

Oracle GoldenGate对接FusionInsight

适用场景

Oracle GoldenGate 12.2 ↔ FusionInsight HD V100R002C60U20

Oracle GoldenGate 12.3 ↔ FusionInsight HD V100R002C70SPC200

Oracle GoldenGate 12.3 ↔ FusionInsight HD V100R002C80SPC100

环境信息

软件信息

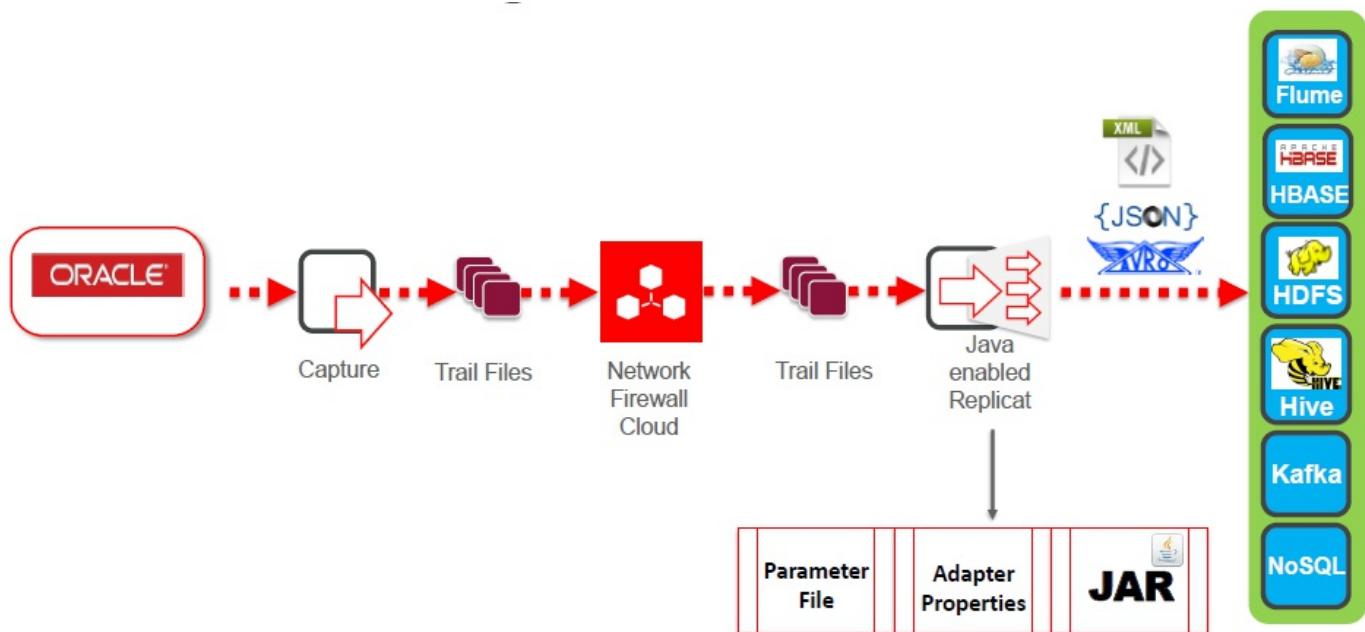
- Oracle GoldenGate 12.2.0.1.1 for Oracle database
- Oracle GoldenGate 12.2.0.1.1 for BigData
- Oracle database 12.1.0.2.0
- jdk-7u71-linux-x64.rpm
- FusionInsight V100R002C60U20

硬件信息

- 源端OGG VM: 162.1.115.68 Redhat6.5 (包含Oracle DB12c的数据库)
- 目标端OGG VM: 162.1.115.69 Redhat6.5 (包含Hadoop的客户端)

拓朴结构

测试拓扑结构如下图所示:



测试表

源端测试表:

在源端Oracle的PDBORCL数据库的test用户下创建test1表，其中ID为主键

```
SQL> desc test1;
Name          Null?    Type
-----        -----   -----
ID           NOT NULL NUMBER
NAME          VARCHAR2(50)

SQL>
```

OGG for Oracle安装

前置条件：完成oracle12c数据库的安装（IP：162.1.115.68）

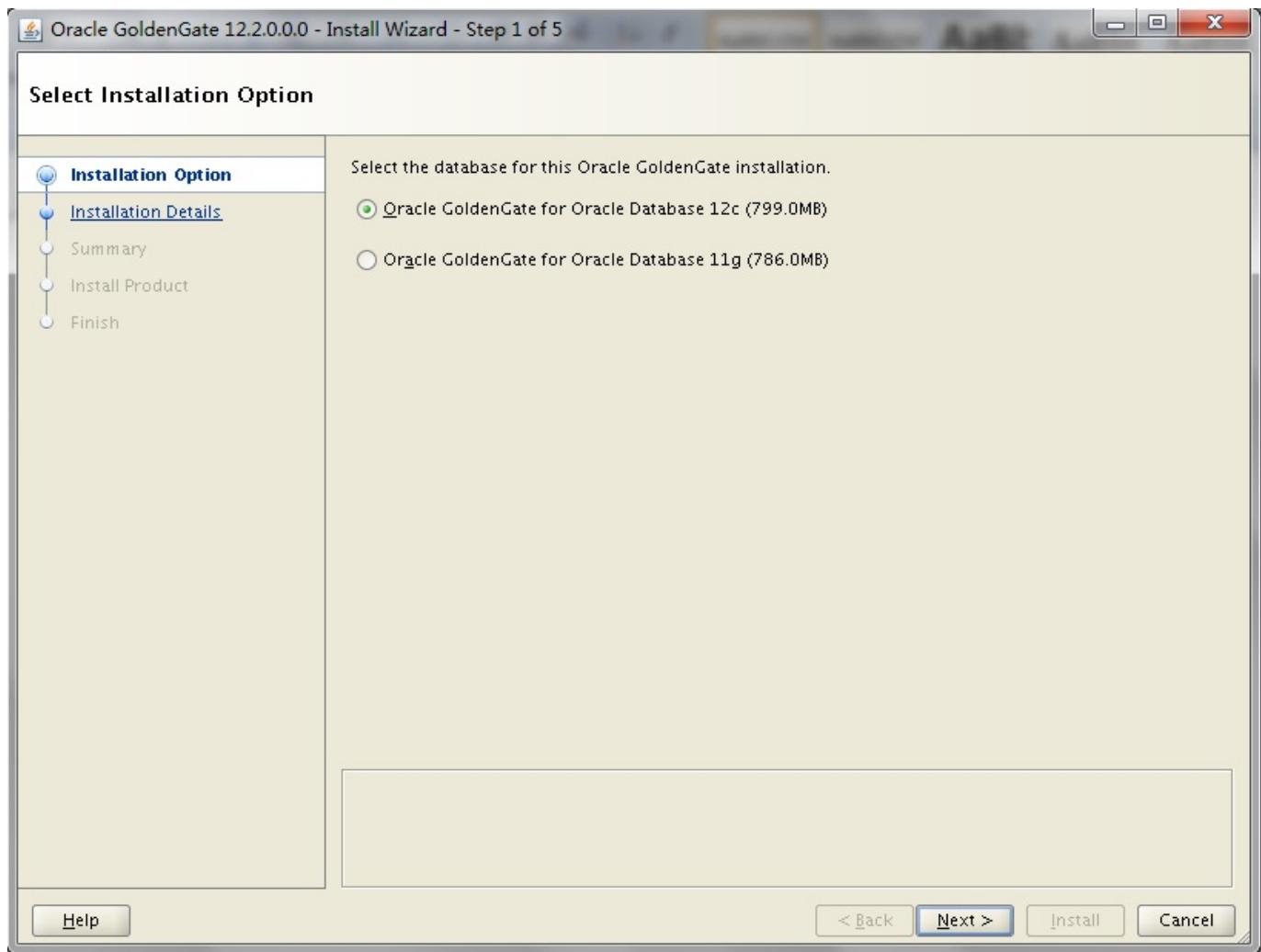
软件版本：linuxamd64_12102_database_1of2.zip, linuxamd64_12102_database_1of2.zip

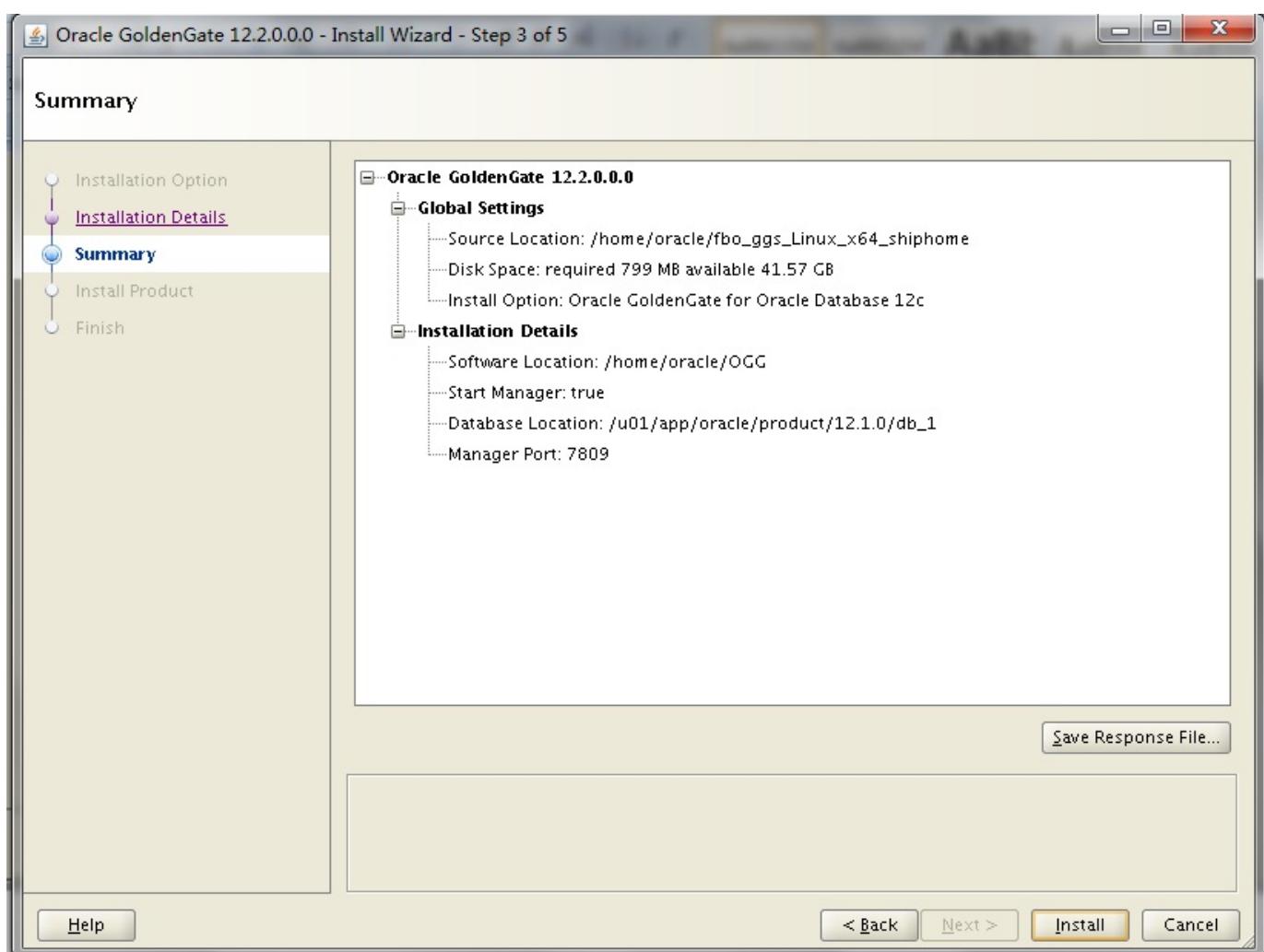
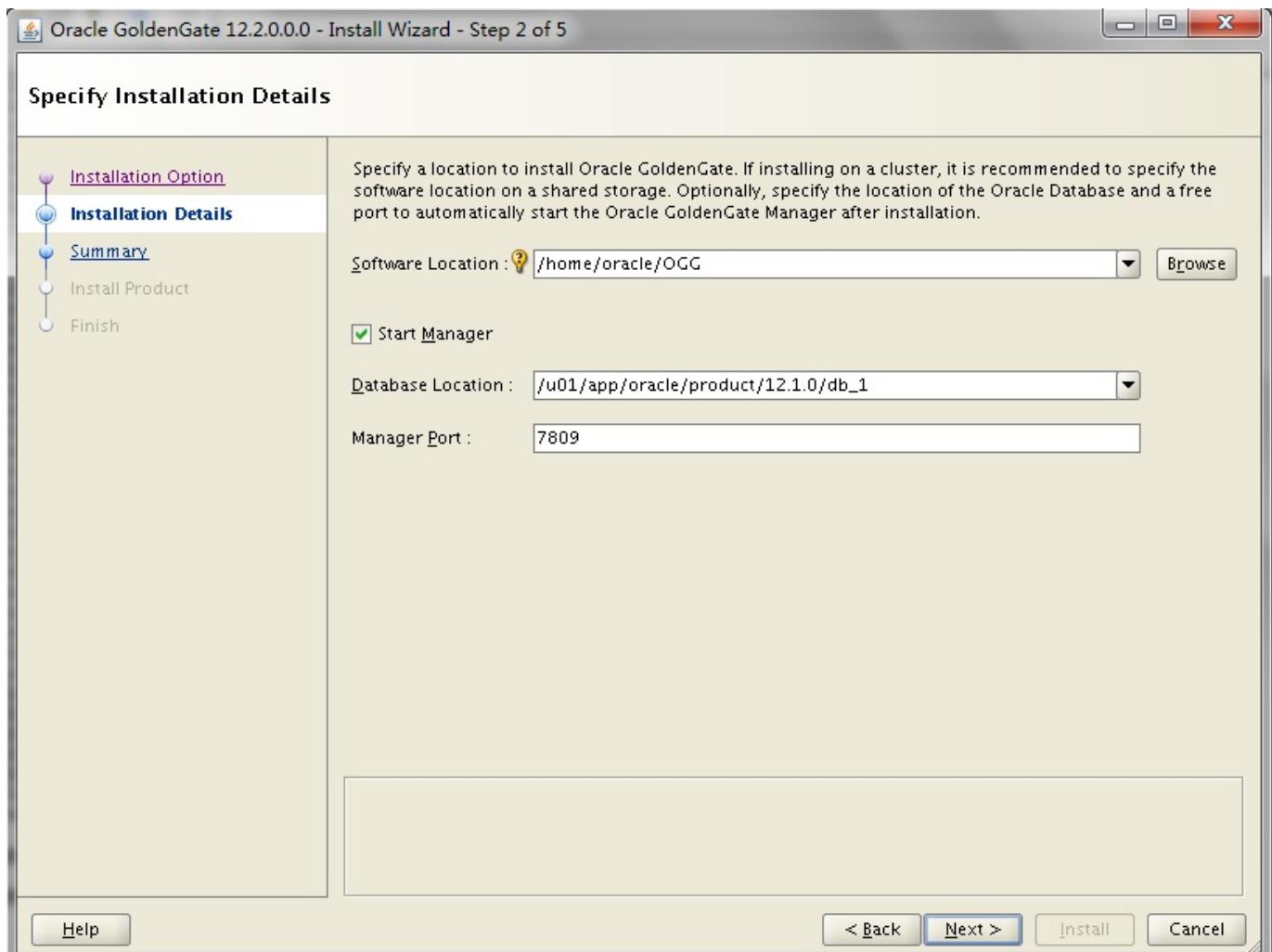
下载并安装OGG for Oracle

