

GetFile/PutFile

绝对路径

GetFTP/PutFTP

注意 Windows 与 Linux 的路径问题，比如 / 与 ./

ExecuteSQL

必须要有输出

Select

Processor Details

SETTINGS SCHEDULING **PROPERTIES** COMMENTS

Required field

Property	Value
Database Connection Pooling Service	DBCPConnectionPool →
SQL select query	select m.id, m.name, s.type, n.nation, m.birthday, m.certtype, ...
Max Wait Time	0 seconds
Normalize Table/Column Names	false
Use Avro Logical Types	false
Default Decimal Precision	10
Default Decimal Scale	0

OK

Controller Service Details

SETTINGS

PROPERTIES

COMMENTS

Required field

Property		Value
Database Connection URL	?	jdbc:oracle:thin:@(DESCRIPTION=(ADDRESS_LIS...
Database Driver Class Name	?	oracle.jdbc.driver.OracleDriver
Database Driver Location(s)	?	./lib/ojdbc6-11.2.0.4.0.jar
Database User	?	dragon
Password	?	Sensitive value set
Max Wait Time	?	3000 millis
Max Total Connections	?	16
Validation query	?	No value set

OK

QueryRecord

Processor Details

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field

Property		Value	
Record Reader	?	AvroReader	→
Record Writer	?	CSVRecordSetWriter	→
Include Zero Record FlowFiles	?	true	
Cache Schema	?	false	
max_oid	?	SELECT MAX(oid) as max_oid FROM FLOWFILE	

OK

ExtractText

Processor Details

SETTINGS SCHEDULING **PROPERTIES** COMMENTS

Required field

Property	Value
Maximum Buffer Size	1 MB
Maximum Capture Group Length	1024
Enable Canonical Equivalence	false
Enable Case-insensitive Matching	false
Permit Whitespace and Comments in Pattern	false
Enable DOTALL Mode	false
Enable Literal Parsing of the Pattern	false
Enable Multiline Mode	false
Enable Unicode-aware Case Folding	false
Enable Unicode Predefined Character Classes	false
Enable Unix Lines Mode	false
Include Capture Group 0	true
Enable repeating capture group	false
max_oid	^[0-9]*\$

OK

UpdateAttribute

Processor Details

SETTINGS SCHEDULING **PROPERTIES** COMMENTS

Required field

Property	Value
Delete Attributes Expression	No value set
Store State	Do not store state
Stateful Variables Initial Value	No value set
filename	\$(max_oid.0.isNull():ifElse(\$(filename), \$(max_oid.0)))

ADVANCED

OK

重点是掌握 EL 语法

splitAvro

Processor Details

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property		Value
Split Strategy	?	Record
Output Size	?	1
Output Strategy	?	Datafile
Transfer Metadata	?	true

OK

注意 output strategy 的选择

Datafile 包含 schema 信息，如果需要 convert 则选择此项。

Bare Record 不包含 schema 信息，如果需要输出到文件或者 kafka 中，选择此项。

ConvertavroSchema

Processor Details

SETTINGS SCHEDULING **PROPERTIES** COMMENTS

Required field

Property	Value
Input Schema	{ "type": "record", "name": "NiFi_ExecuteSQL_Record", "names...
Output Schema	{ "type": "record", "name": "LGY", "namespace": "cn.gov.ytga.av...
Locale	default
Compression type	SNAPPY
ADDRESS	address
AREA	area
BIRTHDAY	birthday
BISCOMPLETE	biscomplete
CERTNO	certno
CERTYPE	certype
CREDITNO	creditno
CREDITYPE	creditype
FLAGDUIB	flagduib
FLAGPIPE	flannine

OK

分别设置输入输出项，然后下面配置输入输出 item 对应关系

ConvertAvroJSON

Processor Details

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field

Property		Value
JSON container options		array
Wrap Single Record		false
Avro schema		{ "type": "record", "name": "LGY", "namespace": "cn.gov.ytga.avro", ...

OK

EvaluateJsonPath

Processor Details

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field

Property		Value
Destination		flowfile-content
Return Type		auto-detect
Path Not Found Behavior		ignore
Null Value Representation		empty string
itime		\$.itime

OK

学习 jsonpath 语法, W3CSchool 有详细课程

SplitXml

注意 depth 参数

EvaluateXPath

Processor Details

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property	Value
Destination	flowfile-attribute
Return Type	auto-detect
Validate DTD	true
dz	*[local-name()='Item']/@DZ
dzqh	*[local-name()='Item']/@DZQH
fzjgmc	*[local-name()='Item']/@FZJGMC
gj	*[local-name()='Item']/@GJ
swkssj	*[local-name()='Item']/@SWKSSJ
swryxm	*[local-name()='Item']/@SWRYXM
swzdh	*[local-name()='Item']/@SWZDH
xwsj	*[local-name()='Item']/@XWSJ
yycsdm	*[local-name()='Item']/@YYCSDM
yycsmc	*[local-name()='Item']/@YYCSMC
zihm	*[local-name()='Item']/@Z_IHM

