Kettle部署

1. Kettle简介

Kettle是一款国外开源免费的ETL工具，纯java编写。提供的数据抽取能力高效稳定。且可以在Window、Linux、Unix环境上运行，依赖少，部署方便。且支持集群搭建，具有良好的扩展能力。

Kettl支持上下版本兼容性，kettle 5版本可可调用最新版本（7.1），且安装部署等都一致。以下以最新版本7.1为例，7.1版本依赖JDK8。

1. 安装依赖NFS

* 安装NFS

说明：正常RedHat、Centos均默认安装，如果版本为Centos Mini则需执行

yum install -y nfs-utils rpcbind

* 配置NFS

vi /etc/sysconfig/nfs

# 文件最后写入

RQUOTAD\_PORT=30001

LOCKD\_TCPPORT=30002

LOCKD\_UDPPORT=30002

MOUNTD\_PORT=30003

STATD\_PORT=30004

wq!

* 创建共享目录

groupadd -g 500 weblogic

useradd -u 500 -g 500 weblogic

mkdir -p /shareDatas

* 创建项目所需共享目录(需要确保拥有者是500)

mkdir -p /shareDatas/recommendDatas

mkdir -p /shareDatas/consumerDatas

mkdir -p /shareDatas/FileRepo

* 查看应用工程的ID(如果未weblogic部署,默认为500)

使用部署用户登录应用系统

如：[weblogic@hastzjapkc01 ~]$ id

uid=500(weblogic) gid=500(weblogic) groups=500(weblogic)

* 配置共享目录权限

说明：

第一行IP为平台(或应用)的IP，rw为读写权限

第二行IP为Kettle部署设备的IP（段），ro为读权限

vi /etc/exports

/ shareDatas 192.168.80.101 (rw,sync, all\_squash,anonuid=500,anongid=500)

/ shareDatas 192.168.80.\*(ro,sync, all\_squash,anonuid=500,anongid=500)

wq!

* 启动与自启动

chkconfig nfs on

chkconfig rpcbind on

service nfs start

service rpcbind start

1. Kettle部署—RedHat

* 安装JDK

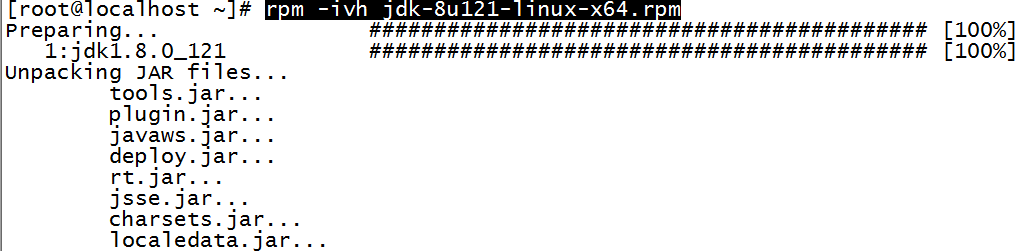
参考: <http://jingyan.baidu.com/article/0964eca26917b18285f53616.html>

1. 官网下载的Windows64位JDK8，选择rpm文件

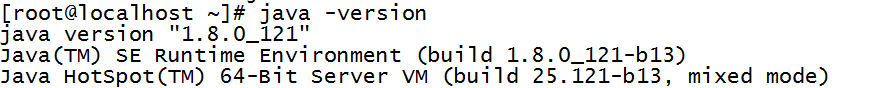
<http://www.oracle.com/technetwork/java/javase/downloads/index.html>



1. 执行rpm -ivh jdk-8u121-linux-x64.rpm



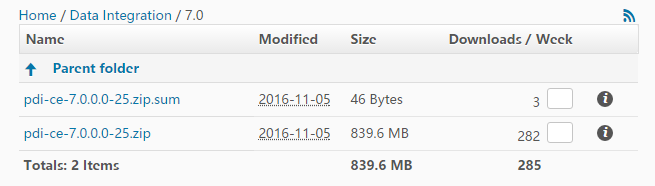
1. 输入java –version，出现java版本则安装成功。



* 安装Kettle

1. 下载7.1版Kettle

<https://sourceforge.net/projects/pentaho/files/Data%20Integration/7.1/>

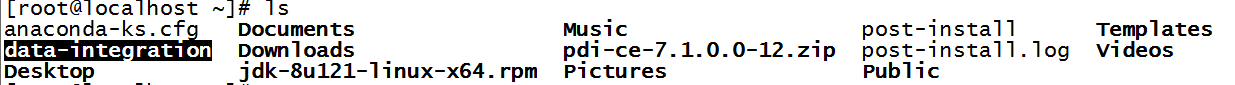


1. 安装nohub、zip、unzip 、NFS

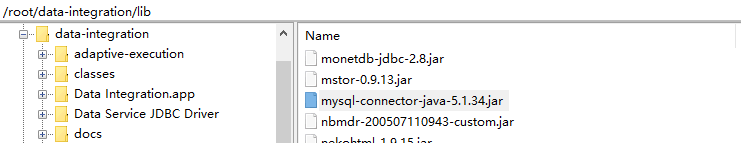
说明：正常RedHat、Centos均默认安装，如果版本为Centos Mini则需执行

yum -y install coreutils zip unzip nfs-utils

1. 执行unzip pdi-ce-7.1.0.0-12.zip解压缩，解压出的文件夹名称为“data-integration”



