

kettle linux 下的使用 - CSDN博客

/*****/

kettle linux 下的使用

一本例 的 linux 无法使用图形，也没有中文环境，故而使用命令行的方式

一将 kettle 安装在一台数据库(目的端或源端) 上可以大幅提高迁移的速度。

/*****/

一 安装

1 将kettle 的 data-integration.rar 文件包解压，重新压缩成ZIP 格式，上传到服务器/home/informix/kettle 下

2 解压

```
# unzip data-integration.zip
```

```
# cd /home/informix/kettle/data-integration
```

```
# chmod +x *.sh
```

二 kjb, ktr文件的更改

1 将中文文件名改为 英文名文件名

如下：

表名称获取 get_tablename.ktr

导出数据 export_data.ktr

设置变量 set_variable.ktr

数据抽取流程 data_extract_flow.kjb

总流程 all_flow.kjb

2 在 WIN 下 打开上面的文件，其中应用其他文件的地方要更改为相应的英文名，各个步骤的作业名称也改为英文

3 数据源 （虽然数据源设为了共享，但是也可能由于某些操作导致 对应文件里的数据源不同，需要检查脚本）

注： 连接 ORACLE RAC 数据库时 ， 需要指定 连接节点的实例名，不能使用 SERVICE_NAME

1) 在 WIN 下设置数据源信息，但是在Linux 下要检查相应的文件的数据源信息(数据库名为test2)

如下:

```
[informix@pt120 data-integration]$ grep test *.kjb
```

```
all_flow.kjb:      test2
```

```
data_extract_flow.kjb:      test2
```

```
[informix@pt120 data-integration]$ grep test *.ktr
```

```
export_data.ktr:      test2
```

```
export_data.ktr:
```

```
export_data.ktr:
```

```
get_tablename.ktr:      test2
```

```
get_tablename.ktr:
```

```
get_tablename.ktr:
```

```
get_tablename.ktr:
```

```
set_variable.ktr:      test2
```

```
set_variable.ktr:
```

```
set_variable.ktr:
```

2) 表名称获取(在WIN 下编辑，此处要 order by TABLE_NAME 排序，可方便的 看到进度和 从中断或失败的操作 继续运行)

```
select TABLE_NAME from user_tables WHERE TABLE_NAME not in  
( 'SYS_LOG', 'DEMO_BUSINESS_TRIP', 'SYS_BPM_HIST_TASK', 'SYS_LOG_1466988417566',
```

```
 'SYS_BPM_HIST_PROCIINST', 'BPM_HIST_PROCIINST', 'BPM_VARIABLE_EXT4', 'BPM_VARIABLE_EXT7', 'BPM_VARIABLE_EXT1', 'BPM_VARIABLE_EXT2' ,
```

```
 'BPM_VARIABLE_EXT3', 'BPM_VARIABLE_EXT5', 'BPM_VARIABLE_EXT6', 'BPM_VARIABLE_EXT8', 'DEMO_BUSINESS_FLOW' )
```

```
order by TABLE_NAME
```

4 此设置可解决TEXT 字段乱码问题 (export LANG=en_us.GB18030 会是乱码)

```
[informix@ptl20 data-integration]$ export LANG=zh_CN.GB18030
```

5 命令行方式运行

1) 运行 ktr

```
[informix@ptl20 data-integration]$ ./pan.sh -file:get_tablename.ktr -logfile=kettle.log
```

2) 运行kjb

注: file 为 kjb 的文件, logfile为操作日志文件

```
[informix@ptl20 data-integration]$ ./kitchen.sh -file:all_flow.kjb -logfile=kettle.log
```

```
[informix@ptl20 ~]$ ps -ef|grep kitchen
```

```
informix 3718 2661 0 13:42 pts/3 00:00:00 /bin/sh ./kitchen.sh -file:all_flow.kjb
```

```
informix 3721 3718 0 13:42 pts/3 00:00:00 /bin/sh /home/informix/kettle/data-  
integration/spoon.sh -main org.pentaho.di.kitchen.Kitchen -file:all_flow.kjb
```

```
informix 3749 3721 55 13:42 pts/3 00:01:09 java -Xmx512m -XX:MaxPermSize=256m -  
Djava.library.path=../libswt/linux/x86_64/ -DKETTLE_HOME= -DKETTLE_REPOSITORY= -  
DKETTLE_USER= -DKETTLE_PASSWORD= -DKETTLE_PLUGIN_PACKAGES= -DKETTLE_LOG_SIZE_LIMIT= -  
DKETTLE_JNDI_ROOT= -jar /home/informix/kettle/data-integration/launcher/pentaho-application-  
launcher-5.4.0.1-130.jar -lib ../libswt/linux/x86_64/ -main org.pentaho.di.kitchen.Kitchen  
-file:all_flow.kjb
```

```
informix 3996 3953 0 13:44 pts/5 00:00:00 grep kitchen
```

--日志信息:

2017/04/19 13:50:13 - Kitchen - Finished!

2017/04/19 13:50:13 - Kitchen - Start=2017/04/19 13:42:13.473, Stop=2017/04/19 13:50:13.281

2017/04/19 13:50:13 - Kitchen - Processing ended after 7 minutes and 59 seconds (479 seconds total).

三 对于中断表的迁移, 从断点继续执行迁移(单表)

1 从8T 端获取 最后一条记录的主键编号(dbid_)

```
> select first 1 dbid_ from BPM_LOB_EXT6 order by rowid desc;
```

dbid_

8a58a64b5b17f8f6015b2272d4ca67ea

1 row(s) retrieved.

2 从ORACLE 找出 该记录的ROWID

```
SQL> select rowid from BPM_LOB_EXT6 where dbid_='8a58a64b5b17f8f6015b2272d4ca67ea';
```

ROWID

AAAEiEAAGAAF135AAD

3 从KETTLE 中输入查询条件，查找 大于该 rowid 的记录

在KETTLE 导出数据.ktr 或(export_data.ktr) table_input 步骤中 加入rowid > 的条件

如下:

```
select * from ${TABLENAME} where rowid>'AAAEiEAAGAAF135AAD'
```

```
/*  
*****  
*****/
```

四 关于在Linux 下 共享数据源的问题。

1 5, 6 版本 shared.xml 文件 不生效 (原因不明)

2 7.0 版本 (JDK 使用 1.8) shared.xml 有效, 这样改变数据源的话, 只需更改这一个文件

```
/*  
*****  
*****/
```

1 安装 JDK 1.8 (D:\jibunn\software\jdk\jdk8\jdk-8u131-linux-x64.tar.gz)

```
# su - root
```

```
cd /file
```

```
gunzip jdk-8u131-linux-x64.tar.gz
```

```
tar -xvf jdk-8u131-linux-x64.tar
```

```
# su - informix
```

```
# vi .bash_profile
```

```
export PATH=/file/jdk1.8.0_131/bin:$INFORMIXDIR/bin:/usr/bin:${PATH}..
```

```
[informix@node1 file]$ java -version
```

```
java version "1.8.0_131"
```

```
Java(TM) SE Runtime Environment (build 1.8.0_131-b11)
```

```
Java HotSpot(TM) 64-Bit Server VM (build 25.131-b11, mixed mode)
```

2 安装 data-integration 7.0

```
[informix@node1 kettle]$ unzip pdi-ce-7.0.0.0-25.zip
```

```
[informix@node1 kettle]$ ll
```

总用量 819896

```
drwxr-xr-x 15 informix informix      4096 11月  5 18:02 data-integration
```

3 创建.kettle 目录及相关文件 (D:\jibunn\software\KETTLE-软件\7.0-data-integration相关配置文件\home-informix.kettle)

```
[informix@node1 ~]$ cd /home/informix
```

```
[informix@node1 ~]$ mkdir .kettle
```

```
[informix@node1 ~]$ cd .kettle
```

