

IBM InfoSphere DataStage对接FusionInsight

适用场景

IBM InfoSphere DataStage 11.3.1.0 ↔ FusionInsight HD V100R002C50

IBM InfoSphere DataStage 11.5.0.2 ↔ FusionInsight HD V100R002C60U20

前提条件

- 已完成IBM InfoSphere DataStage 11.5.0.2的安装部署（本文部署在Centos7.2上）
- 已完成FusionInsight集群的部署，版本FusionInsight HD V100R002C60U20

准备工作

配置域名解析

- 使用 `vi /etc/hosts` 命令修改DataStage Server和Client的hosts文件，添加FI集群节点信息，如：

```
162.1.61.42 FusionInsight2  
162.1.61.41 FusionInsight1  
162.1.61.43 FusionInsight3
```

配置Kerberos认证

- 在FI管理界面创建DataStage对接用户，并赋予该用户所需权限，下载认证凭据

用户名	描述	创建时间	操作
admin	FusionInsight Manager的管理员。	2017-06-16 10:25:00	[key icon] [edit icon] [lock icon] [refresh icon] [remove icon]
kafka		2017-06-22 21:38:11	[key icon] [edit icon] [lock icon] [refresh icon] [remove icon]
test		2017-06-16 17:22:43	[key icon] [edit icon] [lock icon] [refresh icon] [remove icon]

- 解压下载的tar文件，得到Kerberos配置文件krb5.conf和用户的keytab文件。
- 以root登录DataStage Server节点，将FI集群的krb5.conf文件复制到 `/etc` 目录。
- 将用户的user.keytab文件上传到DataStage Server节点的任意目录，如 `/home/dsadm`。

安装FusionInsight客户端

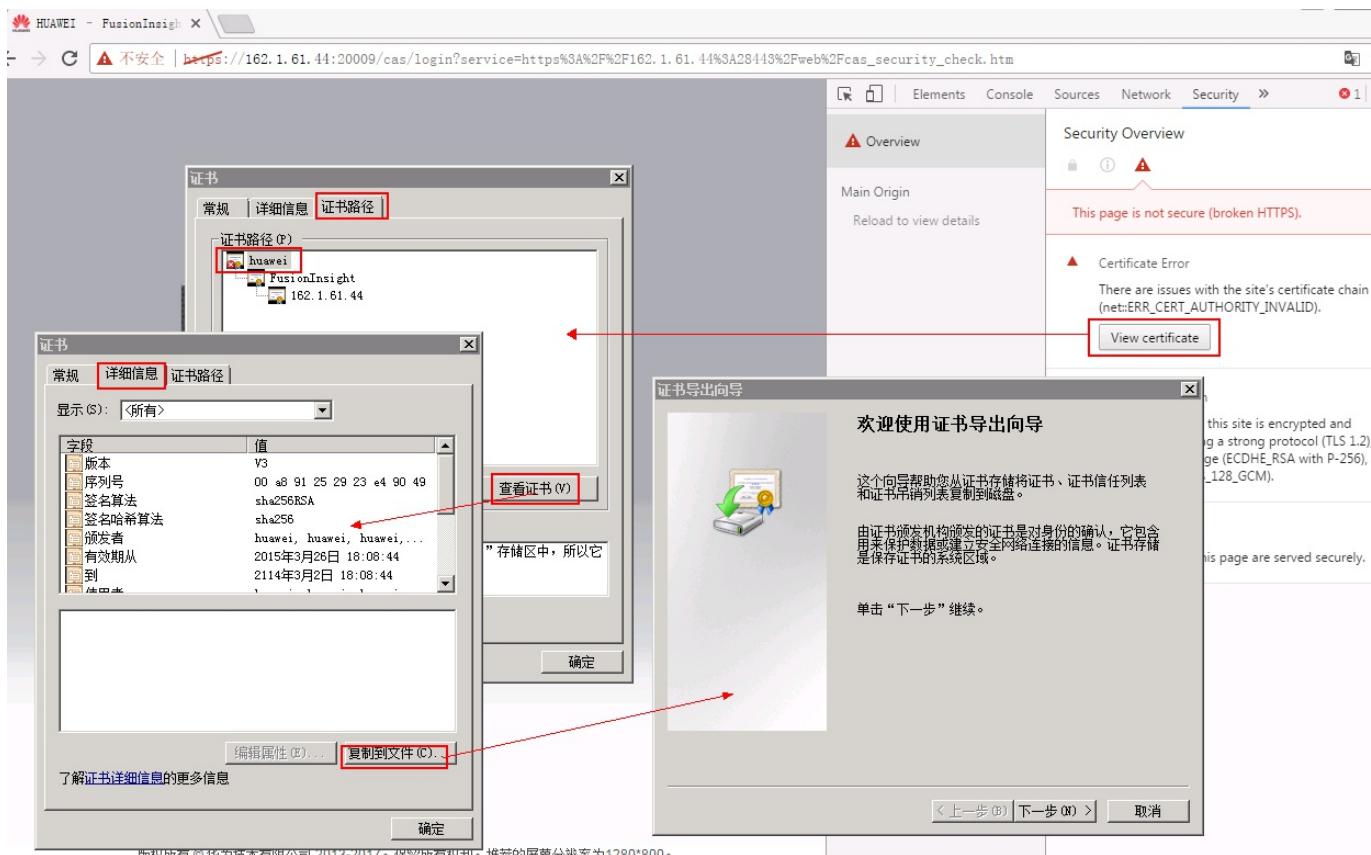
参考FI产品文档，在FI服务管理界面下载完整客户端，上传到DataStageServer，安装至自定义目录，如 `/opt/ficlient`。

对接HDFS

导入FI集群的SSL证书

- 浏览器导出FI集群的根证书

浏览器打开FI管理界面，查看证书，点击“证书路径”页签，选择根路径，查看根证书，在“详细信息”页签下，点击“复制到文件”，导出为cer格式



- 证书导入DataStage的keystore文件

将导出的FI根证书fi-root-ca.cer上传到DataStage服务端，如 `/home/dsadm` 路径下，将证书导入到keystore文件，命令参考：

```
/opt/IBM/InformationServer/jdk/bin/keytool -importcert -file /home/dsadm/fi-root-ca.cer -keystore /home/dsadm/iis-ds-truststore_ssl.jks -alias fi-root-ca.cer -storepass Huawei@123 -trustcacerts -noprompt
chown dsadm:dsstage /home/dsadm/iis-ds-truststore_ssl.jks
```

- 生成并保存加密后的keystore密码

```
[root@datastage11502 dsadm]# /opt/IBM/InformationServer/ASBNode/bin/encrypt.sh
Enter the text to encrypt:
Enter the text again to confirm:
{iisenc}SvtJ2f/uNTrvbuh26XDzag==
[root@datastage11502 dsadm]#
```

使用 `vi /home/dsadm/authenticate.properties` 命令新建配置文件，保存上一步骤生成的密文：

```
password={iisenc}SvtJ2f/uNTrvbuh26XDzag==
```

执行 `chown dsadm:dsstage /home/dsadm/authenticate.properties` 修改配置文件的属主

- 导出truststore环境变量

使用 `vi /opt/IBM/InformationServer/Server/DSEngine/dsenv` 编辑DSEngine的环境变量，在最后添加

```
export DS_TRUSTSTORE_LOCATION=/home/dsadm/iis-ds-truststore_ssl.jks
export DS_TRUSTSTORE_PROPERTIES=/home/dsadm/authenticate.properties
```

- 重启DSEngine，参考命令

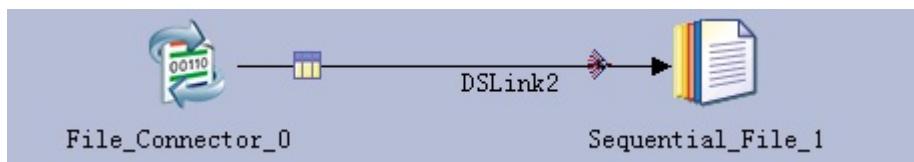
```
su - dsadm
cd $DSHOME
bin/uv -admin -stop
bin/uv -admin -start
```

读取HDFS文件

- 创建作业

新建并行作业，保存为hdfs2sf

添加File_Connector组件和Sequential File组件，以及File_Connector到Sequential File链接



- 参考下图修改配置

File_Connector_0 - File Connector

Stage | Output

Stage name
File_Connector_0

General Properties Advanced

Connection

File system	WebHDFS
Use custom URL	No
Use SSL (HTTPS)	Yes
Use Kerberos	Yes
Use keytab	Yes
Host *	fusioninsight1 NameNode主节点，主机名小写
Port	25003
Service principal	
User name *	test
Password	
Keytab *	/home/dsadm/user.keytab
Custom URL *	

Usage

File name *	/tmp/data.txt
Read mode	Read multiple files
Exclude files	
Reject mode	Continue
File name column	
File format	Comma-separated value (CSV)

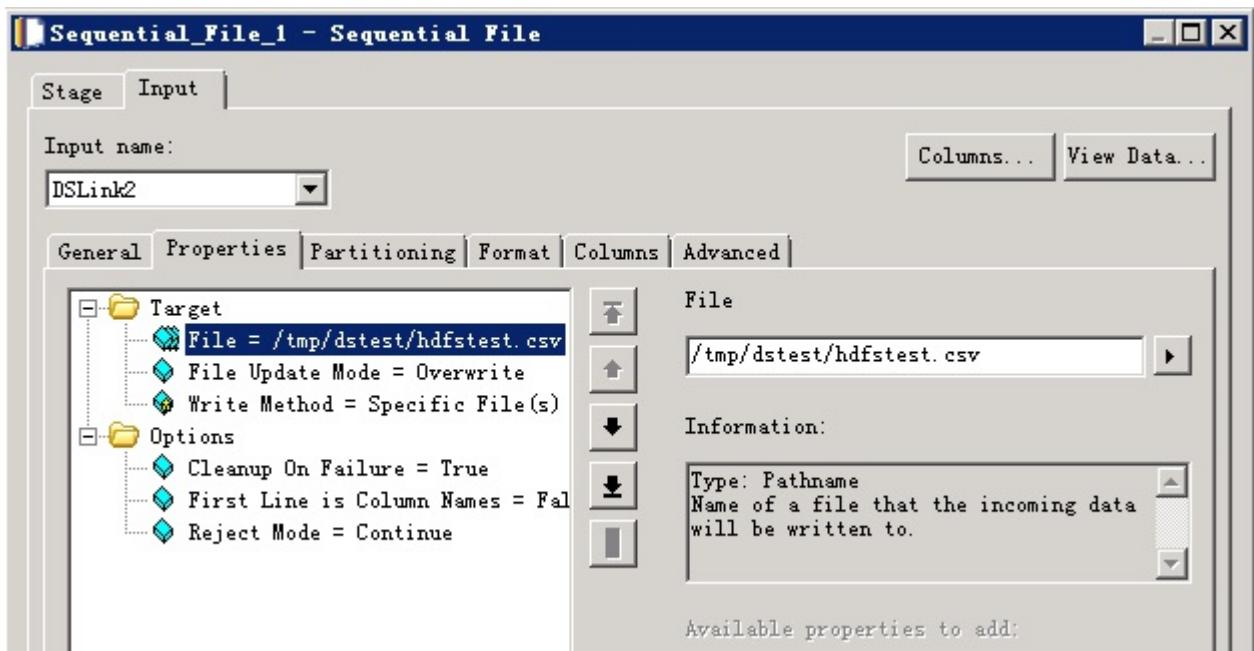
File_Connector_0 - File Connector

Stage | Output

Output name (downstream stage)
DSLink2 (Sequential_File_1)

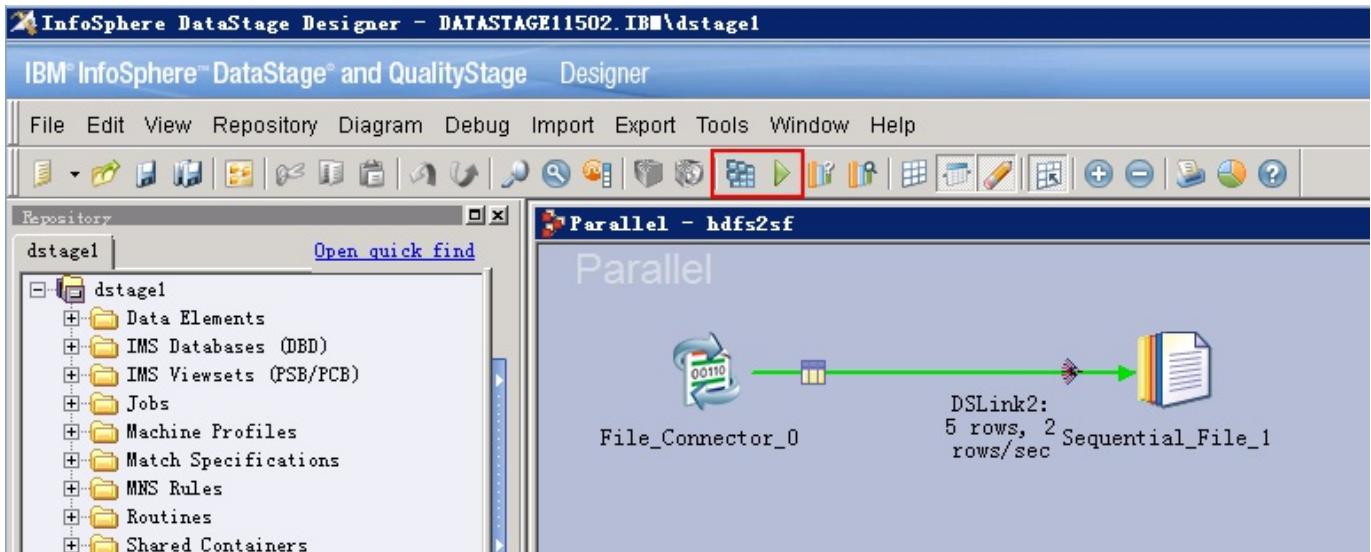
General Properties Link Ordering Advanced Columns

	Column name	Derivation	Key	SQL type	Extended	Length	Scale	Nullable	Data e
1	id		<input type="checkbox"/>	Integer		5		No	
2	name		<input type="checkbox"/>	VarChar		20		No	
3	usd_flag		<input type="checkbox"/>	VarChar		2		No	
4	salary		<input type="checkbox"/>	Double		8		No	
5	address		<input type="checkbox"/>	VarChar		30		No	
6	entrytime		<input type="checkbox"/>	VarChar		4		No	



- 编译运行

保存配置后，编译，运行



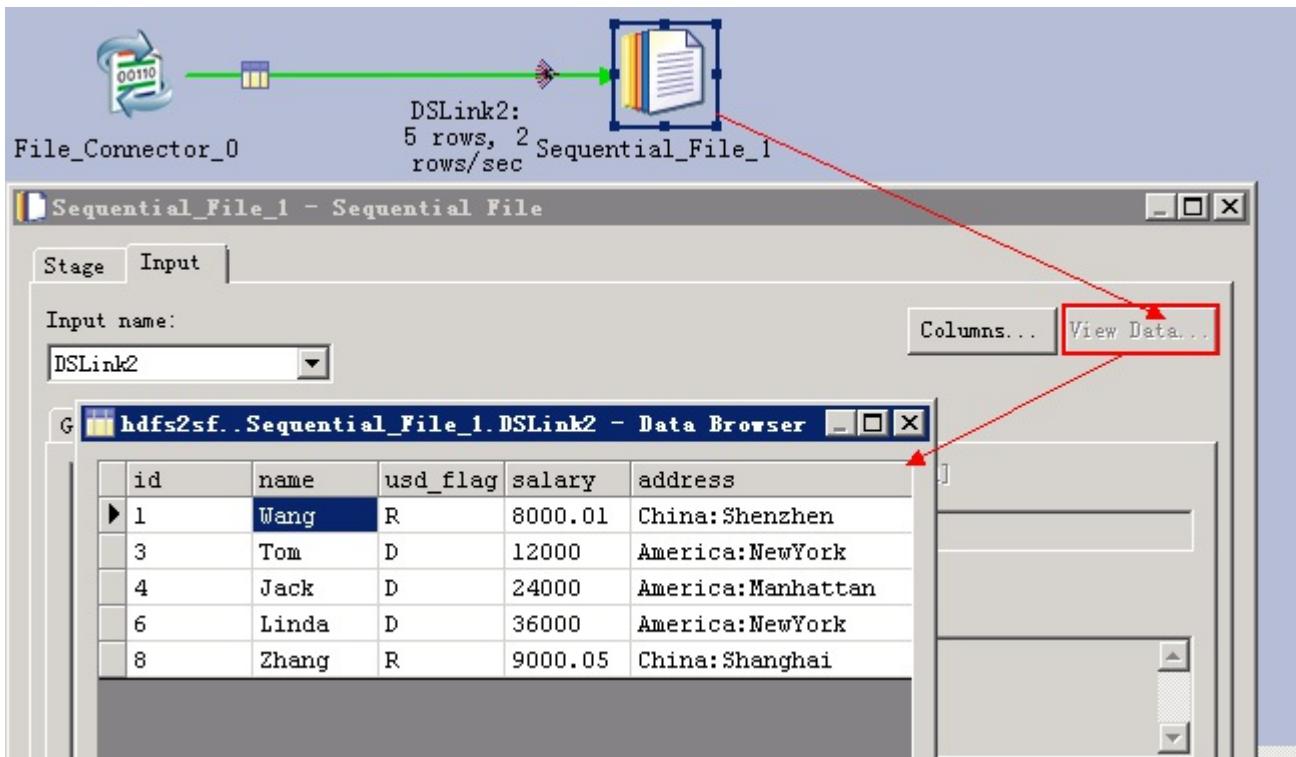
在菜单 Tools -> Run Director 中打开Director客户端，查看作业日志

```

21:07:18 2017/7/6 Info File_Connector_0,1: [JGSS_DBG_CRED] main find keys for test@HADOOP.COM (...)
21:07:18 2017/7/6 Info File_Connector_0,0: [JGSS_DBG_CTX] main initSecContext input buffer: (len=108, offset=0) (...)
21:07:18 2017/7/6 Info File_Connector_0,1: [JGSS_DBG_CTX] main initSecContext input buffer: (len=108, offset=0) (...)
21:07:18 2017/7/6 Info File_Connector_0,0: The file encoding was not specified. The default encoding of UTF-8 shall be assumed.
21:07:18 2017/7/6 Info File_Connector_0,0: 3 rows were read from the /tmp/data.txt file.
21:07:18 2017/7/6 Info File_Connector_0,1: The file encoding was not specified. The default encoding of UTF-8 shall be assumed.
21:07:18 2017/7/6 Info File_Connector_0,1: 2 rows were read from the /tmp/data.txt file.
21:07:18 2017/7/6 Info Sequential_File_1,0: Export complete: 5 records exported successfully, 0 rejected.
21:07:18 2017/7/6 Info main_program: Step execution finished with status = OK.
21:07:18 2017/7/6 Info main_program: Startup time, 0:01; production run time, 0:02.
21:07:18 2017/7/6 Info Parallel job reports successful completion
21:07:18 2017/7/6 Control Finished Job hdf2sf.