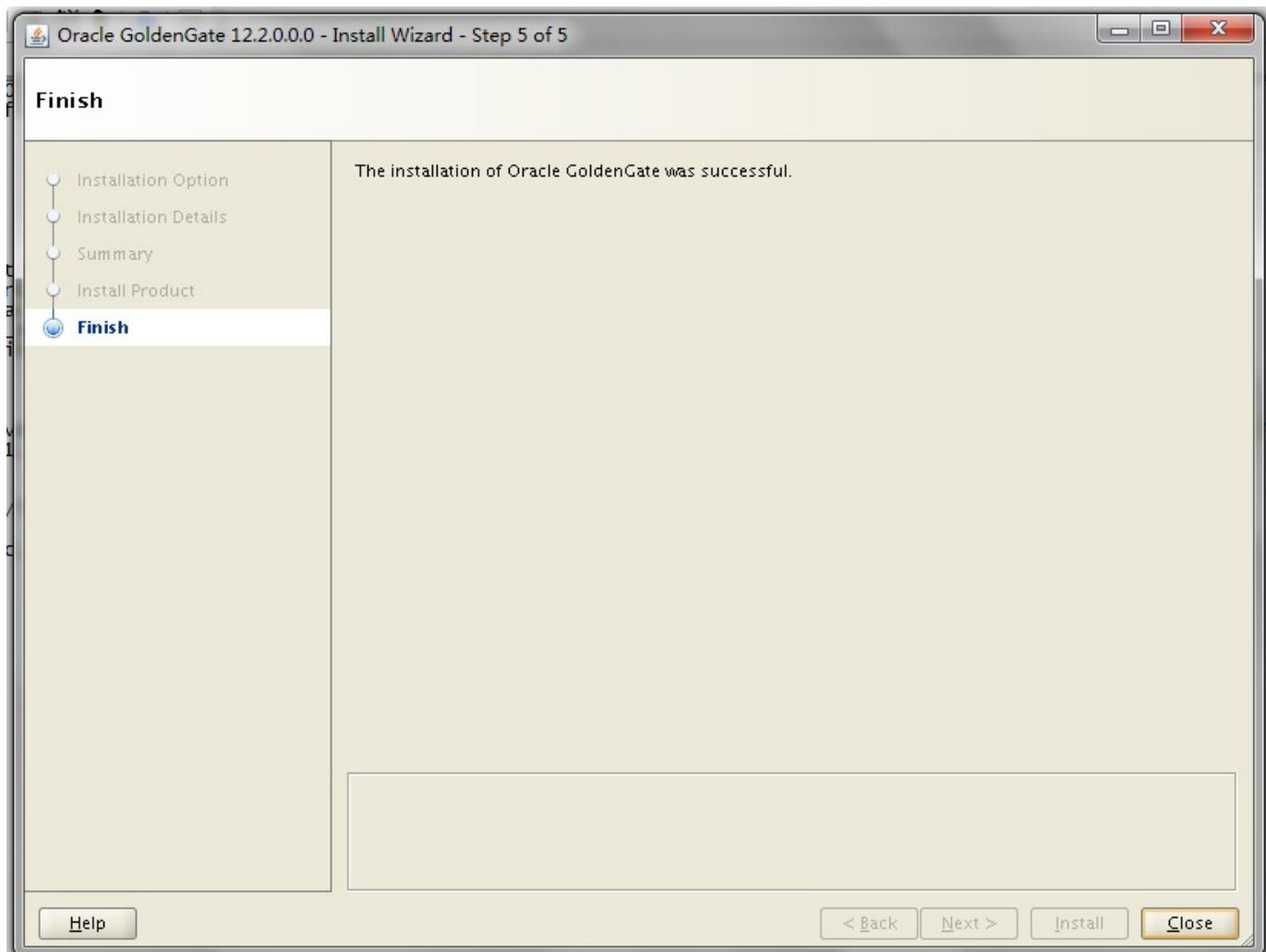
- 将fbo_ggs_Linux_x64_shiphome.zip上传至oracle客户端（ip：162.1.115.68）`/home/oracle`目录下，切换至oracle用户，解压生成fbo_ggs_Linux_x64_shiphome目录。

```
[root@oracle oracle]# cd fbo_ggs_Linux_x64_shiphome
[root@oracle fbo_ggs_Linux_x64_shiphome]# ll
total 4
drwxr-xr-x. 5 oracle oinstall 4096 Dec 12 2015 Disk1
[root@oracle fbo_ggs_Linux_x64_shiphome]# cd Disk1
[root@oracle Disk1]# ll
total 16
drwxr-xr-x. 4 oracle oinstall 4096 Dec 12 2015 install
drwxrwxr-x. 2 oracle oinstall 4096 Dec 12 2015 response
-rwxr-xr-x. 1 oracle oinstall 918 Dec 12 2015 runInstaller
drwxr-xr-x. 11 oracle oinstall 4096 Dec 12 2015 stage
[root@oracle Disk1]#
```

- 在`/home/oracle/fbo_ggs_Linux_x64_shiphome/Disk1`目录下，运行`./runInstaller`







- 安装成功，/home/oracle/OGG/是OGG for Oracle的安装目录。

```
[root@oracle oracle]# ll
total 464760
drwxr-xr-x. 3 oracle oinstall      4096 Dec 12 2015 fbo_ggs_Linux_x64_shiphome
-rw-r--r--. 1 oracle oinstall 475611228 Aug 4 17:14 fbo_ggs_Linux_x64_shiphome_12.2.0.1.1.zip
drwxr-xr-x. 28 oracle oinstall     4096 Nov 9 17:21 OGG
-rw-r--r--. 1 oracle oinstall   282294 Jan 19 2016 OGG-12.2.0.1.1-ReleaseNotes.pdf
-rw-r--r--. 1 oracle oinstall    1559 Jan 19 2016 OGG-12.2.0.1-README.txt
drwxr-xr-x. 2 oracle oinstall     4096 Nov 10 20:23 ora2gs_dir
[root@oracle oracle]# cd OGG
[root@oracle OGG]# ll
total 631812
-rw-r-----. 1 oracle oinstall      426 Oct 15 2010 bcpfmt.tpl
-rw-r-----. 1 oracle oinstall    1725 Oct 15 2010 bcrypt.txt
drwxr-x---. 4 oracle oinstall     4096 Nov 9 11:17 BR
-rwxrwxr-x. 1 oracle oinstall 8607365 Dec 11 2015 cachefiledump
drwxr-----. 4 oracle oinstall     4096 Nov 8 09:15 cfgtoollogs
-rwxrwxr-x. 1 oracle oinstall 11215946 Dec 11 2015 checkprm
-rw-r-----. 1 oracle oinstall    1327 Nov 15 2013 chkpt_ora_create.sql
-rwxr-x---. 1 oracle oinstall 12015022 Dec 11 2015 convchk
-rwxrwxr-x. 1 oracle oinstall 17510186 Dec 11 2015 convprm
-rw-r-----. 1 oracle oinstall      159 Oct 15 2010 db2cntl.tpl
-rw-r-----. 1 oracle oinstall    1059 Oct 15 2010 ddl_cleartrace.sql
-rw-r-----. 1 oracle oinstall    9338 Mar 13 2015 ddl_create.sql
-rw-r-----. 1 oracle oinstall    4189 Oct 15 2010 ddl_ddl2file.sql
-rw-r-----. 1 oracle oinstall      746 Oct 15 2010 ddl_disable.sql
-rw-r-----. 1 oracle oinstall      692 Oct 15 2010 ddl_enable.sql
-rw-r-----. 1 oracle oinstall    3644 Jan 30 2015 ddl_filter.sql
-rw-r-----. 1 oracle oinstall    14688 May 11 2015 ddl_ora10.sql
-rw-r-----. 1 oracle oinstall    3167 Aug 28 2011 ddl_ora10upCommon.sql
-rw-r-----. 1 oracle oinstall    15690 May 11 2015 ddl_ora11.sql
-rw-r-----. 1 oracle oinstall    14930 May 11 2015 ddl_ora9.sql
-rw-r-----. 1 oracle oinstall    1026 Oct 15 2010 ddl_pin.sql
-rw-r-----. 1 oracle oinstall    4461 Apr 11 2012 ddl_remove.sql
-rw-r-----. 1 oracle oinstall      425 Oct 15 2010 ddl_session1.sql
-rw-r-----. 1 oracle oinstall    1053 Oct 15 2010 ddl_session.sql
-rw-r-----. 1 oracle oinstall 308085 Apr 29 2015 ddl_setup.sql
-rw-r-----. 1 oracle oinstall    10683 Jun 13 2012 ddl_status.sql
-rw-r-----. 1 oracle oinstall    2506 Oct 15 2010 ddl_staymetadata_off.sql
-rw-r-----. 1 oracle oinstall    2501 Oct 15 2010 ddl_staymetadata_on.sql
-rw-r-----. 1 oracle oinstall    2955 Oct 15 2010 ddl_tracelevel.sql
-rw-r-----. 1 oracle oinstall    2721 Dec 28 2010 ddl_trace_off.sql
-rw-r-----. 1 oracle oinstall    2862 Oct 15 2010 ddl_trace_on.sql
```

配置环境变量

- 切换到oracle用户

```
su - oracle
vi .bash_profile
```

- 文件.bash_profile内容如下：

```
# .bash_profile
# Get the aliases and functions
if [ -f ~/.bashrc ]; then
    . ~/.bashrc
fi

# User specific environment and startup programs

PATH=$PATH:$HOME/bin

export PATH

PATH=$PATH:$HOME/bin:/u01/app/oracle/product/12.1.0/db_1/bin
export PATH
umask 022
export ORACLE_BASE=/u01/app/oracle
export ORACLE_HOME=/u01/app/oracle/product/12.1.0/db_1
export ORACLE_SID=orcl

export LD_LIBRARY_PATH=$ORACLE_HOME/lib
```

```

[root@oracle ~]# su - oracle
[oracle@oracle ~]$ vi .bash_profile

# .bash_profile

# Get the aliases and functions
if [ -f ~/.bashrc ]; then
    . ~/.bashrc
fi

# User specific environment and startup programs

PATH=$PATH:$HOME/bin

export PATH

PATH=$PATH:$HOME/bin:/u01/app/oracle/product/12.1.0/db_1/bin
export PATH
umask 022
export ORACLE_BASE=/u01/app/oracle
export ORACLE_HOME=/u01/app/oracle/product/12.1.0/db_1
export ORACLE_SID=orcl
export LD_LIBRARY_PATH=$ORACLE_HOME/lib
~
```

- 运行OGG

```

[oracle@oracle OGG]$ cd /home/oracle/OGG
[oracle@oracle OGG]$ ./ggsci

Oracle GoldenGate Command Interpreter for Oracle
Version 12.2.0.1.1 OGGCORE_12.2.0.1.0_PLATFORMS_151211.1401_FBO
Linux, x64, 64bit (optimized), Oracle 12c on Dec 12 2015 02:56:48
Operating system character set identified as UTF-8.

Copyright (c) 1995, 2015, oracle and/or its affiliates. All rights reserved.
```

```

GGSCI (oracle) 1> info all
Program      Status      Group      Lag at Chkpt  Time since chkpt
MANAGER      RUNNING

GGSCI (oracle) 2> ■
```

打开数据库归档及开启最小附加日志

- 使用Sqlplus / as sysdba登陆Oracle源端数据库后打开Archive Log:

```

shutdown immediate;
startup mount;
alter database archivelog;
alter database open;
archive log list;
```

```
[oracle@oracledb ~]$ sqlplus / as sysdba
SQL*Plus: Release 12.1.0.2.0 Production on Fri Feb 17 10:54:00 2017
Copyright (c) 1982, 2014, Oracle. All rights reserved.

Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit Production
with the Partitioning, OLAP, Advanced Analytics and Real Application Testing options

SQL> shutdown immediate;
Database closed.
Database dismounted.
ORACLE instance shut down.
SQL> startup mount;
ORACLE instance started.

Total System Global Area 2483027968 bytes
Fixed Size          3713864 bytes
Variable Size       721421496 bytes
Database Buffers   1744830464 bytes
Redo Buffers        13062144 bytes
Database mounted.
SQL> alter database archivelog;

Database altered.

SQL> alter database open;

Database altered.

SQL> archive log list;
Database log mode           Archive Mode
Automatic archival          Enabled
Archive destination          USE_DB_RECOVERY_FILE_DEST
Oldest online log sequence   14
Next log sequence to archive 16
Current log sequence         16
SQL>
```

- 源端数据库打开数据库级最小附加日志及force logging:

```
SELECT supplemental_log_data_min, force_logging FROM v$database;
alter database add supplemental log data;
alter database force logging;
```

- 切换日志以使附加日志生效:

```
ALTER SYSTEM switch logfile;

SQL> SELECT supplemental_log_data_min, force_logging FROM v$database;
SUPPLEMENT FORCE_LOGGING
-----
NO      NO

SQL> alter database add supplemental log data;
Database altered.

SQL> alter database force logging;
Database altered.

SQL> ALTER SYSTEM switch logfile;
System altered.

SQL> SELECT supplemental_log_data_min, force_logging FROM v$database;
SUPPLEMENT FORCE_LOGGING
-----
YES     YES

SQL>
```

- Enabling Oracle GoldenGate in the Database:

```
show parameter enable_goldengate_replication;
alter system set enable_goldengate_replication = true scope=both;
```

```

SQL> show parameter enable_goldengate_replication;
NAME                                     TYPE        VALUE
enable_goldengate_replication          boolean     FALSE
SQL> alter system set enable_goldengate_replication = true scope=both;
System altered.

SQL> show parameter enable_goldengate_replication;
NAME                                     TYPE        VALUE
enable_goldengate_replication          boolean     TRUE
SQL>

```

- 配置DB12c PDB的tnsname信息 vi \$ORACLE_HOME/network/admin/tnsnames.ora :

```

# tnsnames.ora Network Configuration File: /u01/app/oracle/product/12.1.0/db_1/network/admin,
# Generated by Oracle configuration tools.

ORCL =
  (DESCRIPTION =
    (ADDRESS = (PROTOCOL = TCP)(HOST = oracle)(PORT = 1521))
    (CONNECT_DATA =
      (SERVER = DEDICATED)
      (SERVICE_NAME = orcl)
    )
  )

PDBORCL =
  (DESCRIPTION =
    (ADDRESS = (PROTOCOL = TCP)(HOST = oracle)(PORT = 1521))
    (CONNECT_DATA =
      (SERVER = DEDICATED)
      (SERVICE_NAME = pdborcl)
    )
  )

```

在数据库中创建ogg用户并赋予权限

- 使用 sqlplus / as sysdba 登陆数据库后创建ogg用户并赋予权限

```

create user c##ogg identified by welcome1;
grant dba to c##ogg container=all;
grant create session, connect, resource to c##ogg container=all;
grant alter any table to c##ogg container=all;
grant alter system to c##ogg container=all;
exec dbms_goldengate_auth.grant_admin_privilege('c##ogg',container=>'all');

```

```

SQL> create user c##ogg identified by welcome1;
User created.

SQL> grant dba to c##ogg container=all;
Grant succeeded.

SQL> grant create session, connect, resource to c##ogg container=all;
Grant succeeded.

SQL> grant alter any table to c##ogg container=all;
Grant succeeded.

SQL> grant alter system to c##ogg container=all;
Grant succeeded.

SQL> exec dbms_goldengate_auth.grant_admin_privilege('c##ogg',container=>'all');

PL/SQL procedure successfully completed.

SQL>

```

配置GoldenGate 登陆数据库的别名

- 在GoldenGate中创建用户别名，用于登录Oracle数据库读取数据库日志：

```
add credentialstore  
ALTER CREDENTIALSTORE ADD USER c##ogg PASSWORD welcome1 ALIAS ogg_src
```

- 这样就可以用别名ogg_src登陆数据库了：

```
dblogin useralias ogg_src
```

```
GGSCI (oracle) 3> dblogin useralias ogg_src  
Successfully logged into database CDB$ROOT.
```

- C##ogg是Oracle DB12c的普通用户，可以访问多个数据库实例。

创建test用户和test1表

test用户是基于pdbocl数据库实例的：

- 登陆数据库

```
Sqlplus / as sysdba
```

- 创建用户

```
alter session set container=pdbocl;  
alter database open;  
create user test identified by welcome1;  
grant resource, connect to test;  
CREATE TABLESPACE test DATAFILE '/u01/app/oracle/oradata/orcl/pdbocl/test01.dbf' SIZE 500M UNIFORM SIZE 128K;  
alter user test quota unlimited on test;  
alter user test quota unlimited on users;
```

- 创建测试表

```
conn test/welcome1@pdbocl;  
create table test1(id number primary key, name varchar2(50));
```

```
SQL> alter session set container=pdbocl;  
Session altered.  
SQL> create user test identified by welcome1;  
create user test identified by welcome1  
*  
ERROR at line 1:  
ORA-01109: database not open
```

```
SQL> alter database open;  
Database altered.  
SQL> create user test identified by welcome1;  
User created.  
SQL> grant resource, connect to test;  
Grant succeeded.
```

```
SQL> CREATE TABLESPACE test DATAFILE '/u01/app/oracle/oradata/orcl/pdbocl/test01.dbf' SIZE 500M UNIFORM SIZE 128K;  
Tablespace created.  
SQL> alter user test quota unlimited on test;  
User altered.  
SQL> conn test/welcome1@pdbocl;  
Connected.  
SQL> create table test1(id number primary key, name varchar2(50));  
Table created.
```

配置GoldenGate捕获进程

- 编辑eora.prm，在GGSCI命令行下运行edit param eora命令：

```
GGSCI> edit param eora
```

```
extract eora
useridalias ogg_src
ddl include objname pdborcl.test.*
logallsupcols
updaterecordformat full
NOCOMPRESSUPDATES
getupdatebefores
exttrail dirdat/eo
sourcecatalog pdborcl
table test.*;
```

```
GGSCI> edit param mgr
```

```
PORT 7809
```

```
GGSCI> edit param phdfs
```

```
extract phdfs
passthru
rmthost 162.1.115.69, mgrport 7809

rmttrail dirdat/rs
sourcecatalog pdborcl
table *.*;
```

```
GGSCI> edit param phbase
```

```
extract phbase
passthru
rmthost 162.1.115.69, mgrport 7809
rmttrail dirdat/se
sourcecatalog pdborcl
table *.*;
```

```
GGSCI> edit param pkafka
```

```
extract pkafka
passthru
rmthost 162.1.115.69, mgrport 7809

rmttrail dirdat/rk
sourcecatalog pdborcl
table *.*;
```

```
GGSCI> edit param pflume
```

```
extract pflume
passthru
rmthost 162.1.115.69, mgrport 7809

rmttrail dirdat/rf
sourcecatalog pdborcl
table *.*;
```

- 编辑 `diroby/eora.oby` 文件，在GGSCI命令行下运行 `shell vi diroby/eora.oby` 命令：(shell之后接操作系统命令)

- 使用oracle用户创建diroby目录:

```
cd /home/oracle/OGG/
mkdir diroby
```

```
GGSCI> shell vi diroby/eora.oby
```

```
delete extract eora
delete exttrail dirdat/eo
add extract eora, integrated tranlog, begin now
add exttrail dirdat/eo, extract eora, megabytes 20
info extract eora
~
```

注意进程名eora和数据文件dirdat/eo的对应关系

- 在GGSCI命令行下运行obey diroby/eora.oby命令, 把捕获进程eora加入到管理者进程中:

```
GGSCI> obey diroby/eora.oby
```

```
GGSCI (oracledb) 2> obey diroby/eora.oby
GGSCI (oracledb) 3> delete extract eora
ERROR: EXTRACT EORA does not exist.

GGSCI (oracledb) 4> delete exttrail dirdat/eo
No match for specified extract target 'DIRDAT/EO'.
GGSCI (oracledb) 5> add extract eora, integrated tranlog, begin now
EXTRACT (Integrated) added.

GGSCI (oracledb) 6> add exttrail dirdat/eo, extract eora, megabytes 20
EXTTRAIL added.

GGSCI (oracledb) 7> info extract eora

EXTRACT      EORA      Initialized 2017-02-19 11:04      Status STOPPED
Checkpoint Lag    00:00:00 (updated 00:00:02 ago)
Log Read Checkpoint  Oracle Integrated Redo Logs
                           2017-02-19 11:04:03
                           SCN 0.0 (0)
```

- 把捕获进程eora注册到pdborcl数据库中:

```
GGSCI> dblogin useralias ogg_src
GGSCI> register extract eora database container(pdborcl)
```

```
GGSCI (oracledb) 9> dblogin useralias ogg_src
Successfully logged into database CDB$ROOT.
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 10> register extract eora database container(pdborcl)
```

```
2017-02-19 11:06:05 INFO OGG-02003 Extract EORA successfully registered with database at SCN 1880599.
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 11>
```

- 为pdborcl.test下的所有表添加表级附加日志:

```
GGSCI> add schematradata pdborcl.test allcols
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 11> add schematrandata pdborcl.test allcols  
2017-02-19 11:07:26 INFO OGG-01788 SCHEMATRANDATA has been added on schema test.  
2017-02-19 11:07:26 INFO OGG-01976 SCHEMATRANDATA for scheduling columns has been added on schema test.  
2017-02-19 11:07:26 INFO OGG-01977 SCHEMATRANDATA for all columns has been added on schema test.
```

- 启动GoldenGate捕获进程eora:

```
GGSCI> start eora
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 12> start eora  
Sending START request to MANAGER ...  
EXTRACT EORA starting
```

配置GoldenGate传输进程phdfs

配置GoldenGate传输进程phdfs，将OGG生成的数据文件传递给目标端GoldenGate HDFS处理。

- 编辑phdfs.prm，在GGSCI命令行下运行 `edit param phdfs` 命令：

```
extract phdfs  
passthru  
rmthost 162.1.115.69, mgrport 7809  
  
rmttrail dirdat/rs  
sourcecatalog pdborcl  
table *.*;  
-
```

- 编辑 `diroby/phdfs.oby` 文件，在GGSCI命令行下运行 `shell vi diroby/phdfs.oby` 命令：(shell之后接操作系统命令)

```
GGSCI> shell vi diroby/phdfs.oby
```

```
-----  
delete extract phdfs  
add extract phdfs, exttrailsource dirdat/eo  
add rmttrail dirdat/rs, extract phdfs, megabytes 20
```

注意进程名phdfs和数据文件dirdat/rs的对应关系

- 在GGSCI命令行下运行 `obey diroby/phdfs.oby` 命令，把捕获进程phdfs加入到管理者进程中：

```
GGSCI> obey diroby/phdfs.oby
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 15> obey diroby/phdfs.oby
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 16> delete extract phdfs
```

```
ERROR: EXTRACT PHDFS does not exist.
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 17> add extract phdfs, exttrailsource dirdat/eo  
EXTRACT added.
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 18> add rmttrail dirdat/rs, extract phdfs, megabytes 20  
RMTTRAIL added.
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 19>
```

- 启动GoldenGate捕获进程phdfs：

```
GGSCI> start phdfs
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 19> start phdfs
```

```
Sending START request to MANAGER ...
EXTRACT PHDFS starting
```

配置GoldenGate传输进程phbase

配置GoldenGate传输进程phbase，将OGG生成的数据文件传递给目标端GoldenGate HBASE处理。

- 编辑phbase.prm，在GGSCI命令行下运行 `edit param phbase` 命令：

```
extract phbase
passthru
rmthost 162.1.115.69, mgrport 7809
rmttrail dirdat/se
sourcecatalog pdborcl
table *.*;
~
```

编辑 `diroby/phbase.oby` 文件，在GGSCI命令行下运行 `shell vi diroby/phbase.oby` 命令：(shell之后接操作系统命令)

```
GGSCI> shell vi diroby/phbase.oby
```

```
delete extract phbase
add extract phbase, exttrailsource dirdat/eo
add rmttrail dirdat/se, extract phbase, megabytes 20
```

注意进程名phbase和数据文件dirdat/se的对应关系

- 在GGSCI命令行下运行 `obey diroby/phbase.oby` 命令，把捕获进程phbase加入到管理者进程中：

```
GGSCI> obey diroby/phbase.oby
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 23> obey diroby/phbase.oby
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 24> delete extract phbase
```

```
ERROR: EXTRACT PHBASE does not exist.
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 25> add extract phbase, exttrailsource dirdat/eo
EXTRACT added.
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 26> add rmttrail dirdat/se, extract phbase, megabytes 20
RMTTRAIL added.
```

- 启动GoldenGate捕获进程phbase：

```
GGSCI> start phbase
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 27> start phbase
```

```
Sending START request to MANAGER ...
EXTRACT PHBASE starting
```

配置GoldenGate传输进程pflume

配置GoldenGate传输进程pflume，将OGG生成的数据文件传递给目标端GoldenGate FLUME处理。

- 编辑pflume.prm，在GGSCI命令行下运行 `edit param pflume` 命令：

```
extract pflume
passthru
rmthost 162.1.115.69, mgrport 7809

rmttrail dirdat/rf
sourcecatalog pdborcl
table *.*;
```

- 编辑 `diroby/pflume.oby` 文件，在GGSCI命令行下运行 `shell vi diroby/pflume.oby` 命令：(shell之后接操作系统命令)

```
GGSCI> shell vi diroby/pflume.oby
```

```
delete extract pflume
add extract pflume, exttrailsource dirdat/eo
add rmttrail dirdat/rf, extract pflume, megabytes 20
```

注意进程名pflume和数据文件dirdat/rf的对应关系

- 在GGSCI命令行下运行 `obey diroby/pflume.oby` 命令，把捕获进程pflume加入到管理者进程中：

```
GGSCI> obey diroby/pflume.oby
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 30> obey diroby/pflume.oby
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 31> delete extract pflume
ERROR: EXTRACT PFLUME does not exist.

GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 32> add extract pflume, exttrailsource dirdat/eo
EXTRACT added.

GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 33> add rmttrail dirdat/rf, extract pflume, megabytes 20
RMTTRAIL added.
```

- 启动GoldenGate捕获进程pflume：

```
GGSCI> start pflume
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 34> start pflume
Sending START request to MANAGER ...
EXTRACT PFLUME starting
```

配置GoldenGate传输进程pkafka

配置GoldenGate传输进程pkafka，将OGG生成的数据文件传递给目标端GoldenGate Kafka处理。

- 编辑pkafka.prm，在GGSCI命令行下运行 `edit param pkafka` 命令：

```
extract pkafka
passthru
rmthost 162.1.115.69, mgrport 7809

rmttrail dirdat/rk
sourcecatalog pdborcl
table *.*;
```

- 编辑 `diroby/pkafka.oby` 文件，在GGSCI命令行下运行 `shell vi diroby/pkafka.oby` 命令：(shell之后接操作系统命令)

```
GGSCI> shell vi diroby/pkafka.oby
```

```
delete extract pkafka
add extract pkafka, exttrailsource dirdat/eo
add rmttrail dirdat/rk, extract pkafka, megabytes 20
```

注意进程名pkafka和数据文件dirdat/rk的对应关系

- 在GGSCI命令行下运行 `obey diroby/pkafka.oby` 命令，把捕获进程pkafka加入到管理者进程中：

```
GGSCI> obey diroby/ pkafka.oby
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 37> obey diroby/pkafka.oby
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 38> delete extract pkafka
```

```
ERROR: EXTRACT PKAFKA does not exist.
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 39> add extract pkafka, exttrailsource dirdat/eo
```

```
EXTRACT added.
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 40> add rmttrail dirdat/rk, extract pkafka, megabytes 20
RMTTRAIL added.
```

- 启动GoldenGate捕获进程pkafka：

```
GGSCI> start pkafka
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 41> start pkafka
```

```
Sending START request to MANAGER ...
EXTRACT PKAFKA starting
```

查看GoldenGate进程运行状态

- 查看GoldenGate进程状态：(EORCL是与ELK对接的进程)

```
GGSCI> info all
```

```
GGSCI (oracledb as c##ogg@orcl/CDB$ROOT) 42> info all
```

Program	Status	Group	Lag at Chkpt	Time since Chkpt
MANAGER	RUNNING			
EXTRACT	RUNNING	EORA	00:00:10	00:00:07
EXTRACT	ABENDED	PFLUME	00:00:00	00:03:39
EXTRACT	ABENDED	PHBASE	00:00:00	00:12:51
EXTRACT	ABENDED	PHDFS	00:00:00	00:16:56
EXTRACT	ABENDED	PKAFKA	00:00:00	00:00:48

- 查看某个进程的详细信息：

```
GGSCI> info eora detail
```

```
GGSCI (oracle as c##ogg@orcl/CDB$ROOT) 14> info eora detail
EXTRACT      EORA      Last Started 2016-11-11 09:17    STATUS RUNNING
Checkpoint Lag          00:00:09 (updated 00:00:00 ago)
Process ID       31203
Log Read Checkpoint   Oracle Integrated Redo Logs
                      2016-11-11 14:17:34
                      SCN 0.2118599 (2118599)
```

Target Extract Trails:

Trail Name	Seqno	RBA	Max MB	Trail Type
dirdat/eo	2	186068	20	EXTTRAIL

Integrated Extract outbound server first scn: 0.1668196 (1668196)

Integrated Extract outbound server filtering start scn: 0.1669095 (1669095)

Extract Source	Begin	End
Not Available	2016-11-08 10:07	2016-11-11 14:17
Not Available	2016-11-08 10:07	2016-11-11 09:17
Not Available	2016-11-08 10:07	2016-11-11 09:17
Not Available	2016-11-08 10:07	2016-11-09 17:41
Not Available	2016-11-08 10:07	2016-11-09 17:41
Not Available	* Initialized *	2016-11-08 10:07
Not Available	* Initialized *	2016-11-08 10:07

Current directory /home/oracle/OGG

Report file	/home/oracle/OGG/dir rpt/EORA.rpt
Parameter file	/home/oracle/OGG/dir prm/eora.prm
Checkpoint file	/home/oracle/OGG/dir chk/EORA.cpe
Process file	/home/oracle/OGG/dir pcs/EORA.pce
Error log	/home/oracle/OGG/ggserr.log

