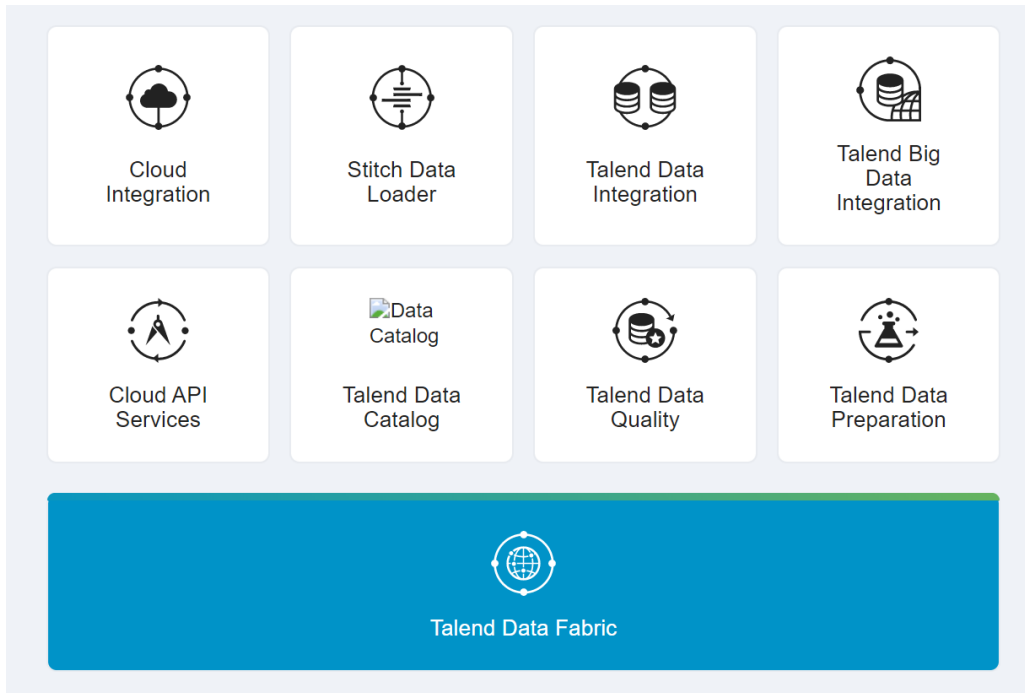


Workshop



- Talend Data Integration



Workshop

- Import sql server database: info7370_input
- Create info7370_output database for:
 - Microsoft SQL Server
 - MySQL
 - PostgreSQL
 - Oracle database
- Create and load Chinook database for:
 - Microsoft SQL Server
 - MySQL
 - PostgreSQL
 - Oracle database

dbName	SchemaName	TableName	Table_Rows
info7370_input	dbo	BuyingGroups	2
info7370_input	dbo	Cities	37,940
info7370_input	dbo	ColdRoomTemperatures	33,510,516
info7370_input	dbo	Colors	36
info7370_input	dbo	Countries	190
info7370_input	dbo	CustomerCategories	8
info7370_input	dbo	Customers	719
info7370_input	dbo	CustomerTransactions	254,256
info7370_input	dbo	DeliveryMethods	10
info7370_input	dbo	OrderLines	588,331
info7370_input	dbo	Orders	192,244
info7370_input	dbo	PackageTypes	14
info7370_input	dbo	PaymentMethods	4
info7370_input	dbo	People	1,167
info7370_input	dbo	SpecialDeals	2
info7370_input	dbo	StateProvinces	53
info7370_input	dbo	StockGroups	10
info7370_input	dbo	StockItemHoldings	227
info7370_input	dbo	StockItems	227
info7370_input	dbo	StockItemStockGroups	442
info7370_input	dbo	StockItemTransactions	598,736
info7370_input	dbo	SupplierCategories	9
info7370_input	dbo	Suppliers	13
info7370_input	dbo	SupplierTransactions	5,413
info7370_input	dbo	TransactionTypes	13

Data Integration Tasks - Introduction

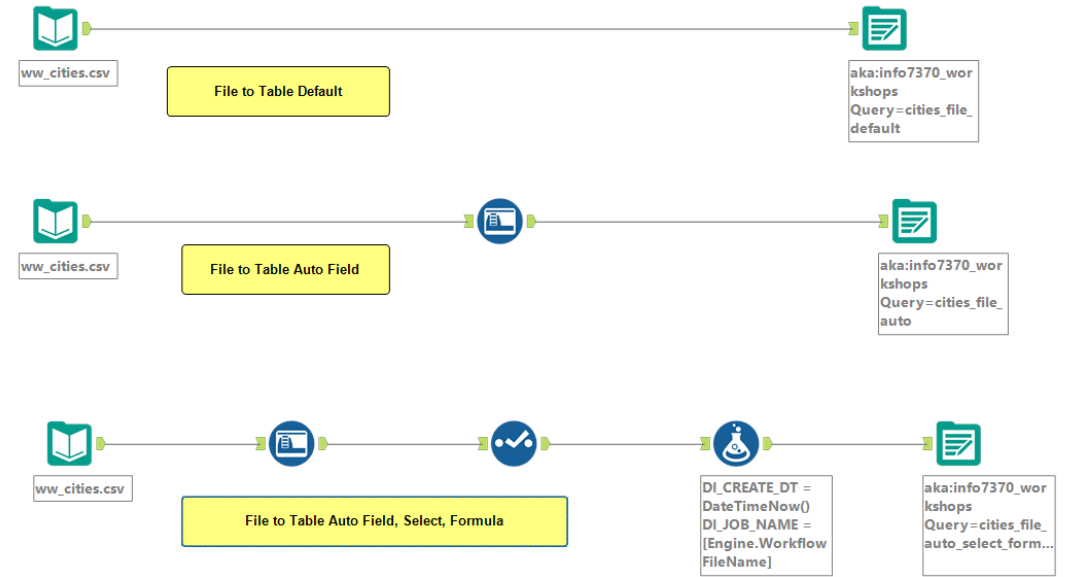
- Read data source(s) & write to target(s)
 - Relational tables/views
 - csv files
 - Excel
- Data Source database(s) – “Chinook” on various dbms
 - SQL Server, MySQL, Oracle, PostgreSQL
- Data Target database(s) – “info7370_output” on various dbms
 - SQL Server, MySQL, Oracle, PostgreSQL
- Primary data integration activity read data, filter data, join and/or lookup data and write
 - ER data model to “flattened” or denormalized files