```

- 查看读取的数据



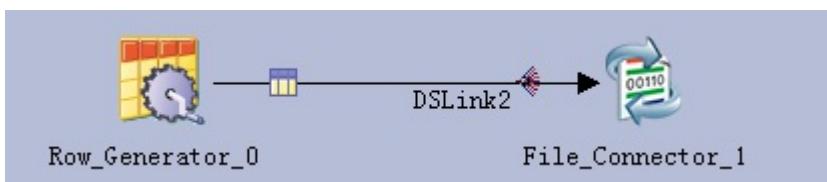
```
[root@datastage11502 tmp]# cd /tmp/dstest/
[root@datastage11502 dstest]# ll
total 4
-rw-rw-r-- 1 dsadm dstage 331 Jul  6 21:07 hdfstest.csv
[root@datastage11502 dstest]# cat hdfstest.csv
"1","Wang","R","8.0000100000000022E+03","China:Shenzhen","2014"
"3","Tom","D","1.2000020000000004E+04","America:NewYork","2014"
"4","Jack","D","2.4000029999999988E+04","America:Manhattan","2014"
"6","Linda","D","3.6000040000000009E+04","America:NewYork","2014"
"8","Zhang","R","9.0000499999999927E+03","China:Shanghai","2014"
```

写入HDFS文件

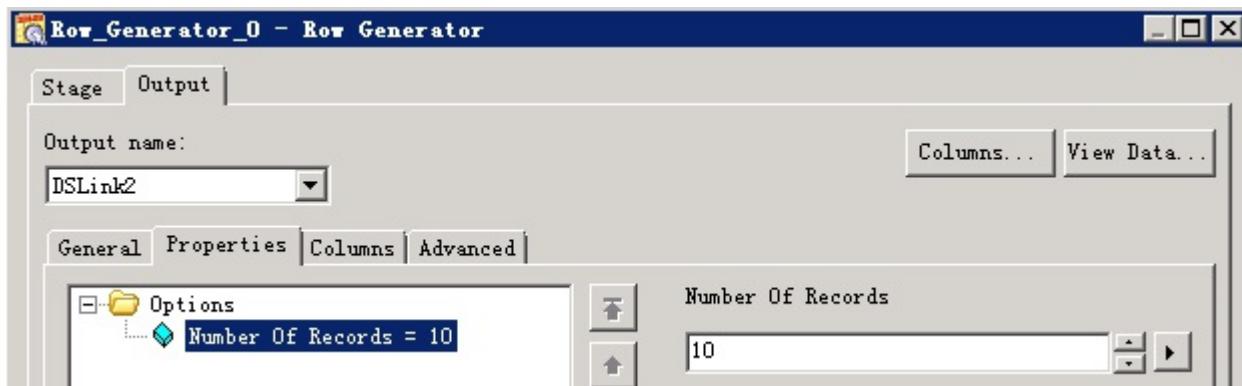
- 创建作业

新建并行作业，保存为hdfswrite

添加Row Generator组件和File Connector组件，以及Row Generator到File Connector链接



- 参考下图修改配置

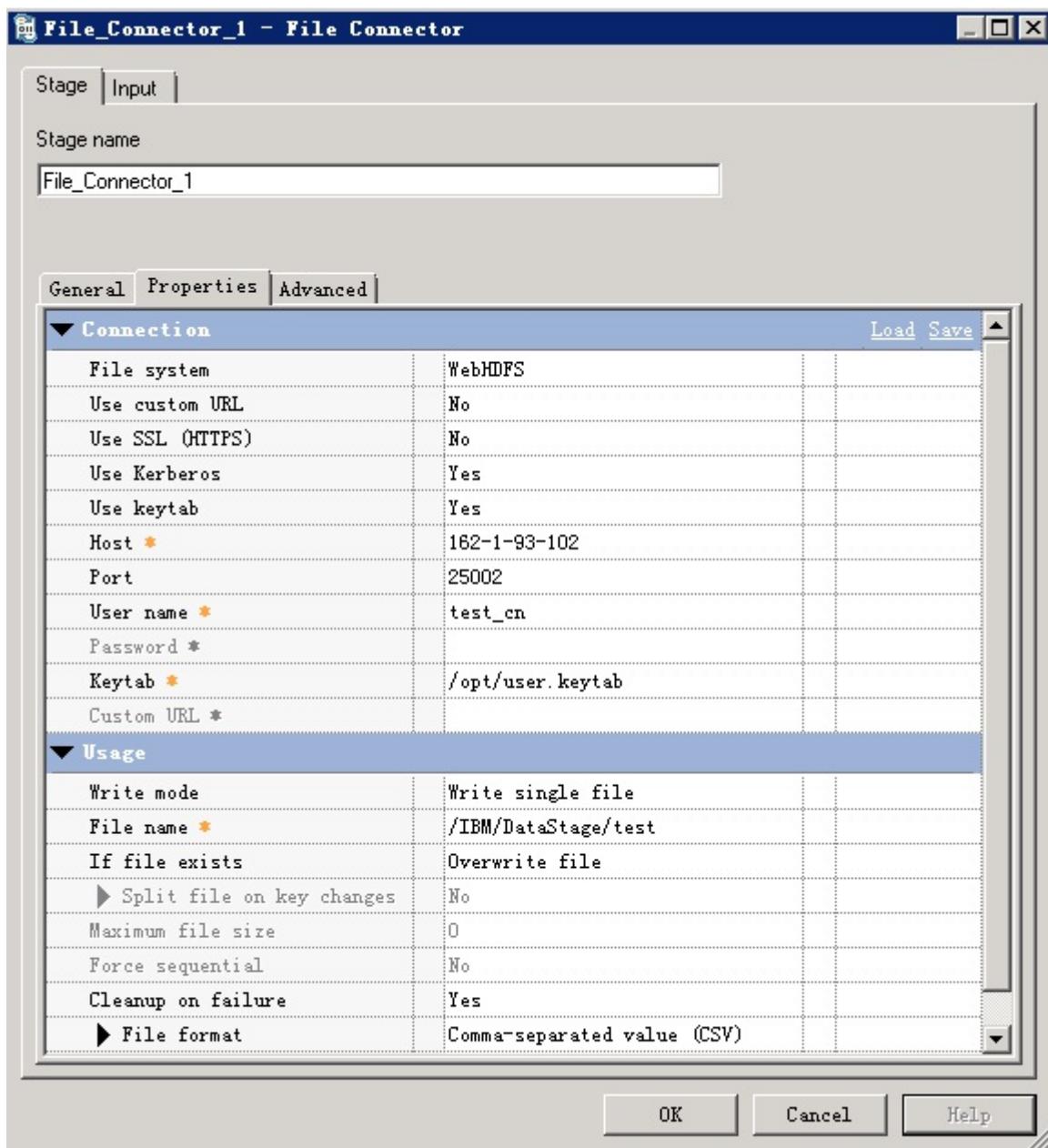


Row_Generator_0 - Row Generator

Stage		Output				
Output name:		<input type="button" value="Columns..."/> <input type="button" value="View Data..."/>				
DSLink2						
<input type="button" value="General"/> <input type="button" value="Properties"/> <input type="button" value="Columns"/> <input type="button" value="Advanced"/>						
1	Column name	Key	SQL type			
1	id	<input type="checkbox"/>	Integer	5	No	
2	name	<input type="checkbox"/>	VarChar	Unicode	20	No
3	usd_flag	<input type="checkbox"/>	VarChar	Unicode	2	No
4	salary	<input type="checkbox"/>	Double		8	No
5	address	<input type="checkbox"/>	VarChar	Unicode	30	No
6	entrytime	<input type="checkbox"/>	VarChar	Unicode	4	No

File_Connector_1 - File Connector

Stage		Input																									
Stage name																											
<input type="text" value="File_Connector_1"/>																											
<input type="button" value="General"/> <input type="button" value="Properties"/> <input type="button" value="Advanced"/>		<input type="button" value="Load"/> <input type="button" value="Save"/>																									
Connection <table border="1"> <tr> <td>File system</td> <td>WebHDFS</td> </tr> <tr> <td>Use custom URL</td> <td>No</td> </tr> <tr> <td>Use SSL (HTTPS)</td> <td>Yes</td> </tr> <tr> <td>Use Kerberos</td> <td>Yes</td> </tr> <tr> <td>Use keytab</td> <td>Yes</td> </tr> <tr> <td>Host *</td> <td>fusioninsight1</td> </tr> <tr> <td>Port</td> <td>25003</td> </tr> <tr> <td>Service principal</td> <td></td> </tr> <tr> <td>User name *</td> <td>test</td> </tr> <tr> <td>Password</td> <td></td> </tr> <tr> <td>Keytab *</td> <td>/home/dsadm/user.keytab</td> </tr> <tr> <td>Custom URL *</td> <td></td> </tr> </table>				File system	WebHDFS	Use custom URL	No	Use SSL (HTTPS)	Yes	Use Kerberos	Yes	Use keytab	Yes	Host *	fusioninsight1	Port	25003	Service principal		User name *	test	Password		Keytab *	/home/dsadm/user.keytab	Custom URL *	
File system	WebHDFS																										
Use custom URL	No																										
Use SSL (HTTPS)	Yes																										
Use Kerberos	Yes																										
Use keytab	Yes																										
Host *	fusioninsight1																										
Port	25003																										
Service principal																											
User name *	test																										
Password																											
Keytab *	/home/dsadm/user.keytab																										
Custom URL *																											
Usage <table border="1"> <tr> <td>File name *</td> <td>/tmp/dstest.csv</td> </tr> <tr> <td>Write mode</td> <td>Write single file</td> </tr> <tr> <td>If file exists</td> <td>Overwrite file</td> </tr> <tr> <td>► Split file on key changes</td> <td>No</td> </tr> <tr> <td>Maximum file size</td> <td>0</td> </tr> <tr> <td>Force sequential</td> <td>No</td> </tr> <tr> <td>Cleanup on failure</td> <td>Yes</td> </tr> <tr> <td>▼ File format</td> <td>Delimited</td> </tr> </table>				File name *	/tmp/dstest.csv	Write mode	Write single file	If file exists	Overwrite file	► Split file on key changes	No	Maximum file size	0	Force sequential	No	Cleanup on failure	Yes	▼ File format	Delimited								
File name *	/tmp/dstest.csv																										
Write mode	Write single file																										
If file exists	Overwrite file																										
► Split file on key changes	No																										
Maximum file size	0																										
Force sequential	No																										
Cleanup on failure	Yes																										
▼ File format	Delimited																										



- 编译运行

保存 — 编译 — 运行，查看作业日志：

```
21:33:32 2017/7/6 Info File_Connector_1,0: Accessing file via WebHDFS file system.  
21:33:33 2017/7/6 Info File_Connector_1,0: IBMJGSSProvider Build-Level: -20160808 (...)  
21:33:33 2017/7/6 Info File_Connector_1,0: The /tmp/dstest.csv file, which contains 10 rows, was written.  
21:33:33 2017/7/6 Info main_program: Step execution finished with status = OK.  
21:33:33 2017/7/6 Info main_program: Startup time, 0:01; production run time, 0:02.  
21:33:33 2017/7/6 Info Parallel job reports successful completion  
21:33:33 2017/7/6 Control Finished Job hdfswrite.
```

- 查看写入数据

对接Hive

使用Hive Connector

说明： Hive Connector官方认证过的Hive JDBC Driver只有DataDirect Hive Driver(ISHive.jar)，用DataStage 11.5.0.2中自带的ISHive.jar连接FusionInsight的hive时，会有thrift protocol报错，需要咨询IBM技术支持提供的最新的ISHive.jar

设置JDBC Driver配置文件

- 在\$DSHOME路径下创建isjdbc.config文件，CLASSPATH变量中添加DataDirect Hive Driver (ISHive.jar)的路径，CLASS_NAMES变量中添加com.ibm.isf.jdbc.hive.HiveDriver，参考命令：

```
su - dsadm  
cd $DSHOME  
vi isjdbc.config
```

在isjdbc.config中添加如下信息：

```
CLASSPATH=/opt/IBM/InformationServer/ASBNode/lib/java/ISHive.jar  
CLASS_NAMES=com.ibm.isf.jdbc.hive.HiveDriver
```

- 配置Kerberos认证信息：

在ISHive.jar所在目录下创建JDBCDriverLogin.conf

```
cd /opt/IBM/InformationServer/ASBNode/lib/java/  
vi JDBCDriverLogin.conf
```

文件内容如下：

```
JDBC_DRIVER_test_cache{  
    com.ibm.security.auth.module.Krb5LoginModule required  
    credsType=initiator  
    principal="test@HADOOP.COM"  
    useCcache="FILE:/tmp/krb5cc_1004";  
};  
JDBC_DRIVER_test_keytab{  
    com.ibm.security.auth.module.Krb5LoginModule required  
    credsType=both  
    principal="test@HADOOP.COM"  
    useKeytab="/home/dsadm/user.keytab";  
};
```

读取Hive数据

- 创建作业

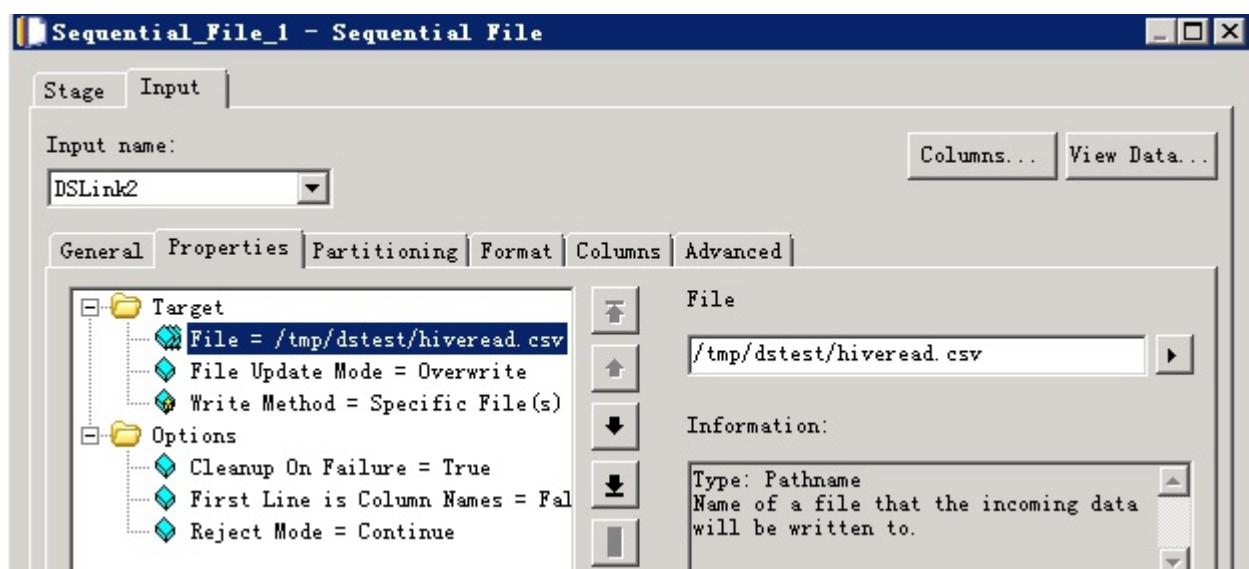
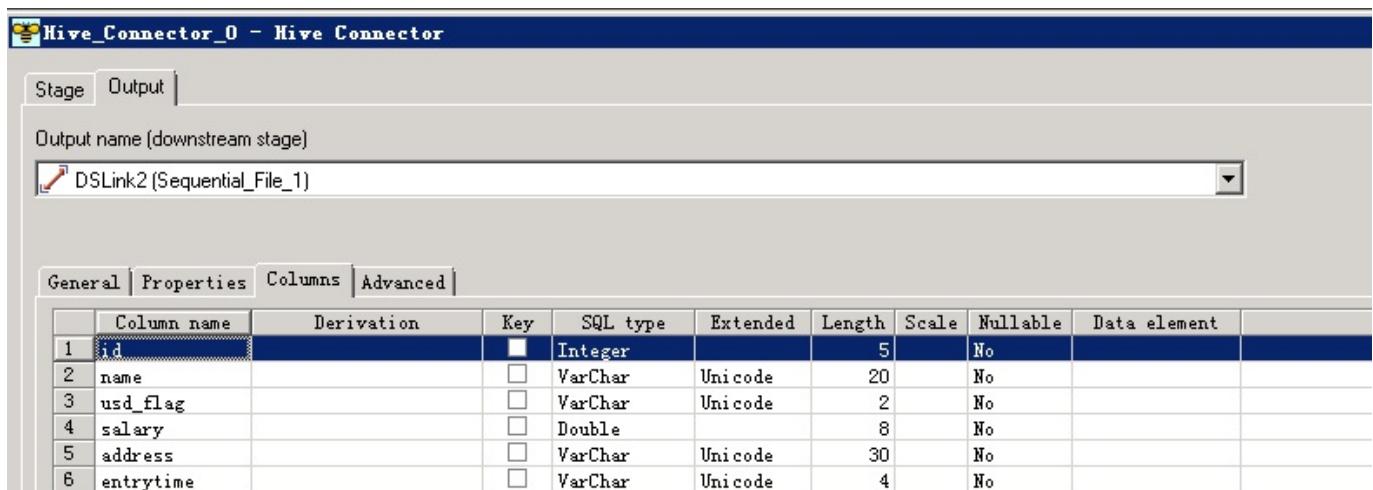
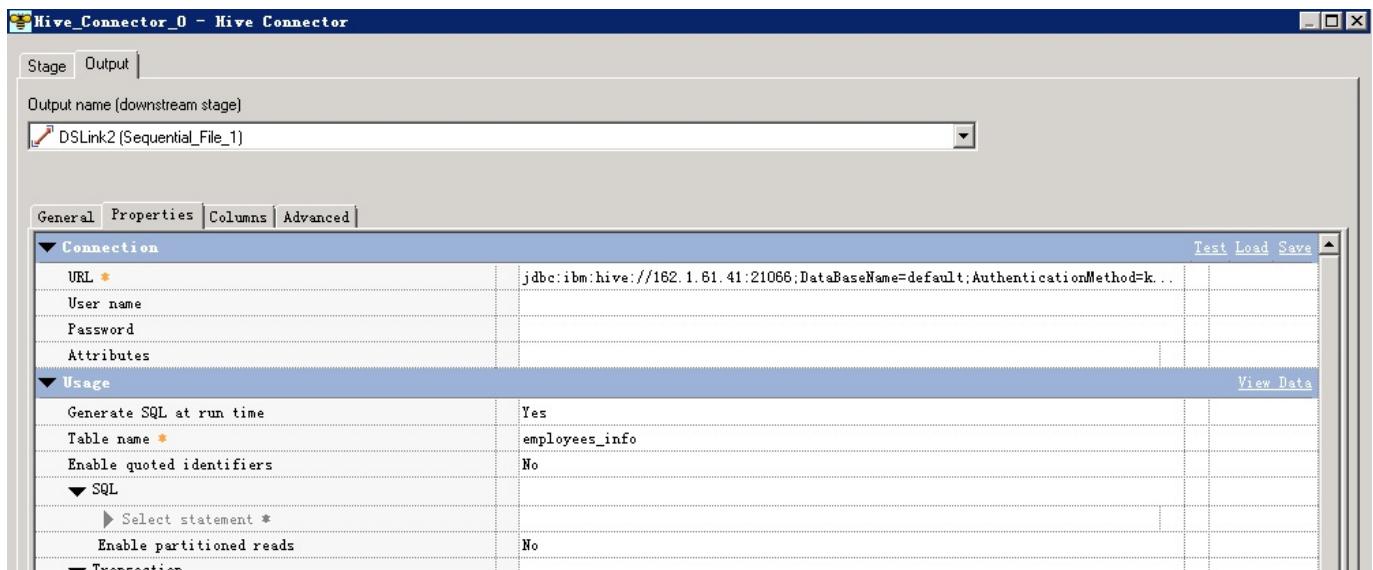


- 修改配置

URL参考如下进行配置：

```
jdbc:ibm:hive://162.1.61.41:21066;DataBaseName=default;AuthenticationMethod=kerberos;ServicePrincipalName=hive/hadoop.hadoop.com@HADOOP.COM;loginConfigName=JDBC_DRIVER_test_keytab;
```

