**CDH安装及升级步骤**

目录

[1.1 背景 3](#_Toc457894348)

[1.2 环境准备 3](#_Toc457894349)

[1.2.1 离线仓库下载 3](#_Toc457894350)

[1.2.2 离线仓库配置 3](#_Toc457894351)

[1.2.3 重新配置主机yum源 4](#_Toc457894352)

[1.3 Cloudera 升级 4](#_Toc457894353)

[1.3.1 Cloudera Manager 升级 4](#_Toc457894354)

[1.3.2 Parcel 升级 9](#_Toc457894355)

[1.4 Cloudera 安装 12](#_Toc457894356)

[1.4.1 卸载集群中旧的Cloudera 12](#_Toc457894357)

[1.4.2 安装Cloudera 12](#_Toc457894358)

[1.5 集群优化 15](#_Toc457894359)

[1.5.1 关闭 Linux THG 服务 15](#_Toc457894360)

[1.5.2 设置 linux 内核参数 15](#_Toc457894361)

[1.5.3 配置 HDFS 15](#_Toc457894362)

[1.5.4 配置 Yarn 资源 15](#_Toc457894363)

[1.5.5 配置 oozie 16](#_Toc457894364)

[1.6 FAQ 17](#_Toc457894365)

## 背景

本次使用离线 Cloudrea 仓库安装方式，预安装版本：CDH 5.8.0，当前集群版本，CDH 5.2.0。

## 环境准备

### 离线仓库下载

下载离线仓库

CM5 地址

<https://archive.cloudera.com/cm5/repo-as-tarball/5.8.0/cm5.8.0-centos6.tar.gz>

CDH地址

<https://archive.cloudera.com/cdh5/parcels/5.8.0/>

*cdh下载列表：*

*CDH-5.8.0-1.cdh5.8.0.p0.42-el6.parcel*

*CDH-5.8.0-1.cdh5.8.0.p0.42-el6.parcel.sha1*

*manifest.json*

下载验证

地址：<https://archive.cloudera.com/cm5/redhat/6/x86_64/cm/5.8.0/repodata/>

下载安装脚本

cloudera-manager-installer.bin

### 离线仓库配置

选择一台服务器，配置apache服务地址，本次选择192.168.188.55上配置离线仓库。

1. 开启apache服务：service httpd start
2. 在 /var/www/html 目录下创建cloudear 离线安装目录结构，详见附件



1. 配置archive.cloudera.com离线地址

*即离线仓库的apache服务所在的机器ip，进入要升级cloudera的机器：*

*vi /etc/hosts 增加：192.168.188.55 archive.cloudera.com*

### 重新配置主机yum源

由于使用RedHat自带的yum有问题，故重新安装一个yum源，安装yum需要的rpm包位置 /var/www/html/yum/\*.rpm。

1. 卸载redhat自带的yum

*rpm -qa | grep yum | xargs rpm -e --nodeps*

1. 安装下载的rpm包

*rpm -ivh /var/www/html/yum/\*.rpm*

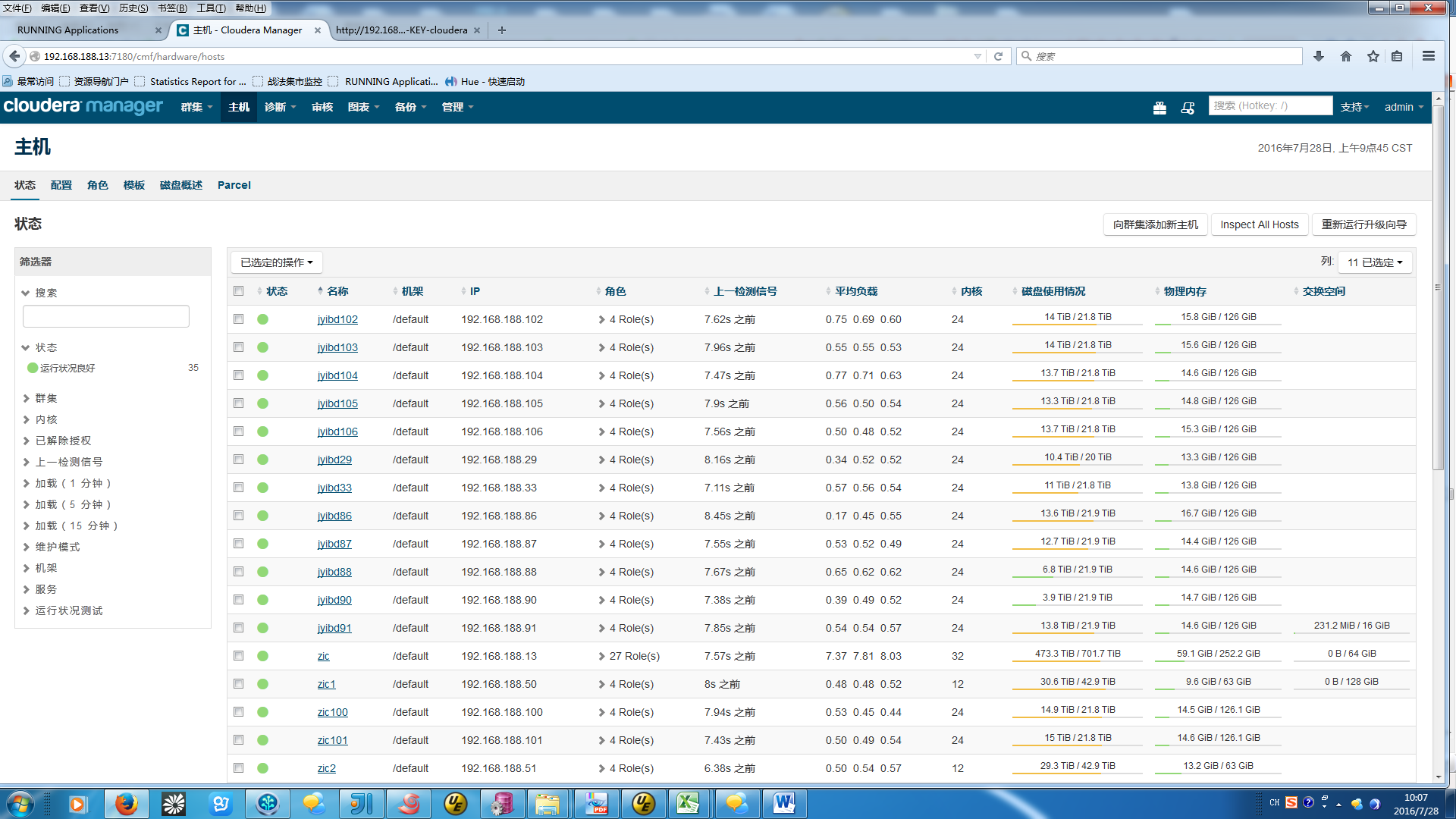
1. #yum clean all
2. #yum makecache
3. 安装成功

