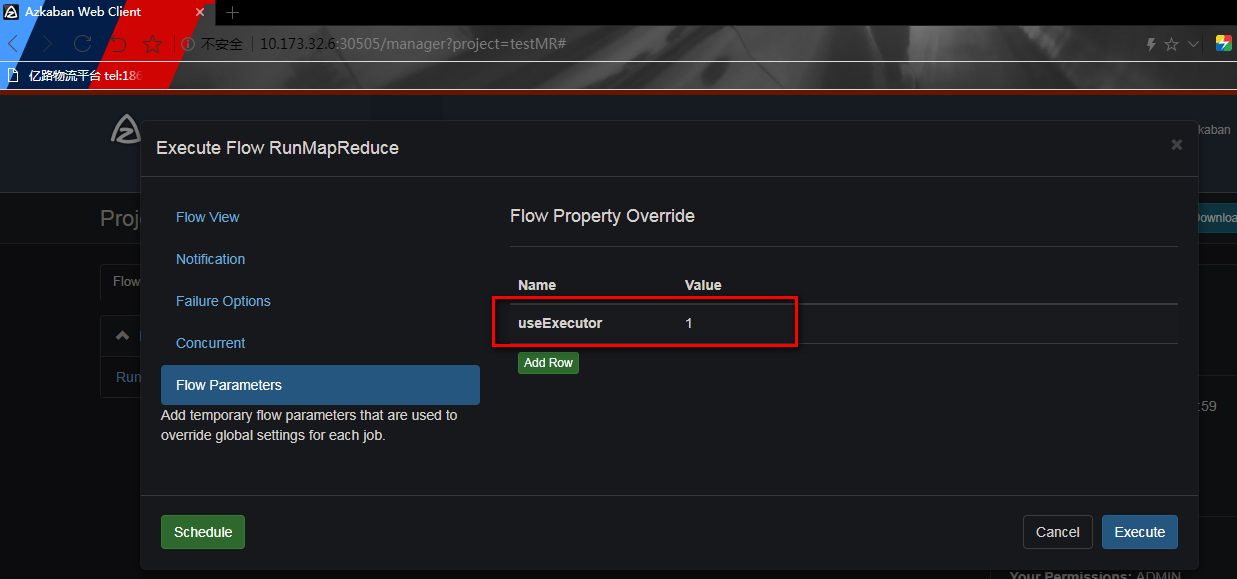
设置job：



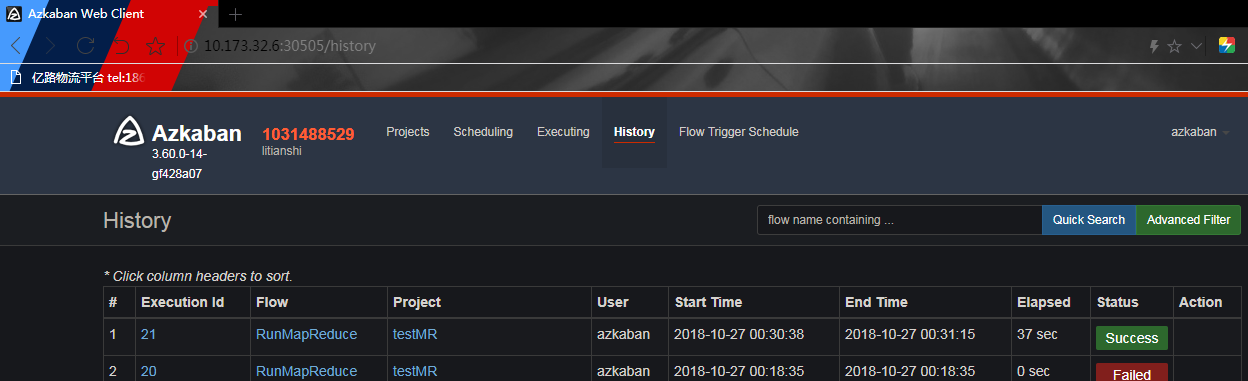
Job文件和jar包一起压缩上传：

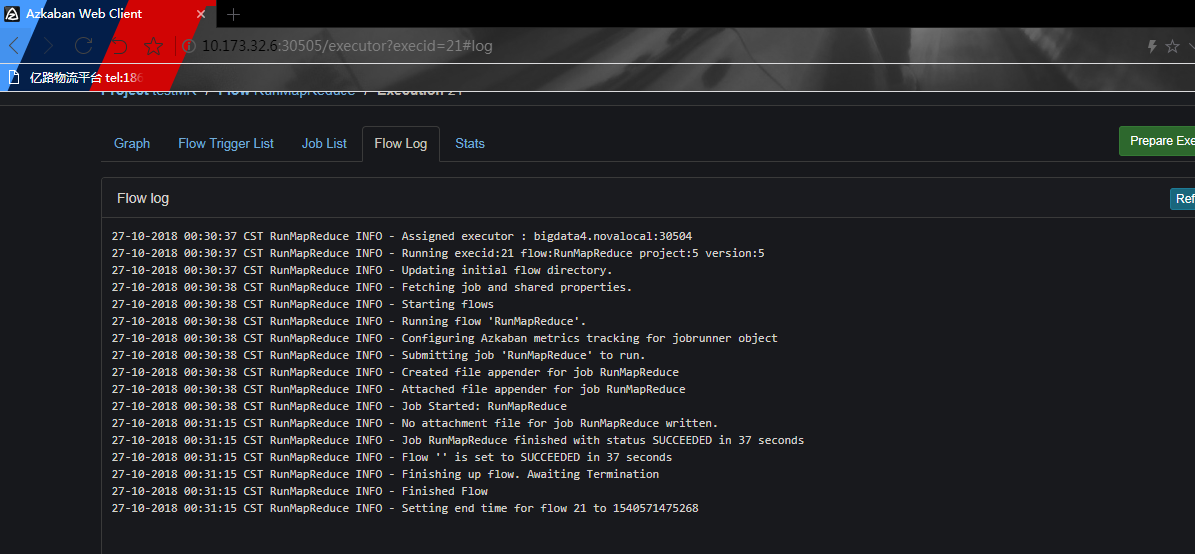


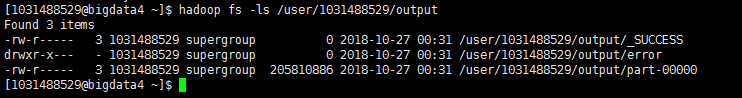
设置参数并运行：



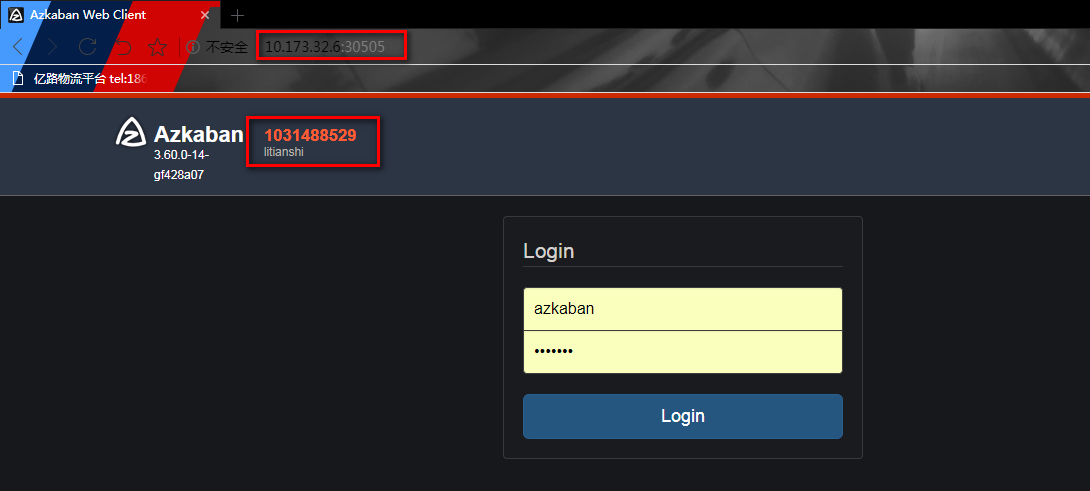
执行任务成功

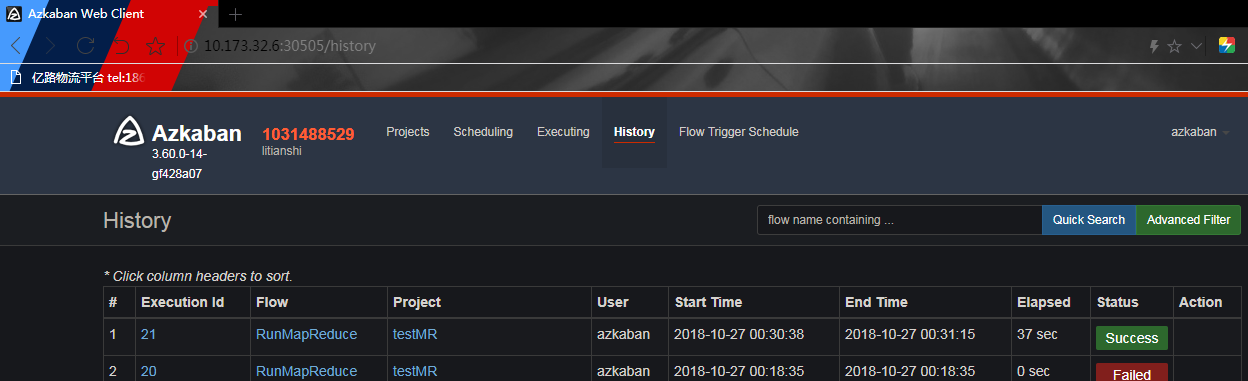


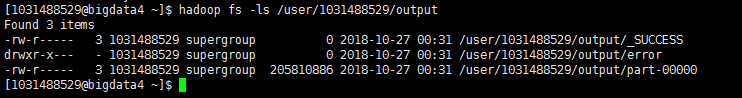




四、实验结果

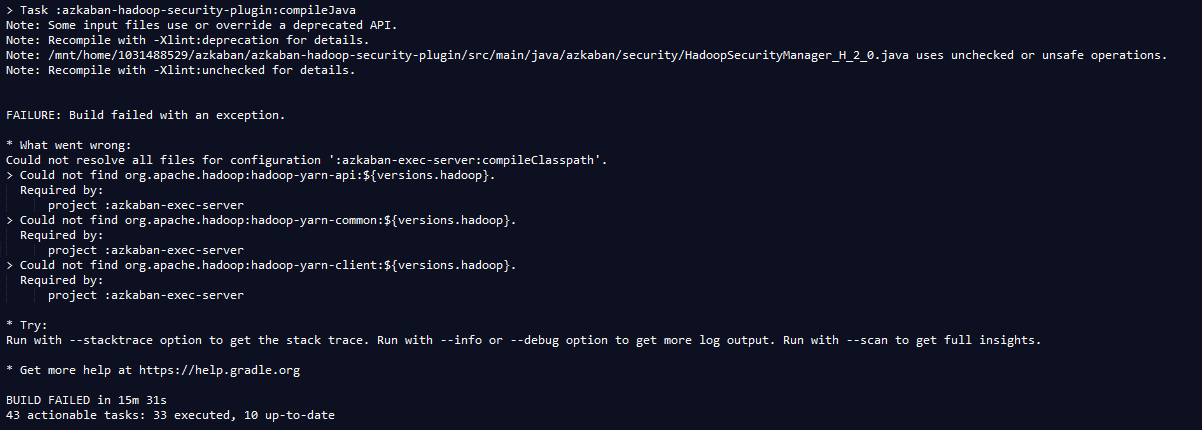