其中JDBC_DRIVER_test_keytab为上一步指定的鉴权信息



- 编译运行

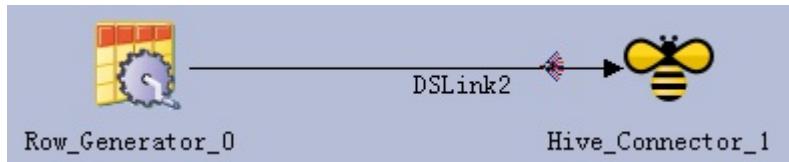
保存 - 编译 - 运行，查看作业日志；

- ## ○ 查看读取的数据

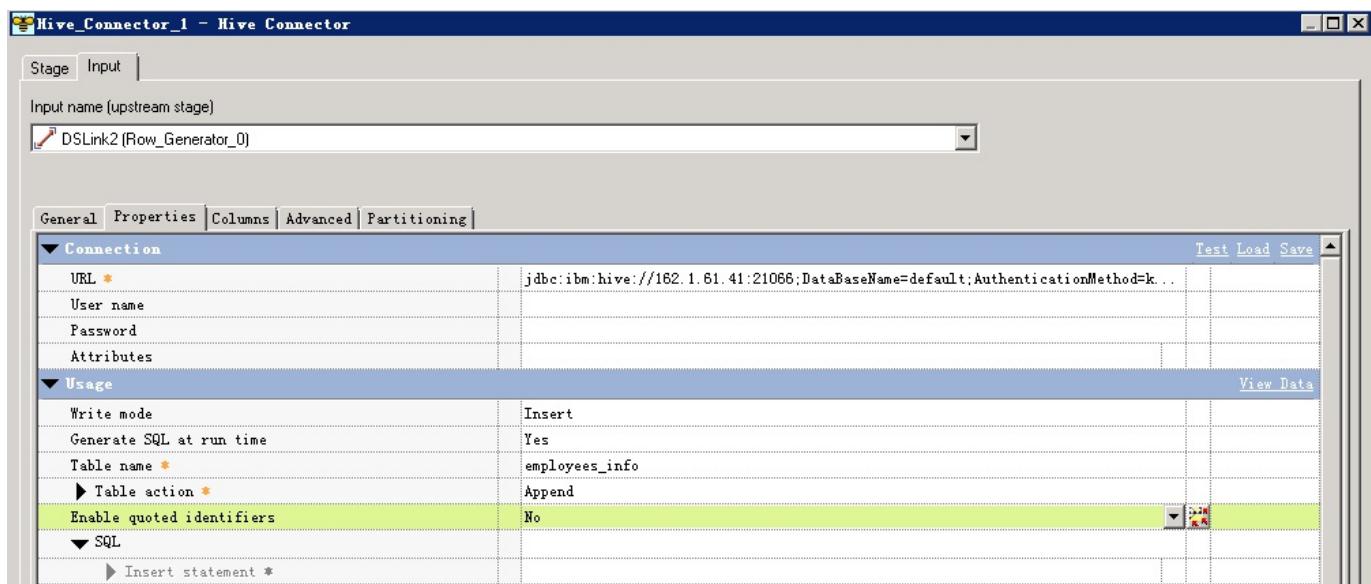
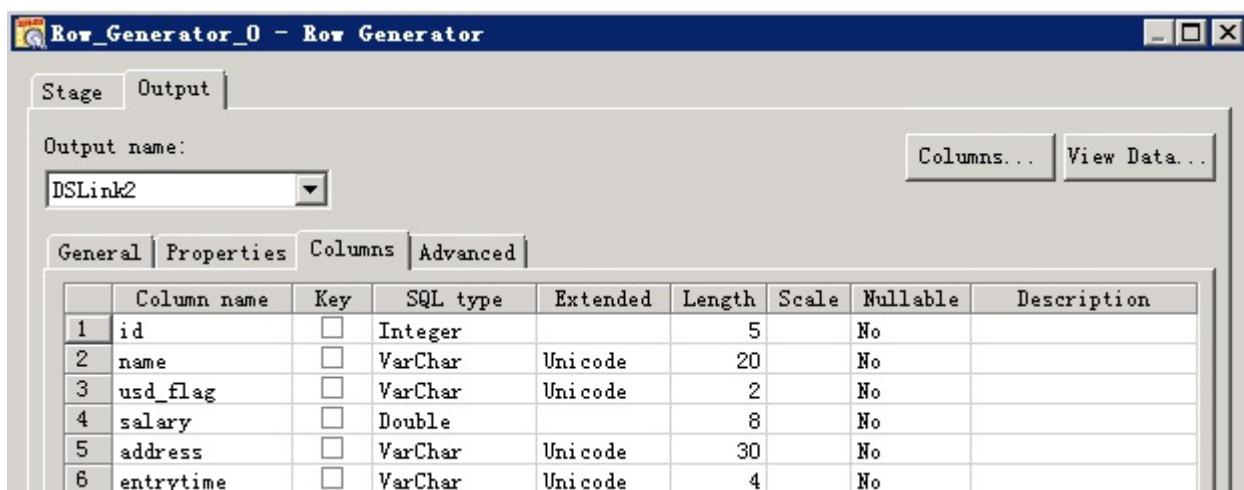
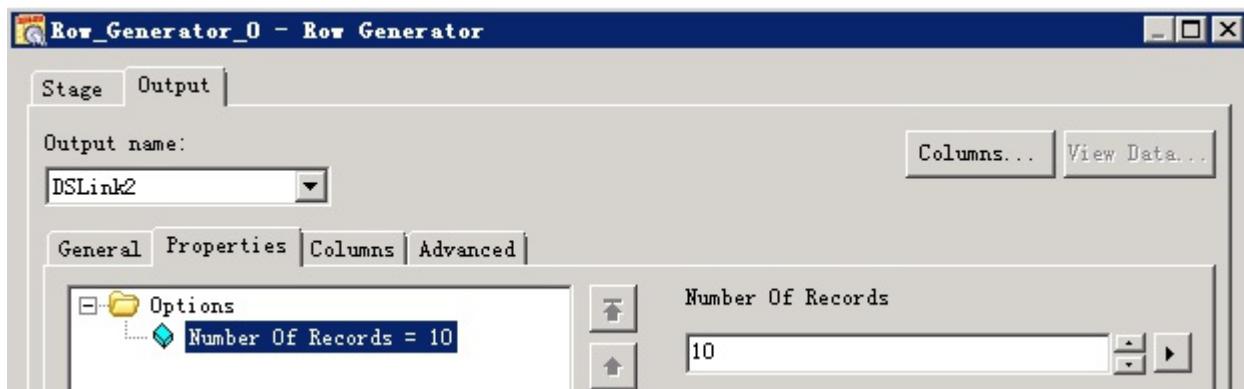
```
[root@datastagell502 dstest]# cat /tmp/dstest/hiveread.csv
"1","Wang","R","8.0000100000000002E+03","China:Shenzhen","2014"
"3","Tom","D","1.2000020000000004E+04","America:NewYork","2014"
"4","Jack","D","2.4000029999999988E+04","America:Manhattan","2014"
"6","Linda","D","3.6000040000000009E+04","America:NewYork","2014"
"8","Zhang","R","9.0000499999999927E+03","China:Shanghai","2014"
```

数据写入Hive表

- 创建作业



- 修改配置



- 编译运行

保存—编译—运行，查看作业日志，写入10条数据，用时2'11"

```
23:51:00 2017/7/6 Info Hive_Connector_1,0: The connector closed the connection to the data source. (...)  
23:51:00 2017/7/6 Info Hive_Connector_1: The connector closed the connection to the data source.  
23:51:00 2017/7/6 Info main_program: Step execution finished with status = OK.  
23:51:00 2017/7/6 Info main_program: Startup time, 0:52; production run time, 2:11.  
23:51:00 2017/7/6 Info Parallel job reports successful completion  
23:51:00 2017/7/6 Control Finished Job hive_write.
```

- 查看Hive表数据:

```
0: jdbc:hive2://162.1.61.42:21066> select * from employees_info;
+-----+-----+-----+-----+-----+-----+-----+
| employees_info.id | employees_info.name | employees_info.usd_flag | employees_info.salary | employees_info.address | employees_info.entrytime |
+-----+-----+-----+-----+-----+-----+
| 1 | 11111 | 11 | 1.0 | 11111111111111111111 | 11111 |
| 3 | 3333333333333333 | 33 | 3.0 | 33333333333333333333 | 33 |
| 5 | 5555555555555555 | 5 | 5.0 | 5555555555555555 | 55 |
| 7 | 77777777 | 77 | 7.0 | 77777777777777777777 | 7 |
| 9 | 999999 | 9 | 9.0 | 99999999999999999999 | 9 |
| 0 | 000000000000000000 | 0.0 | | | |
| 2 | 2 | 22 | 2.0 | 22222222222222 | 2 |
| 4 | 4444444 | 44 | 4.0 | 444444 | 4 |
| 6 | 666666 | 6.0 | 6.0 | 666666 | |
| 8 | 8888888888888888 | 8 | 8.0 | 8888888888888888 | 888 |
| 1 | Wang | R | 8000.01 | China:Shenzhen | 2014 |
| 3 | Tom | D | 12000.02 | America:NewYork | 2014 |
| 4 | Jack | D | 24000.03 | America:Manhattan | 2014 |
| 6 | Linda | D | 36000.04 | America:NewYork | 2014 |
| 8 | Zhang | R | 9000.05 | China:Shanghai | 2014 |
+-----+-----+-----+-----+-----+-----+
15 rows selected (0.488 seconds)
```

Hive Connector向Hive表写数据使用Insert语句，每插入一条数据会起一个MR任务，效率特别低，不推荐使用这种方式。可以将数据直接写入HDFS文件。

使用JDBC Connector

如果要使用FusionInsight的Hive JDBC驱动，用isjdbc.config文件CLASSPATH中添加jdbc驱动和依赖包的方式，在运行作业时会有如下报错，此时需要用导出CLASSPATH环境变量的方式加载

```
JDBC_Connector_0: The connector encountered a Java exception:  
java.sql.SQLException: Create jdbc client failed.  
    at org.apache.hive.jdbc.HiveConnection.<init>(HiveConnection.java:191)  
    at org.apache.hive.jdbc.HiveDriver.connect(HiveDriver.java:105)  
    at com.ibm.is.cc.jdbc.CC_JDBCConnection.openConnection(CC_JDBCConnection.java:1027)  
    at com.ibm.is.cc.jdbc.CC_JDBCConnection.connect(CC_JDBCConnection.java:838)  
Caused by: java.lang.RuntimeException: Illegal Hadoop Version: Unknown (expected A.B.* format)  
    at org.apache.hadoop.hive.shims.ShimLoader.getMajorVersion(ShimLoader.java:167)  
    at org.apache.hadoop.hive.shims.ShimLoader.loadShims(ShimLoader.java:139)  
    at org.apache.hadoop.hive.shims.ShimLoader.getHadoopThriftAuthBridge(ShimLoader.java:125)  
    at org.apache.hive.service.auth.KerberosSaslHelper.getKerberosTransport(KerberosSaslHelper.java:55)  
    at org.apache.hive.jdbc.HiveConnection.createBinaryTransport(HiveConnection.java:507)  
    at org.apache.hive.jdbc.HiveConnection.openTransport(HiveConnection.java:258)  
    at org.apache.hive.jdbc.HiveConnection.createClient(HiveConnection.java:214)  
    at org.apache.hive.jdbc.HiveConnection.access$000(HiveConnection.java:112)  
    at org.apache.hive.jdbc.HiveConnection$1.run(HiveConnection.java:185)  
    at org.apache.hive.jdbc.HiveConnection$1.run(HiveConnection.java:182)  
    at java.security.AccessController.doPrivileged(AccessController.java:488)  
    at javax.security.auth.Subject.doAs(Subject.java:572)  
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1738)  
    at org.apache.hive.jdbc.HiveConnection.<init>(HiveConnection.java:182)  
... 3 more
```

而且只能用JDBC Connector，不能用Hive Connector，否则会有如下报错

```
java.sql.SQLException: Method not supported
    at org.apache.hive.jdbc.HiveConnection.commit(HiveConnection.java:827)
    at com.ibm.is.cc.hive.HiveConnection.commitDirect
(HiveConnection.java:578)
    at com.ibm.is.cc.hive.HiveAdapter.preRun(HiveAdapter.java:776)
```

设置CLASSPATH环境变量

- Hive jdbc驱动包及依赖包位于Hive客户端lib目录下 `/opt/ficlient/Hive/Beeline/lib`，若未安装客户端，也可单独上传这些jar包到任意目录。
 - 设置CLASSPATH环境变量，添加上述jar包的完整路径，参考命令：

```
su - dsadm  
vi $DSHOME/dsenv
```

- 文件最后添加相关的jar包（具体路径根据实际环境调整）

```
export CLASSPATH=/opt/ficlient/Hive/Beeline/lib/commons-cli-1.2.jar:/opt/ficlient/Hive/Beeline/lib/commons-collections-
```

```

3.2.1.jar:/opt/ficlient/Hive/Beeline/lib/commons-configuration-1.6.jar:/opt/ficlient/Hive/Beeline/lib/commons-lang-2.6.jar:/opt/ficlient/Hive/Beeline/lib/commons-logging-1.1.3.jar:/opt/ficlient/Hive/Beeline/lib/curator-client-2.7.1.jar:/opt/ficlient/Hive/Beeline/lib/curator-framework-2.7.1.jar:/opt/ficlient/Hive/Beeline/lib/curator-recipes-2.7.1.jar:/opt/ficlient/Hive/Beeline/lib/guava-14.0.1.jar:/opt/ficlient/Hive/Beeline/lib/hadoop-auth-2.7.2.jar:/opt/ficlient/Hive/Beeline/lib/hadoop-common-2.7.2.jar:/opt/ficlient/Hive/Beeline/lib/hadoop-mapreduce-client-core-2.7.2.jar:/opt/ficlient/Hive/Beeline/lib/hive-beeline-1.3.0.jar:/opt/ficlient/Hive/Beeline/lib/hive-cli-1.3.0.jar:/opt/ficlient/Hive/Beeline/lib/hive-common-1.3.0.jar:/opt/ficlient/Hive/Beeline/lib/hive-exec-1.3.0.jar:/opt/ficlient/Hive/Beeline/lib/hive-jdbc-1.3.0.jar:/opt/ficlient/Hive/Beeline/lib/hive-metastore-1.3.0.jar:/opt/ficlient/Hive/Beeline/lib/hive-serde-1.3.0.jar:/opt/ficlient/Hive/Beeline/lib/hive-service-1.3.0.jar:/opt/ficlient/Hive/Beeline/lib/hive-shims-0.23-1.3.0.jar:/opt/ficlient/Hive/Beeline/lib/hive-shims-common-1.3.0.jar:/opt/ficlient/Hive/Beeline/lib/httpclient-4.5.2.jar:/opt/ficlient/Hive/Beeline/lib/httpcore-4.4.jar:/opt/ficlient/Hive/Beeline/lib/jline-2.12.jar:/opt/ficlient/Hive/Beeline/lib/libfb303-0.9.3.jar:/opt/ficlient/Hive/Beeline/lib/libthrift-0.9.3.jar:/opt/ficlient/Hive/Beeline/lib/log4j-1.2.17.jar:/opt/ficlient/Hive/Beeline/lib/slf4j-api-1.7.5.jar:/opt/ficlient/Hive/Beeline/lib/slf4j-log4j12-1.7.5.jar:/opt/ficlient/Hive/Beeline/lib/super-csv-2.2.0.jar:/opt/ficlient/Hive/Beeline/lib/xercesImpl-2.9.1.jar:/opt/ficlient/Hive/Beeline/lib/zookeeper-3.5.1.jar

```

- 导入环境变量

```
source $DSHOME/dsenv
```

- 重启DSEngine

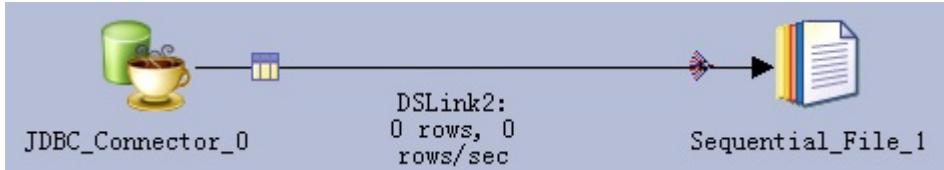
```

cd $DSHOME
bin/uv -admin -stop
bin/uv -admin -start

```

读取Hive数据

- 创建作业



- 修改配置

General	
URL *	jdbc:hive2://162.1.61.41:21066/default;sasl.qop=auth-conf;auth=KERBEROS;principal=hive/hadoop.hadoop.com@HADOOP.COM;user.principal=test@HADOOP.COM;user.keytab=/home/dsadm/user.keytab;
User name	
Password	
Attributes	
Usage	
Generate SQL at run time	Yes
Table name *	employees_info
Enable quoted identifiers	No

其中URL为：

```
jdbc:hive2://162.1.61.41:21066/default;sasl.qop=auth-conf;auth=KERBEROS;principal=hive/hadoop.hadoop.com@HADOOP.COM;user.principal=test@HADOOP.COM;user.keytab=/home/dsadm/user.keytab;
```

JDBC_Connector_0 - JDBC Connector

Stage Output

Output name (downstream stage)

DSLink2 (Sequential_File_1)

General Properties Columns Advanced

	Column name	Derivation	Key	SQL type	Extended	Length	Scale	Nullable	Data element
1	id		<input type="checkbox"/>	Integer		5		No	
2	name		<input type="checkbox"/>	VarChar	Unicode	20		No	
3	usd_flag		<input type="checkbox"/>	VarChar	Unicode	2		No	
4	salary		<input type="checkbox"/>	Double		8		No	
5	address		<input type="checkbox"/>	VarChar	Unicode	30		No	
6	entrytime		<input type="checkbox"/>	VarChar	Unicode	4		No	

Sequential_File_1 - Sequential File

Stage Input

Input name:

DSLink2

Columns... View Data...

General Properties Partitioning Format Columns Advanced

Target

- File = /tmp/dstest/hive_jdbc_res
- File Update Mode = Overwrite
- Write Method = Specific File(s)

File: /tmp/dstest/hive_jdbc_read.csv

Information:

- 编译运行

```

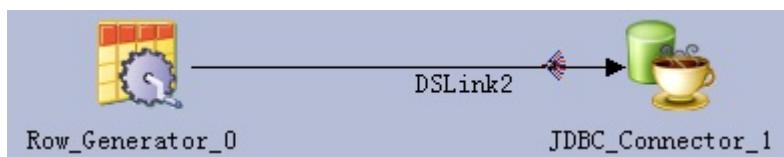
23:03:57 2017/7/7 Info JDBC_Connector_0,0: The connector closed the connection to the data source.
23:03:57 2017/7/7 Info Sequential_File_1,0: Export complete; 5 records exported successfully, 0 rejected. (...)
23:03:57 2017/7/7 Info JDBC_Connector_0: The connector closed the connection to the data source.
23:03:57 2017/7/7 Info main_program: Step execution finished with status = OK.
23:03:57 2017/7/7 Info main_program: Startup time, 0:03; production run time, 0:02.
23:03:57 2017/7/7 Info Parallel job reports successful completion
23:03:57 2017/7/7 Control Finished Job hive_jdbc_connector.

```

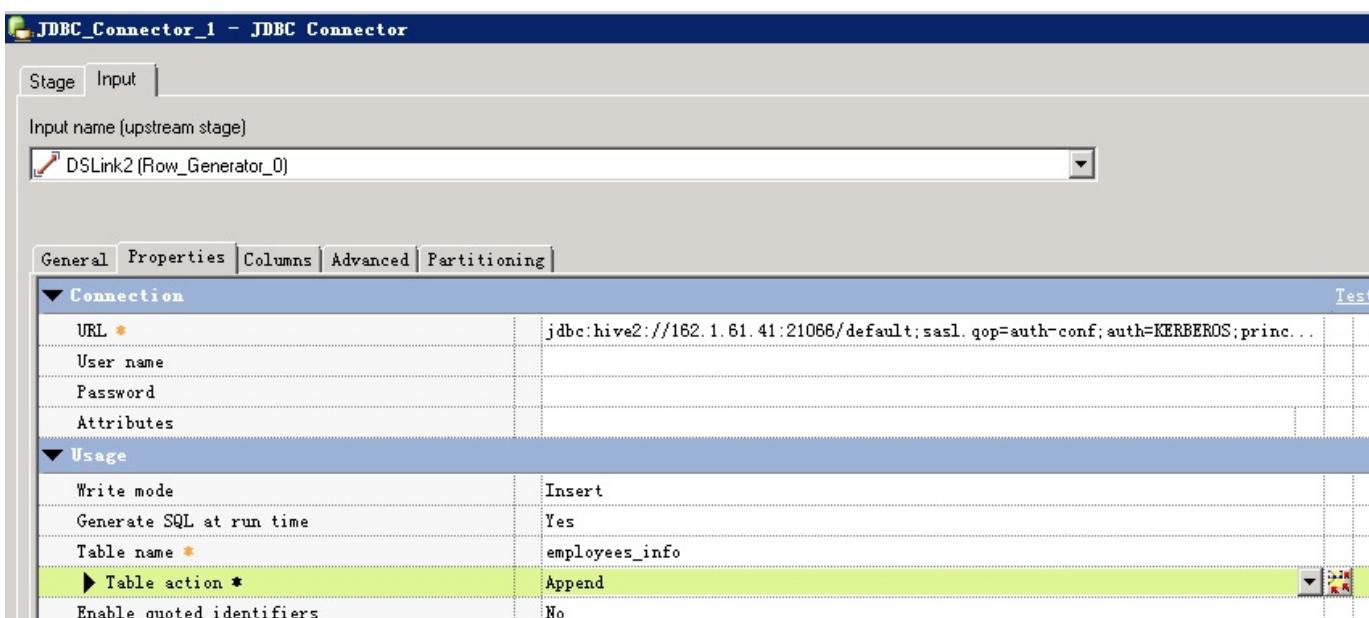
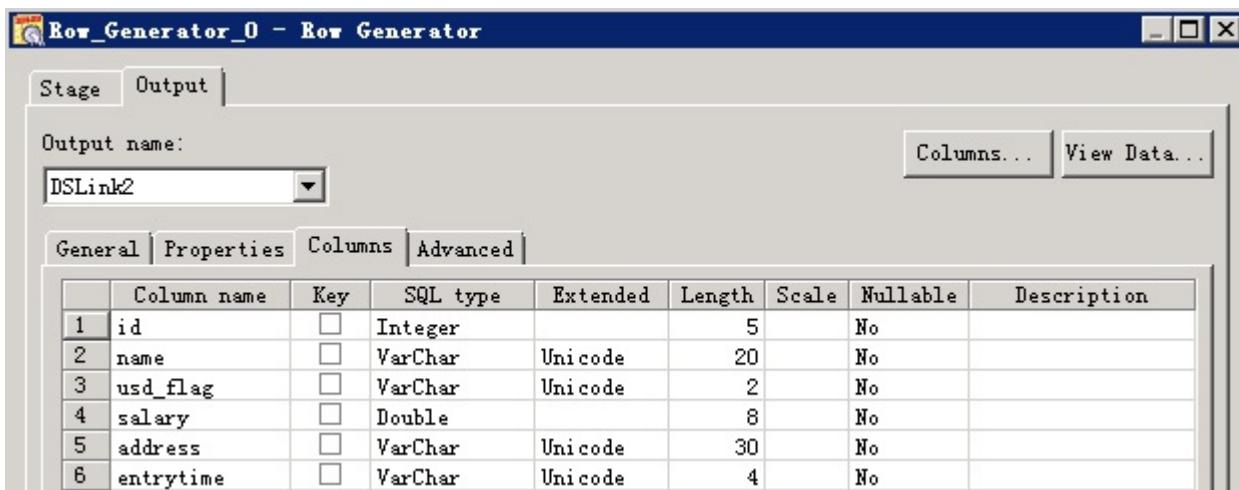
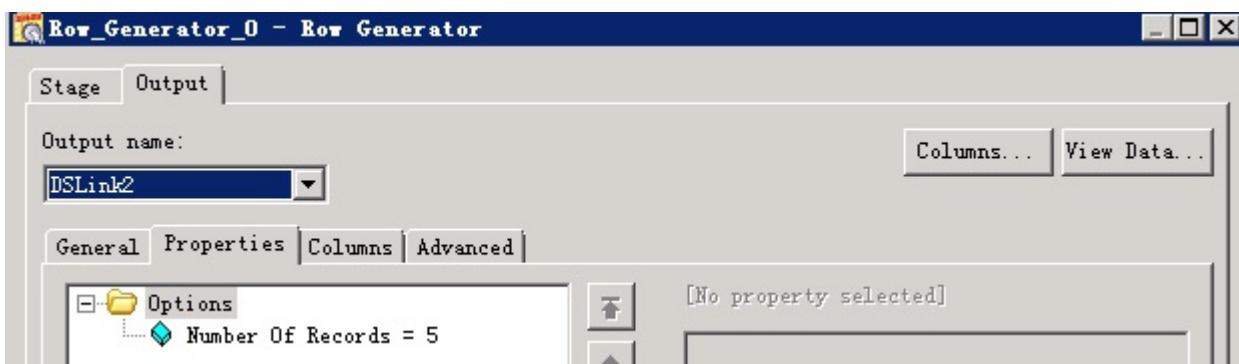
```
[root@datastage11502 dstest]# cat hive_jdbc_read.csv
"1","Linda","D","3.60000400000000009E+04","America:NewYork","2014"
"3","Tom","D","1.20000200000000004E+04","America:NewYork","2014"
"4","Jack","D","2.4000029999999988E+04","America:Manhattan","2014"
"6","Wang","R","8.0000100000000022E+03","China:Shenzhen","2014"
"8","Zhang","R","9.0000499999999927E+03","China:Shanghai","2014"
```

数据写入Hive表

- 创建作业



- 修改配置



- 编译运行

写入5条数据，用时1'49"

```

23:15:59 2017/7/7 Info JDBC_Connector_1: The connector closed the connection to the data source.
23:15:59 2017/7/7 Info main_program: Step execution finished with status = OK.
23:15:59 2017/7/7 Info main_program: Startup time, 0:02; production run time, 1:49.
23:15:59 2017/7/7 Info Parallel job reports successful completion
23:15:59 2017/7/7 Control Finished Job hive_jdbc_write.

```

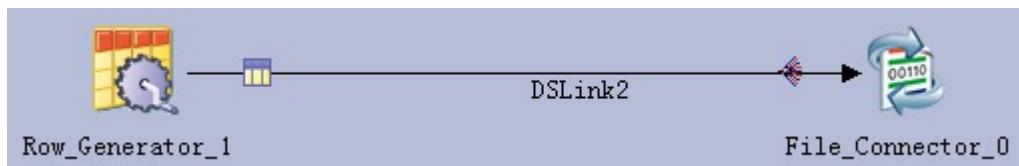
```

0: jdbc:hive2://162.1.61.41:21066/> select * from employees_info;
+-----+-----+-----+-----+-----+-----+-----+
| employees_info.id | employees_info.name | employees_info.usd_flag | employees_info.salary | employees_info.address | employees_info.entrytime |
+-----+-----+-----+-----+-----+-----+
| 1 | 1111 | 11 | 1.0 | 111111111111111111 | 1111 |
| 0 | 000000000000000000 | D | 0.0 | 333333333333333333333333 | 33 |
| 3 | 3333333333333333 | 33 | 3.0 | 2222222222222222 | 2 |
| 2 | 2 | 22 | 2.0 | 444444 | 4 |
| 4 | 4444444 | 44 | 4.0 | 36000.04 | 2014 |
| 1 | Linda | D | 12000.02 | America:NewYork | 2014 |
| 3 | Tom | D | 24000.03 | America:NewYork | 2014 |
| 4 | Jack | D | 8000.01 | America:Manhattan | 2014 |
| 6 | Wang | R | 9000.05 | China:Shenzhen | 2014 |
| 8 | Zhang | R | 9000.05 | China:Shanghai | 2014 |
+-----+-----+-----+-----+-----+-----+
10 rows selected (0.228 seconds)

```

数据导入Hive表的HDFS文件

- 创建作业



- 修改配置

Row_Generator_1 - Row Generator

Stage Output

Output name: DSLink2 Columns... View Data...

General Properties Columns Advanced

Number Of Records

Number Of Records = 100

File_Connector_0 - File Connector

Stage Input

Stage name: File_Connector_0

General Properties Advanced

Connection

File system	WebHDFS
Use custom URL	No
Use SSL (HTTPS)	Yes
Use Kerberos	Yes
Use keytab	Yes
Host *	fusioninsight1
Port	25003
Service principal	
User name *	test
Password	
Keytab *	/home/dsadm/user.keytab
Custom URL *	

Usage

File name *	/user/hive/warehouse/employees_info/data_ds.txt
Write mode	Write single file
If file exists	Overwrite file
► Split file on key changes	No
Maximum file size	0
Force sequential	No
Cleanup on failure	Yes
► File format	Delimited
► Avro format properties	

- 编译运行

```
0:16:07 2017/7/7 Info File_Connector_0:0: Accessing file via WebHDFS file system.
0:16:08 2017/7/7 Info File_Connector_0:0: IBMJGSSProvider Build-Level: -20160808 (...)
0:16:08 2017/7/7 Info File_Connector_0:0: The /user/hive/warehouse/employees_info/data_ds.txt file, which contains 100 rows, was written.
0:16:08 2017/7/7 Info main_program: Step execution finished with status = OK.
0:16:08 2017/7/7 Info main_program: Startup time, 0:03; production run time, 0:02.
0:16:08 2017/7/7 Info Parallel job reports successful completion
0:16:08 2017/7/7 Control Finished Job hive_write_hdfs.
```

- 查看写入数据

```
[root@datastage11502 ~]# hdfs dfs -ls /user/hive/warehouse/employees_info
Found 4 items
-rwxrwx---+ 3 test hive      223 2017-07-06 23:50 /user/hive/warehouse/employees_info/000000_0
-rwxrwx---+ 3 test hive      146 2017-07-06 23:51 /user/hive/warehouse/employees_info/000000_0_copy_1499356264319
-rwxrwx---+ 3 test hive      194 2017-06-20 11:42 /user/hive/warehouse/employees_info/data.txt
-rw-r-x---+ 3 test hive     2934 2017-07-07 00:16 /user/hive/warehouse/employees_info/data_ds.txt
[root@datastage11502 ~]# hdfs dfs -cat /user/hive/warehouse/employees_info/data_ds.txt
0,aa,a,0,aaaaaaaaaaaaaaaaaaaa,
1,b,b,1.0,bbbbbb,bbbb
2,cccc,c,2.0,c,c
3,ddddddddd,d,3.0,ddddddddd,dd
4,eeee,e,4.0,eeeeeee,e
5,ffff,f,5.0,ffffffffffff,ff
6,,g,6.0,gggggg,
7,hhhhhhhh,h,7.0,hhhhhhhh,h
8,iiiiii,i,8.0,iiiiiiiiii,i,iii
9,jjjjjjjj,j,9.0,jjjjjjj,j
10,kkkkkkkkkk,k,10.0,kkkkkkkkkkkkkkkkkk,kkkk
11,,l,11.0,,ll
12,mmmm,m,12.0,mmmmmmmmmmmmmmmmmm,mmm
13,nnnnnnnn,n,13.0,,
14,0.00.0.14.0,0.00000000.000
```

hive表数据增量100

```

0: jdbc:hive2://162.1.61.42:21066> select * from employees_info limit 10;
+-----+-----+-----+-----+-----+-----+
| employees_info.id | employees_info.name | employees_info.usd_flag | employees_info.salary | employees_info.address | employees_info.entrytime |
+-----+-----+-----+-----+-----+
| 1                | 11111             | 11                  | 1.0                 | 11111111111111111111 | 11111             |
| 3                | 3333333333333333 | 33                  | 3.0                 | 33333333333333333333 | 33                |
| 5                | 5555555555555555 | 5                   | 5.0                 | 5555555555555555      | 55                |
| 7                | 7777777777777777 | 77                  | 7.0                 | 77777777777777777777 | 7                  |
| 9                | 999999             | 9                   | 9.0                 | 99999999999999999999 | 9                  |
| 0                | 000000000000000000 |                    | 0.0                 | 00000000000000000000 | 0.0                |
| 2                | 2                  | 22                  | 2.0                 | 2222222222222222      | 2                  |
| 4                | 4444444           | 44                  | 4.0                 | 4444444               | 4                  |
| 6                | 666666             |                     | 6.0                 | 666666                | 666666             |
| 8                | 8888888888888888 | 8                   | 8.0                 | 8888888888888888      | 888               |
+-----+-----+-----+-----+-----+
10 rows selected (0.34 seconds)
0: jdbc:hive2://162.1.61.42:21066> select count(*) from employees_info;
INFO : Number of reduce tasks determined at compile time: 1
INFO : In order to change the average load for a reducer (in bytes):
INFO :   set hive.exec.reducers.bytes.per.reducer=<number>
INFO : In order to limit the maximum number of reducers:
INFO :   set hive.exec.reducers.max=<number>
INFO : In order to set a constant number of reducers:
INFO :   set mapreduce.job.reduces=<number>
INFO : number of splits:1
INFO : Submitting tokens for job: job_1498704716735_0035
INFO : Kind: HDFS_DELEGATION_TOKEN, Service: ha-hdfs:hacluster
INFO : Kind: HIVE_DELEGATION_TOKEN, Service: HiveServer2ImpersonationToken
INFO : The url to track the job: https://fusioninsight1:26001/proxy/application_1498704716735_0035/
INFO : Starting Job = job_1498704716735_0035, Tracking URL = https://fusioninsight1:26001/proxy/application_1498704716735_0035/
INFO : Kill Command = /opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Hive-1.3.0/hive-1.3.0/bin/..../hadoop/bin/hadoop job -kill job_1498704716735_0035
INFO : Hadoop command information for Stage-1: number of mappers: 1; number of reducers: 1
INFO : 2017-07-07 00:22:41,092 Stage-1 map = 0%,  reduce = 0%
INFO : 2017-07-07 00:22:48,590 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.93 sec
INFO : 2017-07-07 00:22:54,009 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.93 sec
INFO : MapReduce Total cumulative CPU time: 3 seconds 930 msec
INFO : Ended Job = job_1498704716735_0035
+-----+
| _c0 |
+-----+
| 115 |
+-----+
1 row selected (23.059 seconds)

```

增量数据定期自动导入Hive表的HDFS文件

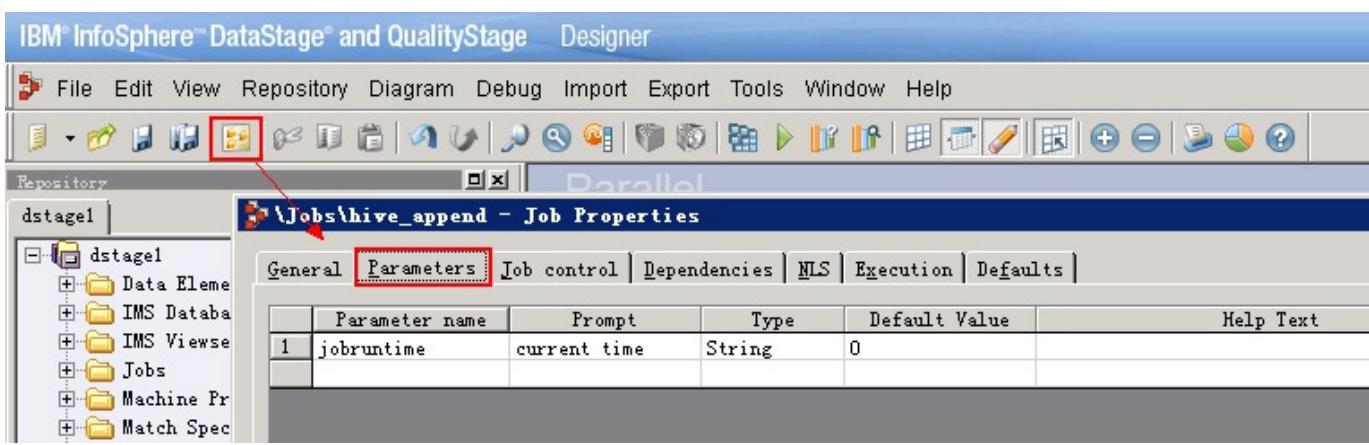
增量数据可以新增HDFS文件的方式导入hive，如果要定期自动化执行，导入的文件名中需要包含可变参数进行设置和区分，然后以命令或脚本方式运行作业，给该参数赋值。

- ## ● 创建作业

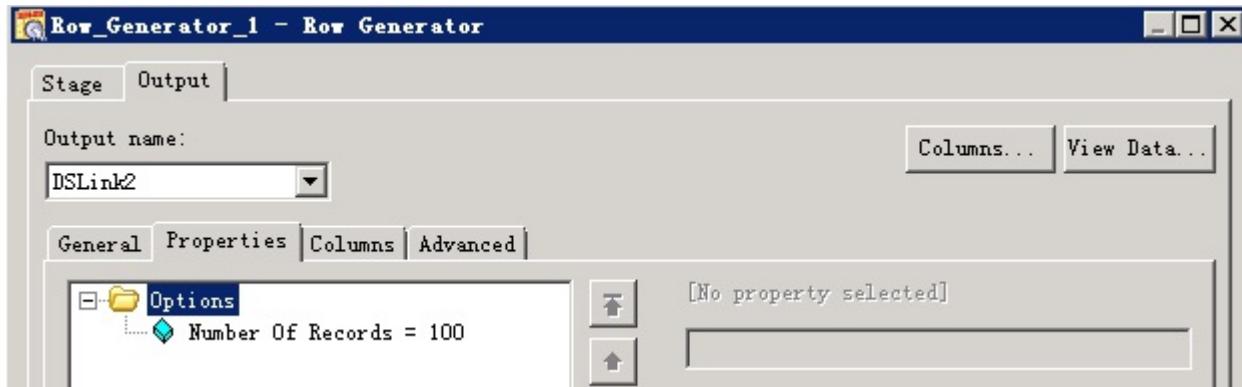


- 设置作业参数

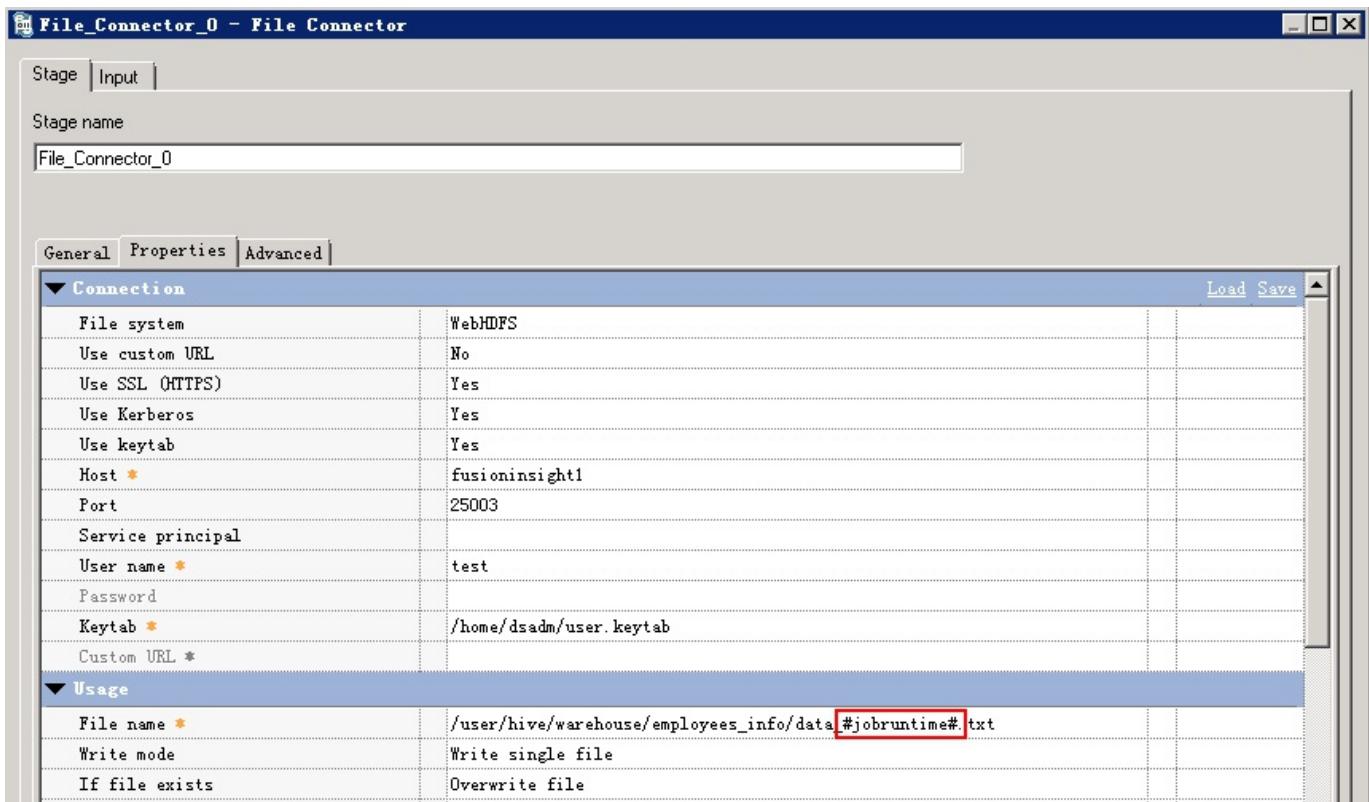
点击“job properties”按钮，设置参数如下



- 修改配置



File Connector配置导出文件的名称，以“#”引用设置的参数



- dsjob命令运行作业

保存编译作业，在DataStage Server上执行dsjob -run命令，格式为：

```
dsjob -run [-mode <NORMAL | RESET | VALIDATE>] -param = -jobstatus PROJECT_NAME JOB_NAME
```

命令参考：

```
su - dsadm
cd $DSHOME/bin
```

```
./dsjob -run -param jobruntime=`date +'%Y-%m-%d-%H-%M-%S'` -jobstatus dstage1 hive_append
```

```
[dsadm@datastage11502 bin]$ ./dsjob -run -param jobruntime=`date +'%Y-%m-%d-%H-%M-%S'` -jobstatus dstage1 hive_append
Waiting for job...
Finished waiting for job
Job Status      : (1)

Status code = 1
[dsadm@datastage11502 bin]$ ./dsjob -run -param jobruntime=`date +'%Y-%m-%d-%H-%M-%S'` -jobstatus dstage1 hive_append
Waiting for job...
Finished waiting for job
Job Status      : (1)

Status code = 1
[dsadm@datastage11502 bin]$
```

- 查看HDFS文件:

```
[root@datastage11502 ~]# hdfs dfs -ls /user/hive/warehouse/employees_info
Found 6 items
-rwxrwx---+ 3 test hive      223 2017-07-06 23:50 /user/hive/warehouse/employees_info/000000_0
-rwxrwx---+ 3 test hive      146 2017-07-06 23:51 /user/hive/warehouse/employees_info/000000_0_copy_1499356264319
-rwxrwx---+ 3 test hive      194 2017-06-20 11:42 /user/hive/warehouse/employees_info/data.txt
-rw-r-x---+ 3 test hive    2934 2017-07-07 00:50 /user/hive/warehouse/employees_info/data_2017-07-07-00-50-03.txt
-rw-r-x---+ 3 test hive    2934 2017-07-07 00:51 /user/hive/warehouse/employees_info/data_2017-07-07-00-51-32.txt
-rw-r-x---+ 3 test hive    2934 2017-07-07 00:16 /user/hive/warehouse/employees_info/data_ds.txt
[root@datastage11502 ~]#
```

- 查看Hive数据增量为200条

```
0: jdbc:hive2://162.1.61.42:21066> select count(*) from employees_info;
INFO  : Number of reduce tasks determined at compile time: 1
INFO  : In order to change the average load for a reducer (in bytes):
INFO  :   set hive.exec.reducers.bytes.per.reducer=<number>
INFO  : In order to limit the maximum number of reducers:
INFO  :   set hive.exec.reducers.max=<number>
INFO  : In order to set a constant number of reducers:
INFO  :   set mapreduce.job.reduces=<number>
INFO  : number of splits:1
INFO  : Submitting tokens for job: job_1498704716735_0036
INFO  : Kind: HDFS_DELEGATION_TOKEN, Service: ha-hdfs:hacluster
INFO  : Kind: HIVE_DELEGATION_TOKEN, Service: HiveServer2ImpersonationToken
INFO  : The url to track the job: https://FusionInsight1:26001/proxy/application_1498704716735_0036/
INFO  : Starting Job = job_1498704716735_0036, Tracking URL = https://FusionInsight1:26001/proxy/application_1498704716735_0036/
INFO  : Kill Command = /opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Hive-1.3.0/hive-1.3.0/bin/../../hadoop/bin/
INFO  : Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
INFO  : 2017-07-07 00:54:12,877 Stage-1 map = 0%, reduce = 0%
INFO  : 2017-07-07 00:54:20,343 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.15 sec
INFO  : 2017-07-07 00:54:26,752 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.22 sec
INFO  : MapReduce Total cumulative CPU time: 4 seconds 220 msec
INFO  : Ended Job = job_1498704716735_0036
+-----+
| _c0 |
+-----+
| 315 |
+-----+
1 row selected (23.953 seconds)
```

对接SparkSQL

与使用FI Hive JDBC驱动类似，可以用SparkSQL JDBC驱动连接Hive，同样需要导出CLASSPATH环境变量来加载驱动包及依赖包。

SparkSQL jdbc不支持insert into语句，只能用来读hive数据，不能插入数据到hive表。

设置CLASSPATH环境变量

- SparkSQL jdbc驱动包及依赖包位于Spark客户端lib目录下 [/opt/ficlient/Spark/spark/lib/](#)，若未安装客户端，也可单独上传所需jar包到任意目录。
- 设置CLASSPATH环境变量，添加上述jar包的完整路径，以及spark客户端配置文件路径（SparkSQL jdbc连接hive时需要读取hive-site.xml中的配置）：

```
su - dsadm
vi $DSHOME/dsenv
```