## Cloudera 升级

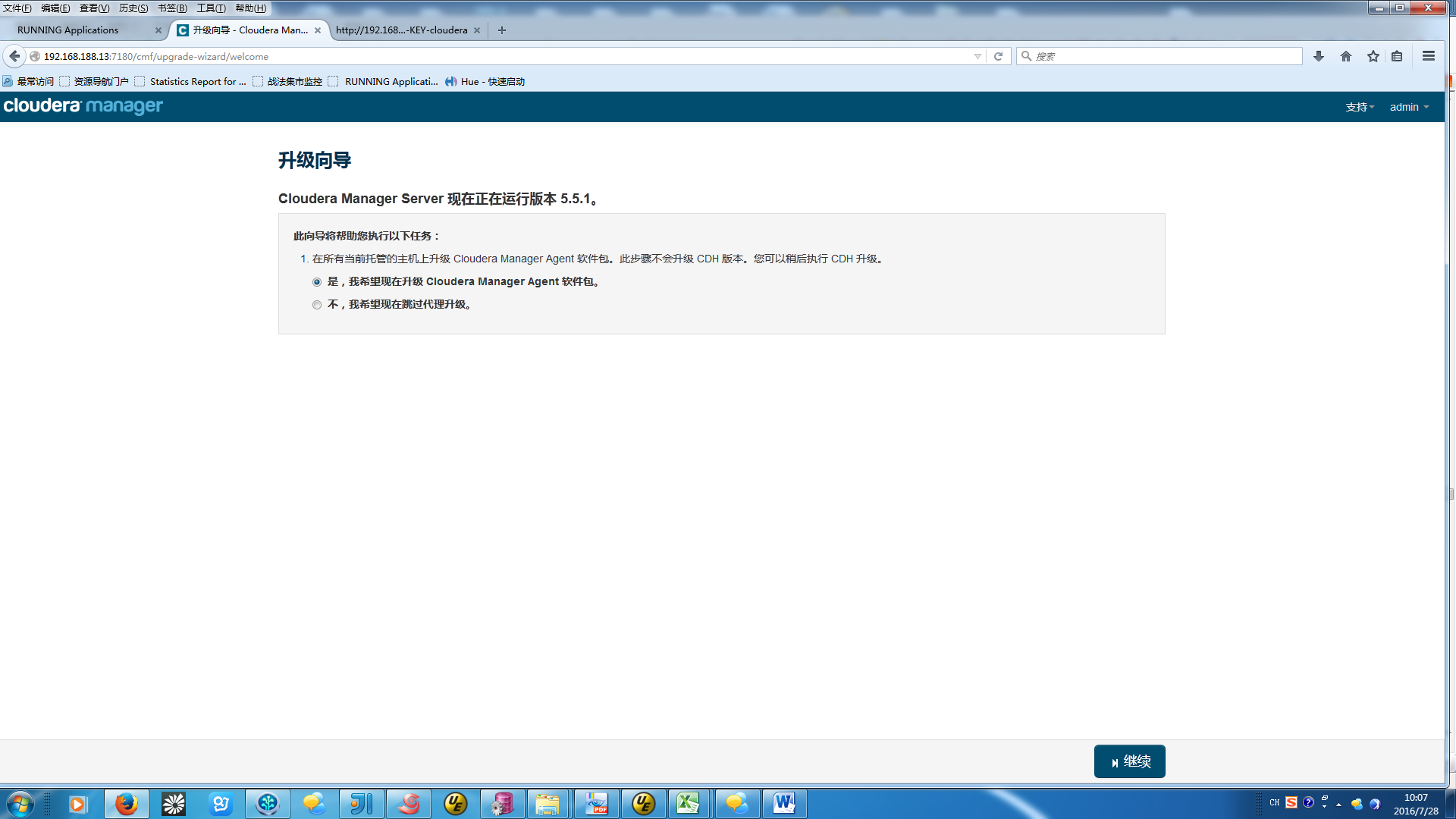
### Cloudera Manager 升级

为确保数据安全，建议升级前先备份HDFS上的数据。

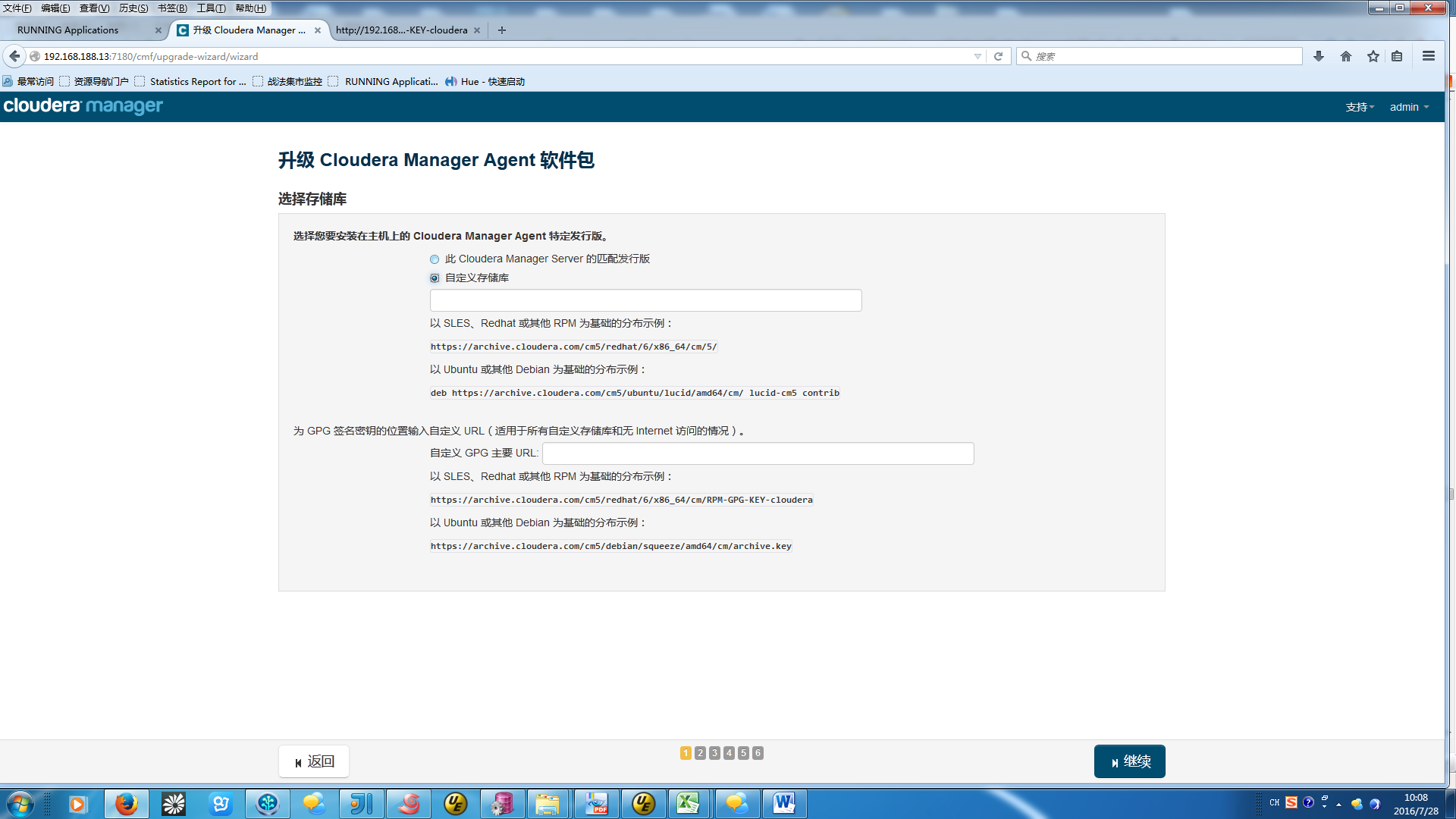
1. 登录cloudera manage
2. 点击 主机->所有主机 进入如下界面



