

kettle部署说明书

修订历史：

修订版本号	修订日期	修订人	说明
v1.0	2015-09-21	杨立平	创建文档，实现kettle的shell调度

一、下载解压kettle	2
二、jdk部署说明	2
三、mysql驱动部署说明	3
四、部署ktr和kjb	3
(1) 部署kettle项目	3
(2) 部署执行	3
五、附录	4
(1) kettle调度shell脚本	4

一、下载解压kettle

从官网上下载的kettle（我的是pdi-ce-5.4）解压后放在该文件夹下。在解压的路径下找到kitchen.sh所在的路径，进入该目录，通过chmod +x *.sh赋予.sh文件执行权限。

```
[root@123qwe ~]# ls -l
```

总用量 8

```
drwxr-xr-x. 12 root root 4096 9月 15 13:02 data-integration
drwxr-xr-x. 15 root root 4096 9月 15 08:49 tools
```

```
[root@123qwe data-integration]# ls -l *.sh
```

```
-rwxr-xr-x. 1 root root 405 6月 14 14:35 carte.sh
-rwxr-xr-x. 1 root root 140 6月 14 14:35 encr.sh
-rwxr-xr-x. 1 root root 275 6月 14 14:35 import.sh
-rwxr-xr-x. 1 root root 308 6月 14 14:35 kitchen.sh
-rwxr-xr-x. 1 root root 274 6月 14 14:35 pan.sh
-rwxr-xr-x. 1 root root 305 6月 14 14:35 runSamples.sh
-rwxr-xr-x. 1 root root 3710 6月 14 14:35 set-pentaho-env.sh
-rwxr-xr-x. 1 root root 5341 6月 14 14:35 spoon.sh
-rwxr-xr-x. 1 root root 1628 6月 14 14:35 yarn.sh
```

键入./kitchen.sh，测试，可以得到输入信息说明部署成功。

```
[root@123qwe data-integration]# ./kitchen.sh
```

Options:

-rep	= Repository name
-user	= Repository username
-pass	= Repository password
-job	= The name of the job to launch
-dir	= The directory (dont forget the leading /)
-file	= The filename (Job XML) to launch
-level	= The logging level (Basic, Detailed, Debug, Rowlevel)
-logfile	= The logging file to write to
-listdir	= List the directories in the repository
-listjobs	= List the jobs in the specified directory
-listrep	= List the available repositories
-norep	= Do not log into the repository
-version	= show the version, revision and build date
-param	= Set a named parameter <NAME>=<VALUE>.

查看Kettle版本号，如下：

```
[root@123qwe data-integration]# ./pan.sh -version
```

```
2015/09/21 14:59:38 - Pan - Kettle 版本5.4.0.1-130, 编译1, 编译日期 : 2015-06-14_12-34-55
```

```
[root@123qwe data-integration]# ./kitchen.sh -version
```

```
2015/09/21 14:59:55 - Kitchen - Kettle version 5.4.0.1-130, build 1,
build date : 2015-06-14_12-34-55
```

二、jdk部署说明

确认jdk以及安装，如果没有安装，进行安装完即可。

```
[root@123qwe ~]# java -version
```

```
java version "1.7.0_79"
```

OpenJDK Runtime Environment (rhel-2.5.5.3.el6_6-x86_64 u79-b14)
OpenJDK 64-Bit Server VM (build 24.79-b02, mixed mode)

三、mysql驱动部署说明

这里使用的mysql版本如下：

```
[root@123qwe ~]# mysql -v
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 783
Server version: 5.6.26 MySQL Community Server (GPL)
```

将mysql-connector-java-commercial-5.1.25-bin.jar复制到kettle解压主目录的lib下：

```
[root@123qwe ~]# cp mysql-connector-java-commercial-5.1.25-bin.jar r
data-integration/lib/
```

注意：这里的mysql一对要和kettle的版本匹配，不然读取Mysql的时候会出错。

四、部署ktr和kjb

(1) 部署kettle项目

将kettle所有的文件复制到kettle服务器，如下：

```
[root@123qwe hb]# ll
总用量 144
-rwx-----. 1 root root 16635 9月 21 13:43 HBCL_STAT_DAILY_SLEEP.kjb
-rwx-----. 1 root root 11237 9月 21 13:43 HBCL_STAT_DAILY_SLEEP.ktr
-rwx-----. 1 root root 22913 9月 21 13:43 HBML_RC_RECORD.ktr
-rwx-----. 1 root root 14959 9月 21 13:43 HBML_TP_TEMPLATE.ktr
-rwx-----. 1 root root 26270 9月 21 13:43 HBML_UR_USERDEVICE_RELA_H.ktr
-rwx-----. 1 root root 16717 9月 21 13:43 HBML_UR_USERINFO.ktr
-rwx-----. 1 root root 15164 9月 21 13:43 HBML_UR_USERTEPA_RELA.ktr
-rwxr-xr-x. 1 root root 694 9月 21 15:22 kettle.sh
drwxr-xr-x. 2 root root 4096 9月 21 15:40 log
```

(2) 部署执行

然后执行kettle.sh脚本运行kettle的job任务。详细的kettle调度脚本参见本文档附录，部署后，手动进行运行，脚本如果没有添加参数，默认当天的日期，如果带了日期，则使用参数日期，如下：

```
[root@123qwe hb]# ./kettle.sh
没有输入日期参数，取当前日期：2015-09-21
ETL抽取执行完成，没有出现错误！
```

```
[root@123qwe hb]# ./kettle.sh 2015-09-07
ETL当前执行日期为：2015-09-07
ETL抽取执行完成，没有出现错误！
```

如果日期格式不对，或者非法日期，则会提示以下错误：

```
[root@123qwe hb]# ./kettle.sh 20150921
失败：确认输入日期参数格式<yyyy-mm-dd> 例如：2000-01-01
```

```
[root@123qwe hb]# ./kettle.sh 2015-09-31
date: 无效的日期"2015-09-31"
失败：确认输入日期参数格式<yyyy-mm-dd> 例如：2000-01-01
```

五、附录

(1) kettle调度shell脚本

```
[root@123qwe hb]# cat kettle.sh
#!/bin/bash

#set kettle project path
basepath=$(cd `dirname $0`; pwd)
kettle_path="/opt/pdi-ce-5.4.0.1-130/data-integration"

mkdir -p "${basepath}/log"

if [ $# -eq 1 ]
then
    date -d $1 "+%Y-%m-%d" | grep -q $1
    if [ $? -eq 1 ]
    then
        echo "失败: 确认输入日期参数格式<yyyy-mm-dd> 例如:2000-01-01"
        exit 1
    else
        # 通过的日期即为符合格式的合法日期
        etldate=$1
        echo "ETL当前执行日期为: $etldate"
    fi
else
    etldate=`date "+%Y-%m-%d"`
    echo "没有输入日期参数, 取当前日期: $etldate"
fi

##go to kettle soft dir
cd $kettle_path
./kitchen.sh -file=${basepath}/HBCL_STAT_DAILY_SLEEP.kjb -level=basic
-param:etl_dt=${etldate} >>${basepath}/log/kettle_${etldate}.log

echo "ETL抽取执行完成, 没有出现错误! "
```