配置如下内容：

```
export CLASSPATH= /opt/ficlient/Spark/spark/lib/commons-collections-3.2.2.jar:/opt/ficlient/Spark/spark/lib/commons-configuration-1.6.jar:/opt/ficlient/Spark/spark/lib/commons-lang-2.6.jar:/opt/ficlient/Spark/spark/lib/commons-logging-1.1.3.jar:/opt/ficlient/Spark/spark/lib/curator-client-2.7.1.jar:/opt/ficlient/Spark/spark/lib/curator-framework-2.7.1.jar:/opt/ficlient/Spark/spark/lib/guava-12.0.1.jar:/opt/ficlient/Spark/spark/lib/hadoop-auth-
```

```
2.7.2.jar:/opt/ficlient/Spark/spark/lib/hadoop-common-2.7.2.jar:/opt/ficlient/Spark/spark/lib/hadoop-mapreduce-client-core-2.7.2.jar:/opt/ficlient/Spark/spark/lib/hive-common-1.2.1.spark.jar:/opt/ficlient/Spark/spark/lib/hive-exec-1.2.1.spark.jar:/opt/ficlient/Spark/spark/lib/hive-jdbc-1.2.1.spark.jar:/opt/ficlient/Spark/spark/lib/hive-metastore-1.2.1.spark.jar:/opt/ficlient/Spark/spark/lib/hive-service-1.2.1.spark.jar:/opt/ficlient/Spark/spark/lib/htrace-core-3.1.0-incubating.jar:/opt/ficlient/Spark/spark/lib/httpclient-4.5.2.jar:/opt/ficlient/Spark/spark/lib/httpcore-4.4.4.jar:/opt/ficlient/Spark/spark/lib/libthrift-0.9.3.jar:/opt/ficlient/Spark/spark/lib/log4j-1.2.17.jar:/opt/ficlient/Spark/spark/lib/slf4j-api-1.7.10.jar:/opt/ficlient/Spark/spark/lib/slf4j-log4j12-1.7.10.jar:/opt/ficlient/Spark/spark/lib/xercesImpl-2.9.1.jar:/opt/ficlient/Spark/spark/lib/zookeeper-3.5.1.jar:/opt/ficlient/Spark/spark/conf
```

- 导入环境变量

```
source $DSHOME/dsenv
```

- 重启DSEngine

```
cd $DSHOME  
bin/uv -admin -stop  
bin/uv -admin -start
```

读取Hive表数据

- 创建作业



- 修改配置

JDBC_Connector_0 - JDBC Connector

Stage Output

Output name (downstream stage)
DSLink5 (Sequential_File_7)

General Properties Columns Advanced

Connection

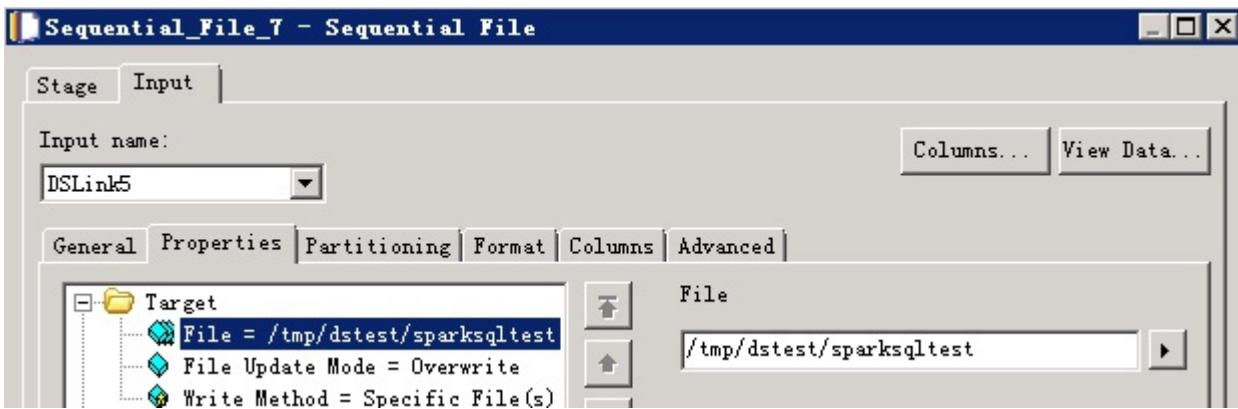
URL *	jdbc:hive2://ha-cluster/default;user.principal=spark/hadoop.hadoop.com@HADOOP...
User name	
Password	
Attributes	

Usage

Generate SQL at run time	Yes
Table name *	employees_info
Enable quoted identifiers	No

URL参考:

```
jdbc:hive2://ha-cluster/default;user.principal=spark/hadoop.hadoop.com@HADOOP.COM;saslQop=auth-conf;auth=KERBEROS;principal=spark/hadoop.hadoop.com@HADOOP.COM;user.principal=test@HADOOP.COM;user.keytab=/home/dsadm/user.keytab;
```



- 编译运行

```
23:51:44 2017/7/7 Info JDBC_Connector_0,0: Number of rows fetched on the current node: 10.  
23:51:44 2017/7/7 Info JDBC_Connector_0,0: [JGSS_DBG_WRAP] main Krb5Context.wrap: set Flag Fields 0 SentByAccept  
23:51:44 2017/7/7 Info JDBC_Connector_0,0: The connector closed the connection to the data source.  
23:51:44 2017/7/7 Info Sequential_File_7,0: Export complete; 10 records exported successfully, 0 rejected. (...)  
23:51:44 2017/7/7 Info JDBC_Connector_0: The connector closed the connection to the data source.  
23:51:44 2017/7/7 Info main_program: Step execution finished with status = OK.  
23:51:44 2017/7/7 Info main_program: Startup time, 0:13; production run time, 0:04.  
23:51:44 2017/7/7 Info Parallel job reports successful completion  
23:51:44 2017/7/7 Control Finished Job sparkidbtest1.
```

对接Phoenix

使用Phoenix以JDBC方式访问HBase表，也需要导出CLASSPATH环境变量来加载驱动包及依赖包。

设置CLASSPATH环境变量

- Phoenix相关的jar包位于HBase客户端lib目录下 `/opt/ficlient/HBase/hbase/lib`，若未安装客户端，也可单独上传所需jar包到任意目录。
 - 设置CLASSPATH环境变量，添加上述jar包的完整路径，以及HBase客户端配置文件路径（Phoenix连接时需要读取hbase-site.xml中的配置）：

```
su - dsadm  
vi $DSHOME/dsenv
```

配置如下内容：

```
export CLASSPATH=
/opt/ficlient/HBase/hbase/lib/commons-cli-1.2.jar:/opt/ficlient/HBase/hbase/lib/commons-codec-1.9.jar:/opt/ficlient/HBase/hbase/lib/commons-collections-3.2.2.jar:/opt/ficlient/HBase/hbase/lib/commons-configuration-1.6.jar:/opt/ficlient/HBase/hbase/lib/commons-io-2.4.jar:/opt/ficlient/HBase/hbase/lib/commons-lang-2.6.jar:/opt/ficlient/HBase/hbase/lib/commons-logging-1.2.jar:/opt/ficlient/HBase/hbase/lib/dynalogger-V100R002C30.jar:/opt/ficlient/HBase/hbase/lib/gson-2.2.4.jar:/opt/ficlient/HBase/hbase/lib/guava-12.0.1.jar:/opt/ficlient/HBase/hbase/lib/hadoop-auth-2.7.2.jar:/opt/ficlient/HBase/hbase/lib/hadoop-common-2.7.2.jar:/opt/ficlient/HBase/hbase/lib/hadoop-hdfs-2.7.2.jar:/opt/ficlient/HBase/hbase/lib/hadoop-hdfs-client-2.7.2.jar:/opt/ficlient/HBase/hbase/lib/hbase-client-1.0.2.jar:/opt/ficlient/HBase/hbase/lib/hbase-common-1.0.2.jar:/opt/ficlient/HBase/hbase/lib/hbaseFilestream-1.0.jar:/opt/ficlient/HBase/hbase/lib/hbase-protocol-1.0.2.jar:/opt/ficlient/HBase/hbase/lib/hbase-secondaryindex-1.0.2.jar:/opt/ficlient/HBase/hbase/lib/hbase-server-1.0.2.jar:/opt/ficlient/HBase/hbase/lib/htrace-core-3.1.0-incubating.jar:/opt/ficlient/HBase/hbase/lib/httpclient-4.5.2.jar:/opt/ficlient/HBase/hbase/lib/httpcore-4.4.4.jar:/opt/ficlient/HBase/hbase/lib/httpmime-4.3.6.jar:/opt/ficlient/HBase/hbase/lib/jackson-core-asl-1.9.13.jar:/opt/ficlient/HBase/hbase/lib/jackson-mapper-asl-1.9.13.jar:/opt/ficlient/HBase/hbase/lib/log4j-1.2.17.jar:/opt/ficlient/HBase/hbase/lib/luna-0.1.jar:/opt/ficlient/HBase/hbase/lib/netty-3.2.4.Final.jar:/opt/ficlient/HBase/hbase/lib/netty-all-4.0.23.Final.jar:/opt/ficlient/HBase/hbase/lib/noggit-0.6.jar:/opt/ficlient/HBase/hbase/lib/phoenix-core-4.4.0-HBase-1.0.jar:/opt/ficlient/HBase/hbase/lib/protobuf-java-2.5.0.jar:/opt/ficlient/HBase/hbase/lib/slf4j-api-1.7.7.jar:/opt/ficlient/HBase/hbase/lib/slf4j-log4j12-1.7.7.jar:/opt/ficlient/HBase/hbase/lib/solr-solrj-5.3.1.jar:/opt/ficlient/HBase/hbase/lib/zookeeper-3.5.1.jar:/opt/ficlient/HBase/hbase/conf
```

- 导入环境变量

```
source $DSHOME/dsenv
```

- 重启DSEngine

```
cd $DSHOME
bin/uv -admin -stop
bin/uv -admin -start
```

创建jaas配置文件

- Phoenix连接需要查询zookeeper，zookeeper的Kerberos认证需要指定jaas配置文件

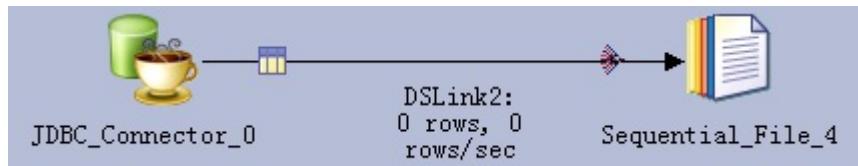
```
su - admin
vi /home/dsadm/jaas.conf
```

文件内容如下：

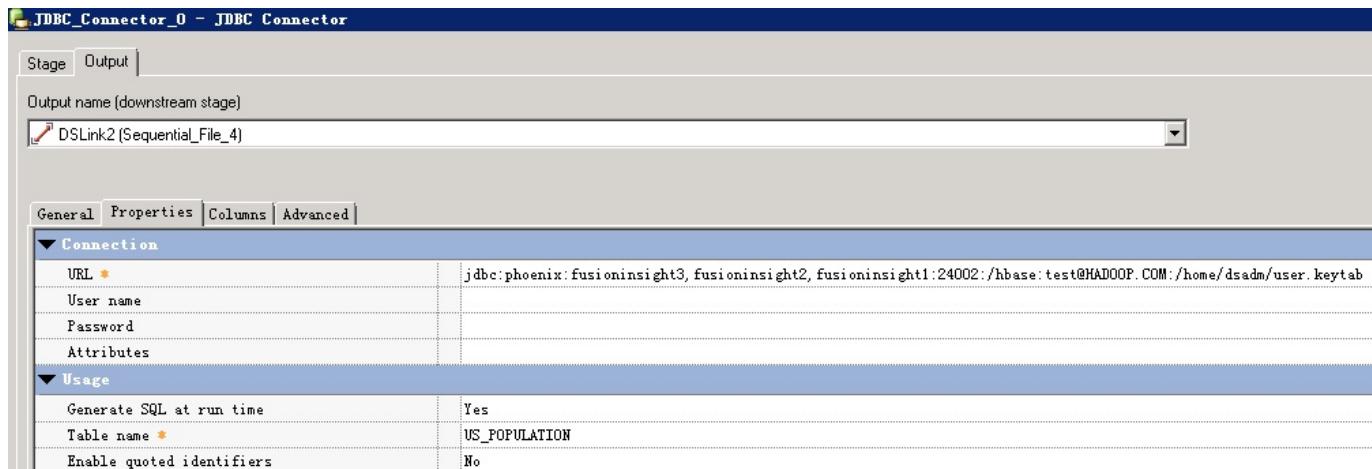
```
Client {
    com.ibm.security.auth.module.Krb5LoginModule required
    credsType=both
    principal="test@HADOOP.COM"
    useKeytab="/home/dsadm/user.keytab";
};
```

读取Phoenix表数据

- 创建作业

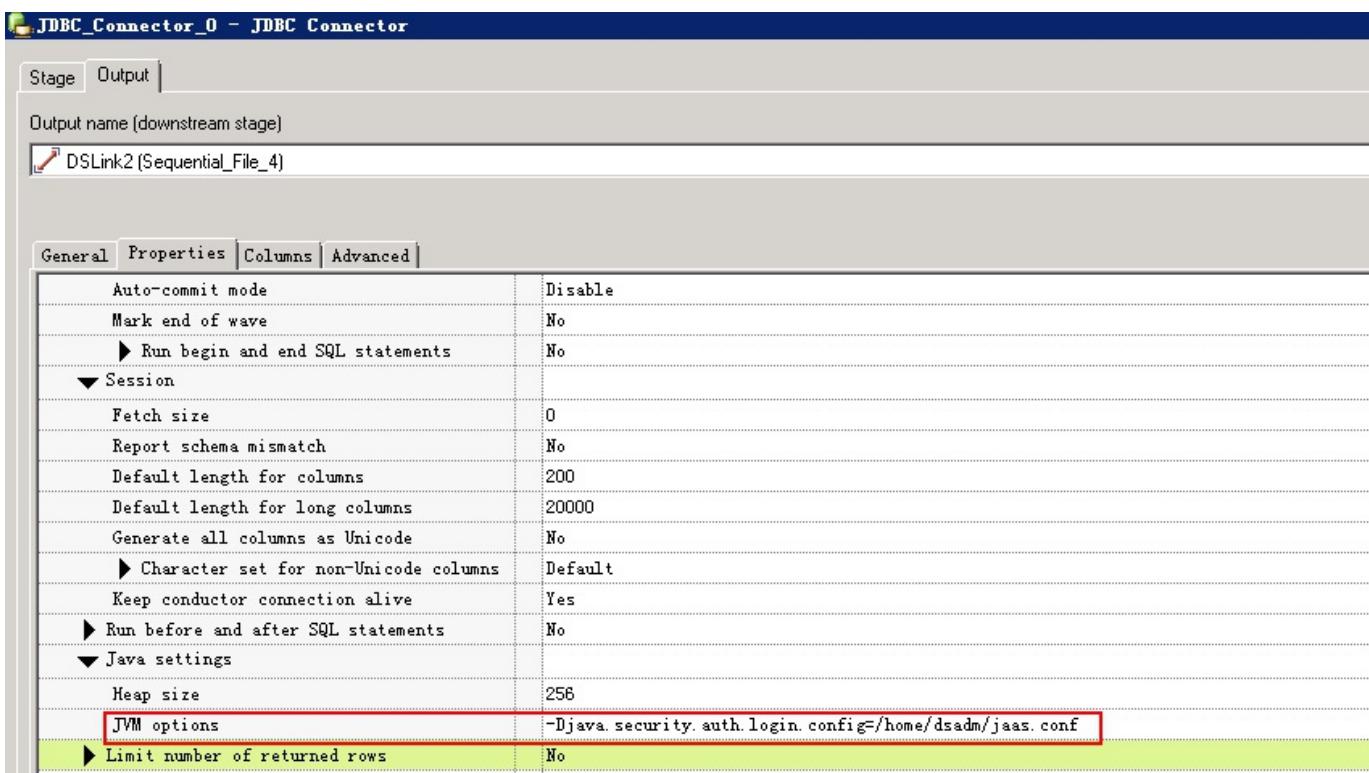


- 修改配置

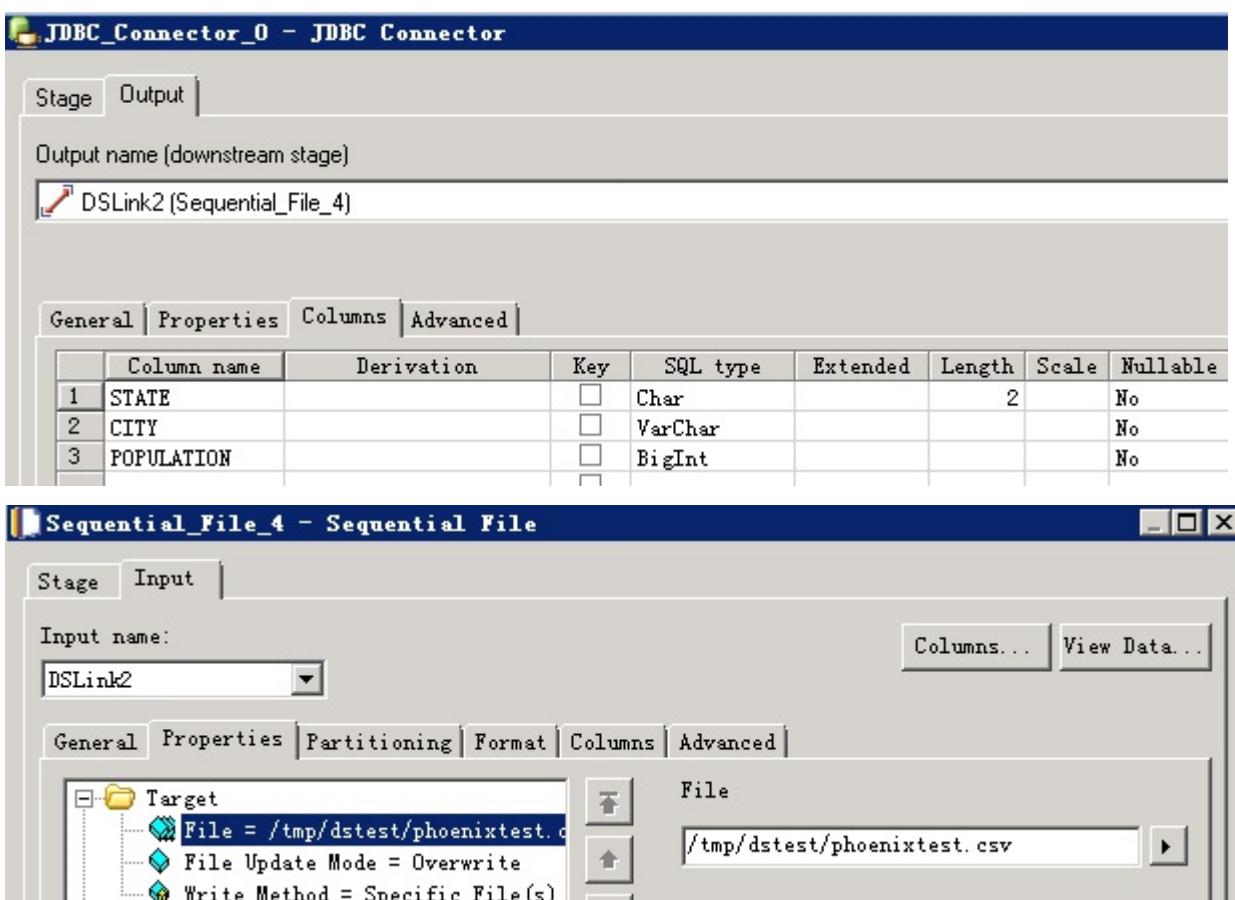


- URL参考：

```
jdbc:phoenix:fusioninsight3,fusioninsight2,fusioninsight1:24002:/hbase:test@test@HADOOP.COM:/home/dsadm/user.keytab
```



- 配置JVM options为 `-Djava.security.auth.login.config=/home/dsadm/jaas.conf`



- 编译运行

```

1:47:06 2017/7/8 Info JDBC_Connector_0,0: Number of rows fetched on the current node: 10.
1:47:06 2017/7/8 Info JDBC_Connector_0,0: The connector closed the connection to the data source.
1:47:06 2017/7/8 Info Sequential_File_4,0: Export complete; 10 records exported successfully, 0 rejected.
1:47:06 2017/7/8 Info JDBC_Connector_0: The connector closed the connection to the data source.
1:47:06 2017/7/8 Info main_program: Step execution finished with status = OK.
1:47:06 2017/7/8 Info main_program: Startup time, 0:03; production run time, 0:05.
1:47:06 2017/7/8 Info Parallel job reports successful completion
1:47:06 2017/7/8 Control Finished Job phoenixtest.