将配置文件上传到 /home/informix/aa 目录

```
[informix@node1 .kettle]$ cp /home/informix/aa/*.* .
```

```
[informix@node1 .kettle]$ ll
```

总用量 16

```
-rw-rw-r-- 1 informix informix 178 5月  3 15:20 db.cache-7.0.0.0-25
```

```
-rw-rw-r-- 1 informix informix 271 5月  3 15:20 kettle.properties
```

```
-rw-rw-r-- 1 informix informix 2611 5月  3 15:20 shared.xml
```

```
-rw-rw-r-- 1 informix informix 2610 5月  3 15:20 shared.xml.backup
```

4 拷贝kjb,ktr 文件

```
[informix@node1 data-integration]$ cp /home/informix/kettle/data-integration5/kjb/*.* .
```

5 拷贝 JDBC JAR 包 (需要重新拷贝这2个文件)

```
[informix@node1 data-integration]$ cd /home/informix/kettle/data-integration/lib
```

```
[informix@node1 lib]$ cp /home/informix/kettle/data-integration5/lib/ojdbc6.jar .
```

```
[informix@node1 lib]$ cp /home/informix/kettle/data-integration5/lib/ixjjdbc.jar .
```

6 修改 shared.xml 数据库连接

gbase8t

192.168.6.200

INFORMIX

Native

eeee

9200

informix

7 运行程序

--设置字符集，否则TEXT 字段会时乱码

```
export LANG=zh_CN.GB18030
```

--运行all_flow.kjb

```
./kitchen.sh -file:all_flow.kjb -logfile=kettle.log
```

--证明 生效的是 shared.xml 文件

```
[informix@node1 data-integration]$ grep pt6 *.k*
```

```
all_flow.kjb:      pt6
```

```
all_flow.kjb:      pt6
```

```
data_extract_flow.kjb:      pt6
```

```
data_extract_flow.kjb:      pt6
```

```
export_data.ktr:      pt6
```

```
export_data.ktr:    pt6
```

```
get_tablename.ktr:    pt6
```

```
get_tablename.ktr:    pt6
```

```
set_variable.ktr:    pt6
```

```
set_variable.ktr:    pt6
```

8 utf8 字符集 数据库

```
export DB_LOCALE=zh_cn.utf8
```

```
export CLIENT_LOCALE=zh_cn.utf8
```

```
> create database utf8db;
```

Database created.

```
> create table t11( a int ,b clob,c text);
```

Table created.

```
--jdbc
```

```
jdbc:informix-
```

```
sqli://192.168.6.200:9200/utf8db:INFORMIXSERVER=gbaseexam;DB_LOCALE=zh_CN.utf8;CLIENT_LOCALE=zh_CN.utf8;NEWCODESET=utf-8,utf8,57372
```

```
--kettle
```

```
gbaseexam;DB_LOCALE=zh_CN.utf8;CLIENT_LOCALE=zh_CN.utf8;NEWCODESET=utf-8,utf8,57372
```

```
--kettle shared.xml
```

gbase8t

192.168.6.200

INFORMIX

Native

utf8db

9200

informix

Encrypted 2be98afc86aa7f2e4a217a87fcc9fa6c2

gbaseexam;DB_LOCALE=zh_CN.utf8;CLIENT_LOCALE=zh_CN.utf8;NEWCODESET=utf-8,utf8,57372

```
[informix@node1 ~]$ echo $LANG
```

en_US.UTF-8

```
[informix@node1 ~]$ cd kettle/data-integration
```

```
[informix@node1 data-integration]$ ./kitchen.sh -file:all_flow.kjb -logfile=kettle.log
```

```
> select * from t11;
```

a 1

b

111W0 我们的

c

w我们

a 1

b

W0w哈哈

c

hha的额

2 row(s) retrieved.