OK

<item a1="v1", b1="v2">

Processor Details

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field

Property		Value	
Destination	?	flowfile-attribute	
Return Type	?	auto-detect	
Validate DTD	?	true	
bc	?	/R1/班次	
cw	?	/R1/仓位	
ddg	?	/R1/到达港	
gpsj	?	/R1/购票时间	
gxsj	?	/R1/更新时间	
hwlx	?	/R1/货物类型	
hwzl	?	/R1/货物重量	
id	?	/R1/系统主键	
lgsj	?	/R1/离港时间	
rksj	?	/R1/入库时间	
sfn	?	/R1/始发港	

OK

<item>

<a>v

v

</item>

重点学习 XPath 语法 W3CSchool XPath

AttributeToJSON

AttributeToJSON Details

SETTINGS | SCHEDULING | **PROPERTIES** | COMMENTS

Required field

Property	Value
Attributes List	yycsdm,yycsmc,dzqh,dz,swryxm,zjlx,zjhm,fzjgmc,gj,swksj,s...
Attributes Regular Expression	No value set
Destination	flowfile-content
Include Core Attributes	true
Null Value	false

OK

QueryDatabasetable

Configure Processor

SETTINGS | SCHEDULING | **PROPERTIES** | COMMENTS

Required field +

Property	Value
Database type	Generic
Table Name	jwkf.jyerp_jcxx_zjxx
Columns to Return	No value set
Maximum-value Columns	zjxx_id
Max Wait Time	0 seconds
Fetch Size	1000
Max Rows Per Flow File	1000
Maximum Number of Fragments	0
Normalize Table/Column Names	false
Use Avro Logical Types	false
Default Decimal Precision	10
Default Decimal Scale	0
Additional WHERE clause	No value set
maxvalue.zjxx_id	2620037 🗑

CANCEL | APPLY

SplitText

SETTINGS SCHEDULING **PROPERTIES** COMMENTS

Required field

Property	Value
Line Split Count	1
Maximum Fragment Size	No value set
Header Line Count	0
Header Line Marker Characters	No value set
Remove Trailing Newlines	true

OK

ConvertcsvtoAvro

SETTINGS SCHEDULING **PROPERTIES** COMMENTS

Required field

Property	Value
Hadoop configuration Resources	No value set
Record schema	{"type": "record", "name": "YIDAO_ORDER", "namespace": "..."}
CSV charset	utf8
CSV delimiter	
CSV quote character	*
CSV escape character	\
Use CSV header line	false
Lines to skip	0
Compression type	SNAPPY

ConsumeAmqp

Processor Details

SETTINGS SCHEDULING **PROPERTIES** COMMENTS

Required field

Property	Value
Queue	qVehiclePassFromYTJJ
Host Name	10.50.129.190
Port	5672
Virtual Host	/
User Name	ucp
Password	Sensitive value set
AMQP Version	0.9.1
SSL Context Service	No value set
Use Certificate Authentication	false
Client Auth	REQUIRED

OK

CaptureChangeMySQL

Processor Details

SETTINGS SCHEDULING **PROPERTIES** COMMENTS

Required field

Property	Value
MySQL Hosts	10.50.129.190:3306
MySQL Driver Class Name	com.mysql.jdbc.Driver
MySQL Driver Location(s)	./lib/mysql/mysql-connector-java-5.1.45.jar
Username	root
Password	Sensitive value set
Server ID	129190
Database/Schema Name Pattern	ucp
Table Name Pattern	ucp_model_control
Max Wait Time	30 seconds
Distributed Map Cache Client	RedisDistributedMapCacheClientService →
Retrieve All Records	true
Include Begin/Commit Events	false
Include DDL Events	false
State Update Interval	0 seconds

OK

State Update Interval	?	0 seconds	
Initial Sequence ID	?	No value set	
Initial Binlog Filename	?	binlog	
Initial Binlog Position	?	3678134	

OK

ConsumeKafka/publishKafka

重点 JSON/XML/CSV 与 AVRO 文件格式的转换

HBase Client 的配置

SETTINGS	SCHEDULING	PROPERTIES	COMMENTS
Required field			+
Property		Value	
HBase Client Service	?	HBase_1_1_2_ClientService	→
Table Name	?	dxftest	
Row Identifier	?	\$(uid)	
Row Identifier Field Name	?	No value set	
Row Identifier Encoding Strategy	?	String	
Column Family	?	name	
Timestamp	?	No value set	
Batch Size	?	25	
Complex Field Strategy	?	Text	
Field Encoding Strategy	?	String	

CANCEL APPLY

SETTINGS PROPERTIES COMMENTS

Required field +

Property		Value
Hadoop Configuration Files		/opt/conf.cloudera.hbase/hbase-site.xml
Kerberos Principal		root/admin
Kerberos Keytab		/root/root.keytab
ZooKeeper Quorum		No value set
ZooKeeper Client Port		No value set
ZooKeeper ZNode Parent		No value set
HBase Client Retries		1
Phoenix Client JAR Location		No value set

CANCEL APPLY