```

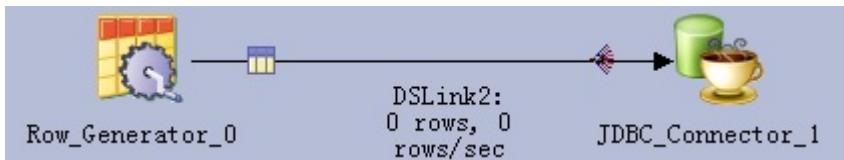
```
[root@datastagell502 dtest]# cat phoenixtest.csv
"CA","Los Angeles","3844829"
"CA","San Diego","1255540"
"CA","San Jose","912332"
"TX","Dallas","1213825"
"AZ","Phoenix","1461575"
"TX","Houston","2016582"
"PA","Philadelphia","1463281"
"IL","Chicago","2842518"
"NY","New York","8143197"
"TX","San Antonio","1256509"
```

写入Phoenix表数据

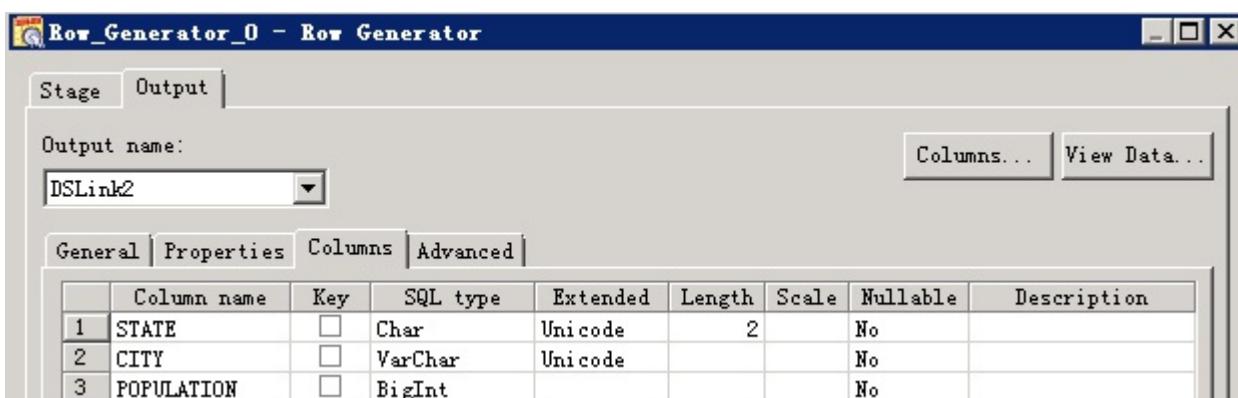
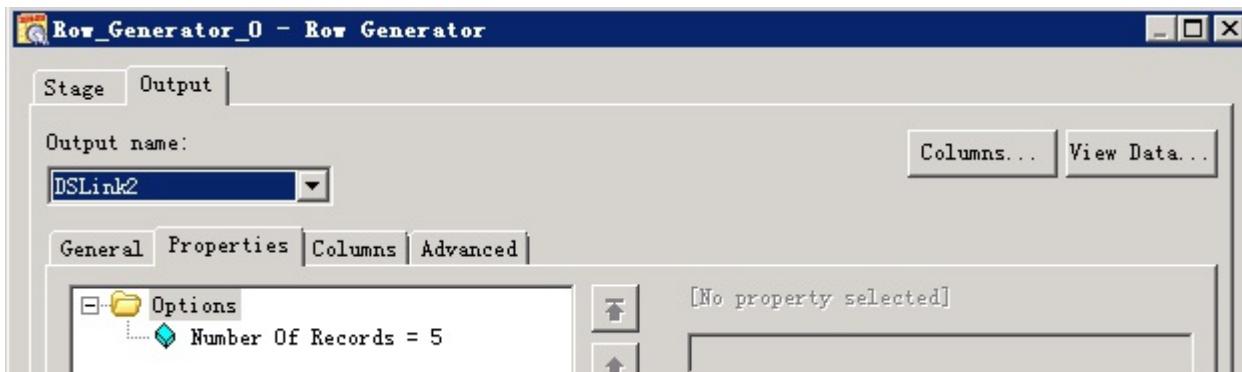
Phoenix插入语句是upsert into，不支持Insert into 语句，所以不能用JDBC Connector在运行时自动生成SQL语句，需要自己填写，否则会报错：

```
main_program: Fatal Error: The connector failed to prepare the statement: INSERT INTO us_population (STATE, CITY, POPULATION) VALUES (?, ?, ?). The reported error is: org.apache.phoenix.exception.PhoenixParserException: ERROR 601 (42P00): Syntax error. Encountered "INSERT" at line 1, column 1..
```

- 创建作业



- 修改配置



JDBC_Connector_1 - JDBC Connector

Stage | Input |

Input name (upstream stage)
DSLink2 (Row_Generator_0)

General Properties Columns Advanced Partitioning

Connection

URL *	jdbc:phoenix:fusioninsight3,fusioninsight2,fusioninsight1:24002:/hbase:test@H...
User name	
Password	
Attributes	

Usage

Write mode	Insert
Generate SQL at run time	No
Table name *	
► Table action *	Append
Enable quoted identifiers	No
▼ SQL	
▼ Insert statement *	UPSERT INTO us_population (STATE, CITY, POPULATION) VALUES (?, ?, ?)
Read insert statement from file	No

JDBC_Connector_1 - JDBC Connector

Stage | Input |

Input name (upstream stage)
DSLink2 (Row_Generator_0)

General Properties Columns Advanced Partitioning

Record count	2000
Isolation level	Default
Auto-commit mode	Disable
► Run begin and end SQL statements	No
▼ Session	
Batch size	2000
Drop unmatched fields	No
Report schema mismatch	No
Default length for columns	200
Default length for long columns	20000
► Character set for non-Unicode columns	Default
Keep conductor connection alive	Yes
► Run before and after SQL statements	No
▼ Java settings	
Heap size	256
JVM options	-Djava.security.auth.login.config=/home/dsadm/jaas..

- 编译运行

```

Info JDBC_Connector_1:1: The connector established connection to the data source.
Info JDBC_Connector_1:1: The character set encoding for the non-Unicode character values on the link is UTF-8. The maximum number of bytes per character in this
Info JDBC_Connector_1:1: The connector will run the following statement on the data source: UPSERT INTO us_population (STATE, CITY, POPULATION) VALUES (?, ?, ?)
Warning JDBC_Connector_1:1: 2017-07-10 14:19:27,914 INFO [hconnection=0x8dcfce8=shared-pool2-t9] ipc.AbstractKpcClient: RPC Server Kerberos principal name for se
Info JDBC_Connector_1:1: [JGSS_DBG_CRED] hconnection=0x8dcfce8=shared-pool2-t9 MN not found; creating one ...
Info JDBC_Connector_1:0: Number of rows inserted on the current node: 3.
Info JDBC_Connector_1:0: The connector closed the connection to the data source.
Info JDBC_Connector_1:1: Number of rows inserted on the current node: 2.
Info JDBC_Connector_1:1: The connector closed the connection to the data source.
Info JDBC_Connector_1:1: The connector closed the connection to the data source.
Info main_program: Step execution finished with status = OK
Info main_program: Startup time, 0:05; production run time, 0:04.
Info Parallel job reports successful completion
Control Finished Job phoenix write.

```

对接Fiber

对接Fiber需要先安装FI客户端

修改JDBC Driver配置文件

- 修改\$DSHOME路径的isjdbc.config文件，CLASSPATH变量中添加Fiber jdbc driver及依赖包的路径，CLASS_NAMES变量中添加com.huawei.fiber.FiberDriver;org.apache.hive.jdbc.HiveDriver; org.apache.phoenix.jdbc.PhoenixDriver

参考命令：

```
su - dsadm  
cd $DSHOME  
vi isidbc.config
```

配置如下：

```
CLASSPATH=/opt/IBM/InformationServer/ASBNode/lib/java/IShive.jar;/opt/mppdb/jdbc/gsjdbc4.jar;/opt/Progress/DataDirect/JDBC\_60/lib/mongodb.jar;/opt/ficlient/Fiber/lib/commons-cli-1.2.jar;/opt/ficlient/Fiber/lib/commons-logging-1.1.3.jar;/opt/ficlient/Fiber/lib/fiber-jdbc-1.0.jar;/opt/ficlient/Fiber/lib/hadoop-common-2.7.2.jar;/opt/ficlient/Fiber/lib/hive-beeline-1.2.1.spark.jar;/opt/ficlient/Fiber/lib/hive-common-1.2.1.spark.jar;/opt/ficlient/Fiber/lib/hive-jdbc-1.2.1.spark.jar;/opt/ficlient/Fiber/lib/jline-2.12.jar;/opt/ficlient/Fiber/lib/log4j-1.2.17.jar;/opt/ficlient/Fiber/lib/slf4j-api-1.7.10.jar;/opt/ficlient/Fiber/lib/slf4j-log4j12-1.7.10.jar;/opt/ficlient/Fiber/lib/super-csv-2.2.0.jar;

CLASS NAMES=com.ibm.isf.jdbc.hive.HiveDriver;org.postgresql.Driver;com.ddtek.jdbc.mongodb.MongoDBDriver;com.huawei.fiber.FiberDrive
```

修改Fiber配置文件

- DataStage使用IBM jdk，需要新建Fiber配置文件给DataStage使用

```
cd /opt/ficlient/Fiber/conf  
cp fiber.xml fiber_ibm.xml
```

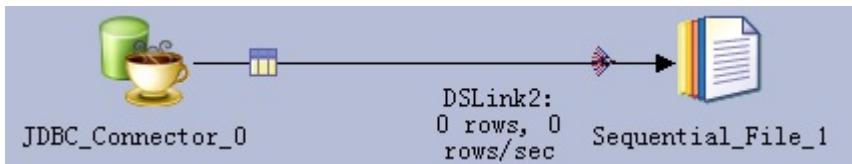
- 修改fiber_ibm.xml中phoenix,hive,spark各driver的以下两个参数：
 - java.security.auth.login.config 修改为 /home/dsadm/jaas.conf
 - zookeeperkinit 修改为 /opt/IBM/InformationServer/jdk/jre/bin/kinit
 - 文件/home/dsadm/jaas.conf的内容如下：

```
Client {
    com.ibm.security.auth.module.Krb5LoginModule required
        credsType=both
        principal="test@HADOOP.COM"
        useKeytab="/home/dsadm/user.keytab";
}.
.
```

- 其它配置项参考FI产品文档Fiber客户端配置指导修改。

使用Hive Driver读取数据

- 创建作业



- 修改配置

JDBC_Connector_0 - JDBC Connector

Stage Output

Output name (downstream stage)
DSLink2 (Sequential_File_1)

General		Properties		Columns		Advanced	
Connection							
URL *	jdbc:fiber://fiberconfig=/opt/ficlient/Fiber/conf/fiber_ibm.xml;defaultDriver=hive						
User name							
Password							
Attributes							
Usage							
Generate SQL at run time	Yes						
Table name *	employees_info						
Enable quoted identifiers	No						

URL参考：

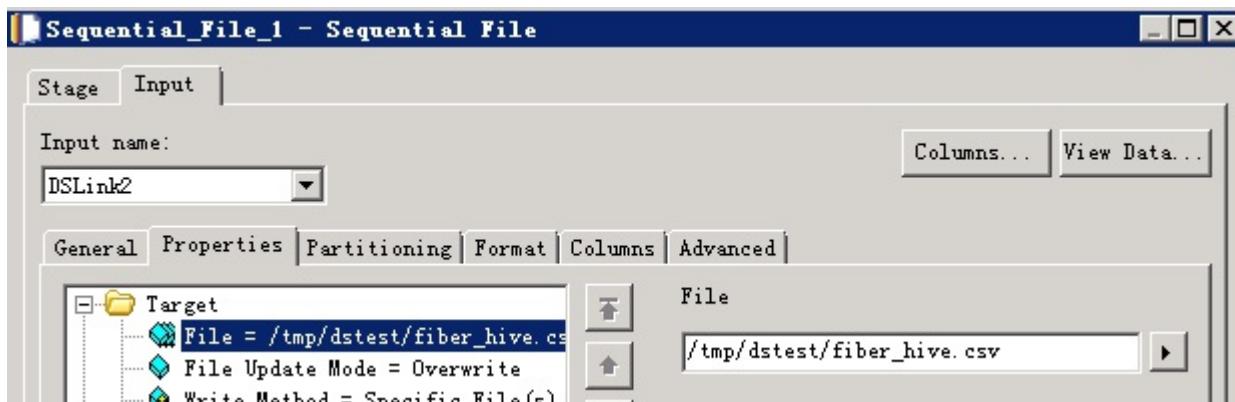
jdbc:fiber://fiberconfig=/opt/ficlient/Fiber/conf/fiber_ibm.xml;defaultDriver=hive

JDBC_Connector_0 - JDBC Connector

Stage Output

Output name (downstream stage)
DSLink2 (Sequential_File_1)

General		Properties		Columns		Advanced	
Column name	Derivation	Key	SQL type	Extended	Length	Scale	Nullable
1 id		<input type="checkbox"/>	Integer		5		No
2 name		<input type="checkbox"/>	VarChar	Unicode	20		No
3 usd_flag		<input type="checkbox"/>	VarChar	Unicode	2		No
4 salary		<input type="checkbox"/>	Double		8		No
5 address		<input type="checkbox"/>	VarChar	Unicode	30		No
6 entrytime		<input type="checkbox"/>	VarChar	Unicode	4		No

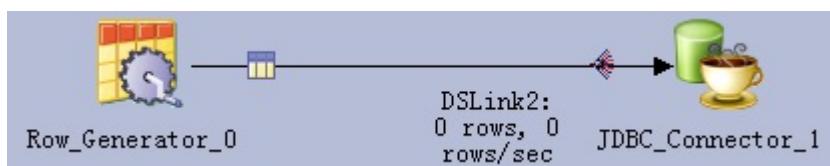


- ## • 编译运行

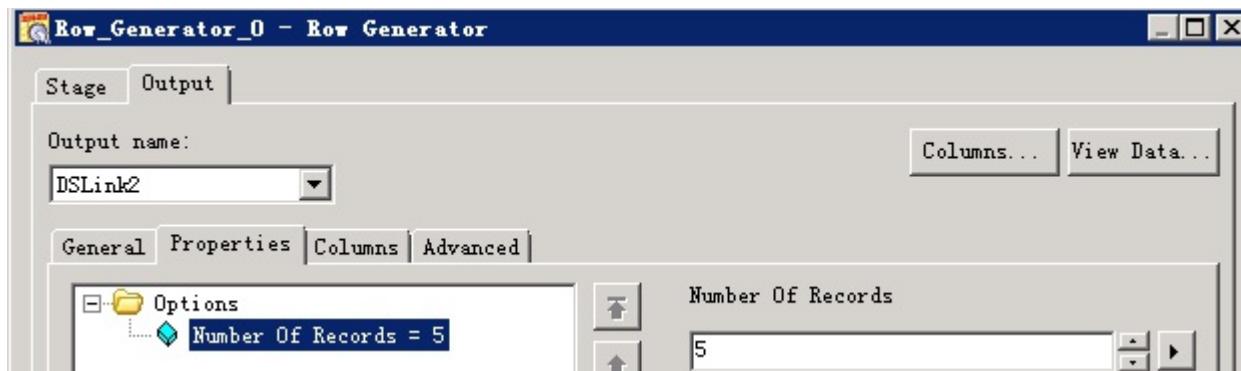
```
12:14:58 2017/7/8 Info JDBC_Connector_0,0: Number of rows fetched on the current node: 10.  
12:14:58 2017/7/8 Info JDBC_Connector_0,0: [JGSS_DBG_WRAP] main Krb5Context.wrap: set Flag Fields 0 SentBy:  
12:14:58 2017/7/8 Info JDBC_Connector_0,0: The connector closed the connection to the data source.  
12:14:58 2017/7/8 Info Sequential_File_1,0: Export complete; 10 records exported successfully, 0 rejected.  
12:14:58 2017/7/8 Info JDBC_Connector_0: The connector closed the connection to the data source.  
12:14:58 2017/7/8 Info main_program: Step execution finished with status = OK.  
12:14:58 2017/7/8 Info main_program: Startup time, 0:04; production run time, 0:02.  
12:14:58 2017/7/8 Info Parallel job reports successful completion  
12:14:58 2017/7/8 Control Finished Job fiber hive read.
```

使用Hive Driver写入数据

- 创建作业



- 修改配置



Row_Generator_0 - Row Generator

Stage Output |

Output name: **DSLink2** | Columns... View Data...

General Properties Columns Advanced |

	Column name	Key	SQL type	Extended	Length	Scale	Nullable	Description
1	id	<input type="checkbox"/>	Integer		5		No	
2	name	<input type="checkbox"/>	VarChar	Unicode	20		No	
3	usd_flag	<input type="checkbox"/>	VarChar	Unicode	2		No	
4	salary	<input type="checkbox"/>	Double		8		No	
5	address	<input type="checkbox"/>	VarChar	Unicode	30		No	
6	entrytime	<input type="checkbox"/>	VarChar	Unicode	4		No	

JDBC_Connector_1 - JDBC Connector

Stage Input |

Input name (upstream stage) **DSLink2 (Row_Generator_0)**

General Properties Columns Advanced Partitioning |

Connection

URL *	<code>jdbc:fiber://fiberconfig=/opt/ficlient/Fiber/conf/fiber_ibm.xml;defaultDriver=hive</code>
User name	
Password	
Attributes	

Usage

Write mode	Insert
Generate SQL at run time	Yes
Table name *	<code>employees_info</code>
► Table action *	Append

- 编译运行

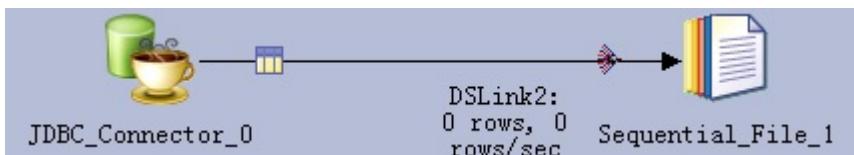
```

12:23:45 2017/7/8 Info JDBC_Connector_1,1: Number of rows inserted on the current node: 2.
12:23:45 2017/7/8 Info JDBC_Connector_1,1: [JGSS_DBG_MARSH] main 0000: 80 01 00 01 00 00 00 0c 43 6c 8f 7c
12:23:48 2017/7/8 Info JDBC_Connector_1,1: The connector closed the connection to the data source.
12:24:06 2017/7/8 Info JDBC_Connector_1,0: [JGSS_DBG_WRAP] main GSSContextImpl.unwrap buffer (len=113, off=
12:24:06 2017/7/8 Info JDBC_Connector_1,0: Number of rows inserted on the current node: 3.
12:24:06 2017/7/8 Info JDBC_Connector_1,0: [KRB_DBG_CRYP] AES128EType:main: AES128>>> AESencrypt result:
12:24:06 2017/7/8 Info JDBC_Connector_1,0: The connector closed the connection to the data source. (...)
12:24:06 2017/7/8 Info JDBC_Connector_1: The connector closed the connection to the data source.
12:24:06 2017/7/8 Info main_program: Step execution finished with status = OK.
12:24:06 2017/7/8 Info main_program: Startup time, 0:02; production run time, 1:52.
12:24:06 2017/7/8 Info Parallel job reports successful completion
12:24:06 2017/7/8 Control Finished Job fiber_hive_write.

```

使用Spark Driver读取数据

- 创建作业



- 修改配置

JDBC_Connector_0 - JDBC Connector

Stage Output

Output name (downstream stage)
DSLink2 (Sequential_File_1)

General Properties Columns Advanced

Connection

URL *	jdbc:fiber://fiberconfig=/opt/ficlient/Fiber/conf/fiber_ibm.xml;defaultDriver=spark
User name	
Password	
Attributes	

Usage

Generate SQL at run time	Yes
Table name *	employees_info
Enable quoted identifiers	No

URL参考:

jdbc:fiber://fiberconfig=/opt/ficlient/Fiber/conf/fiber_ibm.xml;defaultDriver=spark

JDBC_Connector_0 - JDBC Connector

Stage Output

Output name (downstream stage)
DSLink2 (Sequential_File_1)

General Properties Columns Advanced

	Column name	Derivation	Key	SQL type	Extended	Length	Scale	Nullable
1	id		<input type="checkbox"/>	Integer		5		No
2	name		<input type="checkbox"/>	VarChar	Unicode	20		No
3	usd_flag		<input type="checkbox"/>	VarChar	Unicode	2		No
4	salary		<input type="checkbox"/>	Double		8		No
5	address		<input type="checkbox"/>	VarChar	Unicode	30		No
6	entrytime		<input type="checkbox"/>	VarChar	Unicode	4		No

Sequential_File_1 - Sequential File

Stage Input

Input name:
DSLink2

Columns... View Data...

General Properties Partitioning Format Columns Advanced

Target

- File = /tmp/dstest/fiber_spark.csv
- File Update Mode = Overwrite

File
/tmp/dstest/fiber_spark.csv

● 编译运行

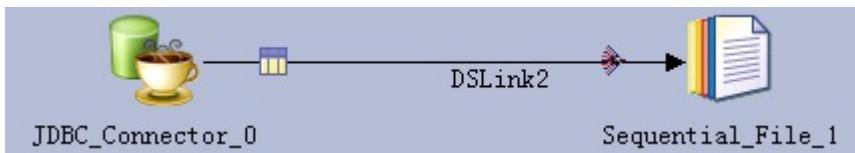
```

12:29:43 2017/7/8 Info JDBC_Connector_0,0: Number of rows fetched on the current node: 15.
12:29:43 2017/7/8 Info JDBC_Connector_0,0: [JGSS_DBG_WRAP] main Krb5Context.wrap: set Flag Fields 0 SentBy
12:29:43 2017/7/8 Info JDBC_Connector_0,0: The connector closed the connection to the data source.
12:29:43 2017/7/8 Info Sequential_File_1,0: Export complete; 15 records exported successfully, 0 rejected.
12:29:43 2017/7/8 Info JDBC_Connector_0: The connector closed the connection to the data source.
12:29:43 2017/7/8 Info main_program: Step execution finished with status = OK.
12:29:43 2017/7/8 Info main_program: Startup time, 0:15; production run time, 0:06.
12:29:43 2017/7/8 Info Parallel job reports successful completion
12:29:43 2017/7/8 Control Finished Job fiber_spark_read.

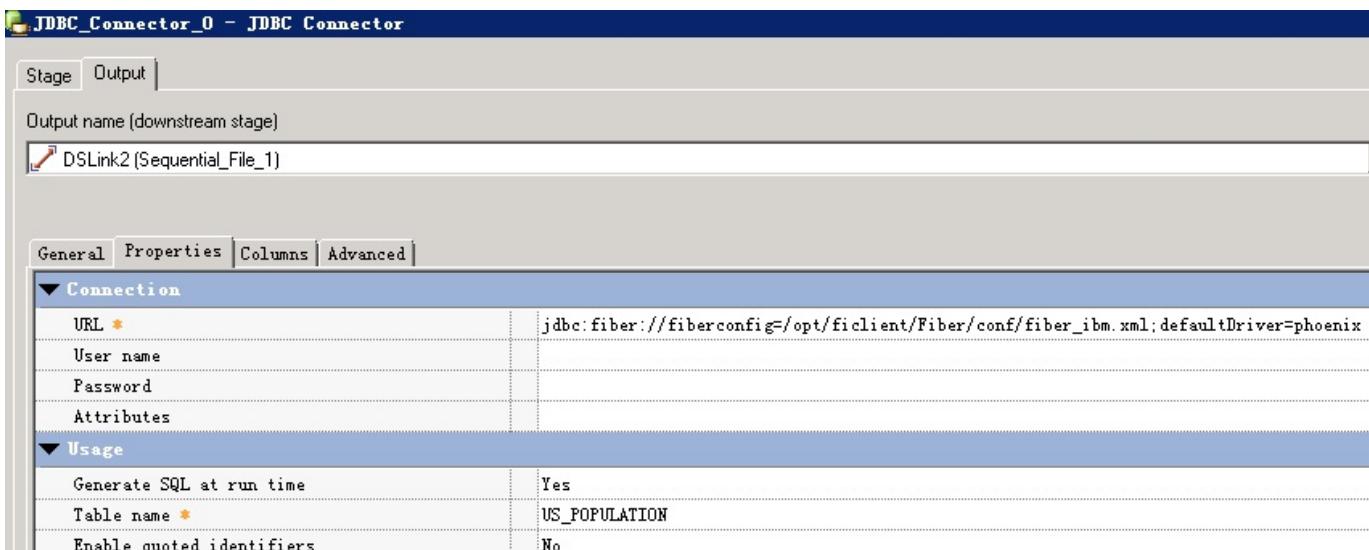
```

使用Phoenix Driver读取数据

- 创建作业

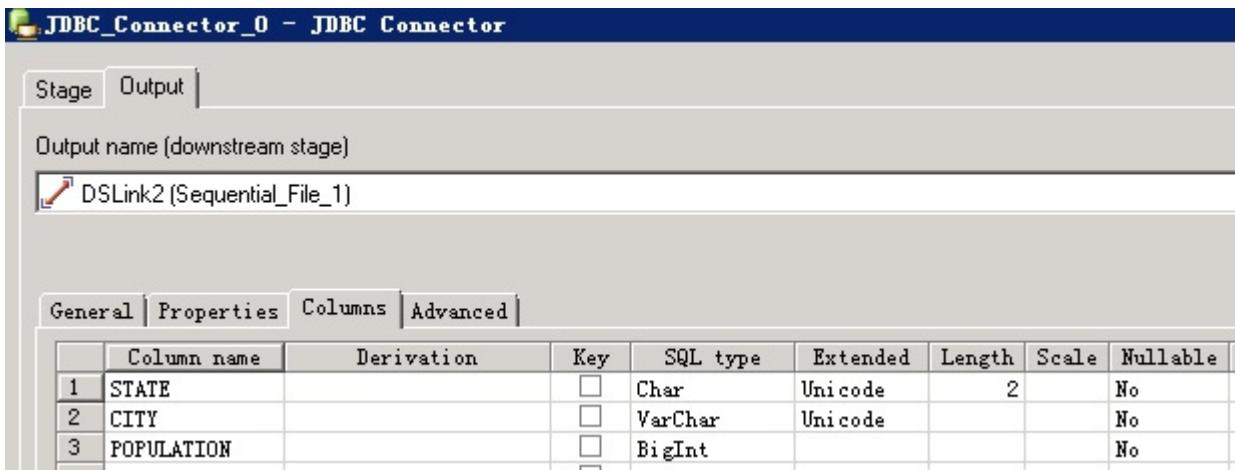


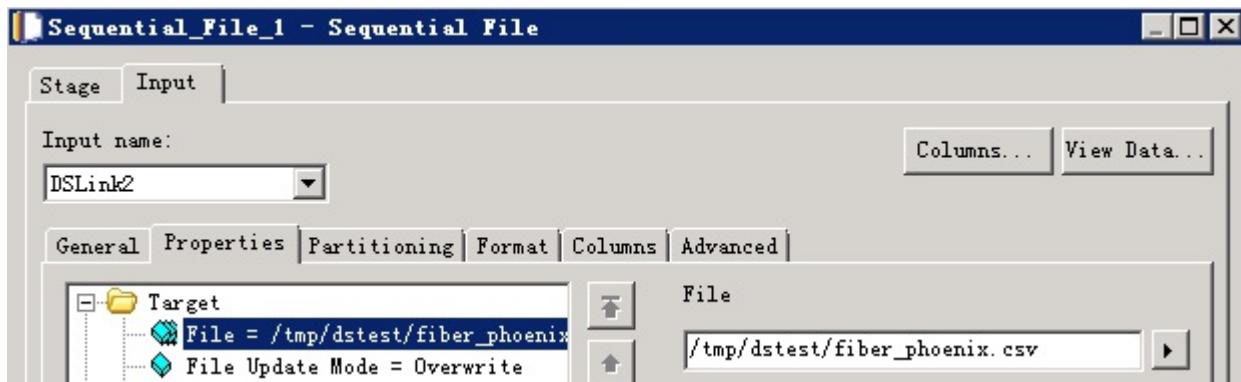
- 修改配置



URL参考：

```
jdbc:fiber://fiberconfig=/opt/ficlient/Fiber/conf/fiber_ibm.xml;defaultDriver=phoenix
```





- 编译运行

目前未能读取到数据，“The connector could not determine the value for the fetch size.”，问题正在确认中

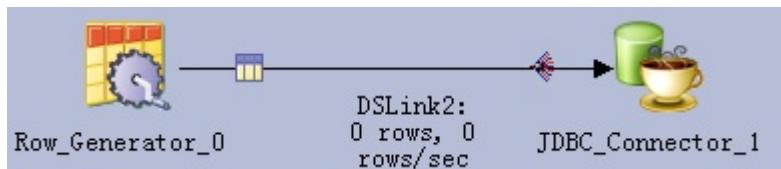
```

12:40:46 2017/7/8 Info JDBC_Connector_0,0: The connector will run the following statement on the data source:
12:40:46 2017/7/8 Info JDBC_Connector_0,0: The connector could not determine the value for the fetch size.
12:40:46 2017/7/8 Info JDBC_Connector_0,0: Number of rows fetched on the current node: 0.
12:40:46 2017/7/8 Info JDBC_Connector_0,0: The connector closed the connection to the data source.
12:40:46 2017/7/8 Info Sequential_File_1,0: Export complete; 0 records exported successfully, 0 rejected.
12:40:46 2017/7/8 Info JDBC_Connector_0: The connector closed the connection to the data source.
12:40:46 2017/7/8 Info main_program: Step execution finished with status = OK.
12:40:46 2017/7/8 Info main_program: Startup time, 0:05; production run time, 0:04.
12:40:46 2017/7/8 Info Parallel job reports successful completion
12:40:46 2017/7/8 Control Finished Job fiber_phoenix_read.

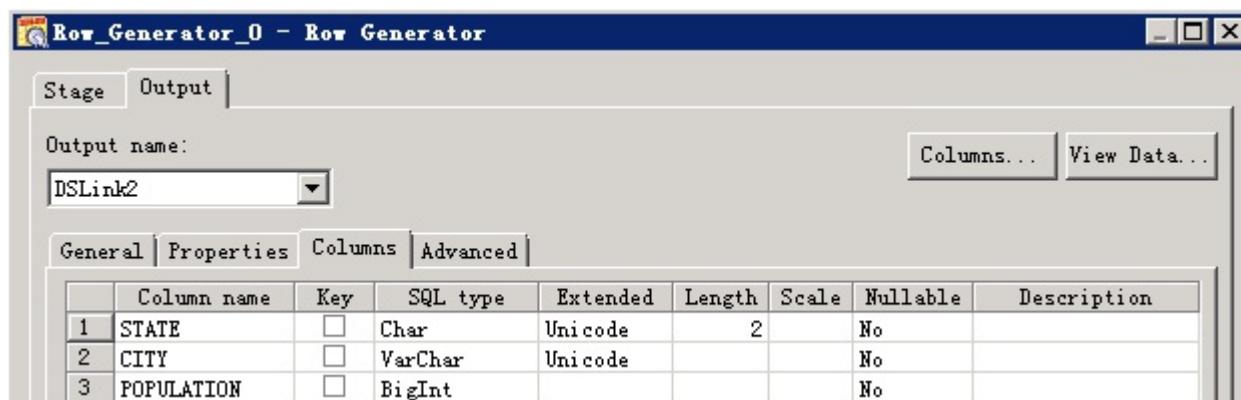
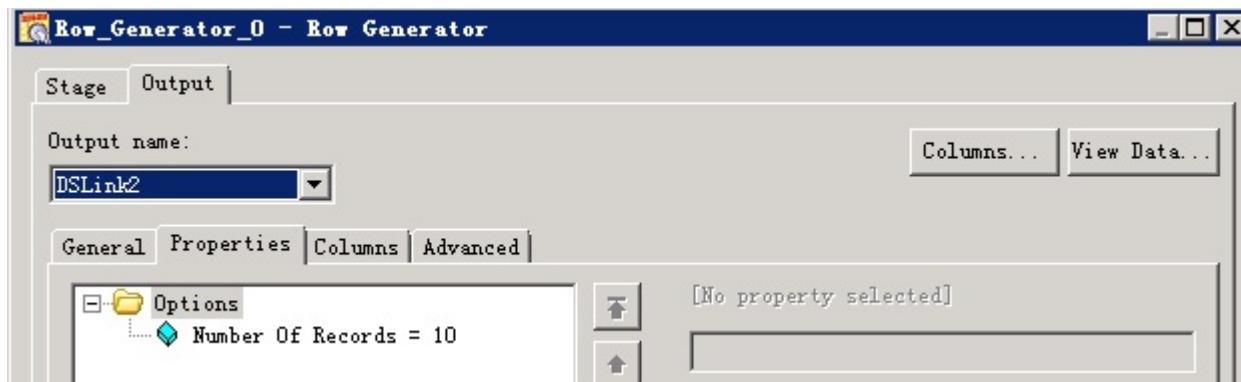
```

使用Phoenix Driver写入数据

- 创建作业



- 修改配置



URL参考:

```
jdbc:fiber://fiberconfig=/opt/ficlient/Fiber/conf/fiber_ibm.xml;defaultDriver=phoenix
```

- 编译运行

写入数据0行，问题正在确认中

```
JDBC_Connector_1,1: Number of rows inserted on the current node: 0.
JDBC_Connector_1,1: The connector closed the connection to the data source.
JDBC_Connector_1,0: The character set encoding for the non-Unicode character values on the link is UTF-8. The maximum number of bytes per character in this
JDBC_Connector_1,0: The connector will run the following statement on the data source: UPSERT INTO us_population (STATE, CITY, POPULATION) VALUES (?, ?, ?)
JDBC_Connector_1,0: Number of rows inserted on the current node: 0.
JDBC_Connector_1,0: The connector closed the connection to the data source.
JDBC_Connector_1: The connector closed the connection to the data source.
main_program: Step execution finished with status = OK.
main_program: Startup time, 0:05; production run time, 0:05.
Parallel job reports successful completion
Finished Job fiber_phoenix_write.
```

对接Kafka

说明： kafka Connector不支持发送或者消费integer, float, double, numeric, decimal等数值类型的字段，需要转换成char, varchar, longvarchar等类型，否则会有如下报错：

```
main_program: APT_PMsectionLeader(2, node2), player 2 - Unexpected termination by Unix signal 9(SIGKILL).
```

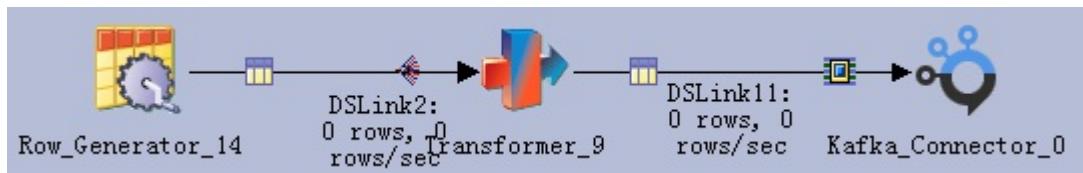
安装kafka客户端

- kafka Connector需要配置Kafka client Classpath，可以在DataStage节点安装kafka客户端来获取kafka-client jar包。安装步骤参考FusionInsight产品文档。
- Kafka Client Classpath 需要提供kafka-client, log4j, slf4j-api 三个jar包的路径，如：

```
/opt/ficlient/Kafka/kafka/libs/kafka-clients-0.10.0.0.jar;/opt/ficlient/Kafka/kafka/libs/log4j-
1.2.17.jar;/opt/ficlient/Kafka/kafka/libs/slf4j-api-1.7.21.jar
```

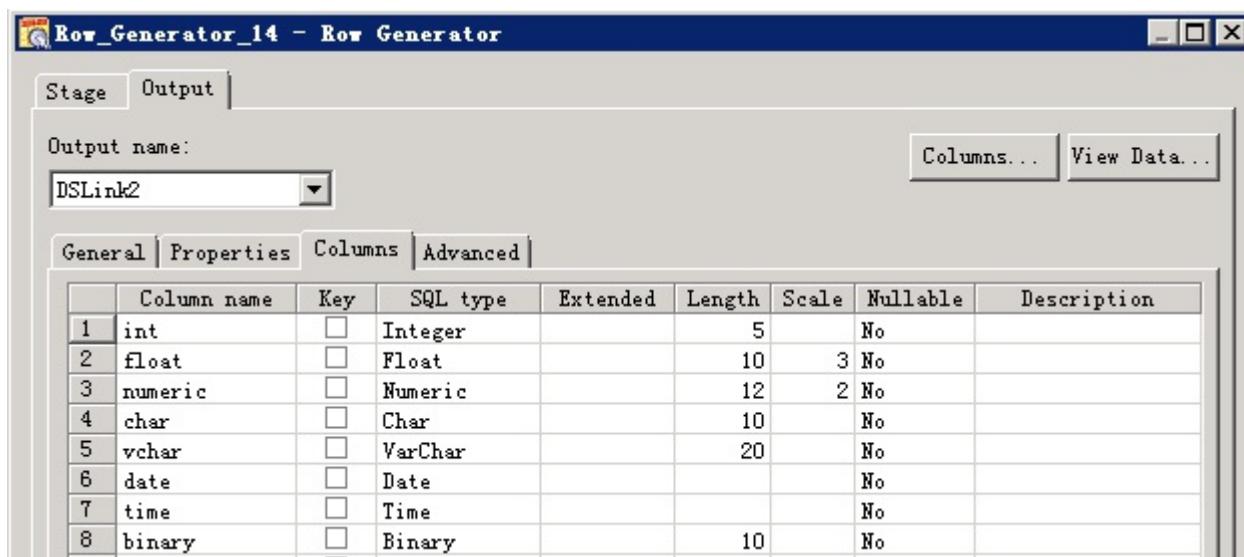
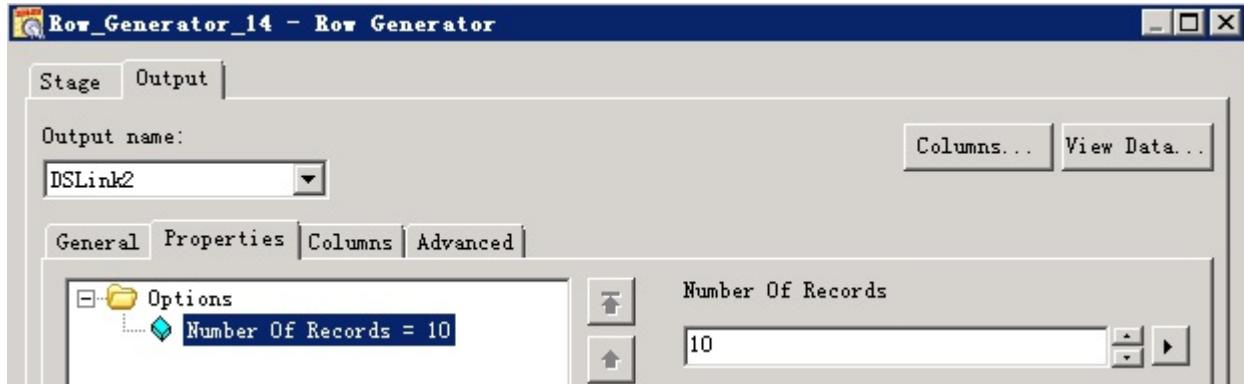
发送消息到kafka

- 创建作业



- 修改配置

RowGenerator 生成数据



transformer数据类型转换:

Transformer_9 - Transformer Stage

DSLink2

int
float
numeric
char
vchar
date
time
binary

Stage Variables

Derivation	Stage Variable
------------	----------------

Loop Condition (No Loop)

DSLink11

Constraint:	
Derivation	Column Name
DSLink2.int	int
DSLink2.float	float
DSLink2.numeric	numeric

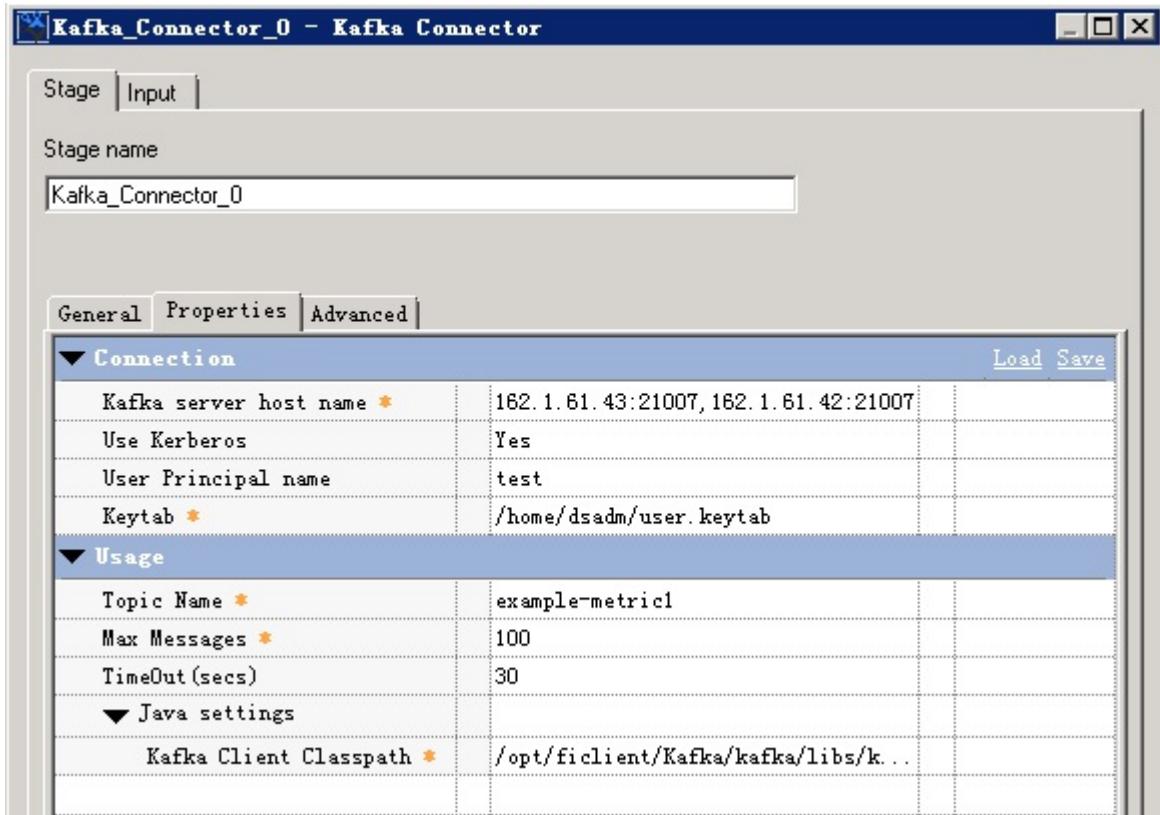
DSLink2

Column name	Key	SQL type	Extended	Length	Scale	Nullable	Script
1 int	<input type="checkbox"/>	Integer		5		No	
2 float	<input type="checkbox"/>	Float		10	3	No	
3 numeric	<input type="checkbox"/>	Numeric		12	2	No	
4 char	<input type="checkbox"/>	Char		10		No	
5 vchar	<input type="checkbox"/>	VarChar		20		No	
6 date	<input type="checkbox"/>	Date				No	
7 time	<input type="checkbox"/>	Time				No	
8 binary	<input type="checkbox"/>	Binary		10		No	

DSLink11

Column name	Key	SQL type	Extended	Length	Scale	Nullable
1 int	<input type="checkbox"/>	VarChar		5		No
2 float	<input type="checkbox"/>	VarChar		10	3	No
3 numeric	<input type="checkbox"/>	VarChar		12	2	No
4 char	<input type="checkbox"/>	Char		10		No
5 vchar	<input type="checkbox"/>	VarChar		20		No
6 date	<input type="checkbox"/>	Date				No
7 time	<input type="checkbox"/>	Time				No
8 binary	<input type="checkbox"/>	Binary		10		No

Kafka配置:

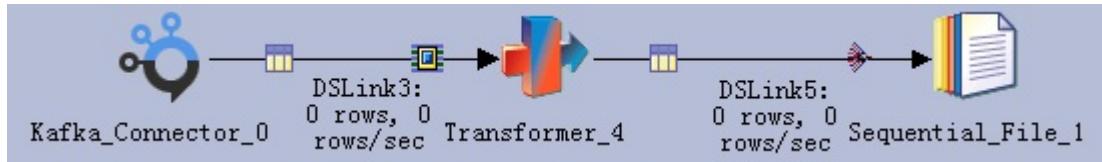


- 编译运行

```
11:28:23 2017/7/7 Info Kafka_Connector_0:1: [JGSS_DBG_CTX] kafka-producer-network-thread
11:28:23 2017/7/7 Info main_program: Step execution finished with status = OK.
11:28:23 2017/7/7 Info main_program: Startup time, 0:00; production run time, 0:01.
11:28:23 2017/7/7 Info Parallel job reports successful completion
11:28:23 2017/7/7 Control Finished Job kafkatest.
```

读取Kafka消息

- 创建作业



- 修改配置

Kafka_Connector_0 - Kafka Connector

Stage | Output |

Stage name
Kafka_Connector_0

General Properties Advanced

Connection

Kafka server host name *	162.1.61.43:21007, 162.1.61.42:21007
Use Kerberos	Yes
User Principal name	test
Keytab *	/home/dsadm/user.keytab

Usage

Topic Name *	example-metric1
Consumer Group *	example-group1
Max Poll Records *	100
Max Messages *	100
Reset Policy	earliest
TimeOut(secs)	300
Java settings	
Kafka Client Classpath *	/opt/ficlient/Kafka/kafka/libs/k...