- 查看GoldenGate的统计信息:

```
GGSCI> stats eora, latest
GGSCI (oracle as c##ogg@orcl/CDB$ROOT) 15> stats eora, latest
Sending STATS request to EXTRACT EORA ...
Start of Statistics at 2016-11-11 14:19:37.
DDL replication statistics (for all trails):
*** Total statistics since extract started ***
Operations                                4.00
Mapped operations                          3.00
Unmapped operations                        0.00
Other operations                           1.00
Excluded operations                        1.00
Output to dirdat/eo:
Extracting from PDBORCL.TEST.ACCOUNT to PDBORCL.TEST.ACCOUNT:
*** Latest statistics since 2016-11-11 09:20:13 ***
Total inserts                            24.00
Total updates                             0.00
Total deletes                            0.00
Total discards                            0.00
Total operations                           24.00
Extracting from PDBORCL.TEST.TEST1 to PDBORCL.TEST.TEST1:
*** Latest statistics since 2016-11-11 09:20:13 ***
Total inserts                            8.00
Total updates                             0.00
Total deletes                            2.00
Total discards                            0.00
Total operations                           10.00
Extracting from PDBORCL.TEST.TEST3 to PDBORCL.TEST.TEST3:
```

- 查看GoldenGate进程报告，用于定位问题：

```
GGSCI> view report eora

GGSCI (oracle as c##ogg@orcl/CDB$ROOT) 16> view report eora

*****
          oracle GoldenGate Capture for oracle
  version 12.2.0.1.1 OGGCORE_12.2.0.1.0_PLATFORMS_151211.1401_FBO
  Linux, x64, 64bit (optimized), Oracle 12c on Dec 12 2015 03:12:27
Copyright (c) 1995, 2015, oracle and/or its affiliates. All rights reserved.

Starting at 2016-11-11 09:17:24
*****

Operating System Version:
Linux
Version #1 SMP Tue Jan 29 11:47:41 EST 2013, Release 2.6.32-358.el6.x86_64
Node: oracle
Machine: x86_64

Address Space Size : soft limit    hard limit
Heap Size          : unlimited     unlimited
File Size          : unlimited     unlimited
CPU Time           : unlimited     unlimited

Process id: 31203

Description:

*****
**          Running with the following parameters          **
*****


2016-11-11 09:17:24 INFO    OGG-03059 Operating system character set identified as US-ASCII.
2016-11-11 09:17:24 INFO    OGG-02695 ANSI SQL parameter syntax is used for parameter parsing.
extract eora
useridalias ogg_src
```

OGG for Bigdata安装

环境准备

- 下载安装FusionInsight客户端
- 在Bigdata客户端机器上（ip: 162.1.115.69）按照FusionInsight产品文档安装FusionInsight客户端。将客户端JDK替换成1.7版本。
- 下载并安装oracle JDK1.7

```
[root@localhost dirprm]# mkdir /usr/java
[root@localhost dirprm]# cd /usr/java
[root@localhost java]# ll
total 0
[root@localhost java]# ll
total 119632
-rw-r--r--. 1 root root 122502805 Sep  6 2013 jdk-7u40-linux-x64.rpm
[root@localhost java]# chmod +x jdk-7u40-linux-x64.rpm
[root@localhost java]# ll
total 119632
-rwxr-xr-x. 1 root root 122502805 Sep  6 2013 jdk-7u40-linux-x64.rpm
[root@localhost java]# rpm -ivh jdk-7u40-linux-x64.rpm
Preparing...                                          ##### [100%]
  1:jdk                                         ##### [100%]
Unpacking JAR files...
  rt.jar...
  jsse.jar...
  charsets.jar...
  tools.jar...
  localedata.jar...
  jfxrt.jar...
[root@localhost java]#
```

将krb5.conf放在/etc/目录下

- 下载并安装OGG for Bigdata

将122011_ggs_Adapters_Linux_x64.zip上传至客户端/opt目录下:

```
unzip 122011_ggs_Adapters_Linux_x64.zip
```

```
-rw-r--r--. 1 root root 209352325 Sep 28 15:48 122011_ggs_Adapters_Linux_x64.zip
drwxr-xr-x. 3 root root 4096 Nov 11 11:14 client
-rw-rw-r--. 1 root root 680028160 May 5 2016 ggs_Adapters_Linux_x64.tar
drwxr-xr-x. 14 root root 4096 Nov 11 11:06 hadoopclient
-rw-r--r--. 1 root root 0 Nov 11 11:16 OGGBD-12.2.0.1.1-README.txt
drwxrwxr-x. 10 94110 42424 4096 Nov 11 11:22 OGG_HADOOP
drwxr-xr-x. 2 root root 4096 Nov 12 2012 rh
```

将解压后的ggs_Adapters_Linux_x64.tar解压到/opt/OGG_HADOOP目录下:

```
[root@localhost opt]# mkdir OGG_HADOOP
[root@localhost opt]# ll
total 868732
-rw-r--r--. 1 root root 209352325 Feb 13 10:19 122011_ggs_Adapters_Linux_x64.zip
drwxr-xr-x. 3 root root 4096 Feb 19 15:36 client
-rw-rw-r--. 1 root root 680028160 May 5 2016 ggs_Adapters_Linux_x64.tar
drwxr-xr-x. 16 root root 4096 Feb 19 15:37 hadoopclient
-rw-r--r--. 1 root root 1443 May 13 2016 OGGBD-12.2.0.1.1-README.txt
-rw-r--r--. 1 root root 170120 May 17 2016 OGG_BigData_12.2.0.1.1_Release_Notes.pdf
drwxr-xr-x. 2 root root 4096 Feb 19 15:50 OGG_HADOOP
drwxr-xr-x. 2 root root 4096 May 17 2013 rh
[root@localhost opt]# tar -xvf ggs_Adapters_Linux_x64.tar -C /opt/OGG_HADOOP
```

- 配置环境变量

更改环境变量，编辑根目录下 `vi .bash_profile`

```
# .bash_profile

# Get the aliases and functions
if [ -f ~/.bashrc ]; then
    . ~/.bashrc
fi

# User specific environment and startup programs
export JAVA_HOME=/usr/java/jdk1.7.0_40
#export JAVA_HOME=/usr/lib/jvm/java-1.7.0-openjdk-1.7.0.9.x86_64/jre
export CLASSPATH=$CLASSPATH:$JAVA_HOME/lib:$JAVA_HOME/jre/lib
PATH=$JAVA_HOME/bin:$PATH:$HOME/bin

export PATH

#export LD_LIBRARY_PATH=/usr/lib/jvm/java-1.7.0-openjdk-1.7.0.9.x86_64/jre/lib/amd64/server/libjvm.so:/usr/lib/jvm/java-1.7.0-openjdk-1.7.0.9.x86_64/jre/lib/amd64/server:/usr/lib/jvm/java-1.7.0-openjdk-1.7.0.9.x86_64/jre/lib/amd64/libjsig.so:/root/OGG_PostgreSQL/lib:$LD_LIBRARY_PATH

export

LD_LIBRARY_PATH=/usr/java/jdk1.7.0_40/jre/lib/amd64/server/libjvm.so:/usr/java/jdk1.7.0_40/jre/lib/amd64/server:/usr/java/jdk1.7.0_40/jre/lib/amd64/libjsig.so:/root/OGG_PostgreSQL/lib:$LD_LIBRARY_PATH

export LD_LIBRARY_PATH=/usr/local/lib:$LD_LIBRARY_PATH
```

```
[root@localhost OGG_HADOOP]# cd
[root@localhost ~]# vi .bash_profile
```

```
.bash_profile

# Get the aliases and functions
if [ -f ~/.bashrc ]; then
    . ~/.bashrc
fi

# User specific environment and startup programs
export JAVA_HOME=/usr/java/jdk1.7.0_40
#export JAVA_HOME=/usr/lib/jvm/java-1.7.0-openjdk-1.7.0.9.x86_64/jre
export CLASSPATH=$CLASSPATH:$JAVA_HOME/lib:$JAVA_HOME/jre/lib
PATH=$JAVA_HOME/bin:$PATH:$HOME/bin

export PATH

#export LD_LIBRARY_PATH=/usr/lib/jvm/java-1.7.0-openjdk-1.7.0.9.x86_64/jre/lib/amd64/server/libjvm.so:/usr/lib/jvm/java-1.7.0-openjdk-1.7.0.9.x86_64/jre/lib/amd64/libjsig.so:/root/OGG_PostgreSQL/lib:$LD_LIBRARY_PATH

export LD_LIBRARY_PATH=/usr/java/jdk1.7.0_40/jre/lib/amd64/server/libjvm.so:/usr/java/jdk1.7.0_40/jre/lib/amd64/server:/usr/java/jdk1.7.0_40/jre/lib/amd64/libjsig.so:/root/OGG_PostgreSQL/lib:$LD_LIBRARY_PATH

export LD_LIBRARY_PATH=/usr/local/lib:$LD_LIBRARY_PATH
```

Source环境变量, `source .bash_profile`.

```
[root@localhost OGG_HADOOP]# ./ggscli  
Oracle GoldenGate Command Interpreter  
Version 12.2.0.1.160419 OGGCORE_12.2.0.1.00GGBP_PLATFORMS_160430.1401  
Linux, x64, 64bit (optimized), Generic on Apr 30 2016 16:21:34  
Operating system character set identified as UTF-8.
```

Copyright (C) 1995, 2016, Oracle and/or its affiliates. All rights reserved.

```
GGSCI (localhost.localdomain) 1> create subdirs
```

Creating subdirectories under current directory /opt/OGG_HADOOP

Parameter files	/opt/OGG_HADOOP/dirprm: created
Report files	/opt/OGG_HADOOP/dirrpt: created
Checkpoint files	/opt/OGG_HADOOP/dirchk: created
Process status files	/opt/OGG_HADOOP/dirpcs: created
SQL script files	/opt/OGG_HADOOP/dirsq1: created
Database definitions files	/opt/OGG_HADOOP/dirdef: created
Extract data files	/opt/OGG_HADOOP/dirdat: created
Temporary files	/opt/OGG_HADOOP/dirtmp: created
Credential store files	/opt/OGG_HADOOP/dircrd: created
Masterkey wallet files	/opt/OGG_HADOOP/dirwlt: created
Dump files	/opt/OGG_HADOOP/dirdmp: created

将 `/opt/OGG_HADOOP/AdapterExamples/big-data` 下的四个目录下的所有文件拷贝到 `/opt/OGG_HADOOP/dirprm` 目录下。

```
[root@localhost AdapterExamples]# cd big-data/  
[root@localhost big-data]# ll  
total 24  
drwxr-xr-x. 2 94110 42424 4096 May 5 2016 flume  
drwxr-xr-x. 2 94110 42424 4096 May 5 2016 hbase  
drwxr-xr-x. 2 94110 42424 4096 May 5 2016 hdfs  
drwxr-xr-x. 2 94110 42424 4096 May 5 2016 kafka  
drwxr-xr-x. 4 94110 42424 4096 May 5 2016 metadata_provider  
-r--r--r--. 1 94110 42424 903 May 5 2016 README.txt
```

```
[root@localhost OGG_HADOOP]# cd dirprm/  
[root@localhost dirprm]# ll  
total 40  
-rw-r--r--. 1 root root 107 May 5 2016 custom-flume-rpc.properties  
-rw-r--r--. 1 root root 288 May 5 2016 custom_kafka_producer.properties  
-rw-r--r--. 1 root root 812 May 5 2016 flume.props  
-rw-r--r--. 1 root root 724 May 5 2016 hbase.props  
-rw-r--r--. 1 root root 899 May 5 2016 hdfs.props  
-rw-r--r--. 1 root root 816 May 5 2016 kafka.props  
-rw-r--r--. 1 root root 332 May 5 2016 rflume.prm  
-rw-r--r--. 1 root root 332 May 5 2016 rhbase.prm  
-rw-r--r--. 1 root root 329 May 5 2016 rhdfs.prm  
-rw-r--r--. 1 root root 332 May 5 2016 rkafka.prm
```

配置GoldenGate管理进程

- 编辑mgr.prm

```
GGSCI> edit param mgr
```

```
port 7809  
~  
~
```

```
GGSCI>start mgr  
GGSCI>info all
```

```
GGSCI (localhost.localdomain) 11> start mgr  
Manager started.
```

```
GGSCI (localhost.localdomain) 12> info all
```

Program	Status	Group	Lag at Chkpt	Time Since Chkpt
MANAGER	RUNNING			
REPLICAT	STOPPED	RHDFS	00:00:00	00:02:24

```
GGSCI (localhost.localdomain) 13> █
```

配置GoldenGate HDFS 复制进程

- 编辑rhdfs.prm，在GGSCI命令行下运行 `edit param rhdfs` 命令：

```
GGSCI> edit param rhdfs
```

```
REPLICAT rhdfs  
-- Trail file for this example is located in "AdapterExamples/trail" directory  
-- Command to add REPLICAT  
-- add replicat rhdfs, extrail AdapterExamples/trail/tr  
TARGETDB LIBFILE libggjava.so SET property=dirprm/hdfs.props  
REPORTCOUNT EVERY 1 MINUTES, RATE  
GROUPTRANSOPS 10000  
MAP pdborcl.test.* , TARGET test1.*;
```

- 编辑hdfs.props，在GGSCI命令行下运行 `shell vi dirprm/hdfs.props` 命令：(shell之后接操作系统命令)

```
GGSCI> shell vi dirprm/hdfs.props
```

```
gg.handlerlist=hdfs  
gg.handler.hdfs.type=hdfs  
gg.handler.hdfs.includeTokens=false  
gg.handler.hdfs.maxFileSize=lg  
gg.handler.hdfs.rootFilePath=/ogg1  
gg.handler.hdfs.fileRollInterval=0  
gg.handler.hdfs.inactivityRollInterval=0  
gg.handler.hdfs.fileSuffix=.txt  
gg.handler.hdfs.partitionByTable=true  
gg.handler.hdfs.rollOnMetadataChange=true  
#gg.handler.hdfs.authType=none  
gg.handler.name.authType=kerberos  
gg.handler.name.kerberosKeytabFile=/opt/OGG_HADOOP/dirprm/hdfs.keytab  
gg.handler.name.kerberosPrincipal=hdfs/hadoop.hadoop.com@HADOOP.COM  
  
gg.handler.hdfs.format=delimitedtext  
gg.handler.hdfs.format.includeColumnNames=true  
  
gg.handler.hdfs.mode=tx  
  
goldengate.userexit.timestamp=utc  
goldengate.userexit.writers=javawriter  
javawriter.stats.display=TRUE  
javawriter.stats.full=TRUE  
  
gg.log=log4j  
gg.log.level=INFO  
  
gg.report.time=30sec  
  
#gg.classpath=/var/lib/hadoop/share/hadoop/common/*:/var/lib/hadoop/share/hadoop/common/lib/*:/var/lib/hadoop/share/hadoop/hdfs/*:/var/lib/hadoop/etc/hadoop/:  
gg.classpath=/opt/hadoopclient/HDFS/hadoop/etc/hadoop:/opt/hadoopclient/HDFS/hadoop/share/hadoop/common/lib/*:/opt/hadoopclient/HDFS/hadoop/share/hadoop/common/*:/opt/hadoopclient/HDFS/hadoop/share/hadoop/common/*:/opt/hadoopclient/HDFS/hadoop/share/hadoop/hdfs/*:/opt/hadoopclient/HDFS/hadoop/share/hadoop/hdfs/lib/*  
  
javawriter.bootoptions=-Xmx512m -Xms32m -Djava.class.path=ggjava/ggjava.jar
```

- 需要在HDFS中创建/ogg1目录。
 - 将hdfs.keytab文件拷贝到/opt/OGG_HADOOP/dirprm目录中:
- ```
[root@localhost ~]# scp -r root@162.1.93.101:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/etc/1_10_DataNode/hdfs.keytab /opt/OGG_HADOOP/dirprm/
```
- 把GoldenGate复制进程rhdfs加入到GoldenGate管理者进程中:

```
GGSCI> add replicat rhdfs, exttrail dirdat/rs
```

```
GGSCI (localhost.localdomain) 10> add replicat rhdfs, exttrail dirdat/rs
REPLICAT added.
```

```
GGSCI>info all
```

```
GGSCI (localhost.localdomain) 11> start mgr
Manager started.
```

```
GGSCI (localhost.localdomain) 12> info all
```

| Program  | Status  | Group | Lag at Chkpt | Time Since Chkpt |
|----------|---------|-------|--------------|------------------|
| MANAGER  | RUNNING |       |              |                  |
| REPLICAT | STOPPED | RHDFS | 00:00:00     | 00:02:24         |

```
GGSCI (localhost.localdomain) 13> █
```

```
GGSCI>start rhdfs
GGSCI>info all
```

```
GGSCI (localhost.localdomain) 15> info all
```

| Program  | Status  | Group | Lag at Chkpt | Time Since Chkpt |
|----------|---------|-------|--------------|------------------|
| MANAGER  | RUNNING |       |              |                  |
| REPLICAT | RUNNING | RHDFS | 00:00:00     | 00:00:07         |

## 配置GoldenGate HBase 复制进程

- 编辑rhbase.prm，在GGSCI命令行下运行 `edit param rhbase` 命令:

```
GGSCI> edit param rhbase
```

```
REPLICAT rhbase
-- Trail file for this example is located in "AdapterExamples/trail" directory
-- Command to add REPLICAT
-- add replicat rhbase, exttrail AdapterExamples/trail/tr
TARGETDB LIBFILE libggjava.so SET property=dirprm/hbase.props
REPORTCOUNT EVERY 1 MINUTES, RATE
GROUPTRANSOPS 10000
MAP pdborcl.test.* , TARGET test1.*;
```

- 编辑hbase.props, 在GGSCI命令行下运行 `shell vi dirprm/hbase.props` 命令: (shell之后接操作系统命令)

```
GGSCI> shell vi dirprm/hbase.props
```

```

gg.handlerlist=hbase
gg.handler.hbase.type=hbase
gg.handler.hbase.hbaseColumnFamilyName=cf
gg.handler.hbase.keyValueDelimiter=CDATA[=]
gg.handler.hbase.keyValuePairDelimiter=CDATA[,]
gg.handler.hbase.encoding=UTF-8
gg.handler.hbase.pkUpdateHandling=aend
gg.handler.hbase.nullValueRepresentation=CDATA[NULL]
#gg.handler.hbase.authType=none

gg.handler.hbase.authType=kerberos
gg.handler.hbase.kerberosKeytabFile=/opt/OGG_HADOOP/dirprm/hbase.keytab
gg.handler.hbase.kerberosPrincipal=hbase/hadoop.hadoop.com@HADOOP.COM

gg.handler.hbase.includeTokens=false

gg.handler.hbase.mode=tx

goldengate.userexit.timestamp=utc
goldengate.userexit.writers=javawriter
javawriter.stats.display=TRUE
javawriter.stats.full=TRUE

gg.log=log4j
gg.log.level=debug

gg.report.time=30sec

#ggs.classpath=/var/lib/hbase/lib/*:/var/lib/hbase/conf:
ggs.classpath=/opt/hadoopclient/HBase/hbase/lib/*:/opt/hadoopclient/HBase/hbase/conf

javawriter.bootoptions=-Xmx512m -Xms32m -Djava.class.path=ggjava/ggjava.jar [-Djava.security.auth.login.config=/opt/OGG_HADOOP/dirprm/jaas.conf]

```

- 拷贝hbase.keytab和jaas.conf到 `/opt/OGG_HADOOP/dirprm/` 下:

```

[root@localhost dirprm]# scp -r root@162.1.93.101:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/etc/1_16_RegionServer/hbase.keytab /opt/OGG_HADOOP/dirprm/
root@162.1.93.101's password:
hbase.keytab
[root@localhost dirprm]# scp -r root@162.1.93.101:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/etc/1_16_RegionServer/jaas.conf /opt/OGG_HADOOP/dirprm/
root@162.1.93.101's password:
jaas.conf

```

- jaas.conf文件

```

Client {
 com.sun.security.auth.module.Krb5LoginModule required
 useKeyTab=true
 keyTab="/opt/OGG_HADOOP/dirprm/hbase.keytab"
 principal="hbase/zkclient.hadoop.com@HADOOP.COM"
 useTicketCache=false
 storeKey=true
 debug=true;
};
~
```

- 把GoldenGate复制进程rhbase加入到GoldenGate管理者进程中:

```
GGSCI> add replicat rhbase, exttrail dirdat/se
```

```
GGSCI (localhost.localdomain) 18> add replicat rhbase, exttrail dirdat/se
REPLICAT added.
```

```
GGSCI>start rhbase
```

```
GGSCI (localhost.localdomain) 19> start rhbase
```

```
Sending START request to MANAGER ...
REPLICAT RHBASE starting
```

```
GGSCI>info all
```

| GGSCI (localhost.localdomain) 1> info all |         |        |              |                  |  |
|-------------------------------------------|---------|--------|--------------|------------------|--|
| Program                                   | Status  | Group  | Lag at Chkpt | Time since Chkpt |  |
| MANAGER                                   | RUNNING |        |              |                  |  |
| REPLICAT                                  | RUNNING | RHBASE | 00:00:00     | 00:00:03         |  |
| REPLICAT                                  | RUNNING | RHDFS  | 00:00:00     | 00:00:00         |  |

## 配置GoldenGate Kafka 复制进程

- 创建 kafka 消息，进入 FusionInsight 客户端 `/opt/hadoopclient/Kafka/kafka/bin`

```
[root@localhost bin]# ll
total 88
-rwxr-xr-x. 1 root root 1052 Aug 11 15:16 connect-distributed.sh
-rwxr-xr-x. 1 root root 1051 Aug 11 15:16 connect-standalone.sh
-rwxr-xr-x. 1 root root 861 Aug 11 15:16 kafka-acls.sh
-rwxr-xr-x. 1 root root 865 Aug 11 15:16 kafka-broker-info.sh
-rwxr-xr-x. 1 root root 864 Aug 11 15:16 kafka-configs.sh
-rwxr-xr-x. 1 root root 945 Aug 11 15:16 kafka-console-consumer.sh
-rwxr-xr-x. 1 root root 944 Aug 11 15:16 kafka-console-producer.sh
-rwxr-xr-x. 1 root root 871 Aug 11 15:16 kafka-consumer-groups.sh
-rwxr-xr-x. 1 root root 872 Aug 11 15:16 kafka-consumer-offset-checker.sh
-rwxr-xr-x. 1 root root 948 Aug 11 15:16 kafka-consumer-perf-test.sh
-rwxr-xr-x. 1 root root 862 Aug 11 15:16 kafka-mirror-maker.sh
-rwxr-xr-x. 1 root root 886 Aug 11 15:16 kafka-preferred-replica-election.sh
-rwxr-xr-x. 1 root root 959 Aug 11 15:16 kafka-producer-perf-test.sh
-rwxr-xr-x. 1 root root 874 Aug 11 15:16 kafka-reassign-partitions.sh
-rwxr-xr-x. 1 root root 868 Aug 11 15:16 kafka-replay-log-producer.sh
-rwxr-xr-x. 1 root root 874 Aug 11 15:16 kafka-replica-verification.sh
-rwxr-xr-x. 1 root root 6443 Nov 11 11:05 kafka-run-class.sh
-rwxr-xr-x. 1 root root 870 Aug 11 15:16 kafka-simple-consumer-shell.sh
-rwxr-xr-x. 1 root root 863 Aug 11 15:16 kafka-topics.sh
-rwxr-xr-x. 1 root root 958 Aug 11 15:16 kafka-verifiable-consumer.sh
-rwxr-xr-x. 1 root root 958 Aug 11 15:16 kafka-verifiable-producer.sh
```

Kafka 创建消息：

```
./kafka-topics.sh --create --zookeeper 162.1.93.101:24002,162.1.93.102:24002,162.1.93.103:24002/kafka --replication-factor 1 --
partitions 1 --topic test
```

```
[root@localhost bin]# ./kafka-topics.sh --create --zookeeper 162.1.93.101:24002,162.1.93.102:24002,162.1.93.103:24002/kafka --replication-factor 1 --partitions 1 --topic test
Created topic "test".
[root@localhost bin]#
```

Kafka 查看消息：

```
./kafka-topics.sh --list --zookeeper 162.1.93.101:24002,162.1.93.102:24002,162.1.93.103:24002/kafka --topic test
```

```
[root@localhost bin]# ./kafka-topics.sh --list --zookeeper 162.1.93.101:24002,162.1.93.102:24002,162.1.93.103:24002/kafka --topic test
test
[root@localhost bin]#
```

Kafka 给消息授权：

```
./kafka-acls.sh --authorizer-properties zookeeper.connect=162.1.93.101:24002,162.1.93.102:24002,162.1.93.103:24002/kafka --add --
operation All --allow-principal User:* --cluster --topic test
```

```
[root@localhost bin]# ./kafka-acls.sh --authorizer-properties zookeeper.connect=162.1.93.101:24002,162.1.93.102:24002,162.1.93.103:24002/kafka --add --operation All --allow-principal User:* --cluster --topic test
Adding ACLs for resource 'Topic:test':
User:* has Allow permission for operations: All from hosts: *
Adding ACLs for resource 'Cluster:kafka-cluster':
User:* has Allow permission for operations: All from hosts: *
Current ACLs for resource 'Topic:test':
User:* has Allow permission for operations: All from hosts: *
Current ACLs for resource 'Cluster:kafka-cluster':
User:* has Allow permission for operations: All from hosts: *
[root@localhost bin]#
```

- 编辑 rkafka.prm，在 GGSCI 命令行下运行 `edit param rkafka` 命令：

```
GGSCI> edit param rkafka
```

```
REPLICAT rkafka
-- Trail file for this example is located in "AdapterExamples/trail" directory
-- Command to add REPLICAT
-- add replicat rkafka, exttrail AdapterExamples/trail/tr
TARGETDB LIBFILE libggjava.so SET property=dirprm/kafka.props
REPORTCOUNT EVERY 1 MINUTES, RATE
GROUPTRANSOPS 10000
MAP pdborcl.test.* , TARGET test1.*;
```

- 编辑kafka.props, 在GGSCI命令行下运行 `shell vi dirprm/kafka.props` 命令: (shell之后接操作系统命令)

```
GGSCI> shell vi dirprm/kafka.props
```

```
gg.handlerlist = kafkahandler
gg.handler.kafkahandler.type = kafka
gg.handler.kafkahandler.KafkaProducerConfigFile=custom_kafka_producer.properties
gg.handler.kafkahandler.TopicName =test
gg.handler.kafkahandler.format =json
#gg.handler.kafkahandler.SchemaTopicName=mySchemaTopic
gg.handler.kafkahandler.BlockingSend =false
gg.handler.kafkahandler.includeTokens=false

gg.handler.kafkahandler.mode =op
#ggs.handler.kafkahandler.maxGroupSize =100, 1Mb
#ggs.handler.kafkahandler.minGroupSize =50, 500Kb

goldengate.userexit.timestamp=utc
goldengate.userexit.writers=javawriter
javawriter.stats.display=TRUE
javawriter.stats.full=TRUE

gg.log=log4j
gg.log.level=INFO

gg.report.time=30sec

#gg.classpath=dirprm/:/var/lib/kafka/libs/*:
gg.classpath=dirprm/:/opt/hadoopclient/Kafka/kafka/libs/*:/opt/hadoopclient/Kafka/kafka/config/*

javawriter.bootoptions=-Xmx512m -Xms32m -Djava.class.path=ggjava/ggjava.jar
```

- 其中 `gg.handler.kafkahandler.BlockingSend` 属性控制同步和异步, 默认false, 异步。

```
GGSCI> shell vi dirprm/custom_kafka_producer.properties
```

```

bootstrap.servers=162.1.93.101:21005,162.1.93.102:21005,162.1.93.103:21005
acks=1
compression.type=gzip
reconnect.backoff.ms=1000

value.serializer=org.apache.kafka.common.serialization.ByteArraySerializer
key.serializer=org.apache.kafka.common.serialization.ByteArraySerializer
100KB per partition
batch.size=102400
linger.ms=10000
max.request.size=5024000
send.buffer.bytes=5024000
~
```

- 修改Kafka里的配置，将如下选项修改为True

| 参数                             | 值                                                                 | 参数文件              |
|--------------------------------|-------------------------------------------------------------------|-------------------|
| allow.everyone.if.no.acl.found | <input type="radio"/> true <input checked="" type="radio"/> false | server.properties |

| 参数                               | 值                                                                                                                                                                             |
|----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| allow.everyone.if.no.acl.found   | <input checked="" type="radio"/> true <input type="radio"/> false                                                                                                             |
| * compression.type               | <input checked="" type="checkbox"/> producer <input type="checkbox"/> uncompressed <input type="checkbox"/> gzip <input type="checkbox"/> snappy <input type="checkbox"/> lz4 |
| * connections.max.idle.ms        | 600000                                                                                                                                                                        |
| delete.topic.enable              | <input checked="" type="radio"/> true <input type="radio"/> false                                                                                                             |
| * max.connections.per.ip         | 2147483647                                                                                                                                                                    |
| max.connections.per.ip.overrides |                                                                                                                                                                               |
| * port                           | 21005                                                                                                                                                                         |
| * sasl-ssl.port                  | 21009                                                                                                                                                                         |
| * sasl.port                      | 21007                                                                                                                                                                         |
| * security.inter.broker.protocol | <input type="checkbox"/> PLAINTEXT <input type="checkbox"/> SSL <input checked="" type="checkbox"/> SASL_PLAINTEXT <input type="checkbox"/> SASL_SSL                          |

- 把GoldenGate复制进程rkafka加入到GoldenGate管理者进程中:

```

GGSCI> add replicat rkafka, exttrail dirdat/rk
GGSCI (localhost.localdomain) 4> add replicat rkafka, exttrail dirdat/rk
REPLICAT added.
```

```

GGSCI>start rkafka
```

```

GGSCI (localhost.localdomain) 5> start rkafka
```

```

Sending START request to MANAGER ...
REPLICAT RKAFKA starting
```

```

GGSCI>info all
```

```
GGSCI (localhost.localdomain) 11> info all
```

| Program  | Status  | Group  | Lag at Chkpt | Time since Chkpt |
|----------|---------|--------|--------------|------------------|
| MANAGER  | RUNNING |        |              |                  |
| REPLICAT | RUNNING | RHBASE | 00:00:00     | 00:00:07         |
| REPLICAT | RUNNING | RHDFS  | 00:00:00     | 00:00:04         |
| REPLICAT | RUNNING | RKAFKA | 00:00:00     | 00:00:00         |