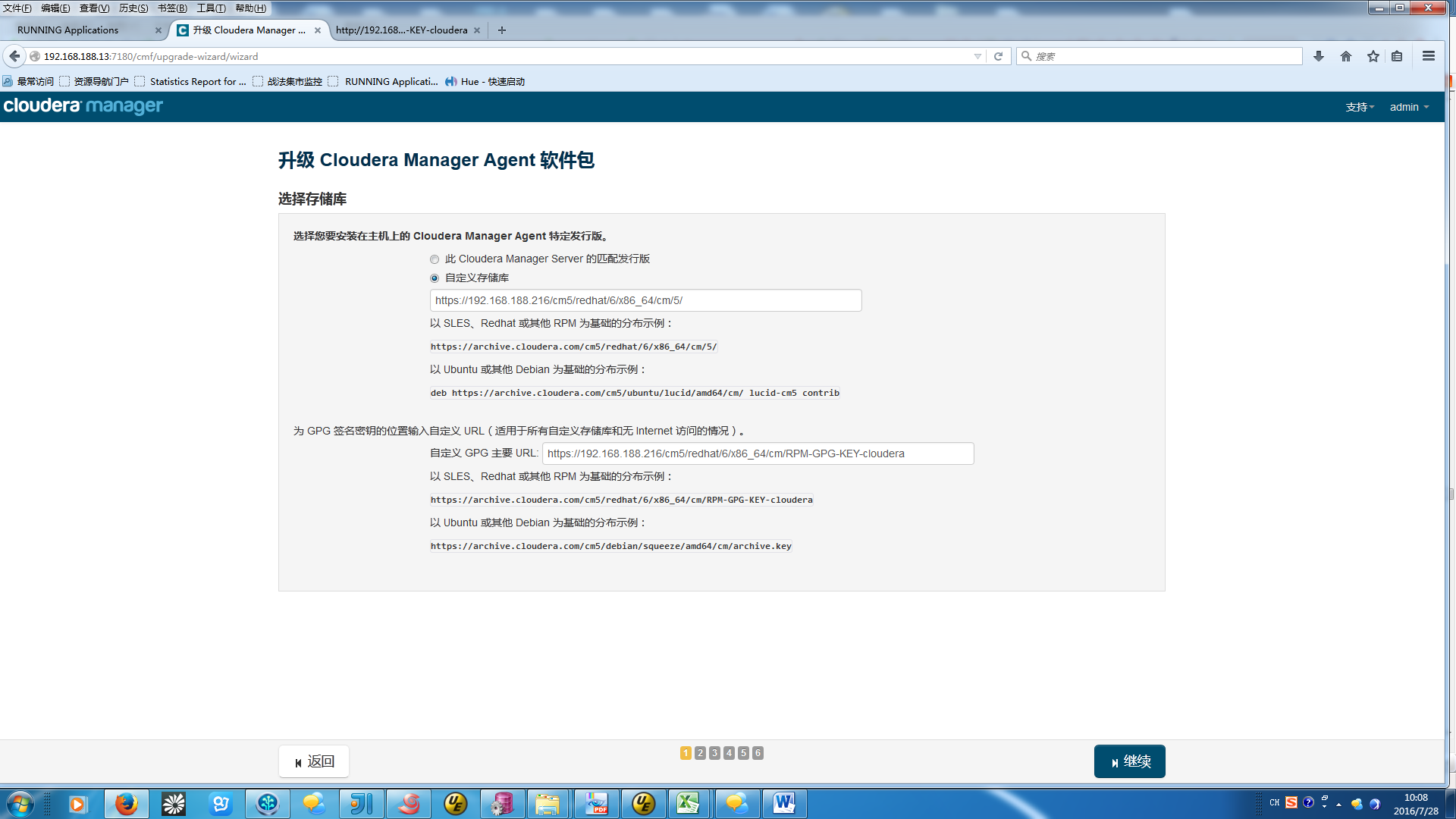
1. 升级向导
   1. 点击重新运行升级向导，点击是，继续



* 1. 点击自定义存储库



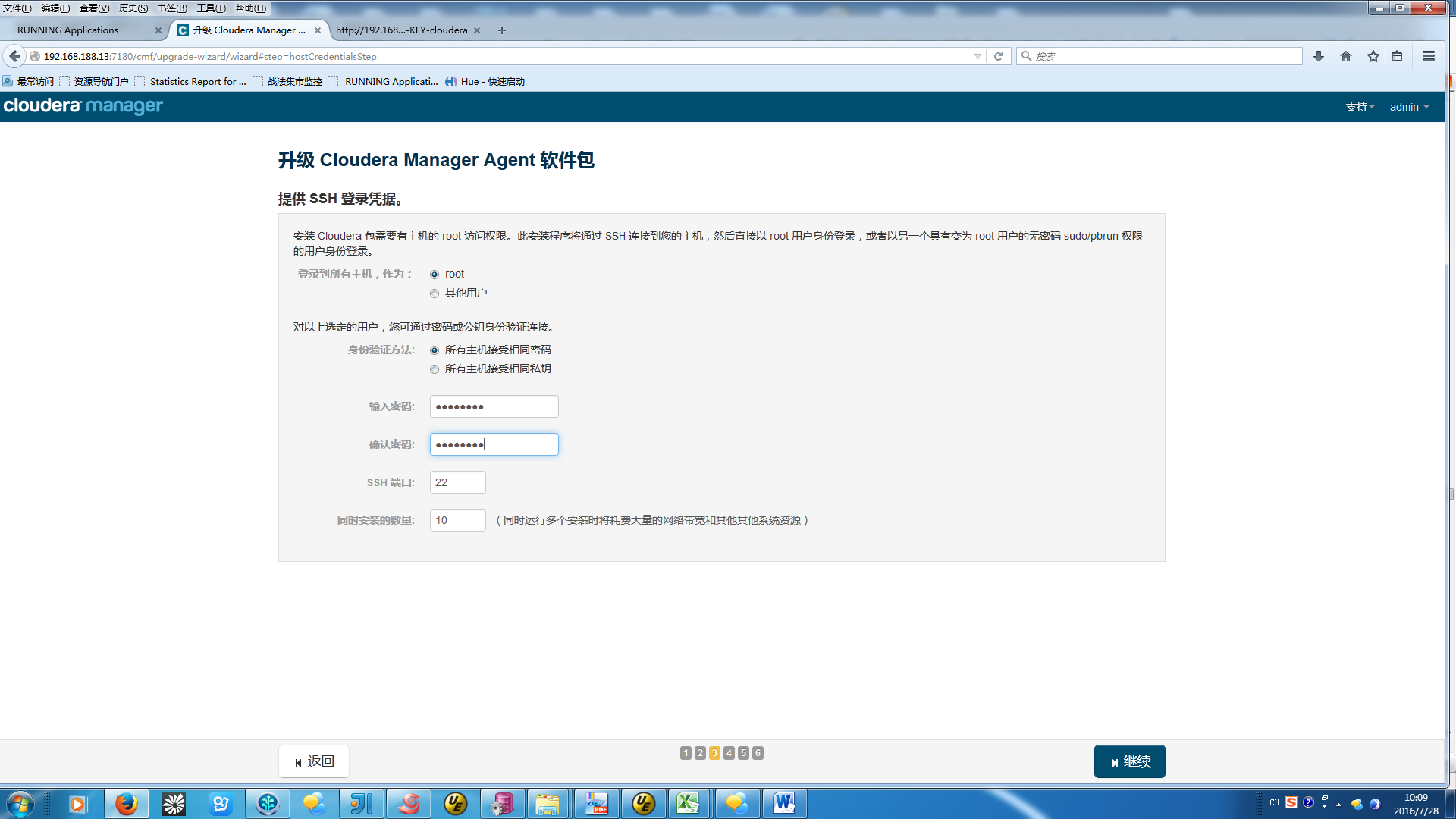
* 1. 将apache服务器上的cm包指定目录填入两个输入框中（注：地址在apache服务器上可以正确访问到），继续