1. 将数据库驱动 jar包放入data-integration\lib中



1. 设置执行权限，在目录data-integration下

chmod +x \*.sh

1. 编辑data-integration/pwd/carte-config-master-8080.xml文件,样例如下:

<slave\_config>

<slaveserver>

<!-- 第一台设备为master1,第二台为master2 -->

<name>master1</name>

<!-- 输入本机IP -->

<hostname>192.168.80.131</hostname>

<port>8080</port>

<master>Y</master>

<username>cluster</username>

<password>cluster</password>

</slaveserver>

<max\_log\_lines>50000</max\_log\_lines>

<max\_log\_timeout\_minutes>720</max\_log\_timeout\_minutes>

<object\_timeout\_minutes>720</object\_timeout\_minutes>

<repository>

<name>KettleFileRepo</name>

<username>guest</username>

<password>guest</password>

</repository>

</slave\_config>

1. 在data-integration目录下执行如下命令

./carte.sh /root/data-integration/pwd/carte-config-master-8080.xml

1. 在浏览器输入<http://IP:8080>, 如果出现输入用户名/密码提示标识启动成功



1. 执行如下命令,关停Kettle

ps -ef|grep carte.sh

kill -9 \*\*\*

1. 在Mysql中创建Kettle的database:kettle,编码为UTF-8(Oracle为创建服务)



1. 在database:kettle中执行如下SQL脚本





1. 挂载共享目录

mkdir -p /data/FileRepo

mount -t nfs <NFS服务IP>:/ shareDatas / shareDatas

1. 配置开机启动共享目录

vi /etc/fstab

增加一条记录:

<NFS服务端IP>:/shareDatas /shareDatas nfs defaults 0 0

wq!

1. 配置本机资源信息

在/data/FileRepo目录中写入master1.ksl文件

说明:

<name>第一台为master1,第二台为master2,与文件名称一致

<hostName>为本机IP

vi /data/FileRepo/master1.ksl

<slaveserver>

<name>master1</name>

<hostname>192.168.80.131</hostname>

<port>8080</port>

<webAppName/>

<username>cluster</username>

<password>Encrypted 2be98afc86aa7f2e4cb1aa265cd86aac8</password>

<proxy\_hostname/>

<proxy\_port/>

<non\_proxy\_hosts/>

<master>Y</master>

<sslMode>N</sslMode>

</slaveserver>

chown -R weblogic:weblogic /data/FileRepo/

chmod -R 777 /data/FileRepo/

1. 进入Kettle的环境目录(root用户为/root/.kettle),编辑repositories.xml



1. 为平台/应用挂载共享目录

参见12~13步骤

1. 修改Java项目中的Kettle配置文件,其配置与Kettle环境一致

修改项目的kettle\_env.properties

修改项如下:

# 平台应用的共享目录,配置参考上边

KETTLE\_FILE\_REPOSITORY\_META\_PATH=/data/FileRepo

# 数据库类型(Oracle数据库填写Oracle)

KETTLE\_RECORD\_DB\_TYPE=MySQL

# 数据库IP

KETTLE\_DATABASE\_REPOSITORY\_DB\_HOST=192.168.80.138

# 数据库DataBase(Oracle对应为服务名)

KETTLE\_DATABASE\_REPOSITORY\_DB\_DATABASENAME=kettle

# 数据库端口(Oracle为1521)

KETTLE\_DATABASE\_REPOSITORY\_DB\_PORT=3306

# 数据库用户名

KETTLE\_DATABASE\_REPOSITORY\_DB\_USER=root

# 数据库密码

KETTLE\_DATABASE\_REPOSITORY\_DB\_PASSWD=123456

1. 设置日志

vi /etc/logrotate.d/kettle

/root/data-integration/log/\*.log {

missingok

notifempty

size 256k

rotate 7

daily

dateext

create 644 root root

}

强制执行logrotate -f /etc/logrotate.d/kettle

测试执行 logrotate -d /etc/logrotate.d/kettle

1. 进入data-integration目录，执行如下命令正式启动

mkdir –p /root/data-integration/log

nohup ./carte.sh pwd/carte-config-master-8080.xml >/root/data-integration/log/console.log 2>/root/data-integration/log/err.log