Transformer_4 - Transformer Stage

DSLLink3

int
float
numeric
char
vchar
date
time
binary

Stage Variables

Derivation	Stage Variable

Loop Condition (No Loop)

DSLLink5

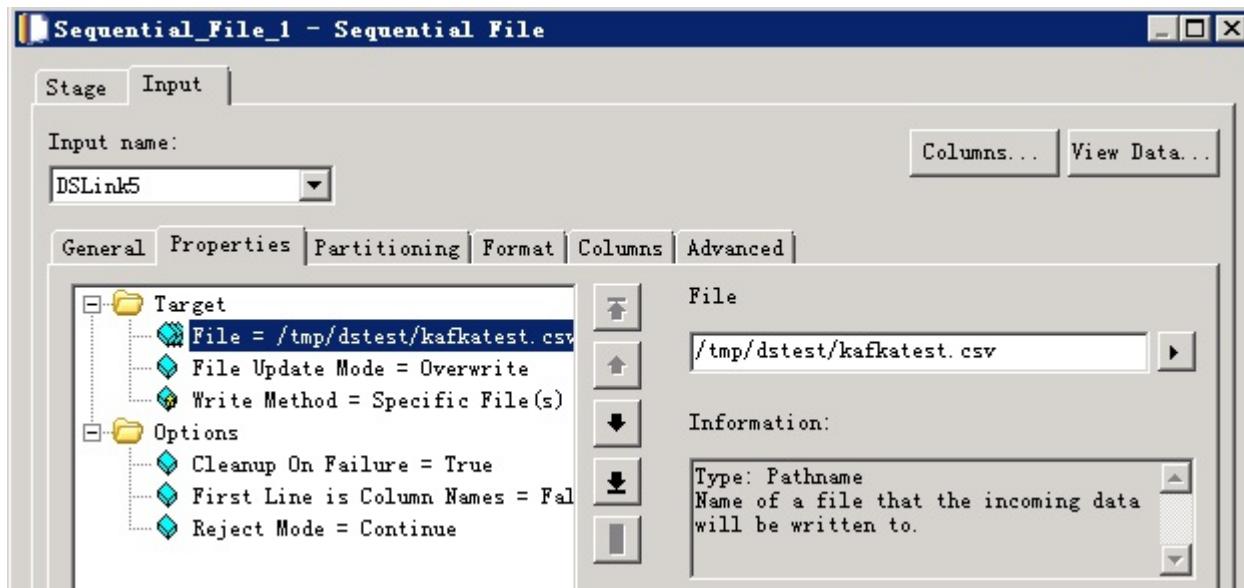
Constraint:	
Derivation	Column Name
DSLLink3.int	int
DSLLink3.float	float
DSLLink3.numeric	numeric

DSLLink3

	Column name	Key	SQL type	Extended	Length	Scale	Nullable	cript
1	int	<input type="checkbox"/>	VarChar		5	No		
2	float	<input type="checkbox"/>	VarChar		10	3	No	
3	numeric	<input type="checkbox"/>	VarChar		12	2	No	
4	char	<input type="checkbox"/>	Char		10	No		
5	vchar	<input type="checkbox"/>	VarChar		20	No		
6	date	<input type="checkbox"/>	Date			No		
7	time	<input type="checkbox"/>	Time			No		
8	binary	<input type="checkbox"/>	Binary		10	No		

DSLLink5

	Column name	Key	SQL type	Extended	Length	Scale	Nullable
1	int	<input type="checkbox"/>	Integer		5	No	
2	float	<input type="checkbox"/>	Float		10	3	No
3	numeric	<input type="checkbox"/>	Numeric		12	2	No
4	char	<input type="checkbox"/>	Char	Unicode	10	No	
5	vchar	<input type="checkbox"/>	VarChar	Unicode	20	No	
6	date	<input type="checkbox"/>	Date			No	
7	time	<input type="checkbox"/>	Time			No	
8	binary	<input type="checkbox"/>	Binary		10	No	



- 编译运行

查看读取的数据

```
[root@datastageell502 dstest]# cat kafkatest.csv
"0","0.0000000E+00"," 0000000000.00","","","0001-01-01","00:00:00","",""
"0","0.0000000E+00"," 0000000000.00","","","0001-01-01","00:00:00","",""
"0","0.0000000E+00"," 0000000000.00","","","0001-01-01","00:00:00","",""
"11","0.0000000E+00"," 0000000000.00","","","0001-01-01","00:00:00","",""
"33","0.0000000E+00"," 0000000000.00","","","0001-01-01","00:00:00","",""
"55","0.0000000E+00"," 0000000000.00","","","0001-01-01","00:00:00","",""
"77","0.0000000E+00"," 0000000000.00","","","0001-01-01","00:00:00","",""
"99","0.0000000E+00"," 0000000000.00","","","0001-01-01","00:00:00","",""
"0","0.0000000E+00"," 0000000000.00","","","0001-01-01","00:00:00","",""
"0","0.0000000E+00"," 0000000000.00","","","0001-01-01","00:00:00","",""
"0","0.0000000E+00"," 0000000000.00","","","0001-01-01","00:00:00","",""
```

对接MPPDB

获取MPPDB JDBC Driver

- 从MPPDB发布包中获取，包名为Gauss200-OLAP-VxxxRxxxCxx-xxxx-64bit-Jdbc.tar.gz
- 解压后得到gsjdbc4.jar，上传到DataStage Server

修改JDBC Driver配置文件

- 修改\$DSHOME路径的isjdbc.config文件，CLASSPATH变量中添加MPPDB Driver 的路径，CLASS_NAMES变量中添加org.postgresql.Driver

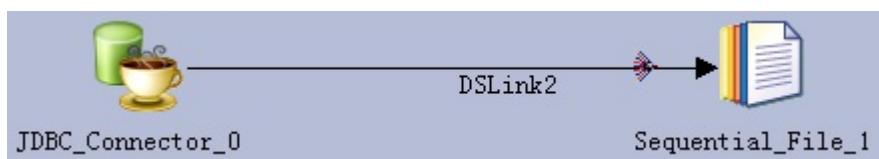
```
su - dsadm
cd $DSHOME
vi isjdbc.config
```

配置：

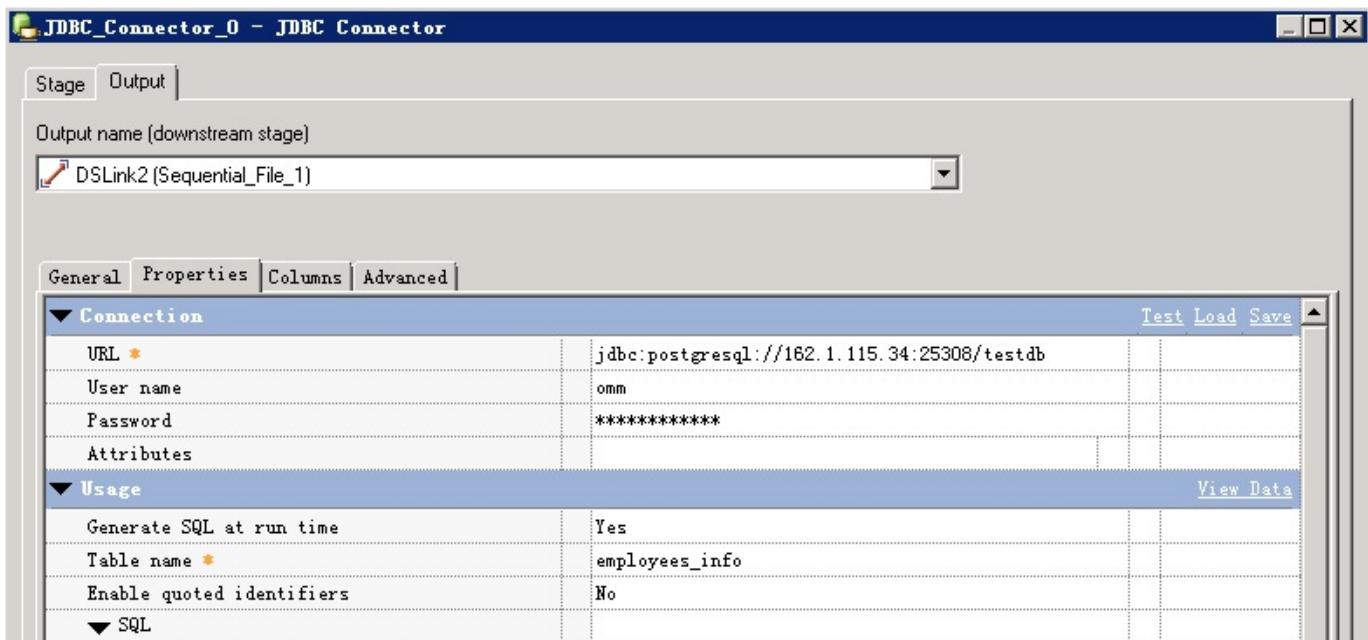
```
CLASSPATH=/opt/IBM/InformationServer/ASBNode/lib/java/IShive.jar;/opt/mppdb/jdbc/gsjdbc4.jar;
CLASS_NAMES=com.ibm.isf.jdbc.hive.HiveDriver;org.postgresql.Driver;
```

读取MPPDB表数据

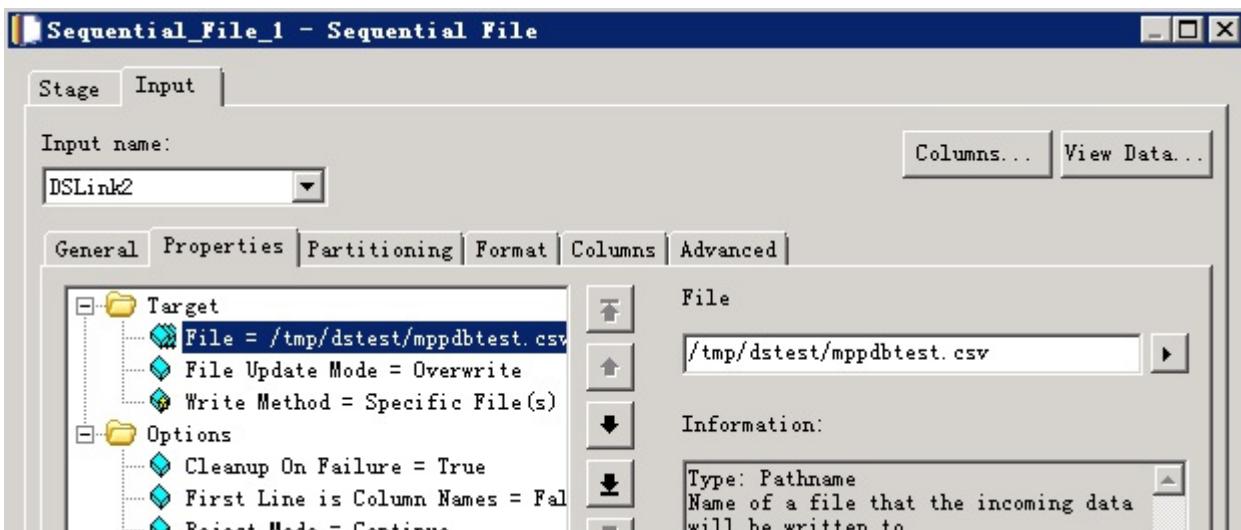
- 创建作业



- 修改配置



URL格式为: jdbc:postgresql://hostport/database



- 编译运行

```

17:19:35 2017/7/7 Info JDBC_Connector_0,0: Number of rows fetched on the current node: 5.
17:19:35 2017/7/7 Info JDBC_Connector_0,0: The connector closed the connection to the data source.
17:19:35 2017/7/7 Info Sequential_File_1,0: Export complete; 5 records exported successfully, 0 rejected.
17:19:35 2017/7/7 Info JDBC_Connector_0: The connector closed the connection to the data source.
17:19:35 2017/7/7 Info main_program: Step execution finished with status = OK.
17:19:35 2017/7/7 Info main_program: Startup time, 0:01; production run time, 0:01.
17:19:35 2017/7/7 Info Parallel job reports successful completion
17:19:35 2017/7/7 Control Finished Job mppdb_read.

```

```

[root@datastagel1502 dstest]# cat mppdbtest.csv
"4","Jack","D","2.4000029999999988E+04","America:Manhattan","2014"
"6","Linda","D","3.6000040000000009E+04","America:NewYork","2014"
"1","Wang","R","8.0000100000000022E+03","China:Shenzhen","2014"
"8","Zhang","R","9.0000499999999927E+03","China:Shanghai","2014"
"3","Tom","D","1.2000020000000004E+04","America:NewYork","2014"

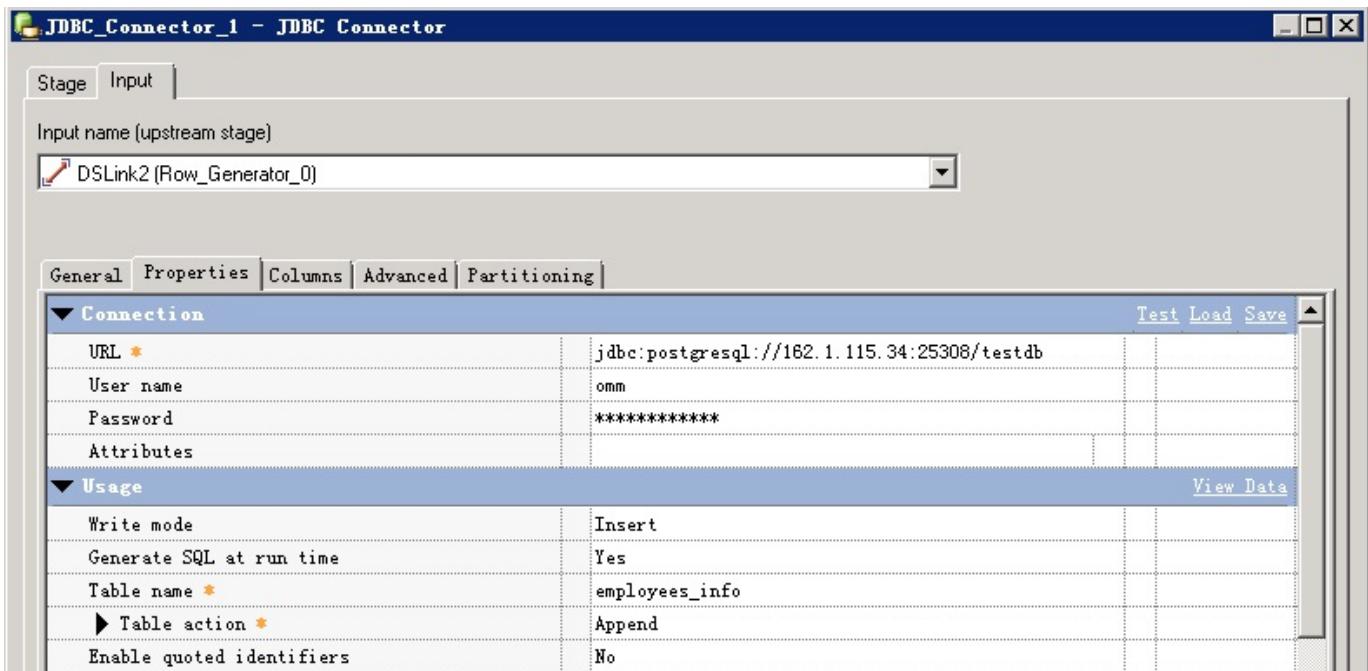
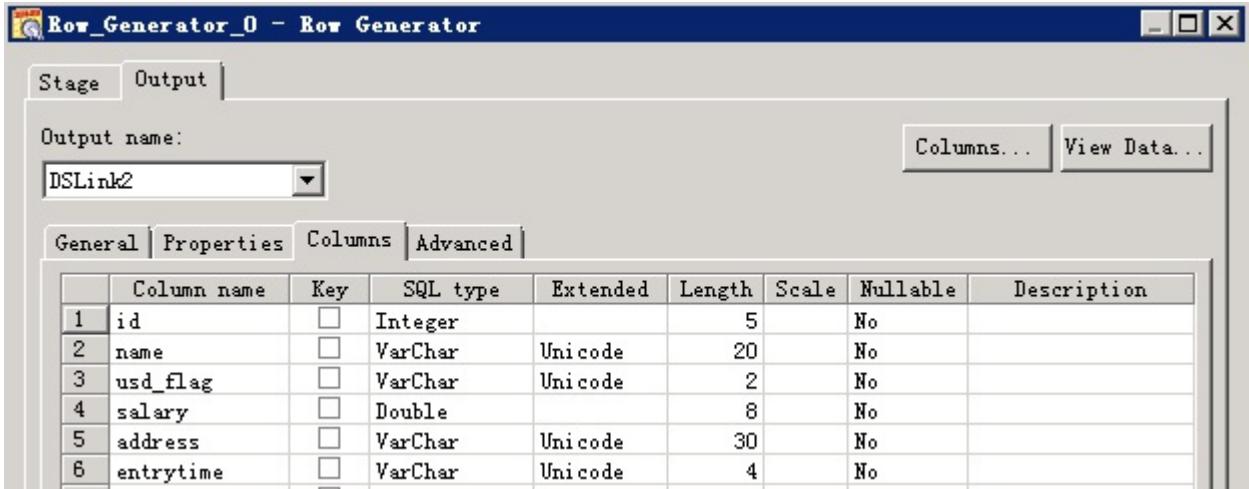
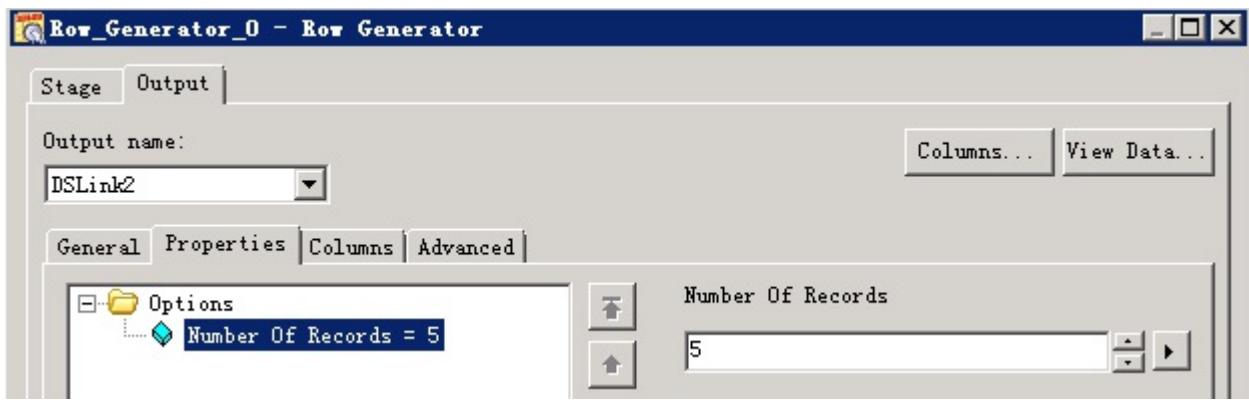
```

数据写入MPPDB表

- 创建作业



- 修改配置



URL格式为: jdbc:postgresql://hostport/database

- 编译运行

```

17:52:16 2017/7/7 Info JDBC_Connector_1,1: Number of rows inserted on the current node: 5.
17:52:16 2017/7/7 Info JDBC_Connector_1,1: The connector closed the connection to the data source.
17:52:16 2017/7/7 Info JDBC_Connector_1,0: Number of rows inserted on the current node: 5.
17:52:16 2017/7/7 Info JDBC_Connector_1,0: The connector closed the connection to the data source.
17:52:16 2017/7/7 Info JDBC_Connector_1: The connector closed the connection to the data source.
17:52:16 2017/7/7 Info main_program: Step execution finished with status = OK.
17:52:16 2017/7/7 Info main_program: Startup time, 0:01; production run time, 0:01.
17:52:16 2017/7/7 Info Parallel job reports successful completion
17:52:16 2017/7/7 Control Finished Job mppdb_write.

```

- 查看MPPDB表数据:

