



Jobs

Generated by Talend Data Fabric

Project Name	zach-talend	GENERATION DATE	Feb 5, 2019 12:10:58 PM
AUTHOR	Lorenzini, Zach	Talend Data Fabric VERSION	7.1.1.20181026_1147

Summary

[Project Description](#)

[Description](#)

[Preview Picture](#)

[Settings](#)

[Context List](#)

[Component List](#)

[Components Description](#)

Project Description

Properties	Values
Name	zach-talend
Language	java
Description	

Description

Properties	Values
Name	To_Hive_Demo
Author	Lorenzini, Zach
Version	0.1
Purpose	
Status	
Description	
Creation	Nov 30, 2018 10:22:02 AM
Modification	Nov 30, 2018 11:27:24 AM

Preview Picture



Settings

Extra settings

Name	Value
COMP_DEFAULT_FILE_DIR	
Multi thread execution	false
Implicit tContextLoad	false

Status & Logs

Name	Value
Use statistics (tStatCatcher)	false
Use logs (tLogCatcher)	false
Use volumetrics (tFlowMeterCatcher)	false
On Console	false
On Files	false
On Databases	false
Catch components statistics	false
Catch runtime errors	true

Catch user errors	true
Catch user warnings	true

Context List

ContextDefault


Name	Prompt	Need Prompt?	Type	Value	Source
db_connection	db_connection?	false	id_String	"db.microsoft.com"	
db_port	db_port?	false	id_String	"1433"	

Component List

Component Name	Component Type
tHiveConnection_1	tHiveConnection
tHiveCreateTable_1	tHiveCreateTable
tRunJob_1	tRunJob

Components Description

Component tHiveConnection

	UNIQUE NAME	tHiveConnection_1	INPUT(S)	none

	LABEL	__UNIQUE_NAME__	OUTPUT(S)	tRunJob_1
--	-------	-----------------	-----------	-----------

Component Parameters:

Properties	Values
Unique Name	tHiveConnection_1
Component Name	tHiveConnection
Version	0.102 (ALPHA)
Family	Databases/DB Specifics/Hive ELT/Connections Big Data/Hive
Start	true
Startable	true
SUBTREE_START	true
END_OF_FLOW	true
Activate	true
DUMMY	false
tStatCatcher Statistics	false
Help	org.talend.help.tHiveConnection
Update components	true
IREPORT_PATH	
JAVA_LIBRARY_PATH	C:\Program Files (x86)\Talend-Studio\studio\configuration\lib\java
Subjob color	
Title color	
Enable parallel execution	false
Number of parallel executions	2

PARALLELIZE_KEEP_EMPTY	false
Property Type	Built-In
!!!HIVE.NAME!!!	
Distribution	MAPR
Version	MAPR600
!!!HIVE.NAME!!!	
!!!HADOOP_CUSTOM_JARS.NAME!!!	
Use Yarn	false
Activate the cross platform application submission	false
Connection mode	STANDALONE
Hive Server	HIVE2
Host	"zmapr1"
!!!TYPE.NAME!!!	Hive
Port	"10000"
Database	"default"
Username	"mapr"
Password	*****
Additional JDBC Settings	""
Inspect the classpath for configurations	false
Use kerberos authentication	false
Hive principal	"hive/_HOST@EXAMPLE.COM"
HiveServer2 local user principal	"/tmp/hiveServer2User.keytab"

HiveServer2 local user keytab	"/tmp/hiveServer2.keytab"
Namenode principal	"nn/_HOST@EXAMPLE.COM"
JobTracker principal	"jt/_HOST@EXAMPLE.COM"
Resource manager principal	"yarn/_HOST@EXAMPLE.COM"
Metastore URL	"jdbc:mysql://"
Driver class	"com.mysql.jdbc.Driver"
Username	"hive"
Password	*****
Use a keytab to authenticate	false
Principal	"hive"
Keytab	"/tmp/hive.keytab"
Use MapR Ticket authentication	false
Username	""
Password	**
Cluster name	"demo.mapr.com"
Ticket duration (in s)	86400L
Use SSL encryption	false
Trust store path	"/home/usr1/ssl/trust_store.jks"
Trust store password	*****
Set Jobtracker URI	true
MAPRED_JT	"localhost:50300"
Set Resource Manager	false


RESOURCE_MANAGER	"localhost:8032"
Set Namenode URI	false
FS_DEFAULT_NAME	"hdfs://localhost:8020"
Set resourcemanager scheduler address	false
RESOURCEMANAGER_SCHEDULER_ADDRESS	"localhost:8033"
Set jobhistory address	false
JOBHISTORY_ADDRESS	"localhost:10020"
Set Hadoop User	false
HADOOP_USER	"hadoop"
Use datanode hostname	false
Hostname	""
Port	"443"
Username	""
Job result folder	""
Username	""
Password	**
Hostname	""
Container	""
Username	""
Password	**
Deployment Blob	""
Project identifier	"my-google-project"

Cluster identifier	"my-cluster-id"
Region	"global"
Google Storage staging bucket	"gs://my-bucket/talend/jars"
Provide Google Credentials in file	false
Path to Google Credentials file	""
Use or register a shared DB Connection	false
Shared DB Connection Name	
Execution engine	mr
Store by HBase	false
Zookeeper quorum	"localhost"
Zookeeper client port	"2181"
Set Zookeeper znode parent	false
ZNODE_PARENT	""
HBase master principal	"hbase/_HOST@EXAMPLE.COM"
HBase regionserver principal	"hbase/_HOST@EXAMPLE.COM"
Define the jars to register for HBase	false
Register jar for HBase	[]
API token	**
Cluster label	false
QUBOLE_CLUSTER_LABEL	""
Change API endpoint	false
QUBOLE_ENDPOINT_URL	"https://api.qubole.com/api"

Tez lib	INSTALL
Install folder in HDFS	"/tmp/{USERNAME}/talend_tez_libs/{custom HIVE_VERSION}"
Group	"supergroup"
!!!TEZ_JARS_NAME.NAME!!!	avro-1.5.4.jar,commons-cli-1.2.jar,commons-collections4-4.0.jar,guava-11.0.2.jar,guice-3.0.jar, had
Lib path(folder or file)	""
Lib jars	[]
Hadoop Properties	[]
Hive Properties	[]
Mapred job map memory mb	"1000"
Mapred job reduce memory mb	"1000"
Set memory	false
Map (in Mb)	"1024"
Reduce (in Mb)	"1024"
ApplicationMaster (in Mb)	"1024"
Path separator in server	":"
Set the MapR home directory	false
MAPR_HOME_DIR	"/opt/mapr"
Specify the hadoop login configuration	false
HADOOP_LOGIN	"kerberos"
Label format	__UNIQUE_NAME__
Hint format	__UNIQUE_NAME__ __COMMENT__
Connection format	row

Show Information	false
Comment	
Use an existing validation rule	false
Validation Rule Type	

Original Function Parameters:
Component tHiveCreateTable

	UNIQUE NAME	tHiveCreateTable_1	INPUT(S)	tRunJob_1
	LABEL	__UNIQUE_NAME__	OUTPUT(S)	none

Component Parameters:

Properties	Values
Unique Name	tHiveCreateTable_1
Component Name	tHiveCreateTable
Version	0.102 (ALPHA)
Family	Databases/DB Specifics/Hive Big Data/Hive
Start	false
Startable	true
SUBTREE_START	true
END_OF_FLOW	true
Activate	true
DUMMY	false
tStatCatcher Statistics	false

Help	org.talend.help.tHiveCreateTable
Update components	true
IREPORT_PATH	
JAVA_LIBRARY_PATH	C:\Program Files (x86)\Talend-Studio\studio\configuration\lib\java
Subjob color	
Title color	
Enable parallel execution	false
Number of parallel executions	2
PARALLELIZE_KEEP_EMPTY	false
Property Type	Built-In
Use an existing connection	true
Component List	tHiveConnection_1
!!!HIVE.NAME!!!	
Distribution	MAPR
Version	MAPR600
!!!HIVE.NAME!!!	
!!!HADOOP_CUSTOM_JARS.NAME!!!	
Use Yarn	false
Activate the cross platform application submission	false
Connection mode	STANDALONE
Hive Server	HIVE2
Host	"zmapr1"

Port	"10000"
Database	"default"
!!!TYPE.NAME!!!	Hive
Username	"mapr"
Password	*****
Additional JDBC Settings	""
Use kerberos authentication	false
Hive principal	"hive/_HOST@EXAMPLE.COM"
HiveServer2 local user principal	"hiveServer2User@EXAMPLE.COM"
HiveServer2 local user keytab	"/tmp/hiveServer2User.keytab"
Namenode principal	"nn/_HOST@EXAMPLE.COM"
JobTracker principal	"jt/_HOST@EXAMPLE.COM"
Resource manager principal	"yarn/_HOST@EXAMPLE.COM"
Metastore URL	"jdbc:mysql://"
Driver class	"com.mysql.jdbc.Driver"
Username	"hive"
Password	*****
Use a keytab to authenticate	false
Principal	"hive"
Keytab	"/tmp/hive.keytab"
Use MapR Ticket authentication	false
Username	""

Password	**
Cluster name	"demo.mapr.com"
Ticket duration	86400L
Use SSL encryption	false
Trust store path	"/home/usr1/ssl/trust_store.jks"
Trust store password	*****
Set Jobtracker URI	false
MAPRED_JT	"localhost:50300"
Set Resource Manager	false
RESOURCE_MANAGER	"localhost:8032"
Set Namenode URI	false
FS_DEFAULT_NAME	"hdfs://localhost:8020"
Set resourcemanager scheduler address	false
RESOURCEMANAGER_SCHEDULER_ADDRESS	"localhost:8025"
Set jobhistory address	false
JOBHISTORY_ADDRESS	"localhost:10020"
Set Hadoop User	false
HADOOP_USER	"hadoop"
Use datanode hostname	true
Hostname	""
Port	"443"
Username	""

Job result folder	""
Username	""
Password	**
Hostname	""
Container	""
Username	""
Password	**
Deployment Blob	""
Project identifier	"my-google-project"
Cluster identifier	"my-cluster-id"
Region	"global"
Google Storage staging bucket	"gs://my-bucket/talend/jars"
Provide Google Credentials in file	false
Path to Google Credentials file	""
API token	**
Cluster label	false
QUBOLE_CLUSTER_LABEL	""
Change API endpoint	false
QUBOLE_ENDPOINT_URL	"https://api.qubole.com/api"
REPOSITORY_ALLOW_AUTO_SWITCH	True
Schema	Built-In
Table Name	"customers"

Action on table	CREATE_IF_NOT_EXIST
Format	TEXTFILE
Storage class	""
Inputformat class	""
Outputformat class	""
Set partitions	false
Schema	Built-In
Set file location	true
Use S3 endpoint	false
URI path	"/user/mapr/customers/"
S3 Bucket	""
Access key	""
Secret key	**
Mapping	hive_id
Set Delimited row format	true
Field	true
FIELD_TERMINATED_BY	"_," ,
Escape	false
FIELD_ESCAPE_BY	""
Collection Item	false
COLLECTION_ITEM_TERMINATED_BY	"_," ,
Map Key	false

MAP_KEY_TERMINATED_BY	"\u0003"
Line	false
LINES_TERMINATED_BY	"\n"
Set Serde row format	false
Serde class name	""
Die on error	false
Like table	false
LIKE_TABLE_NAME	""
Create an external table	false
Table comment	""
As select	false
SELECT	"select id, name from employee"
Set clustered_by or skewed_by statement	false
CLUSTERED_BY_AND_SKEWED_BY	"CLUSTERED BY (col_name, col_name, ...) [SORTED BY (col_name ASC DESC), ...] INTO num_buckets BUCKETS] SKEWED BY (col_name, col_name, ...) ON (((col_value, col_value, ...), ... col_value, col_value, ...)) "
Serde Properties	[]
Table properties	[]
Set temporary path	false
TEMP_PATH	""
Hadoop Properties	[]
Hive Properties	[]
Mapred job map memory mb	"1000"


Mapred job reduce memory mb	"1000"
Set memory	false
Map (in Mb)	"1024"
Reduce (in Mb)	"1024"
ApplicationMaster (in Mb)	"1024"
Path separator in server	":"
Set the MapR home directory	false
MAPR_HOME_DIR	"/opt/mapr"
Specify the hadoop login configuration	false
HADOOP_LOGIN	"kerberos"
Label format	__UNIQUE_NAME__
Hint format	__UNIQUE_NAME__ __COMMENT__
Connection format	row
Show Information	false
Comment	
Use an existing validation rule	false
Validation Rule Type	

Schema for tHiveCreateTable_1 :

Column	Key	Type	Length	Precision	Nullable	Comment
custid	false	Integer			true	
lastname	false	String			true	

firstname	false	String			true	
-----------	-------	--------	--	--	------	--

Original Function Parameters: Component tRunJob

	UNIQUE NAME	tRunJob_1	INPUT(S)	tHiveConnection_1
	LABEL	__UNIQUE_NAME__	OUTPUT(S)	tHiveCreateTable_1

Component Parameters:

Properties	Values
Unique Name	tRunJob_1
Component Name	tRunJob
Version	0.101 (ALPHA)
Family	System Orchestration
Start	false
Startable	true
SUBTREE_START	true
END_OF_FLOW	true
Activate	true
DUMMY	false
tStatCatcher Statistics	false
Help	org.talend.help.tRunJob
Update components	true
IREPORT_PATH	

JAVA_LIBRARY_PATH	C:\Program Files (x86)\Talend-Studio\studio\configuration\lib\java
Subjob color	
Title color	
Enable parallel execution	false
Number of parallel executions	2
PARALLELIZE_KEEP_EMPTY	false
REPOSITORY_ALLOW_AUTO_SWITCH	true
Schema	Built-In
Use dynamic job	false
Context job	
Job	To_Hive_Demo_CF
Context	Default
Use an independent process to run subjob	false
Die on child error	true
Transmit whole context	false
Context Param	[]
Propagate the child result to the output schema	false
Print Parameters	false
Transmit original context value(if unchecked, there will convert context value to string type before transmitting.)	true
Label format	__UNIQUE_NAME__
Hint format	__UNIQUE_NAME__ __COMMENT__

Connection format	row
Show Information	false
Comment	
Use an existing validation rule	false
Validation Rule Type	

Schema for tRunJob_1 :

Column	Key	Type	Length	Precision	Nullable	Comment
--------	-----	------	--------	-----------	----------	---------

Original Function Parameters: