**Flink1.12.3+CDH6.3.1集成指南**

# 环境准备

OS：CentOS7 +

JDK：1.8 +

Scala：2.12

Maven：3.6.3

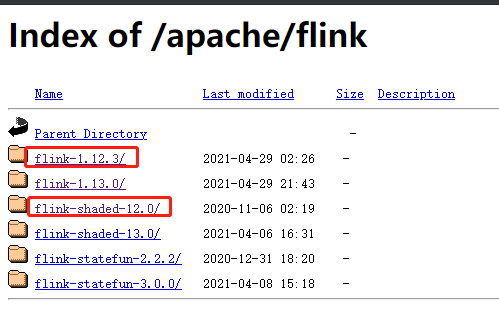
# 资源准备

## 源码包下载

清华源下载<https://mirrors.tuna.tsinghua.edu.cn/apache/flink/>

flink-shaded-12.0-src.tgz

flink-1.12.3-src.tgz



## 解压源码包

tar -zxvf flink-shaded-12.0-src.tgz

tar -zxvf flink-1.12.3-src.tgz

# 源码编译

## 进入shaded目录修改pom文件，新增以下配置

<profile>

<id>vendor-repos</id>

<activation>

<property>

<name>vendor-repos</name>

</property>

</activation>

<!-- Add vendor maven repositories -->

<repositories>

<!-- Cloudera -->

<repository>

<id>cloudera-releases</id>

<url>https://repository.cloudera.com/artifactory/cloudera-repos</url>

<releases>

<enabled>true</enabled>

</releases>

<snapshots>

<enabled>false</enabled>

</snapshots>

</repository>

<!-- Hortonworks -->

<repository>

<id>HDPReleases</id>

<name>HDP Releases</name>

<url>https://repo.hortonworks.com/content/repositories/releases/</url>

<snapshots><enabled>false</enabled></snapshots>

<releases><enabled>true</enabled></releases>

</repository>

<repository>

<id>HortonworksJettyHadoop</id>

<name>HDP Jetty</name>

<url>https://repo.hortonworks.com/content/repositories/jetty-hadoop</url>

<snapshots><enabled>false</enabled></snapshots>

<releases><enabled>true</enabled></releases>

</repository>

<!-- MapR -->

<repository>

<id>mapr-releases</id>

<url>https://repository.mapr.com/maven/</url>

<snapshots><enabled>false</enabled></snapshots>

<releases><enabled>true</enabled></releases>

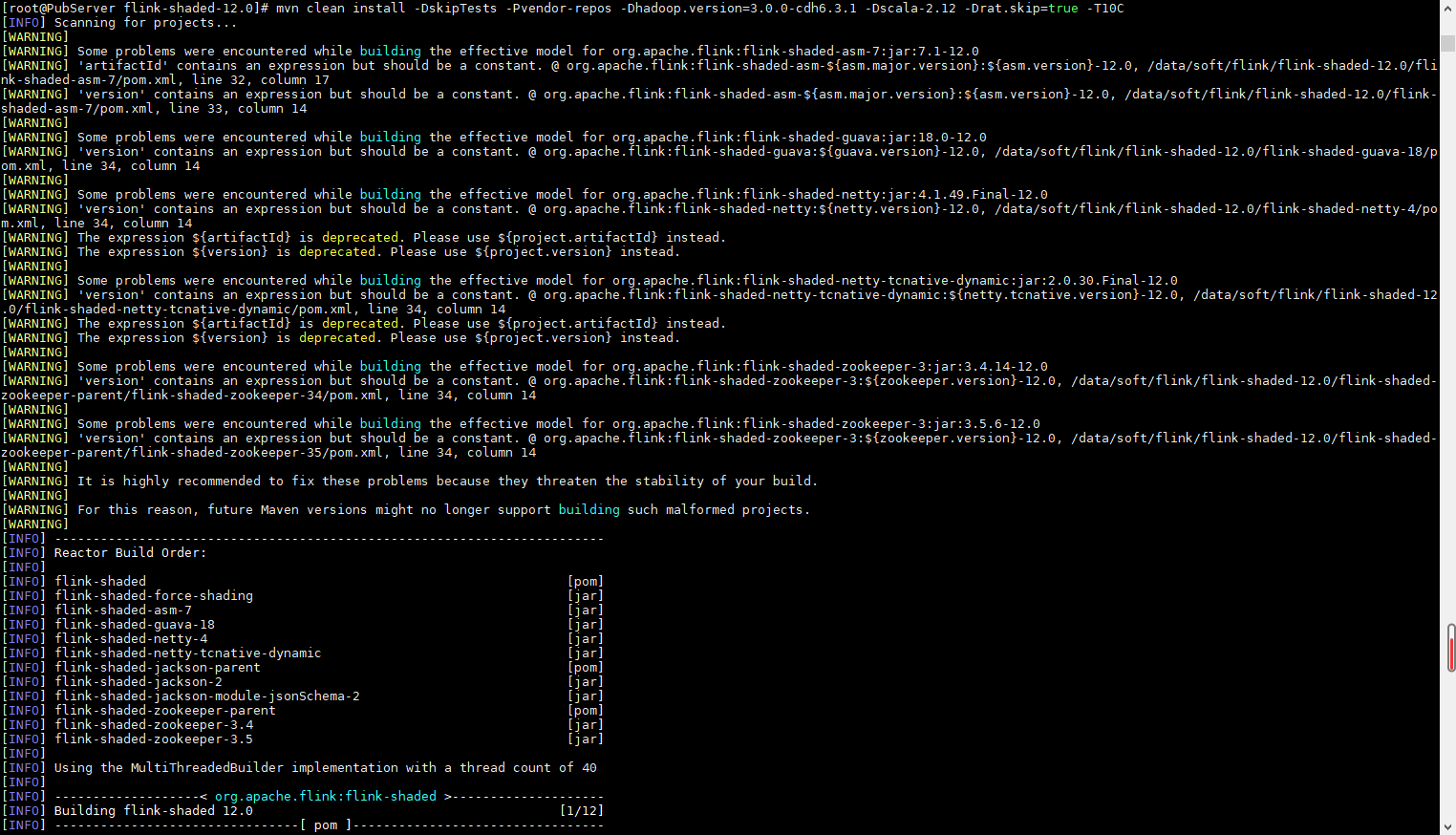
</repository>

</repositories>

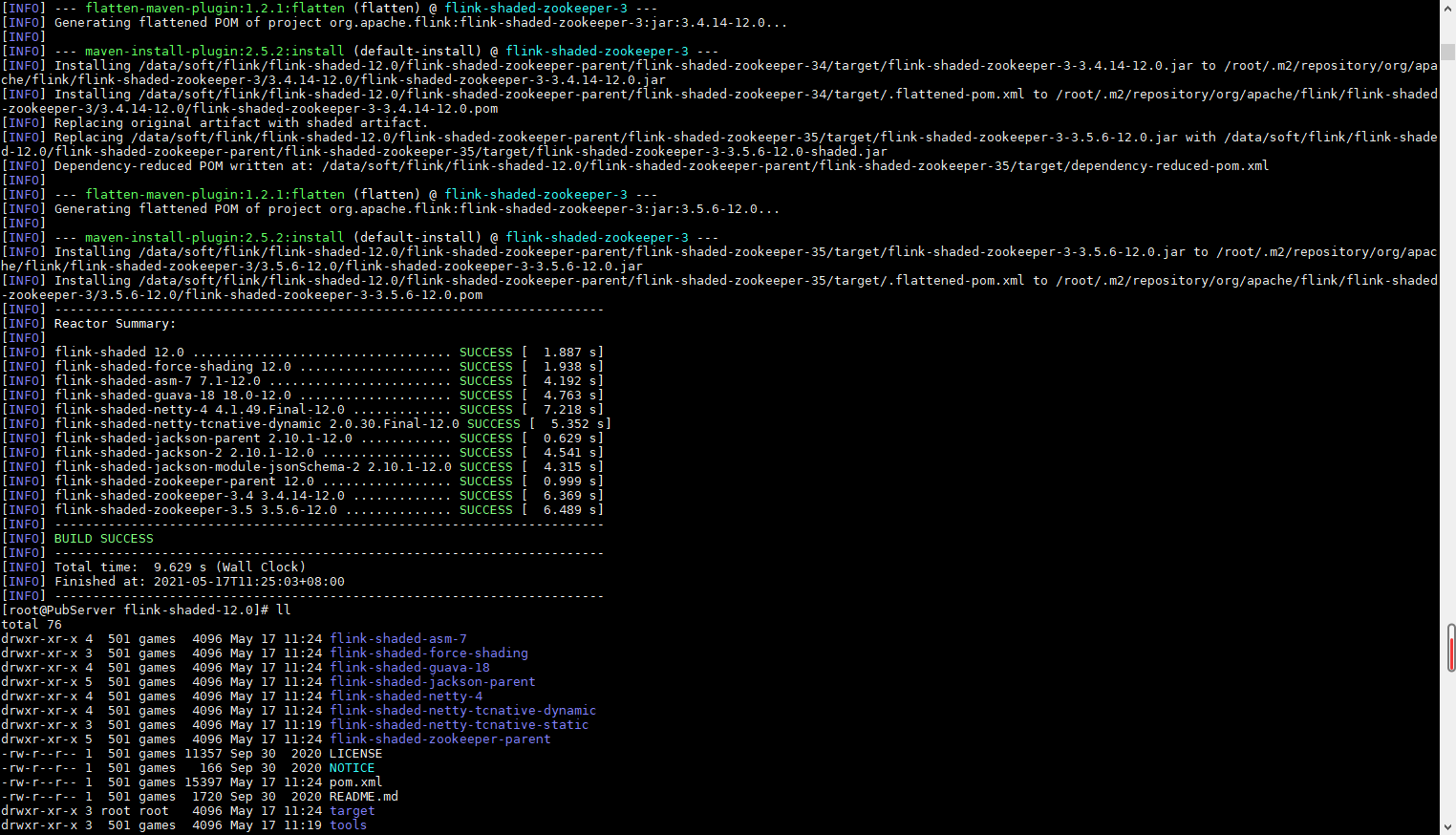
</profile>

## 先编译shaded

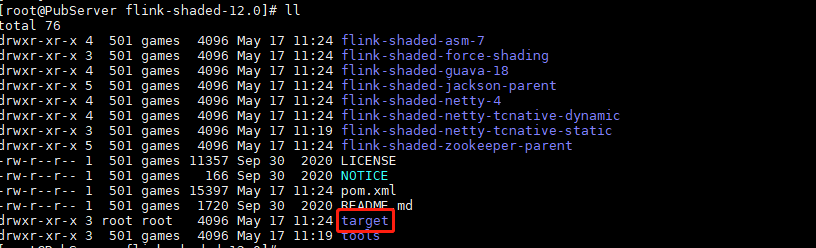
mvn clean install -DskipTests -Pvendor-repos -Dhadoop.version=3.0.0-cdh6.3.1 -Dscala-2.12 -Drat.skip=true -T10C（其中hadoop版本cdh版本和scala版本根据实际情况决定）