* 1. 点击安装Oracle Java SE开发工具包(JDK)，点击继续按钮



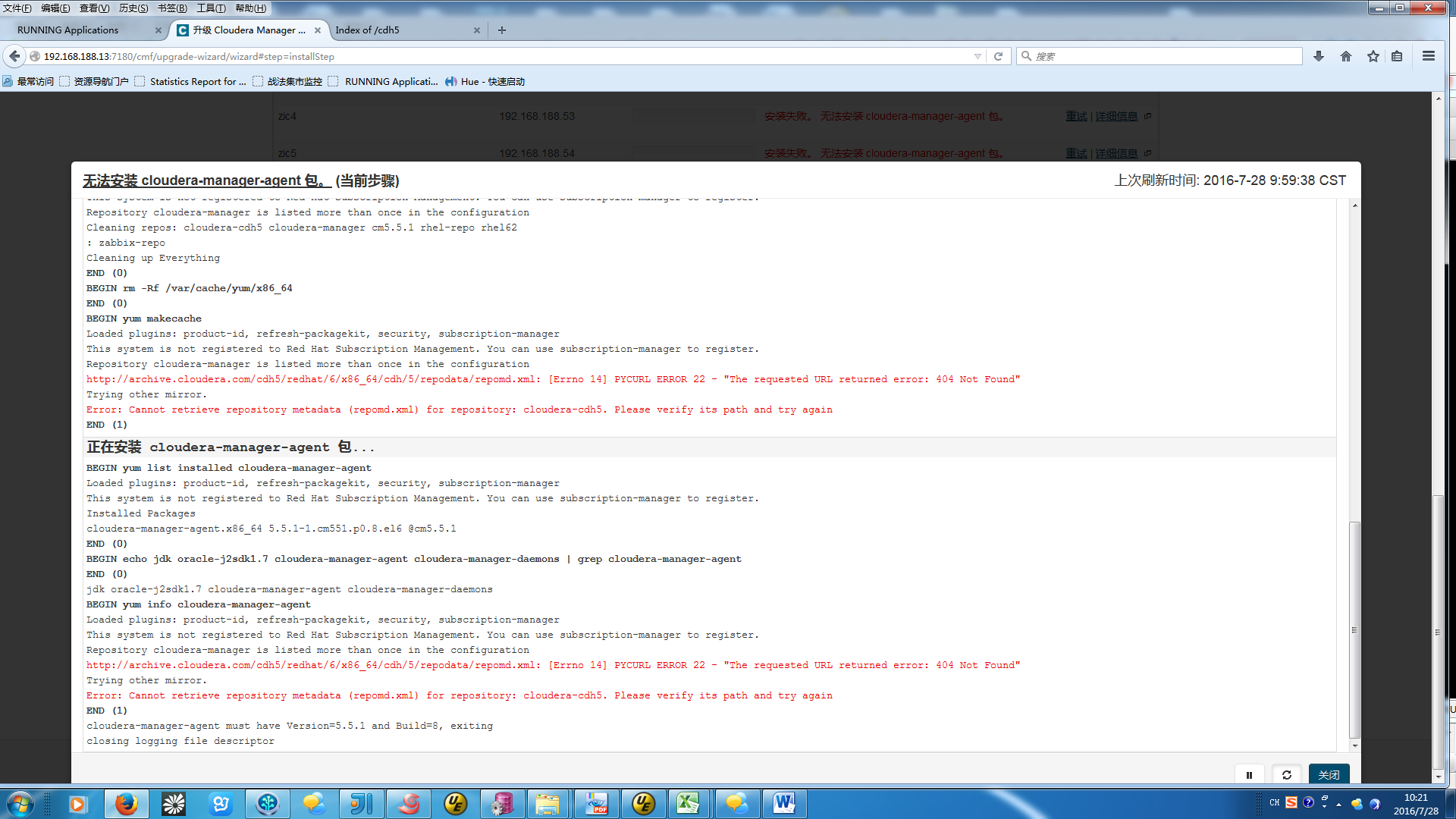
* 1. 输入所有主机的root密码，确定后点击继续



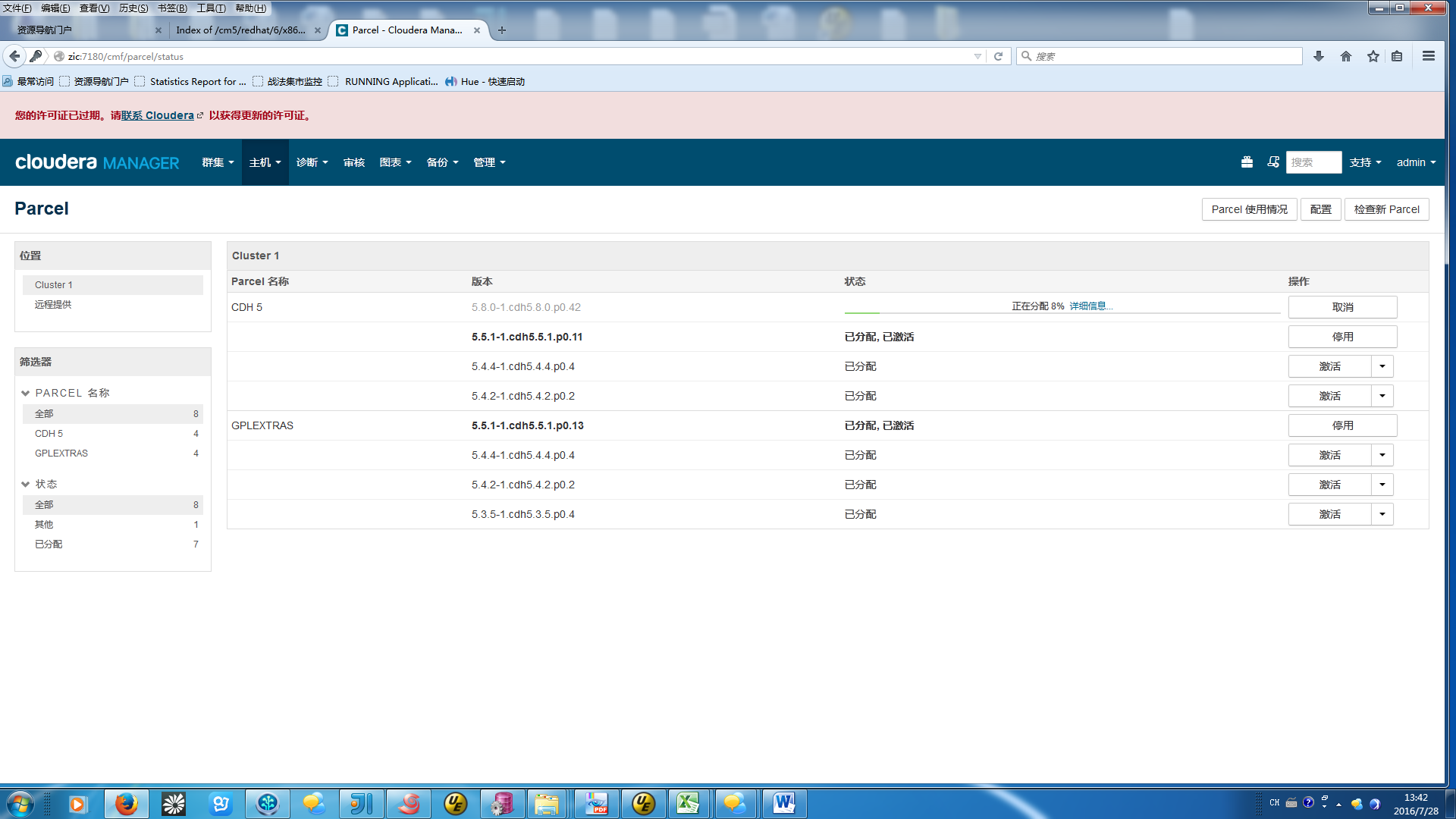
* 1. 接下来就是为所有主机升级cloudera manager操作。

#### 升级过程中遇到的问题

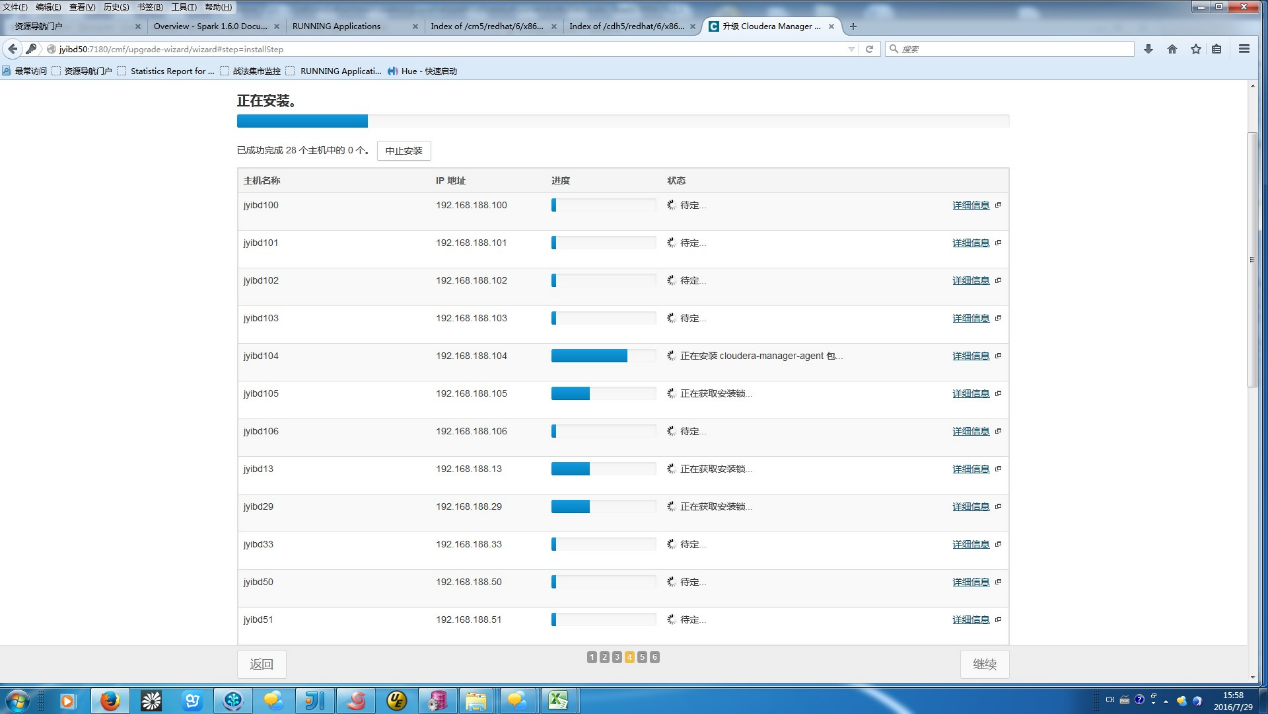
##### 提示Error Cannot retrieve repository metadata [repomod.xml] for cloudera-cdh5.Please verify its path and try again



1. 检查机器的yum及cloudera的yum源配置是否正确
2. 在Cloudera升级步骤(5)中填写的apache上cm5包地址是否正确，协议应该使用http而不是https，不然就会出现这种错误