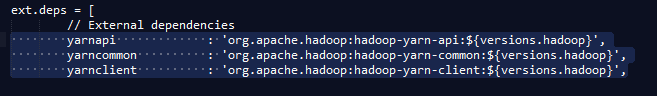




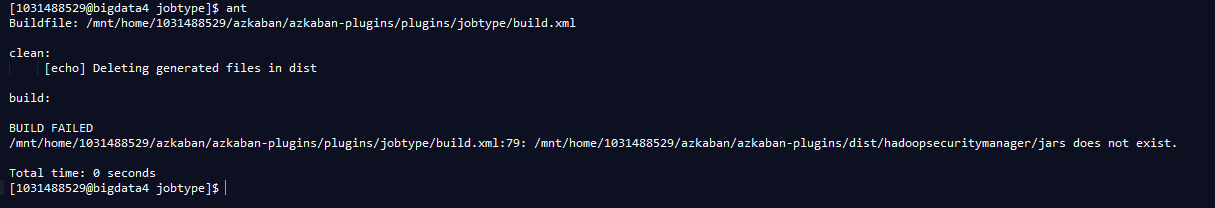
总结反思

1. 编译azkaban报错：

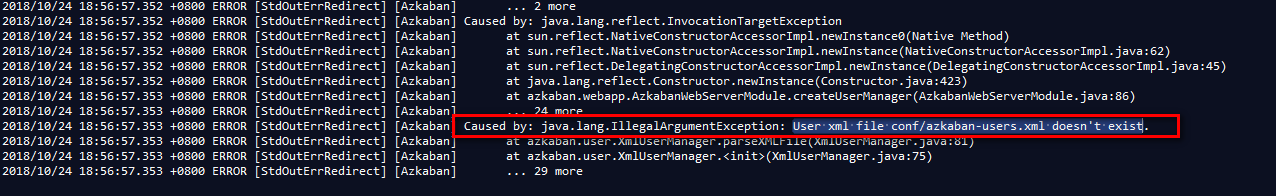


原因是build.gradle里面新家的几个值中用了单引号，改为双引号就解决了 

1. ant编译插件报错：原因是执行路径不对，在azkaban-plugins目录下执行就不报这个错了



1. plugins编译还有一些报错，原因是把azkaban-plugins放在azkaban目录下了，放在和azkaban同级目录就没问题了。
2. azkaban 执行job的时候状态一直是preparing，原因是没有制定executor，解决办法是在job的parameter里面增加参数useExecutor，参数值为1，即上面在mysql的executors表中增加纪录的id；
3. 执行start…启动服务时，必须先进入到web和exec这层目录，再执行bin/start….sh命令，当前路径会作为某个属性传入到配置中，否则会导致一些文件找不到的异常，比如：



1. 端口设置非常关键，启动exec或web服务中遇到过[Azkaban] Address already in use (Bind failed)的异常，可能端口确实被占用了，可尝试换端口；

端口基本有两种：一种是jetty端口，即前段页面访问端口，两个properties文件都用同意的jetty端口，目前自己用的是30505；另一种是executor端口，两个properties文件、mysql表，三个地方都用同一个端口，目前自己用的是30504；

1. 本任务把jar包和job一起打包上传，结果会把jar包也放到executions目录下