## 配置GoldenGate Flume 复制进程

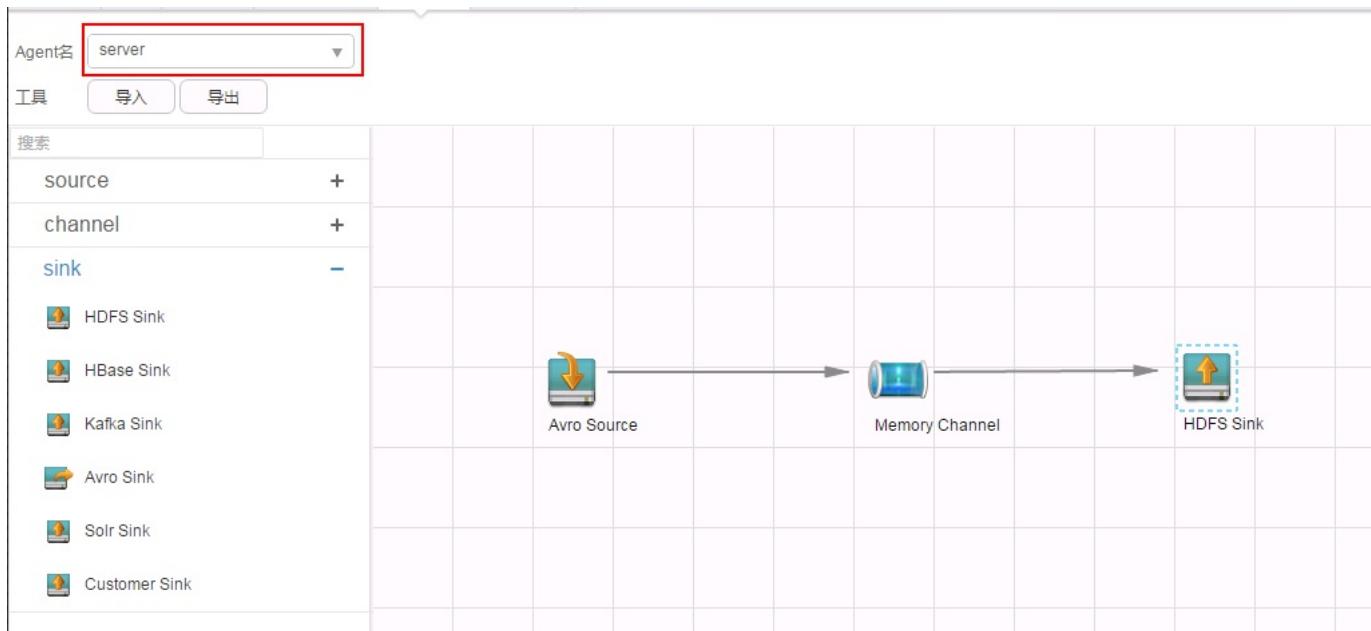
- 安装Flume客户端，配置非加密传输

```
[root@localhost opt]# tar -xvf FusionInsight_V100R002C60U10_Flume_Client.tar
FusionInsight_V100R002C60U10_Flume_ClientConfig.tar
FusionInsight_V100R002C60U10_Flume_ClientConfig.tar.sha256
[root@localhost opt]# sha256sum -c FusionInsight_V100R002C60U10_Flume_ClientConfig.tar.sha256

FusionInsight_V100R002C60U10_Flume_ClientConfig.tar: OK
[root@localhost opt]#
[root@localhost opt]#
[root@localhost opt]# tar -xvf FusionInsight_V100R002C60U10_Flume_ClientConfig.tar
FusionInsight_V100R002C60U10_Flume_ClientConfig/
FusionInsight_V100R002C60U10_Flume_ClientConfig/application.properties
FusionInsight_V100R002C60U10_Flume_ClientConfig/switchuser.py
FusionInsight_V100R002C60U10_Flume_ClientConfig/jythonLib.jar
FusionInsight_V100R002C60U10_Flume_ClientConfig/bigdata_env.sample
FusionInsight_V100R002C60U10_Flume_ClientConfig/README
FusionInsight_V100R002C60U10_Flume_ClientConfig/Flume/
FusionInsight_V100R002C60U10_Flume_ClientConfig/install.sh
FusionInsight_V100R002C60U10_Flume_ClientConfig/refreshConfig.sh
FusionInsight_V100R002C60U10_Flume_ClientConfig/ca.crt
FusionInsight_V100R002C60U10_Flume_ClientConfig/hosts
FusionInsight_V100R002C60U10_Flume_ClientConfig/conf.py
FusionInsight_V100R002C60U10_Flume_ClientConfig/install.bat
FusionInsight_V100R002C60U10_Flume_ClientConfig/JDK/
FusionInsight_V100R002C60U10_Flume_ClientConfig/KrbClient/
FusionInsight_V100R002C60U10_Flume_ClientConfig/log4j.properties
FusionInsight_V100R002C60U10_Flume_ClientConfig/KrbClient/hadoop_security_func.sh
FusionInsight_V100R002C60U10_Flume_ClientConfig/KrbClient/FusionInsight-kerberos-1.10.7.tar.gz
FusionInsight_V100R002C60U10_Flume_ClientConfig/KrbClient/component_env.sample
FusionInsight_V100R002C60U10_Flume_ClientConfig/KrbClient/install.sh
FusionInsight_V100R002C60U10_Flume_ClientConfig/KrbClient/refreshConfig.sh
FusionInsight_V100R002C60U10_Flume_ClientConfig/KrbClient/config/
FusionInsight_V100R002C60U10_Flume_ClientConfig/KrbClient/security_krb5_init_from_pwd.sh
FusionInsight_V100R002C60U10_Flume_ClientConfig/KrbClient/config/kdc.conf
FusionInsight_V100R002C60U10_Flume_ClientConfig/KrbClient/config/.gitignore
FusionInsight_V100R002C60U10_Flume_ClientConfig/KrbClient/config/krb5.conf
FusionInsight_V100R002C60U10_Flume_ClientConfig/KrbClient/config/zoo.cfg
FusionInsight_V100R002C60U10_Flume_ClientConfig/JDK/jdk.tar.gz
FusionInsight_V100R002C60U10_Flume_ClientConfig/JDK/component_env.sample
FusionInsight_V100R002C60U10_Flume_ClientConfig/JDK/install.sh
FusionInsight_V100R002C60U10_Flume_ClientConfig/Flume/FusionInsight-Flume-1.6.0.tar.gz
[root@localhost opt]#
```

```
[root@localhost ~]# cd /opt/client/FusionInsight_V100R002C60U20_Flume_ClientConfig/Flume
[root@localhost Flume]# ll
total 350964
-rw----- 1 root root 359386948 Feb 19 17:28 FusionInsight-Flume-1.6.0.tar.gz
[root@localhost Flume]# tar -xvf FusionInsight-Flume-1.6.0.tar.gz
```

- 配置Server的配置文件properties.properties



### Avro Source-Avro Source

\* 名称: Avro Source

\* bind: 162.1.93.101

\* port: 4141

threads:

compression-type: none

ssl:  true  false

additional-items:

**确认** **取消**

- 导出的properties.properties文件，增加如下配置：

```

server.sources = Avro_Source
server.channels = Memory_Channel
server.sinks = HDFS_Sink

server.sources.Avro_Source.type = avro
server.sources.Avro_Source.bind = 162.1.93.101
server.sources.Avro_Source.port = 4141
server.sources.Avro_Source.threads =
server.sources.Avro_Source.compression-type = none
server.sources.Avro_Source.compression-level = 6
server.sources.Avro_Source.ssl = false
server.sources.Avro_Source.keystore-type = JKS
server.sources.Avro_Source.keystore =
server.sources.Avro_Source.keystore-password =
server.sources.Avro_Source.truststore-type = JKS
server.sources.Avro_Source.truststore =
server.sources.Avro_Source.truststore-password =
server.sources.Avro_Source.channels = Memory_Channel

server.channels.Memory_Channel.type = memory
server.channels.Memory_Channel.capacity = 10000
server.channels.Memory_Channel.transactionCapacity = 1000
server.channels.Memory_Channel.channelfullcount = 10
server.channels.Memory_Channel.keep-alive = 3
server.channels.Memory_Channel.byteCapacity =
server.channels.Memory_Channel.byteCapacityBufferPercentage = 20

server.sinks.HDFS_Sink.type = hdfs
server.sinks.HDFS_Sink.hdfs.path = hdfs://hacluster/ogg/flume
server.sinks.HDFS_Sink.montime =
server.sinks.HDFS_Sink.hdfs.filePrefix = ggflume_
server.sinks.HDFS_Sink.hdfs.fileSuffix =
server.sinks.HDFS_Sink.hdfs.inUsePrefix =
server.sinks.HDFS_Sink.hdfs.inUseSuffix = [tmp]
server.sinks.HDFS_Sink.hdfs.idleTimeout = 0
server.sinks.HDFS_Sink.hdfs.batchSize = 1000
server.sinks.HDFS_Sink.hdfs.codeC =
server.sinks.HDFS_Sink.hdfs.fileType = DataStream
server.sinks.HDFS_Sink.hdfs.maxOpenFiles = 5000
server.sinks.HDFS_Sink.hdfs.writeFormat = Writable
server.sinks.HDFS_Sink.hdfs.callTimeout = 10000
server.sinks.HDFS_Sink.hdfs.threadsPoolSize = 10
server.sinks.HDFS_Sink.hdfs.rollTimerPoolSize = 1
server.sinks.HDFS_Sink.hdfs.kerberosPrincipal = hdfs/hadoop.hadoop.com@HADOOP.COM
server.sinks.HDFS_Sink.hdfs.kerberosKeytab = /opt/huawei/Bigdata/FusionInsight_V100R002C60U20/etc/1_10_DataNode/hdfs.keytab
server.sinks.HDFS_Sink.hdfs.round = false
server.sinks.HDFS_Sink.hdfs.roundUnit = second
server.sinks.HDFS_Sink.hdfs.useLocalTimeStamp = false
server.sinks.HDFS_Sink.hdfs.failcount = 10
server.sinks.HDFS_Sink.hdfs.fileCloseByEndEvent = true
server.sinks.HDFS_Sink.hdfs.rollInterval = 30
server.sinks.HDFS_Sink.hdfs.rollSize = 1024
server.sinks.HDFS_Sink.hdfs.rollCount = 10
server.sinks.HDFS_Sink.hdfs.batchCallTimeout = 0

```

- 可以在HDFS中增加/ogg/flume目录
- 将此properties.properties文件上传至FusionInsight。

服务 > Flume 服务配置

服务状态 实例 服务配置 资源贡献排名 配置工具 Flume管理

保存配置 导入服务配置 导出服务配置 参数类别: 基础配置 角色: 主机:

| 参数                           | 值                                                                                                                                                                   |
|------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Flume->Flume                 |                                                                                                                                                                     |
| flume.config.file            | properties.properties(0730eb6b94aabf723e4d26dad0b872ffbb4291) <span>上传文件</span> <span>下载文件</span>                                                                   |
| FLUME_LOG_MAX_FILE_SIZE      | 50MB                                                                                                                                                                |
| FLUME_LOGFILE_MAXBACKUPINDEX | 20                                                                                                                                                                  |
| GC_OPTS                      | -Xms2G -Xmx4G -XX:CMSFullGCsBeforeCompaction=1 -XX:+UseConcMarkSweepGC -XX:+CMSParallelRefinementGC                                                                 |
| LOG_LEVEL                    | <input type="checkbox"/> FATAL <input type="checkbox"/> ERROR <input type="checkbox"/> WARN <input checked="" type="checkbox"/> INFO <input type="checkbox"/> DEBUG |
| Flume->MonitorServer         |                                                                                                                                                                     |
| LOG_LEVEL                    | <input type="checkbox"/> FATAL <input type="checkbox"/> ERROR <input type="checkbox"/> WARN <input checked="" type="checkbox"/> INFO <input type="checkbox"/> DEBUG |
| MS_LOG_MAX_FILE_SIZE         | 10MB                                                                                                                                                                |
| MS_LOGFILE_MAXBACKUPINDEX    | 10                                                                                                                                                                  |

当前页: 1 总页数: 1 | < < > > | 跳转到: 确定

- 编辑rflume prm，在GGSCI命令行下运行 edit param rflume 命令：

```
GGSCI> edit param rflume
```

```
REPLICAT rflume
-- Trail file for this example is located in "AdapterExamples/trail" directory
-- Command to add REPLICAT
-- add replicat rflume, exttrail AdapterExamples/trail/tr
TARGETDB LIBFILE libggjava.so SET property=dirprm/flume.props
REPORTCOUNT EVERY 1 MINUTES, RATE
GROUPTRANSOPS 10000
MAP pdborcl.test.* , TARGET test1.*;
```

- 编辑flume.props, 在GGSCI命令行下运行 `shell vi dirprm/flume.props` 命令: (shell之后接操作系统命令)

```
GGSCI> shell vi dirprm/flume.props
```

```
gg.handlerlist = flumehandler
gg.handler.flumehandler.type=flume
gg.handler.flumehandler.RpcClientPropertiesFile=custom-flume-rpc.properties
gg.handler.flumehandler.format=avro_op
gg.handler.flumehandler.mode=tx
#gg.handler.flumehandler.maxGroupSize=100, 1Mb
#gg.handler.flumehandler.minGroupSize=50, 500 Kb
gg.handler.flumehandler.EventMapsTo=tx
gg.handler.flumehandler.PropagateSchema=false
gg.handler.flumehandler.includeTokens=false
gg.handler.flumehandler.format.WrapMessageInGenericAvroMessage=false

goldengate.userexit.timestamp=utc
goldengate.userexit.writers=javawriter
javawriter.stats.display=TRUE
javawriter.stats.full=TRUE

gg.log=log4j
gg.log.level=INFO

gg.report.time=30sec

#gg.classpath=dirprm/:/var/lib/flume/lib/*:
gg.classpath=dirprm/:/opt/client/FusionInsight_V100R002C60U20_Flume_ClientConfig/Flume/flume/lib/*:/opt/client/FusionInsight_V100R002C60U20_Flume_ClientConfig/Flume/flume/conf:

javawriter.bootoptions=-Xmx512m -Xms32m -Djava.class.path=ggjava/ggjava.jar

gg.handler.flumehandler.PropagateSchema=false 控制DDL
gg.handler.flumehandler.format.WrapMessageInGenericAvroMessage=false 相同SCHAME打包
```

```
GGSCI> shell vi dirprm/custom-flume-rpc.properties
```

```
client.type=default
hosts=h1
hosts.h1=162.1.93.101:4141
batch-size=100
connect-timeout=20000
request-timeout=20000
kerberos=true
client-principal=flume/hadoop.hadoop.com@HADOOP.COM
client-keytab=/opt/OGG_HADOOP/dirprm/flume.keytab
server-principal=flume/hadoop.hadoop.com@HADOOP.COM
```

- 拷贝flume.keytab文件到 `/opt/OGG_HADOOP/dirprm/` 目录下

```
[root@localhost dirprm]# scp -r root@162.1.93.101:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Flume-1.6.0/flume/conf/flume.keytab /opt/OGG_HADOOP/d
irprm/
root@162.1.93.101's password:
flume.keytab
```

```
100% 126 0.1KB/s 00:00
```