##### 升级阶段主机显示正在获取安全锁



*解决：删除临时目录中的锁，执行以下命令*

*rm -rf /tmp/scm\_prepare\_node.\**

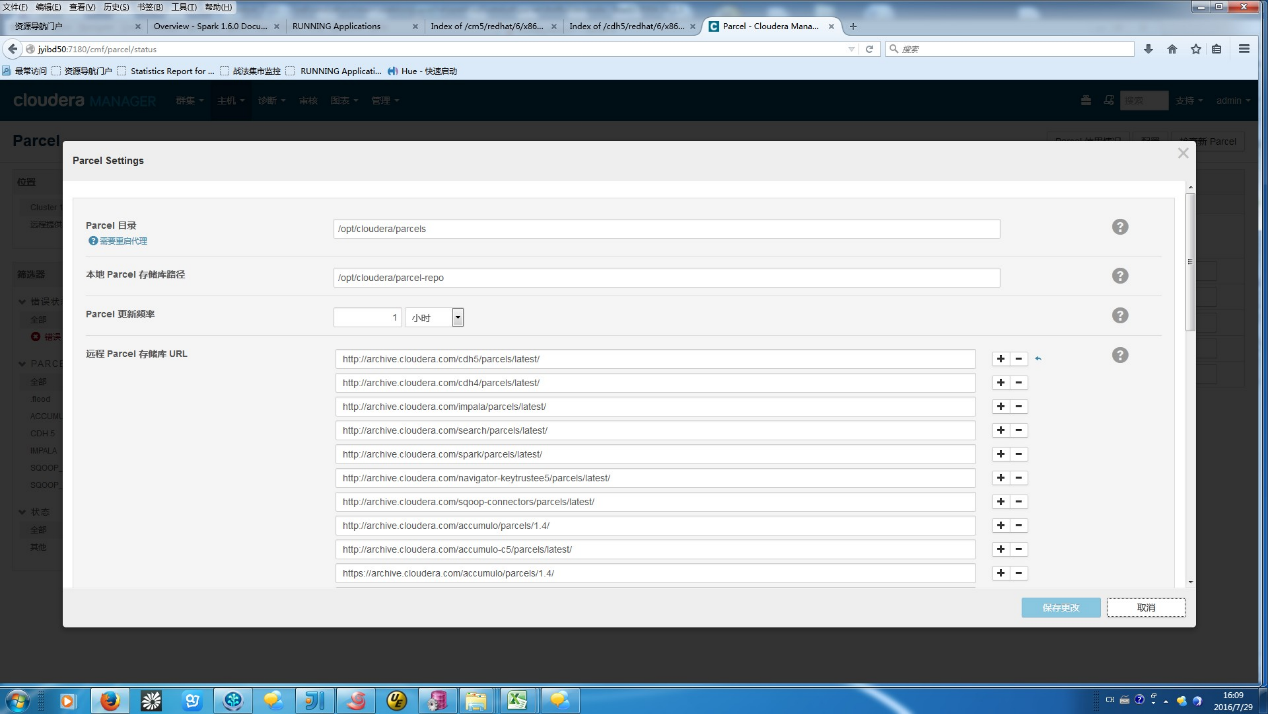
*rm -rf /tmp/.scm\_prepare\_node.lock*

### Parcel 升级

Parcel 包括 hadoop生态圈中组件，如：hive、hbase、hue、spark等。

#### 升级Parcel

1. 配置Parcel，点击主机->Parcel->配置，将仓库中的cdh加入到列表中：http://192.168.188.216/cdh5/parcels/



1. 在配置好Parcel源后，点击下载按钮，下载完毕后直接点击升级按钮，之后根据向导下一步即完成安装。



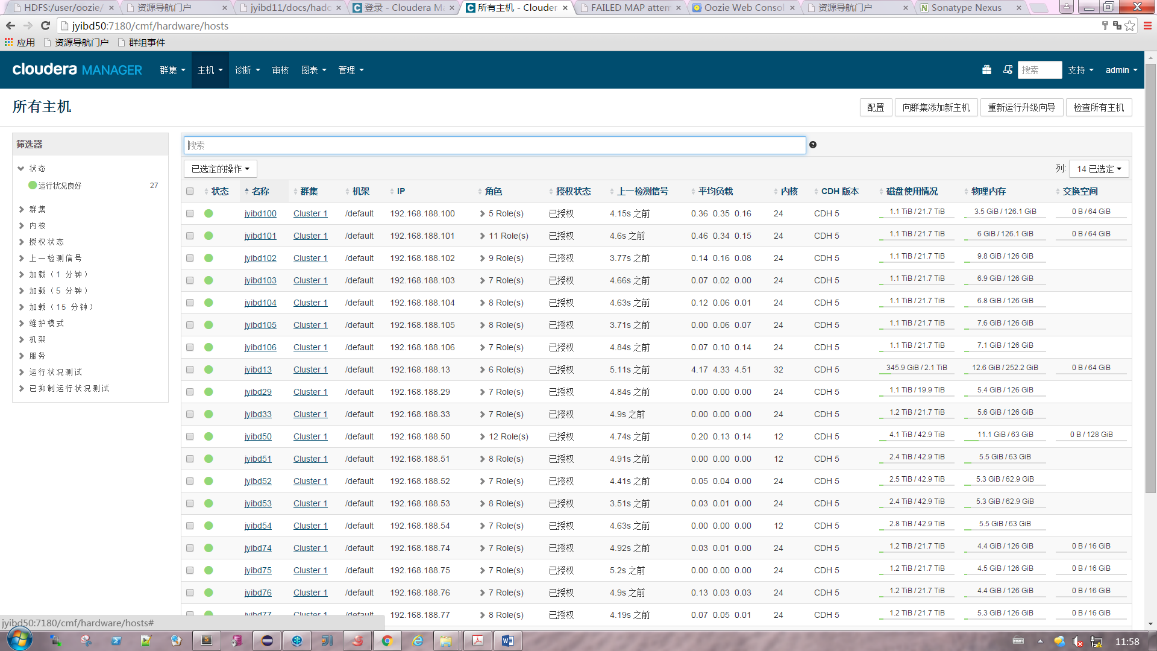
#### 升级Parcel遇到的问题

##### 升级阶段oozie上传共享工具包时超时

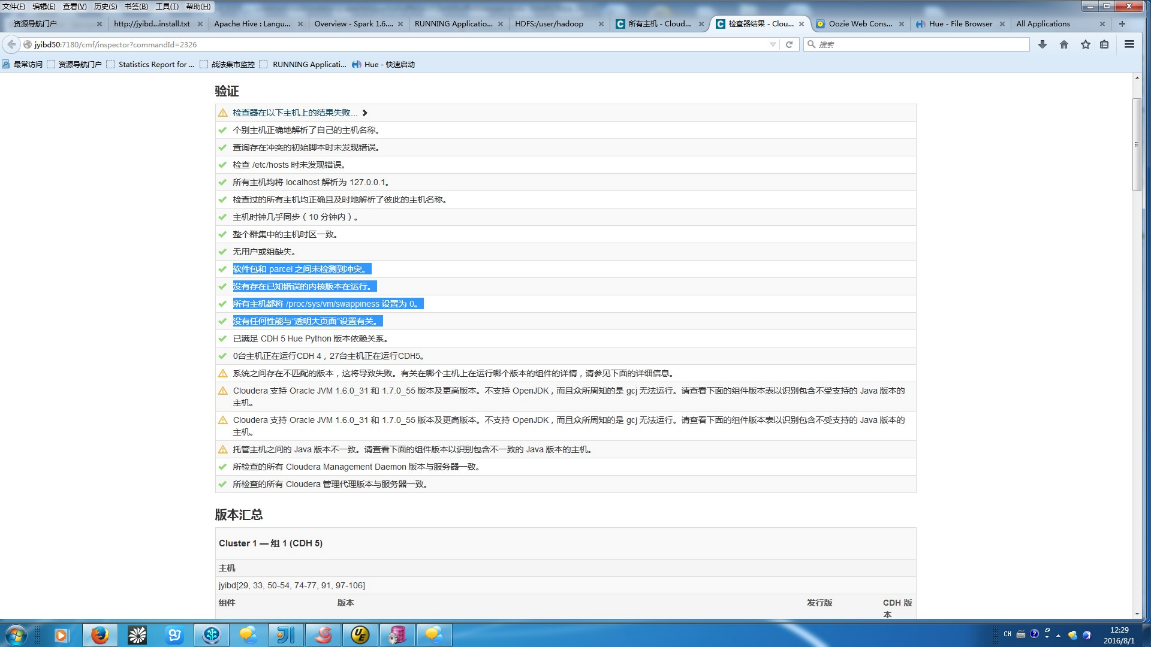
解决：停止除HDFS以外的其它服务，关闭后点击重试按钮，即可完成安装。

##### 版本跨度较大时，比如CDH5.2.0升级到CDH5.8.0，服务器上存放的目录结构已经发生改变，而升级成功后cloudera并没有将命令的软链等重新指定到新的CDH5.8.0，需手动检查解决并重新创建软链

1. 点击 主机 -> 所有主机 -> 检查所有主机



1. 点击显示检查器结果



选中的几项为可能产生的警告。需要进行解决，软件包和 parcel 之间未检测到冲突：若出现警告，则卸载各个机器上的非最新版本的包。

卸载命令：

*#yum remove –y \*cdhX.Y.Z\**

重新创建软链命令：

*#ls /etc/alternatives/ | xargs -i rm -rf /usr/bin/{} && ls /etc/alternatives/ | xargs -i ln -s /etc/alternatives/{} /usr/bin/{}*

##### 运行pig，hive等时报错：

报错信息：

java.lang.RuntimeException: native snappy library not available: this version of libhadoop was built without snappy support

解决办法：

点击Yarn -> 配置 –> 类别 -> 高级, hadoop-env.sh 的 Gateway 客户端环境高级配置代码段（安全阀）中增加环境变量：JAVA\_LIBRARY\_PATH={{JAVA\_LIBRARY\_PATH}}，即将目录$HADOOP\_HOME/lib/native加入到环境变量中。

## Cloudera 安装

### 卸载集群中旧的Cloudera

若机器上安装过Cloudera，安装新版Cloudera需卸载机器上Cloudera相关目录及文件，参考：



### 安装Cloudera

1. 选择一台机器作为Cloudera Manager服务器，参考环境准备步骤中准备好离线仓库。
2. 修改cloudera-manager-installer.bin为可执行脚本：
3. #chmod u+x cloudera-manager-installer.bin
4. 确保主机的yum源配置成功
5. 执行脚本cloudera-manager-installer.bin
6. #./cloudera-manager-installer.bin --skip\_repo\_package=1

*注：由于使用本地仓库，需加--skip\_repo\_package=1参数，不让脚本执行使用在线仓库源，而使用本地仓库地址。*

1. 按照向导提示操作即完成Cloudera Manager安装，等待几分钟后访问host:7180进入管理界面，完成向集群中添加主机、容量分配等操作。

CM主界面如图所示  
 

添加服务



选择需要的服务，点击下一步进行安装下一步即可

例：安装hive服务指定数据库界面



## 集群优化

### 关闭 Linux THG 服务

检查集群中的各个主机的THG（对虚拟化等的内存资源分配是有好处的，但是对hadoop离线计算IO密集型操作是没有优势的，关闭THG可加快处理速度）。

1. 查看THG

*#cat /sys/kernel/mm/redhat\_transparent\_hugepage/defrag*

1. 关闭THG

*#echo never > /sys/kernel/mm/redhat\_transparent\_hugepage/defrag*

### 设置 linux 内核参数

vm.swappiness值的范围为0~100，作用是控制应用数据在物理内存和虚拟内存之间的交换，值越低，交换的越少。默认值为60。

查看集群各个主机的此参数值：

*cat /proc/sys/vm/swappiness*

建议调整值为1：

*#sysctl -w vm.swappiness=1*

### 配置 HDFS

点击HDFS -> 配置 -> 高级：hdfs-site.xml 的 HDFS 服务高级配置代码段（安全阀），加入配置使用公平队列。

<property>

<name>ipc.8020.callqueue.impl</name>

<value>org.apache.hadoop.ipc.FairCallQueue</value>

</property>

### 配置 Yarn 资源

点击Yarn -> 资源管理：

1. 设置ApplicationMaster Java最大堆栈：800M(AM内存默认1G)
2. 容器内存yarn.nodemanager.resource.memory-mb

*计算一个节点需要分配的容器内存方法：*

*主机内存-操作系统预留内存(12G) - Cloudera Manager Agent(1G) - HDFS DN(1G) – Yarn NM(1G)*

*= 主机内存-15G*

*如果安装了hive.需减掉12G左右内存*

*如果安装了hbase.还需减掉12-16G内存。*

*如果安装impala.还需减掉至少16G内存。*

*例：64G内存主机，如果安装了hbase,hive，则建议分配的容器内存大约为：25~30G*

1. 容器虚拟CPU内核yarn.nodemanager.resource.cpu-vcores
2. 计算一个节点需要分配的容器虚拟内核方法：

*(主机cpu核数 – 系统预留1 – Cloudera1 – HDFS1 – Yarn NN 1) \* 4*

*Hbase : -1*

*例：24核机器，为yarn分配可用cpu核数大约20核左右，按照 核数:处理任务数=1:4(比例可酌情调整)，建议分配为80。由于本次集群CPU计算能力没达到官网建议的比例的要求，大约分配的比例为1:2，分配的核数为30核左右。*

1. 高级配置中：mapred-site.xml 的 MapReduce 客户端高级配置代码段（安全阀）

<property>

<name>mapreduce.tasktracker.outofband.heartbeat</name>

<value>true</value>

</property>

### 配置 oozie

点击oozie –> 配置 -> 高级 ： oozie-site.xml 的 Oozie Server 高级配置代码段（安全阀），增加配置：

<property>

<name>oozie.launcher.fs.hdfs.impl.disable.cache</name>

<value>true</value>

</property>

<property>

<name>oozie.action.max.output.data</name>

<value>5000000</value>

</property>

## FAQ

在升级过程中需要删除等批量操作时，执行如下脚本。



使用方法:

#./roundRobinHostCommand.sh autohosts.txt ${command}

注：${command}需要在个节点上执行的命令