## Shaded编译完成

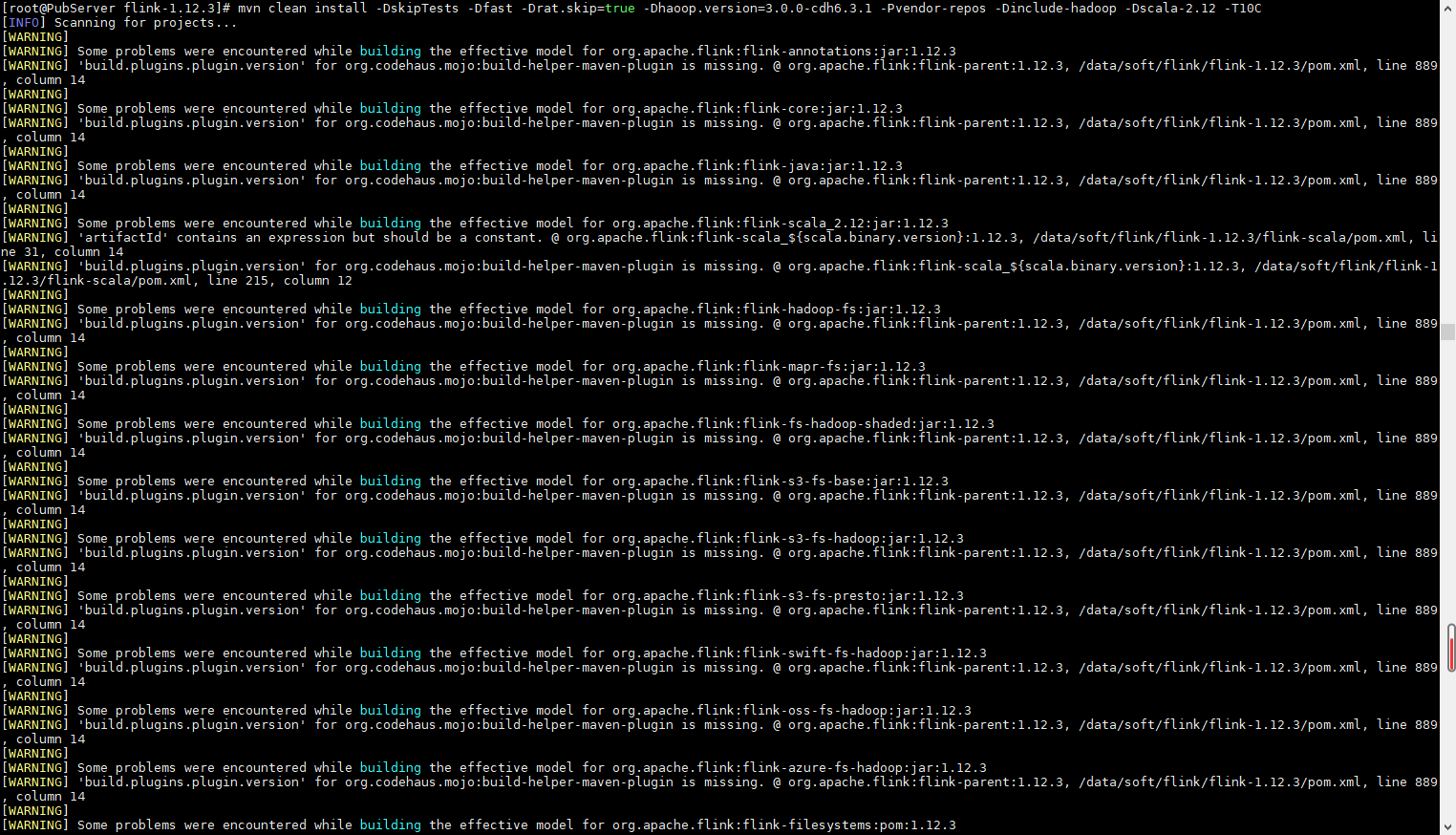


包在target目录



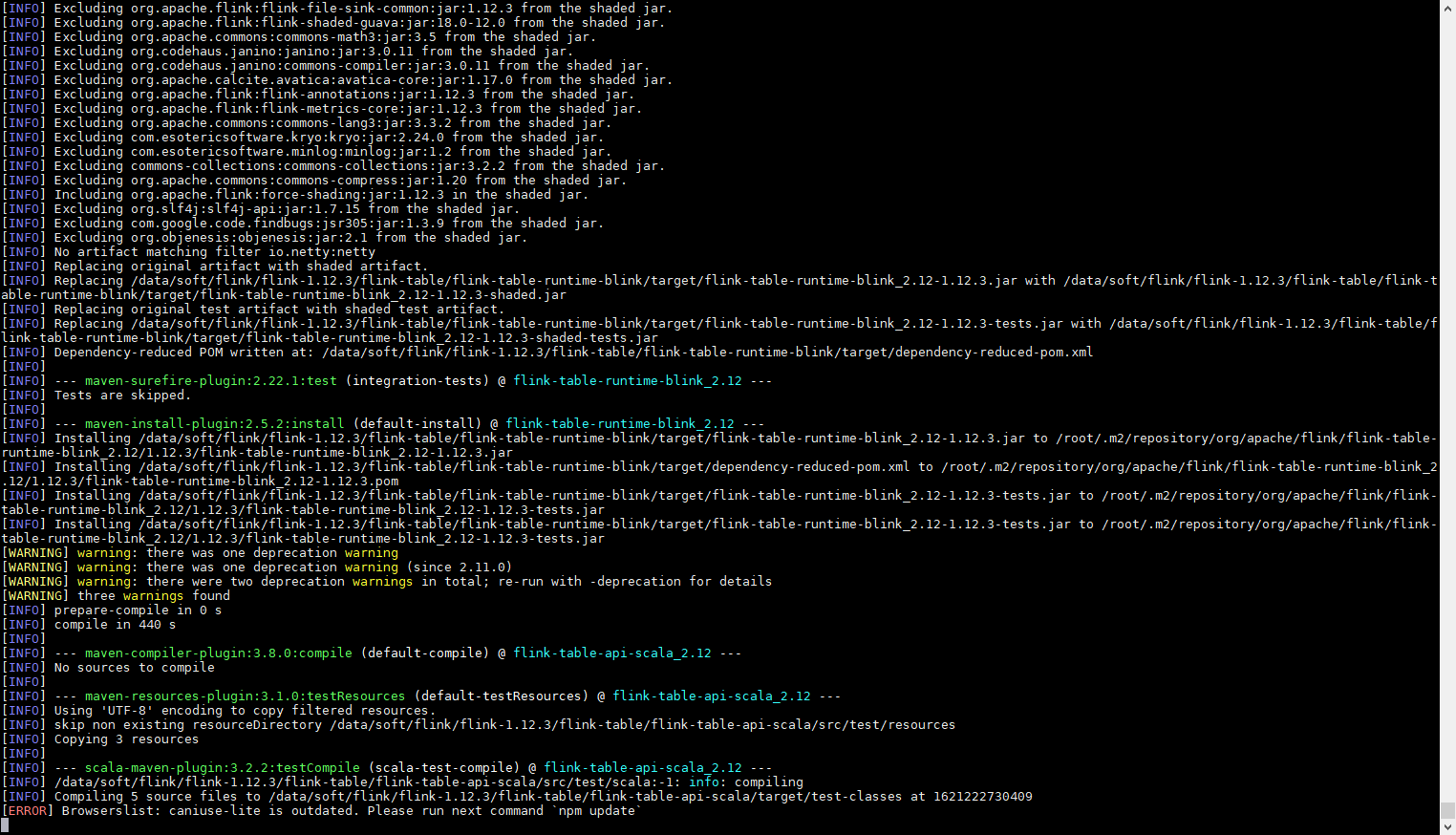
## 再编译Flink 1.12.3

mvn clean install -DskipTests -Dfast -Drat.skip=true -Dhaoop.version=3.0.0-cdh6.3.1 -Pvendor-repos -Dinclude-hadoop -Dscala-2.12 -T10C（其中hadoop版本cdh版本和scala版本根据实际情况决定）

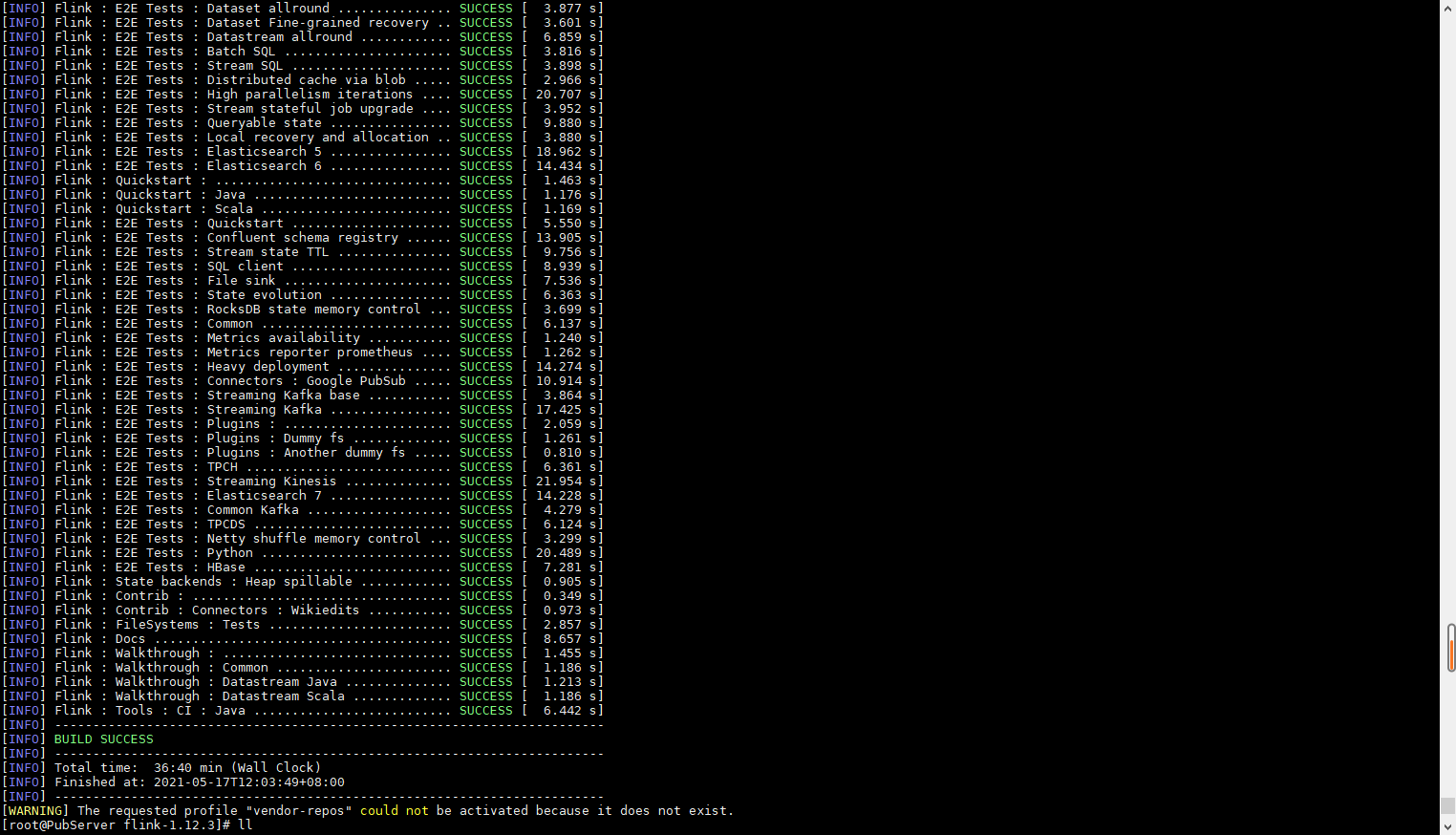


出现提示

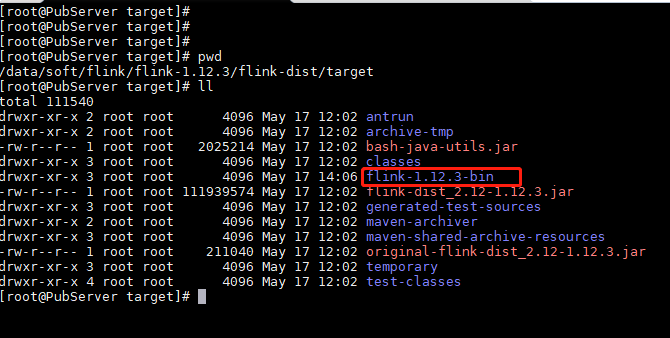
Browserslist: caniuse-lite is outdated. Please run next command `npm update`，好像没有影响，可以继续



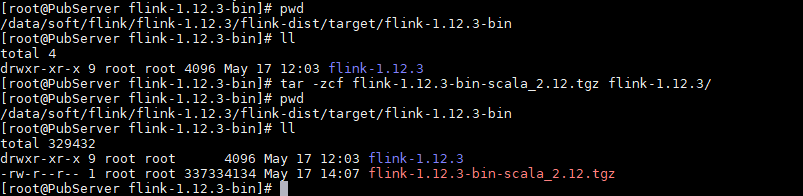
## Flink1.12.3编译完成



包在/data/soft/flink/flink-1.12.3/flink-dist/target/flink-1.12.3-bin目录



## 进入Flink编译完成目录打包



# 制作Parcel包

## 下载Parcel包制作工具

git clone https://github.com/pkeropen/flink-parcel.git

## 将编译完的flink包copy至工具目录



## 下载cm\_ext并copy至工具目录

手动下载cm\_ext\_master（在线下载较慢容易中断）

<https://github.com/cloudera/cm_ext.git>

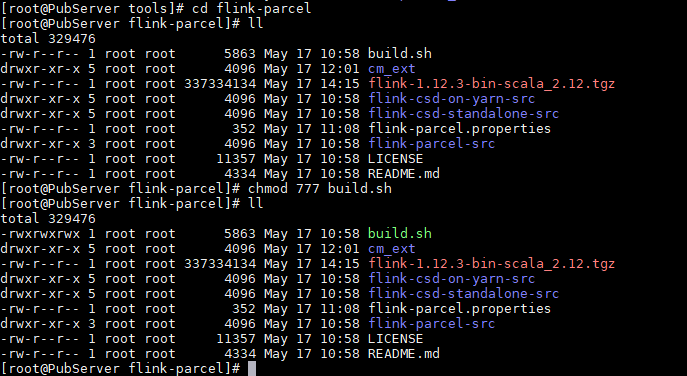
cp -r cm\_ext\_master /data/soft/flink/tools/flink-parcel/cm\_ext

## 给制作脚本赋予执行权限

本次直接给予所有操作权限

chmod 777 build.sh

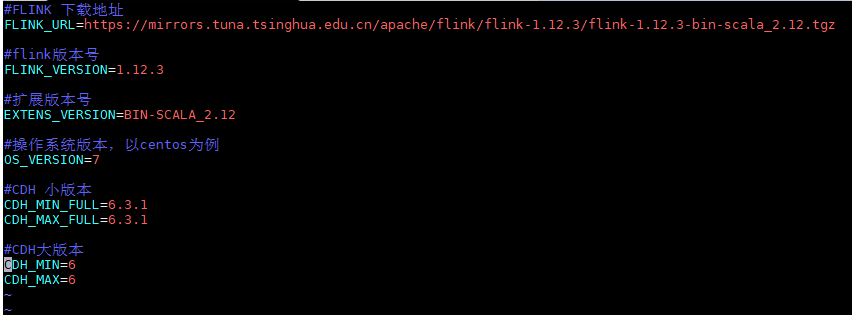
进入工具目录，当前目录内容如下，包含脚本+flink包+cm\_ext文件夹



## 修改配置文件

vim flink-parcel.properties

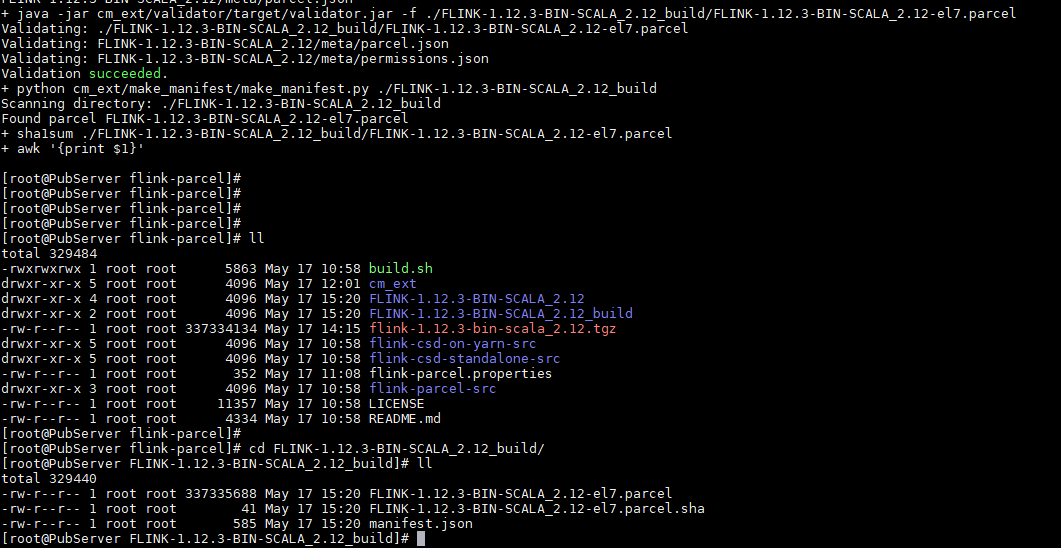
根据实际版本填写



## 制作Parcel包

./build.sh parcel

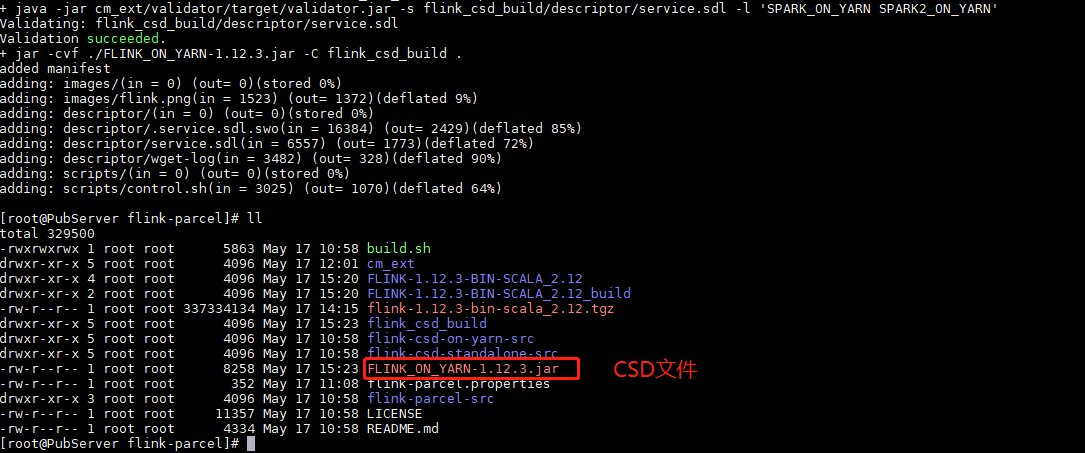
制作完成，parcel包存放路径如下



## 生成CSD文件

### 自动打包

./build.sh csd\_on\_yarn



### 手动打包

修改描述性文件

cd /data/soft/flink/tools/flink-parcel/flink\_csd\_build/descriptor

vim service.sdl

cd /data/soft/flink/tools/flink-parcel/flink\_csd\_build/

jar -cvf FLINK-1.12.3-SCALA\_2.12-el7-CDH\_6.3.1.jar \*

# CDH6.3.1集成

## 将安装包放入CM主节点指定目录

FLINK-1.9.0-csa1.0.0.0-cdh6.3.0.jar -> /opt/cloudera/csd

FLINK-1.9.0-csa1.0.0.0-cdh6.3.0-el7.parcel -> /opt/cloudera/parcel-repo

FLINK-1.9.0-csa1.0.0.0-cdh6.3.0-el7.parcel.sha -> /opt/cloudera/parcel-repo

## 在所有节点重启CDH服务

主节点：

systemctl stop cloudera-scm-server

systemctl stop cloudera-scm-agent

systemctl start cloudera-scm-server

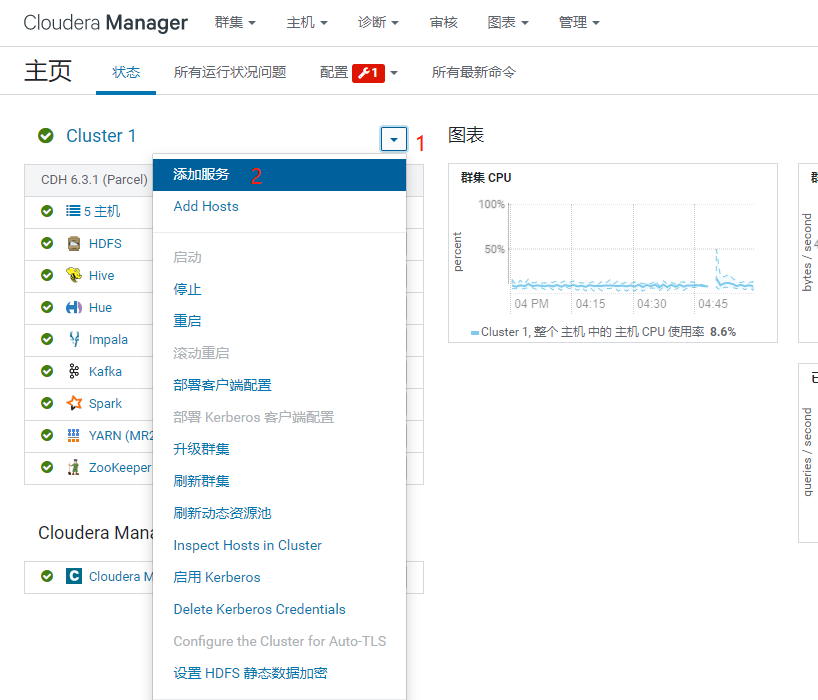
systemctl start cloudera-scm-agent

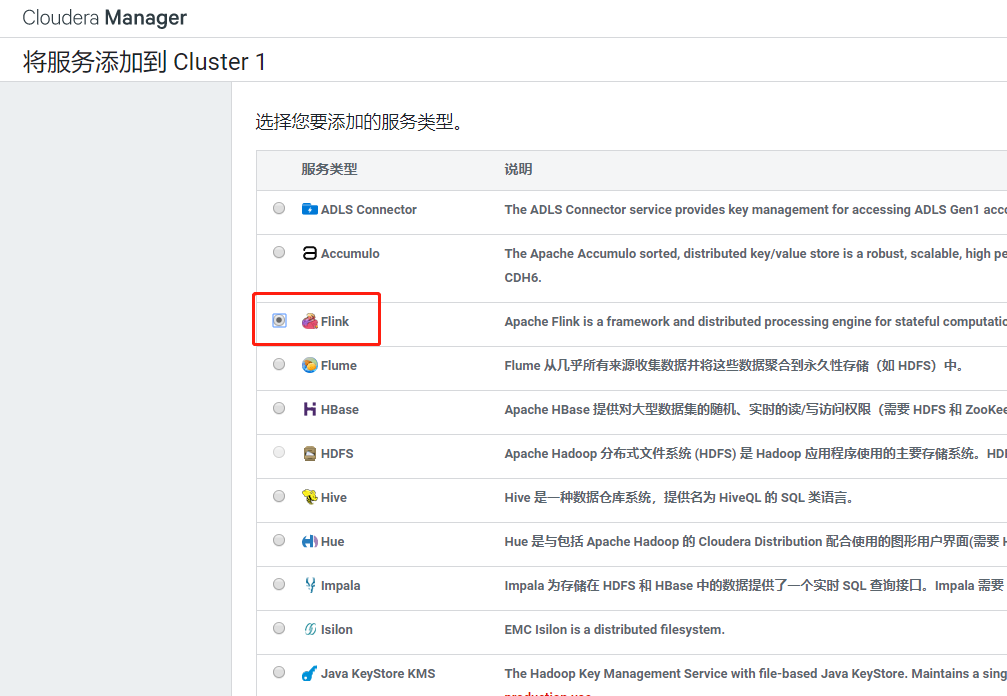
其他节点：

systemctl stop cloudera-scm-agent

systemctl start cloudera-scm-agent

## 添加Flink服务

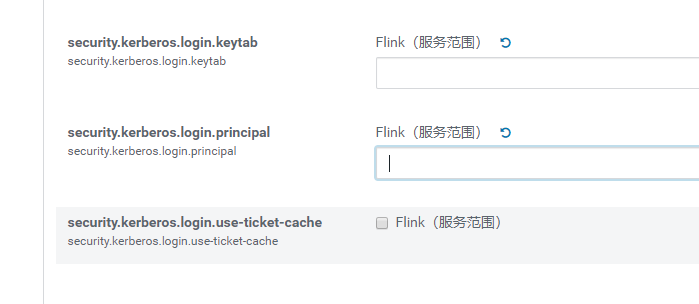








审核更改中去掉kerberos相关3项配置



直接下一步，到结束，可能会启动失败。

## 可能遇到的问题

### 启动服务失败

#### Flink-yarn.sh启动失败



解决方案：注释17行，该行主要用于日志处理，不影响服务运行。（据说添加flink-shaded-hadoop-2-uber-2.8.3-10.0.jar包后可以解决该问题，暂未测试）

#### 缺少hadoop jar包

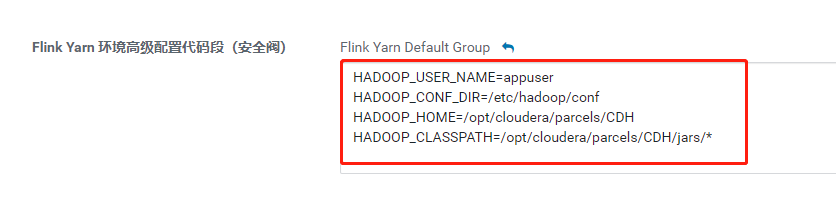
解决方案：在Flink服务-配置-Flink yarn高级配置中增加以下信息：

HADOOP\_USER\_NAME=appuser

HADOOP\_CONF\_DIR=/etc/hadoop/conf

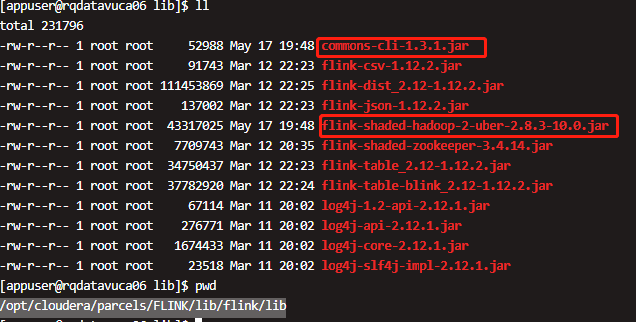
HADOOP\_HOME=/opt/cloudera/parcels/CDH

HADOOP\_CLASSPATH=/opt/cloudera/parcels/CDH/jars/\*



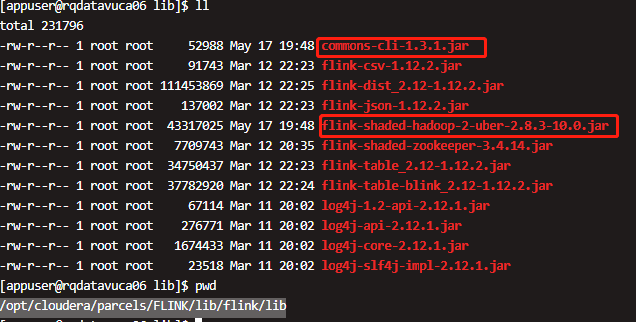
#### 缺少shaded hadoop jar包

Maven仓库下载flink-shaded-hadoop-2-uber-2.8.3-10.0.jar放到主节点和flink部署节点的/opt/cloudera/parcels/FLINK/lib/flink/lib目录



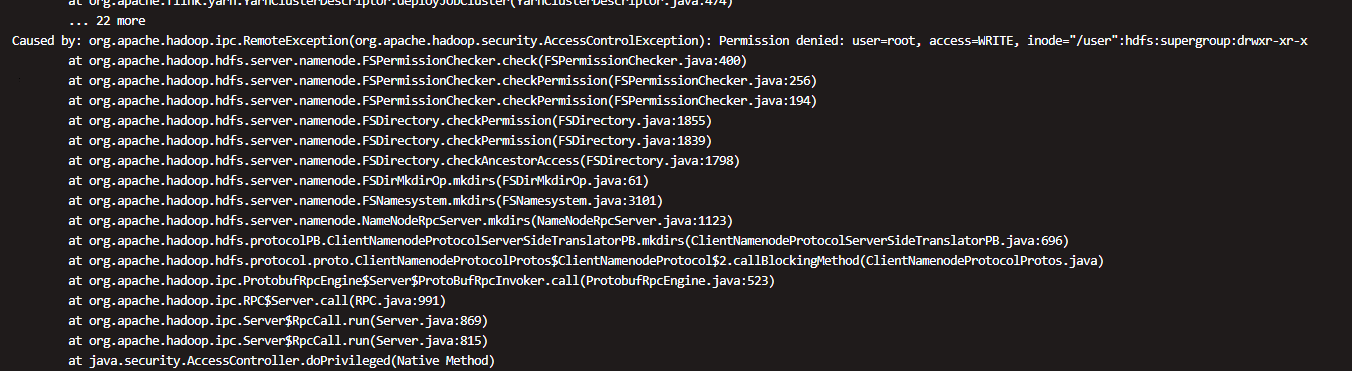
#### 缺少commons-cli jar包

Maven仓库下载commons-cli-1.3.1.jar放到主节点和flink部署节点的/opt/cloudera/parcels/FLINK/lib/flink/lib目录

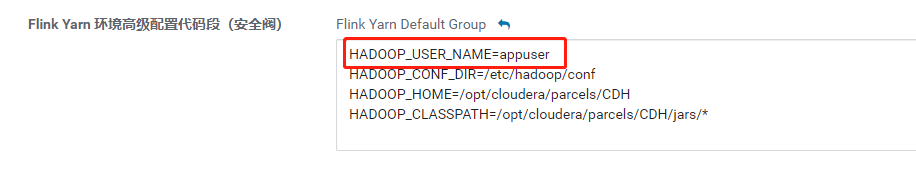


### 任务提交异常

#### root/flink用户提交任务失败



解决方案：使用appuser用户提交任务

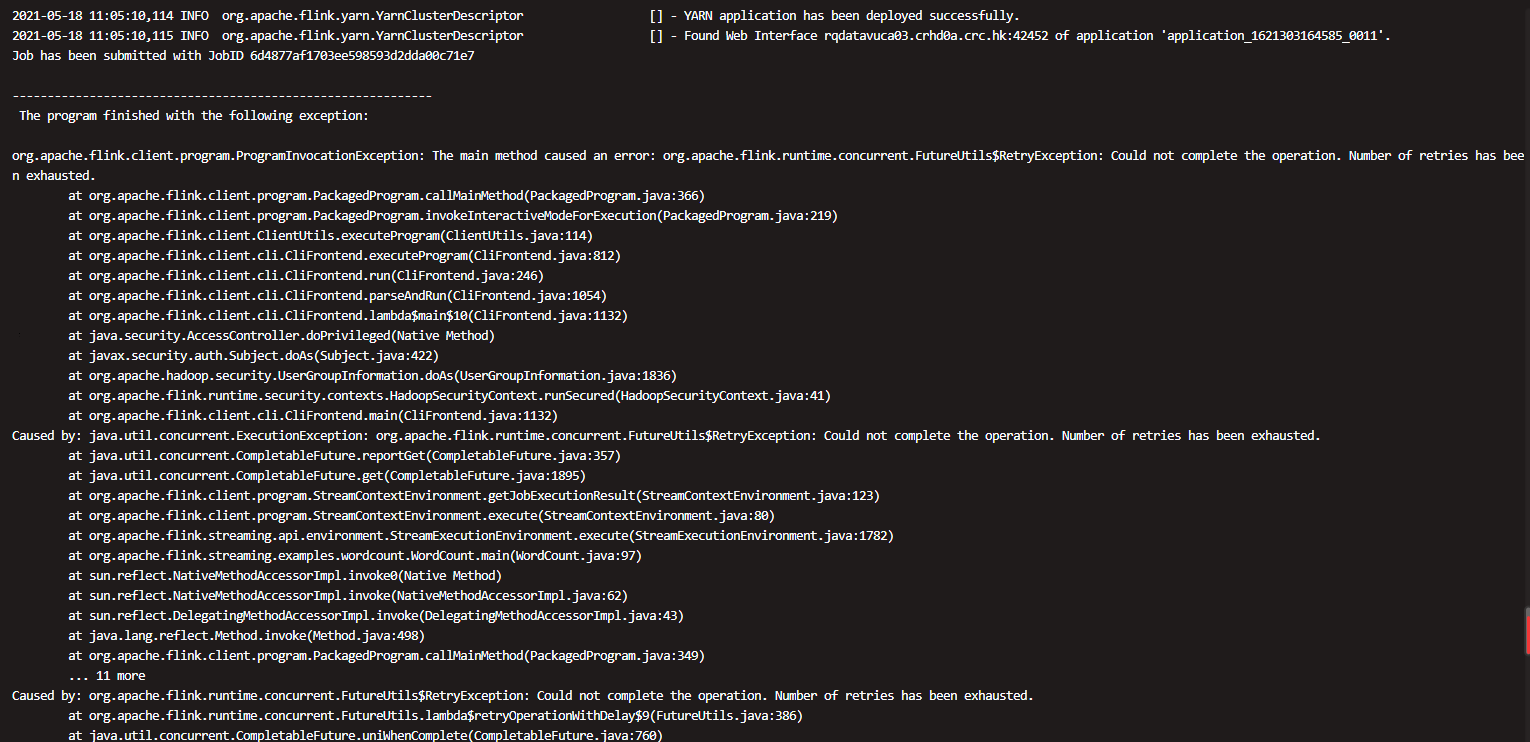


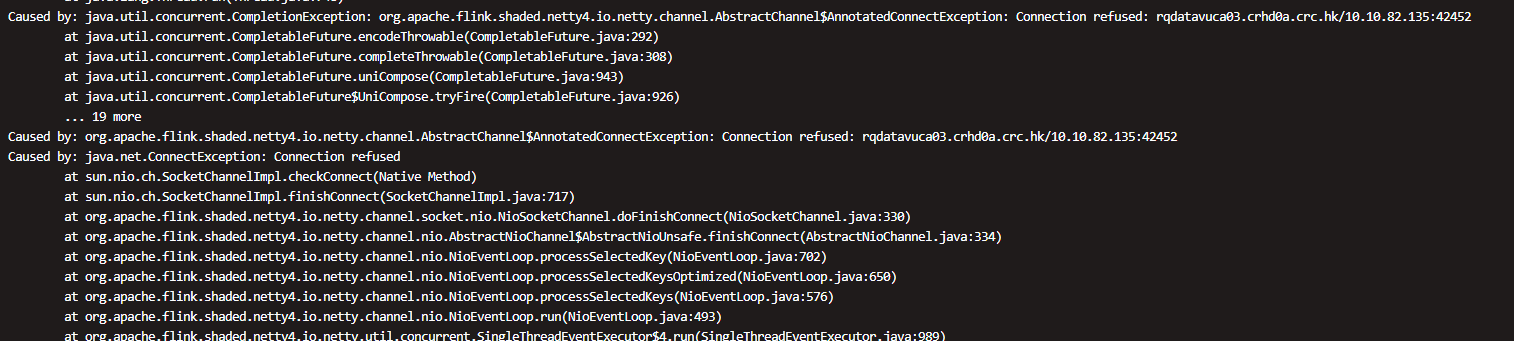
#### Yarn提交任务失败

解决方案：为节点增加nodemanager角色。

#### Hdfs写入失败

The main method caused an error: org.apache.flink.runtime.concurrent.FutureUtils$RetryException: Could not complete the operation. Number of retries has been exhausted.





解决办法：提交到namenode（rqdatavuca06）节点

flink run -m yarn-cluster /opt/cloudera/parcels/FLINK/lib/flink/examples/streaming/WordCount.jar --input hdfs://rqdatavuca06.crhd0a.crc.hk:8020/tmp/word.txt --output hdfs://rqdatavuca06.crhd0a.crc.hk:8020/home/appuser/result1.txt

## 最后：正常的服务状态图