- 把GoldenGate复制进程rflume加入到GoldenGate管理者进程中：

```
GGSCI> add replicat rflume, exttrail dirdat/rf
```

```
GGSCI (localhost.localdomain) 12> add replicat rflume,exttrail dirdat/rf
REPLICAT added.
```

```
GGSCI>start rflume
```

```
GGSCI (localhost.localdomain) 13> start rflume
Sending START request to MANAGER ...
REPLICAT RFLUME starting
```

```
GGSCI>info all
```

```
GGSCI (localhost.localdomain) 19> info all
```

| Program  | Status  | Group  | Lag at Chkpt | Time since Chkpt |
|----------|---------|--------|--------------|------------------|
| MANAGER  | RUNNING |        |              |                  |
| REPLICAT | RUNNING | RFLUME | 00:00:00     | 00:00:02         |
| REPLICAT | RUNNING | RHBASE | 00:00:00     | 00:00:03         |
| REPLICAT | RUNNING | RHDFS  | 00:00:00     | 00:00:00         |
| REPLICAT | RUNNING | RKAFKA | 00:00:00     | 00:00:06         |

## 测试结果

### Oracle端启动所有的传输进程

- 确保所有传输进程均已经正常启动

```
GGSCI (oracledb) 14> info all
```

| Program | Status  | Group  | Lag at Chkpt | Time since Chkpt |
|---------|---------|--------|--------------|------------------|
| MANAGER | RUNNING |        |              |                  |
| EXTRACT | RUNNING | EORA   | 00:00:09     | 00:00:04         |
| EXTRACT | RUNNING | PFLUME | 00:00:00     | 07:07:45         |
| EXTRACT | RUNNING | PHBASE | 00:00:00     | 00:00:09         |
| EXTRACT | RUNNING | PHDFS  | 00:00:00     | 00:00:06         |
| EXTRACT | RUNNING | PKAFKA | 00:00:00     | 00:00:06         |

### 在Oracle数据库源端做Insert操作

```
su - oracle
source .bash_profile
sqlplus test/welcome1@pdborcl
```

```

[root@oracledb ~]# su - oracle
[oracle@oracledb ~]$ source .bash_profile
[oracle@oracledb ~]$ sqlplus test/welcome1@pdborcl

SQL*Plus: Release 12.1.0.2.0 Production on Sun Feb 19 18:45:20 2017

Copyright (c) 1982, 2014, Oracle. All rights reserved.

Last Successful login time: Fri Feb 17 2017 17:33:47 +08:00

Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit Production
with the Partitioning, OLAP, Advanced Analytics and Real Application Testing options

SQL> select * from test1;

SQL> insert into test1 values(18,'qqq');

1 row created.

SQL> commit;

Commit complete.

```

- 查看HDFS同步情况, `hadoop fs -ls /ogg1`

```

[root@localhost ~]# hdfs dfs -ls /ogg1
Found 1 items
drwxr-xr-x+ - cky supergroup 0 2017-02-21 20:17 /ogg1/test1.test1
[root@localhost ~]# hdfs dfs -ls /ogg1/test1.test1
Found 2 items
-rw-r--r--+ 3 cky supergroup 515 2017-02-21 11:50 /ogg1/test1.test1/test1_2017-02-20_10-52-11.822.txt
-rw-r--r--+ 3 cky supergroup 103 2017-02-21 20:17 /ogg1/test1.test1/test1_2017-02-21_20-17-14.650.txt
ITEST1.TEST12017-02-21 12:14:38.1185562017-02-21T20:17:14.648000000000000040000002220ID8NAMEhhh
ITEST1.TEST12017-02-22 02:31:48.3452182017-02-22T10:31:53.853000000000000040000002350ID9NAMEiji
ITEST1.TEST12017-02-22 03:03:42.3436372017-02-22T11:03:46.571000000000000040000002481ID10NAMEjjj
ITEST1.TEST12017-02-22 03:15:45.0979792017-02-22T11:15:50.599000000000000040000002611ID11NAMEkkk
ITEST1.TEST12017-02-22 06:12:15.0898312017-02-22T14:12:19.521000000000000040000002742ID12NAMElll
ITEST1.TEST12017-02-22 07:26:46.0864202017-02-22T15:26:51.183000000000000040000002873ID13NAMEiii
ITEST1.TEST12017-02-22 07:43:47.0855792017-02-22T15:43:53.565000000000000040000003004ID14NAMEmmm
ITEST1.TEST12017-02-22 07:55:23.08502202017-02-22T15:55:28.827000000000000040000003135ID15NAMEennn
ITEST1.TEST12017-02-22 07:58:48.0849892017-02-22T15:58:51.910000000000000040000003265ID16NAMEooo
ITEST1.TEST12017-02-22 08:35:32.0832512017-02-22T16:35:37.732000000000000040000003396ID17NAMEppp
ITEST1.TEST12017-02-22 10:19:53.0785822017-02-22T18:19:57.031000000000000040000003526ID18NAMEqqq

```

- 查看HBase同步情况

hbase shell

```

hbase(main):001:0> list
TABLE
2017-02-22 20:20:58,016 INFO [main] ipc.AbstractRpcClient: RPC Server Kerberos principal name for service=MasterService is hbase/hadoop.hadoop.com@HADOOP.COM
TEST1:TEST1
1 row(s) in 0.2430 seconds

=> [TEST1:TEST1]
hbase(main):002:0> scan "TEST1:TEST1"
ROW COLUMN+CELL
2017-02-22 20:21:10,741 INFO [hconnection-0x6537c964-shared--pool2-t1] ipc.AbstractRpcClient: RPC Server Kerberos principal name for service=ClientService is hbase/hado
op.hadoop.com@HADOOP.COM
2017-02-22 20:21:10,769 INFO [hconnection-0x6537c964-shared--pool1-t1] ipc.AbstractRpcClient: RPC Server Kerberos principal name for service=ClientService is hbase/hado
op.hadoop.com@HADOOP.COM
10
10 column=cf:ID, timestamp=1487732615754, value=10
10 column=cf:NAME, timestamp=1487732615754, value=jjj
11
11 column=cf:ID, timestamp=1487733337026, value=11
11 column=cf:NAME, timestamp=1487733337026, value=kkk
12
12 column=cf:ID, timestamp=1487743926983, value=12
12 column=cf:NAME, timestamp=1487743926983, value=lll
13
13 column=cf:ID, timestamp=1487748398656, value=13
13 column=cf:NAME, timestamp=1487748398656, value=iii
14
14 column=cf:ID, timestamp=1487749421045, value=14
14 column=cf:NAME, timestamp=1487749421045, value=mmm
15
15 column=cf:ID, timestamp=1487750116313, value=15
15 column=cf:NAME, timestamp=1487750116313, value=nnn
16
16 column=cf:ID, timestamp=1487750319395, value=16
16 column=cf:NAME, timestamp=1487750319395, value=ooo
17
17 column=cf:ID, timestamp=1487752526207, value=17
17 column=cf:NAME, timestamp=1487752526207, value=ppp
18
18 column=cf:ID, timestamp=1487758784532, value=18
18 column=cf:NAME, timestamp=1487758784532, value=qqq

```

- 查看kafka结果, 进入kafka客户端 `/opt/hadoopclient/Kafka/kafka/bin`

执行以下命令:

```
./kafka-console-consumer.sh --zookeeper 162.1.93.101:24002,162.1.93.102:24002,162.1.93.103:24002/kafka --topic test --from-beginning
```

```
[root@localhost bin]# ./kafka-console-consumer.sh --zookeeper 162.1.93.101:24002,162.1.93.102:24002,162.1.93.103:24002/kafka --topic test --from-beginning
{"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-19 10:46:09.109769","current_ts":"2017-02-19T18:46:14.413000","pos":"00000000000000001825","after":{"ID":"3","NAME":"ccc"})
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-19 11:27:41.107892","current_ts":"2017-02-19T19:27:45.383000","pos":"00000000000000001959","after":{"ID":"4","NAME":"ddd"))
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-19 12:14:52.105662","current_ts":"2017-02-19T20:14:57.440000","pos":"00000000000000002089","after":{"ID":"5","NAME":"eee"))
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-19 12:24:04.105165","current_ts":"2017-02-19T20:24:09.652000","pos":"00000000000000002218","after":{"ID":"6","NAME":"fff"))
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-20 03:29:42.192251","current_ts":"2017-02-20T11:29:46.911000","pos":"00000000000000002347","after":{"ID":"7","NAME":"ggg"))
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-21 12:14:38.118550","current_ts":"2017-02-21T20:14:43.619000","pos":"00000000000000002477","after":{"ID":"8","NAME":"hhh"))
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 02:31:48.345130","current_ts":"2017-02-22T10:31:55.108000","pos":"00000000000000002607","after":{"ID":"9","NAME":"iii"))
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 03:03:42.343646","current_ts":"2017-02-22T11:03:46.810000","pos":"00000000000000002738","after":{"ID":"10,"NAME":"jjj"))
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 03:15:45.098040","current_ts":"2017-02-22T11:15:50.837000","pos":"00000000000000002868","after":{"ID":"11,"NAME":"kkk"))
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 06:12:15.089840","current_ts":"2017-02-22T14:12:19.719000","pos":"00000000000000002999","after":{"ID":"12,"NAME":"lll"))
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 07:26:46.086558","current_ts":"2017-02-22T15:26:51.359000","pos":"00000000000000003130","after":{"ID":"13,"NAME":"iii"))
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 07:43:47.085649","current_ts":"2017-02-22T15:43:53.736000","pos":"00000000000000003261","after":{"ID":"14,"NAME":"mmm"))
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 07:55:23.085032","current_ts":"2017-02-22T15:55:28.999000","pos":"00000000000000003392","after":{"ID":"15,"NAME":"nnn"))
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 07:58:48.084996","current_ts":"2017-02-22T15:58:52.079000","pos":"00000000000000003522","after":{"ID":"16,"NAME":"ooo"))
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 08:35:32.083167","current_ts":"2017-02-22T16:35:38.874000","pos":"00000000000000003653","after":{"ID":"17,"NAME":"ppp"))
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 10:19:53.078483","current_ts":"2017-02-22T18:19:57.144000","pos":"00000000000000003783","after":{"ID":"18,"NAME":"qqq"))
```

- 在HDFS中查看flume运行结果：查看/ogg/flume/下数据文件：

```
[root@localhost ~]# hdfs dfs -ls /ogg/flume
Found 1 items
-rw-r--r--+ 3 hdfs supergroup 1508 2017-02-22 17:15 /ogg/flume/ggflume_.tmp
[root@localhost ~]# hdfs dfs -cat /ogg/flume/ggflume_.tmp
TEST1.TEST1I42017-02-19 11:27:41.10779442017-02-22T17:15:10.465000(000000002000001825ID4ddd
TEST1.TEST1I42017-02-19 12:14:52.10561442017-02-22T17:15:10.579000(000000002000001959ID5eeeTEST1.TEST1I42017-02-19 1
2:24:04.10516842017-02-22T17:15:10.579001(0000000020000002092ID6ffffTEST1.TEST1I42017-02-20 03:29:42.19220442017-02-22
T17:15:10.579002(0000000020000002221ID7gggTEST1.TEST1I42017-02-21 12:14:38.11858042017-02-22T17:15:10.580000(00000000
0200000002351ID8hhhtEST1.TEST1I42017-02-22 02:31:48.34521842017-02-22T17:15:10.580001(0000000040000001825ID9iiTEST1.T
EST1I42017-02-22 03:03:42.34371142017-02-22T17:15:10.580002(0000000040000001960ID10jjjTEST1.TEST1I42017-02-22 03:15:4
5.09804842017-02-22T17:15:10.580003(0000000060000001825ID11kkkTEST1.TEST1I42017-02-22 06:12:15.08984342017-02-22T17:1
5:10.580004(0000000080000001825ID12111TEST1.TEST1I42017-02-22 07:26:46.08643242017-02-22T17:15:10.581000(000000000800
0001960ID13iiTEST1.TEST1I42017-02-22 07:43:47.08562642017-02-22T17:15:10.581001(0000000080000002095ID14mmmmTEST1.TES
T1I42017-02-22 07:55:23.08502742017-02-22T17:15:10.581002(0000000080000002226ID15nnntEST1.TEST1I42017-02-22 07:58:48.
08499542017-02-22T17:15:10.581003(0000000080000002356ID16000TEST1.TEST1I42017-02-22 08:35:32.08315642017-02-22T17:15:
10.581004(0000000080000002487ID17ppp
TEST1.TEST1I42017-02-22 10:19:53.07848342017-02-22T18:19:57.031000(0000000080000002617ID18qqq
[root@localhost ~]#
```

## 在Oracle数据库源端做Update操作

- 执行以下命令

```
su - oracle
source .bash_profile
sqlplus test/welcome1@pdborcl
```

```
SQL> select * from test1;
```

| ID | NAME |
|----|------|
| 1  | aaa  |
| 2  | bbb  |
| 3  | ccc  |
| 4  | ddd  |
| 5  | eee  |
| 6  | fff  |
| 7  | ggg  |
| 8  | hhh  |
| 9  | iii  |
| 10 | jjj  |
| 11 | kkk  |

| ID | NAME |
|----|------|
| 12 | lll  |
| 13 | iii  |
| 14 | mmm  |
| 15 | nnn  |
| 16 | ooo  |
| 17 | ppp  |
| 18 | qqq  |

18 rows selected.

```
SQL> update test1 set name='test' where id=18;
```

1 row updated.

```
SQL> commit;
```

Commit complete.

```
SQL> update test1 set name='test' where id=18;
```

```
1 row updated.
```

```
SQL> commit;
```

```
Commit complete.
```

```
SQL> select * from test1;
```

| ID | NAME |
|----|------|
| 1  | aaa  |
| 2  | bbb  |
| 3  | ccc  |
| 4  | ddd  |
| 5  | eee  |
| 6  | fff  |
| 7  | ggg  |
| 8  | hhh  |
| 9  | iii  |
| 10 | jjj  |
| 11 | kkk  |

```
[root@localhost bin]# ./kafka-console-consumer.sh --zookeeper 162.1.93.101:24002,162.1.93.102:24002,162.1.93.103:24002/kafka --topic test --from-beginning
{"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-19 10:46:09.109769","current_ts":"2017-02-19T18:46:14.413000","pos":"00000000000000001825","after":{"ID":"3","NAME":"ccc"}}
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-19 11:27:41.107892","current_ts":"2017-02-19T19:27:45.383000","pos":"00000000000000001959","after":{"ID":"4","NAME":"ddd"}}
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-19 12:14:52.105662","current_ts":"2017-02-19T20:14:57.440000","pos":"00000000000000002089","after":{"ID":"5","NAME":"eee"}}
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-19 12:24:04.105165","current_ts":"2017-02-19T20:24:09.652000","pos":"00000000000000002218","after":{"ID":"6","NAME":"fff"}}
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-20 03:29:42.192251","current_ts":"2017-02-20T11:29:46.911000","pos":"00000000000000002347","after":{"ID":"7","NAME":"ggg"}}
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-21 12:14:38.118550","current_ts":"2017-02-21T20:14:43.619000","pos":"00000000000000002477","after":{"ID":"8","NAME":"hhh"}}
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 02:31:48.345130","current_ts":"2017-02-22T10:31:55.108000","pos":"00000000000000002607","after":{"ID":"9","NAME":"iii"}}
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 03:03:42.343646","current_ts":"2017-02-22T11:03:46.810000","pos":"00000000000000002738","after":{"ID":"10","NAME":"jjj"}}
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 03:15:45.098040","current_ts":"2017-02-22T11:15:50.837000","pos":"00000000000000002868","after":{"ID":"11","NAME":"kkk"}}
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 06:12:15.089840","current_ts":"2017-02-22T14:12:19.719000","pos":"00000000000000002999","after":{"ID":"12","NAME":"lll"}}
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 07:26:46.086558","current_ts":"2017-02-22T15:26:51.359000","pos":"00000000000000003130","after":{"ID":"13","NAME":"iii"}}
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 07:43:47.085649","current_ts":"2017-02-22T15:43:53.736000","pos":"00000000000000003261","after":{"ID":"14","NAME":"mmm"}}
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 07:55:23.085032","current_ts":"2017-02-22T15:55:28.999000","pos":"00000000000000003392","after":{"ID":"15","NAME":"nnn"}}
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 07:58:48.084996","current_ts":"2017-02-22T15:58:52.079000","pos":"00000000000000003522","after":{"ID":"16","NAME":"ooo"}}
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 08:35:32.083167","current_ts":"2017-02-22T16:35:38.874000","pos":"00000000000000003653","after":{"ID":"17","NAME":"ppp"}}
 {"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 10:19:53.078483","current_ts":"2017-02-22T18:19:57.144000","pos":"00000000000000003783","after":{"ID":"18","NAME":"qqq"}}
 {"table":"TEST1.TEST1","op_type":"U","op_ts":"2017-02-23 01:24:28.036518","current_ts":"2017-02-23T09:24:34.293000","pos":"00000000000000003914","before":{"ID":"18","NAME":"qqq"},"after":{"ID":"18","NAME":"test"}]
```

- 在HDFS中查看flume运行结果：查看/ogg/flume/下数据文件：

```
[root@localhost ~]# hdfs dfs -cat /ogg/flume/ggflume_.tmp
TEST1.TEST1I42017-02-19 11:27:41.10779442017-02-22T17:15:10.465000(00000000020000001825)D4ddd
TEST1.TEST1I42017-02-19 12:14:52.10561442017-02-22T17:15:10.579000(00000000200000001959)D5eeeTEST1.TEST1I42017-02-19 12:24:04.10516842017-02-22T
17:15:10.579001(00000000200000002092)D6ffffTEST1.TEST1I42017-02-20 03:29:42.19220442017-02-22T17:15:10.579002(00000000200000002221)D799gTEST1.TES
T1I42017-02-21 12:14:38.11858042017-02-22T17:15:10.580000(00000000020000002351)D8hhTEST1.TEST1I42017-02-22 02:31:48.34521842017-02-22T17:15:10.
580001(000000000400000001825)D9iiitEST1.TEST1I42017-02-20 03:03:42.34371142017-02-22T17:15:10.580002(00000000040000001960)D10jjjTEST1.TEST1I42017
-02-21 03:15:45.09804842017-02-22T17:15:10.580003(00000000060000001825)D11kkkTEST1.TEST1I42017-02-22 06:12:15.08984342017-02-22T17:15:10.580004(
000000000800000001825)D12lllTEST1.TEST1I42017-02-22 07:26:46.08643242017-02-22T17:15:10.581000(00000000080000001960)D13iiiTEST1.TEST1I42017-02-22
07:43:47.08562642017-02-22T17:15:10.581001(00000000080000002095)D14mmmTEST1.TEST1I42017-02-22 07:55:23.08502742017-02-22T17:15:10.581002(0000000
000000002226)D15nnnTEST1.TEST1I42017-02-22 07:58:48.08499542017-02-22T17:15:10.581003(00000000080000002356)D16oo0TEST1.TEST1I42017-02-22 08:35
:32.08315642017-02-22T17:15:10.581004(00000000080000002487)D17ppp
TEST1.TEST1I42017-02-22 10:19:53.07848342017-02-22T18:19:57.031000(00000000080000002617)D18qqqa
TEST1.TEST1I42017-02-23 01:24:28.03655042017-02-23T09:24:34.295000(00000000080000002748)D18qqqltest
[root@localhost ~]#
```

## 在Oracle数据库源端做Delete操作

- 执行以下命令

```
su - oracle
source .bash_profile
sqlplus test/welcome1@pdborcl
```

```
17 ppp
18 test
```

18 rows selected.

```
SQL> delete from test1 where id=18;
```

1 row deleted.

```
SQL> commit;
```

Commit complete.

```
SQL> select * from test1;
```

| ID | NAME |
|----|------|
| 1  | aaa  |
| 2  | bbb  |
| 3  | ccc  |
| 4  | ddd  |
| 5  | eee  |
| 6  | fff  |
| 7  | ggg  |
| 8  | hhh  |
| 9  | iii  |
| 10 | jjj  |
| 11 | kkk  |

| ID | NAME |
|----|------|
| 12 | lll  |
| 13 | iii  |
| 14 | mmm  |
| 15 | nnn  |
| 16 | ooo  |
| 17 | ppp  |

```
17 rows selected.
```

```
SQL>
```

- 查看HDFS同步情况，hadoop fs -ls /ogg1

```
[root@localhost ~]# hdfs dfs -ls /ogg1
Found 1 items
drwxr-xr-x+ 3 cyy supergroup 0 2017-02-23 09:26 /ogg1/test1.test1
[root@localhost ~]# hdfs dfs -ls /ogg1/test1.test1
Found 3 items
-rw-r--r--+ 3 cyy supergroup 515 2017-02-21 11:50 /ogg1/test1.test1/test1.test1_2017-02-20_10-52-11.822.txt
-rw-r--r--+ 3 cyy supergroup 1142 2017-02-22 21:16 /ogg1/test1.test1/test1.test1_2017-02-21_20-17-14.650.txt
-rw-r--r--+ 3 cyy supergroup 105 2017-02-23 09:26 /ogg1/test1.test1/test1.test1_2017-02-23_09-26-53.556.txt
[root@localhost ~]# hdfs dfs -cat /ogg1/test1.test1/test1.test1_2017-02-23_09-26-53.556.txt
TEST1.TEST12017-02-23 01:24:28.0365232017-02-23T09:26:53.5540000000000040000003657ID18NAMETest
TEST1.TEST12017-02-23 01:36:28.0359932017-02-23T09:36:32.4450000000000040000003814ID18NAMETest
[root@localhost ~]#
```

- 查看HBase同步情况

```
hbase shell
```

```
hbase(main):003:0> scan "TEST1:TEST1"
ROW
2017-02-23 09:39:42,863 INFO [hconnection-0x776d4580-shared--pool1-t3] ipc.AbstractRpcClient: RPC Server Kerberos principal name for service=ClientService is hbase/hadoop.hadoop.com@HADOOP.COM
10 column=cf:ID, timestamp=1487732615754, value=10
10 column=cf:NAME, timestamp=1487732615754, value=jjj
11 column=cf:ID, timestamp=148773337026, value=11
11 column=cf:NAME, timestamp=148773337026, value=kkk
12 column=cf:ID, timestamp=1487743926983, value=12
12 column=cf:NAME, timestamp=1487743926983, value=lll
13 column=cf:ID, timestamp=1487748398656, value=13
13 column=cf:NAME, timestamp=1487748398656, value=iii
14 column=cf:ID, timestamp=1487749421045, value=14
14 column=cf:NAME, timestamp=1487749421045, value=mmm
15 column=cf:ID, timestamp=1487750116313, value=15
15 column=cf:NAME, timestamp=1487750116313, value=nnn
16 column=cf:ID, timestamp=1487750319395, value=16
16 column=cf:NAME, timestamp=1487750319395, value=ooo
17 column=cf:ID, timestamp=1487752526207, value=17
17 column=cf:NAME, timestamp=1487752526207, value=ppp
3 column=cf:ID, timestamp=1487501166344, value=3
3 column=cf:NAME, timestamp=1487501166344, value=ccc
4 column=cf:ID, timestamp=1487503657237, value=4
4 column=cf:NAME, timestamp=1487503657237, value=ddd
5 column=cf:ID, timestamp=1487506488271, value=5
5 column=cf:NAME, timestamp=1487506488271, value=eee
6 column=cf:ID, timestamp=1487507040483, value=6
6 column=cf:NAME, timestamp=1487507040483, value=fff
7 column=cf:ID, timestamp=1487561376298, value=7
7 column=cf:NAME, timestamp=1487561376298, value=ggg
8 column=cf:ID, timestamp=1487679274113, value=8
8 column=cf:NAME, timestamp=1487679274113, value=hhh
9 column=cf:ID, timestamp=1487730704034, value=9
9 column=cf:NAME, timestamp=1487730704034, value=iii
15 row(s) in 0.0910 seconds
```

- 查看kafka结果，进入kafka客户端 `/opt/hadoopclient/Kafka/kafka/bin`

执行以下命令：

```
./kafka-console-consumer.sh --zookeeper 162.1.93.181:24002,162.1.93.102:24002,162.1.93.103:24002/kafka --topic test --from-beginning
```

```
[root@localhost bin]# ./kafka-console-consumer.sh --zookeeper 162.1.93.101:24002,162.1.93.102:24002,162.1.93.103:24002/kafka --topic test --from-beginning
{"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-19 10:46:09.109769","current_ts":"2017-02-19T18:46:14.413000","pos":"00000000000000001825","after":{"ID":3,"NAME":"ccc"}}
{"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-19 11:27:41.107892","current_ts":"2017-02-19T19:27:45.383000","pos":"00000000000000001959","after":{"ID":4,"NAME":"ddd"}}
{"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-19 12:14:52.105662","current_ts":"2017-02-19T20:14:57.440000","pos":"00000000000000002089","after":{"ID":5,"NAME":"eee"}}
{"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-19 12:24:04.105165","current_ts":"2017-02-19T20:24:09.652000","pos":"00000000000000002218","after":{"ID":6,"NAME":"fff"}}
{"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-20 03:29:42.192251","current_ts":"2017-02-20T11:29:46.911000","pos":"00000000000000002347","after":{"ID":7,"NAME":"ggg"}}
{"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-21 12:14:38.118550","current_ts":"2017-02-21T20:14:43.619000","pos":"00000000000000002477","after":{"ID":8,"NAME":"hhh"}}
{"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 02:31:48.345130","current_ts":"2017-02-22T10:31:55.108000","pos":"00000000000000002607","after":{"ID":9,"NAME":"iii"}}
{"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 03:03:42.343646","current_ts":"2017-02-22T11:03:46.810000","pos":"00000000000000002738","after":{"ID":10,"NAME":"jjj"}}
{"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 03:15:45.098040","current_ts":"2017-02-22T11:15:50.837000","pos":"00000000000000002868","after":{"ID":11,"NAME":"kkk"}}
{"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 06:12:15.089840","current_ts":"2017-02-22T14:12:19.719000","pos":"00000000000000002999","after":{"ID":12,"NAME":"lll"}}
{"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 07:26:46.086558","current_ts":"2017-02-22T15:26:51.359000","pos":"00000000000000003130","after":{"ID":13,"NAME":"iii"}}
{"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 07:43:47.085649","current_ts":"2017-02-22T15:43:53.736000","pos":"00000000000000003261","after":{"ID":14,"NAME":"mmm"}}
{"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 07:55:23.085032","current_ts":"2017-02-22T15:55:28.999000","pos":"00000000000000003392","after":{"ID":15,"NAME":"nnn"}}
{"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 07:58:48.084996","current_ts":"2017-02-22T15:58:52.079000","pos":"00000000000000003522","after":{"ID":16,"NAME":"ooo"}}
{"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 08:35:32.083167","current_ts":"2017-02-22T16:35:38.874000","pos":"00000000000000003653","after":{"ID":17,"NAME":"ppp"}}
{"table":"TEST1.TEST1","op_type":"I","op_ts":"2017-02-22 10:19:53.078483","current_ts":"2017-02-22T18:19:57.144000","pos":"00000000000000003783","after":{"ID":18,"NAME":"qqq"}}
{"table":"TEST1.TEST1","op_type":"U","op_ts":"2017-02-23 01:24:28.036518","current_ts":"2017-02-23T09:24:34.293000","pos":"00000000000000003914","before":{"ID":18,"NAME":"qqq"},"after":{"ID":18,"NAME":"test"}}
{"table":"TEST1.TEST1","op_type":"U","op_ts":"2017-02-23 01:36:28.035996","current_ts":"2017-02-23T09:36:32.558000","pos":"00000000000000004071","before":{"ID":18,"NAME":"test"}}
```

- 在HDFS中查看flume运行结果：查看/ogg/flume/下数据文件：

```
[root@localhost ~]# hdfs dfs -ls /ogg
Found 1 items
drwxrwxr-x+ - hdfs supergroup 0 2017-02-22 17:15 /ogg/flume
[root@localhost ~]# hdfs dfs -ls /ogg/flume
Found 1 items
-rw-r--r--+ 3 hdfs supergroup 1508 2017-02-22 17:15 /ogg/flume/ggflume_.tmp
[root@localhost ~]# hdfs dfs -cat /ogg/flume/ggflume_.tmp
TEST1.TEST142017-02-19 11:27:41.1079442017-02-22T17:15:10.465000(000000000200000018251D4ddd
TEST1.TEST142017-02-19 12:14:52.10561442017-02-22T17:15:10.579000(000000000200000019591D5eeeTest1.TEST1142017-02-19 12:24:04.10516842017-02-22T17:15:10.579001(000000000200000020921D6ffffTest1.TEST1142017-02-02 03:29.42.19220442017-02-22T17:15:10.579002(000000000200000022211D7gggTest1.TEST1142017-02-21 12:14:38.11858042017-02-22T17:15:10.580000(000000000200000023511D8hhTest1.TEST1142017-02-22 03:21:48.34521842017-02-22T17:15:10.580001(000000000400000018251D91iiTest1.TEST1142017-02-22 03:03:42.34371142017-02-22T17:15:10.580002(000000000400000019601D10jjTest1.TEST1142017-02-22 03:15:45.08984842017-02-22T17:15:10.580003(000000000600000018251D11kkkTest1.TEST1142017-02-22 06:12:15.08984342017-02-22T17:15:10.580004(000000000800000018251D2111Test1.TEST1142017-02-22 07:26:46.08634324017-02-22T17:15:10.581000(000000000800000019601D13iiTest1.TEST1142017-02-22 07:43:47.08562642017-02-22T17:15:10.581001(000000000800000020951D14mmmmTest1.TEST1142017-02-22 07:55:23.08502742017-02-22T17:15:10.581002(0000000800000022261D15nnnTest1.TEST1142017-02-22 07:58:48.08499542017-02-22T17:15:10.581003(000000000800000023561D16000Test1.TEST1142017-02-22 08:35:32.08315642017-02-22T17:15:10.581004(000000000800000024871D17ppp
TEST1.TEST1142017-02-22 10:19:53.07848342017-02-22T18:19:57.031000(000000000800000026171D18qqq
TEST1.TEST1142017-02-23 01:24:28.03655042017-02-23T09:24:34.295000(000000000800000027481D18qq1test
TEST1.TEST1142017-02-23 01:36:28.03602142017-02-23T09:36:32.560000(000000000800000029051D1test
[root@localhost ~]#
```