Talend

- DI_Create_DT, DateTime
- DI_Job_Name, varchar(20)

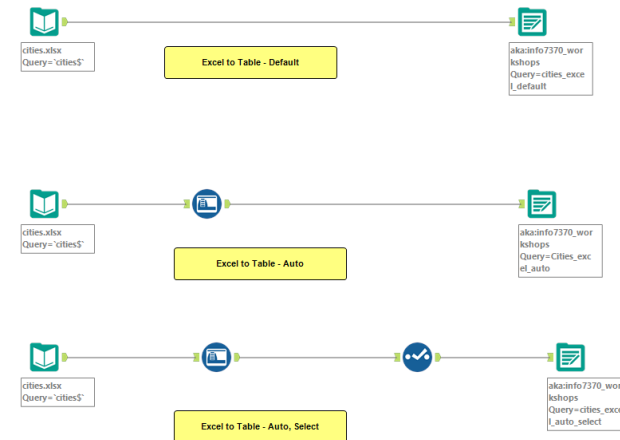
Column	DataType	Talend
di_tool	nvarchar(40)	Talend
di_job_id	nvarchar(40)	
di_job_name	nvarchar(40)	global.jobName
di_create_date	Datetime	getdate()

Alteryx: cities_file_input.yxmd

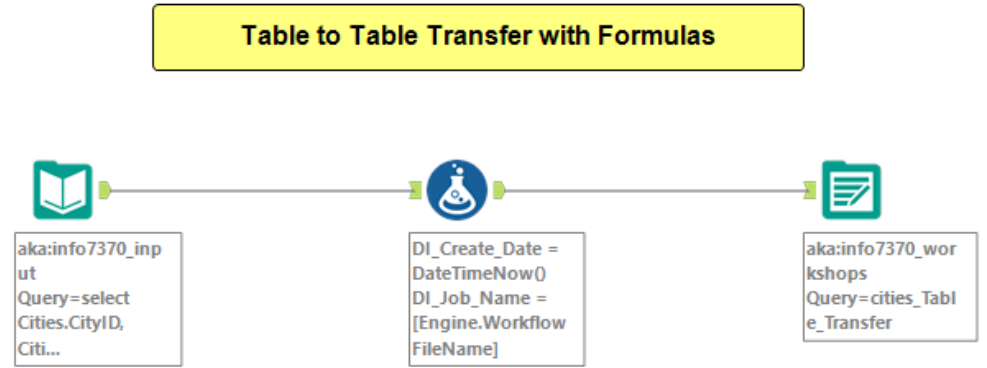


Alteryx: cities_Excel_input.yxmd

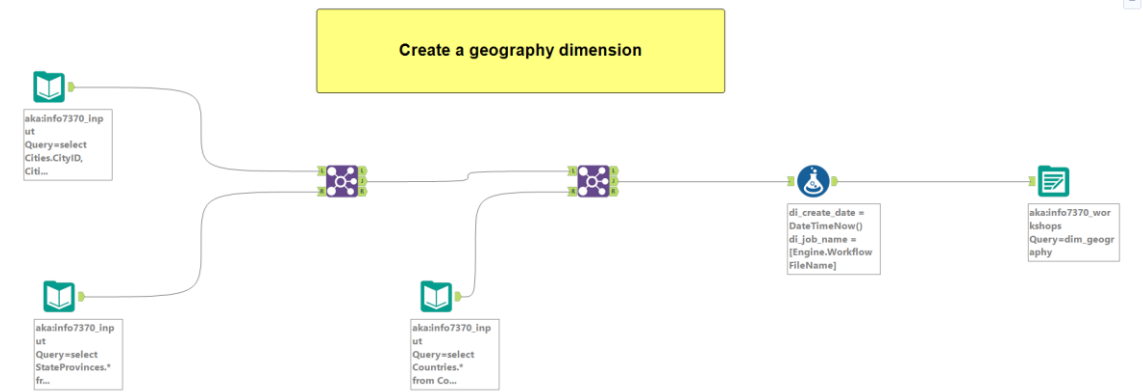
Start Here.yxmd* x customer_in-database_stream_out.yxmd x www_customer_orders_Autosave* x customer_in-database.yxmd* x ww_countries.yxmd x countries_table_to_Table.yxmd x cities_file_input.yxmd* x cities_Excel_input.yxmd x



Alteryx: cities_Table_input.yxmd



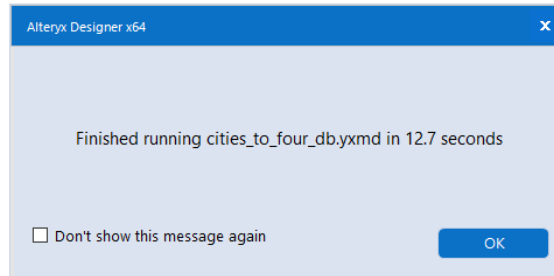
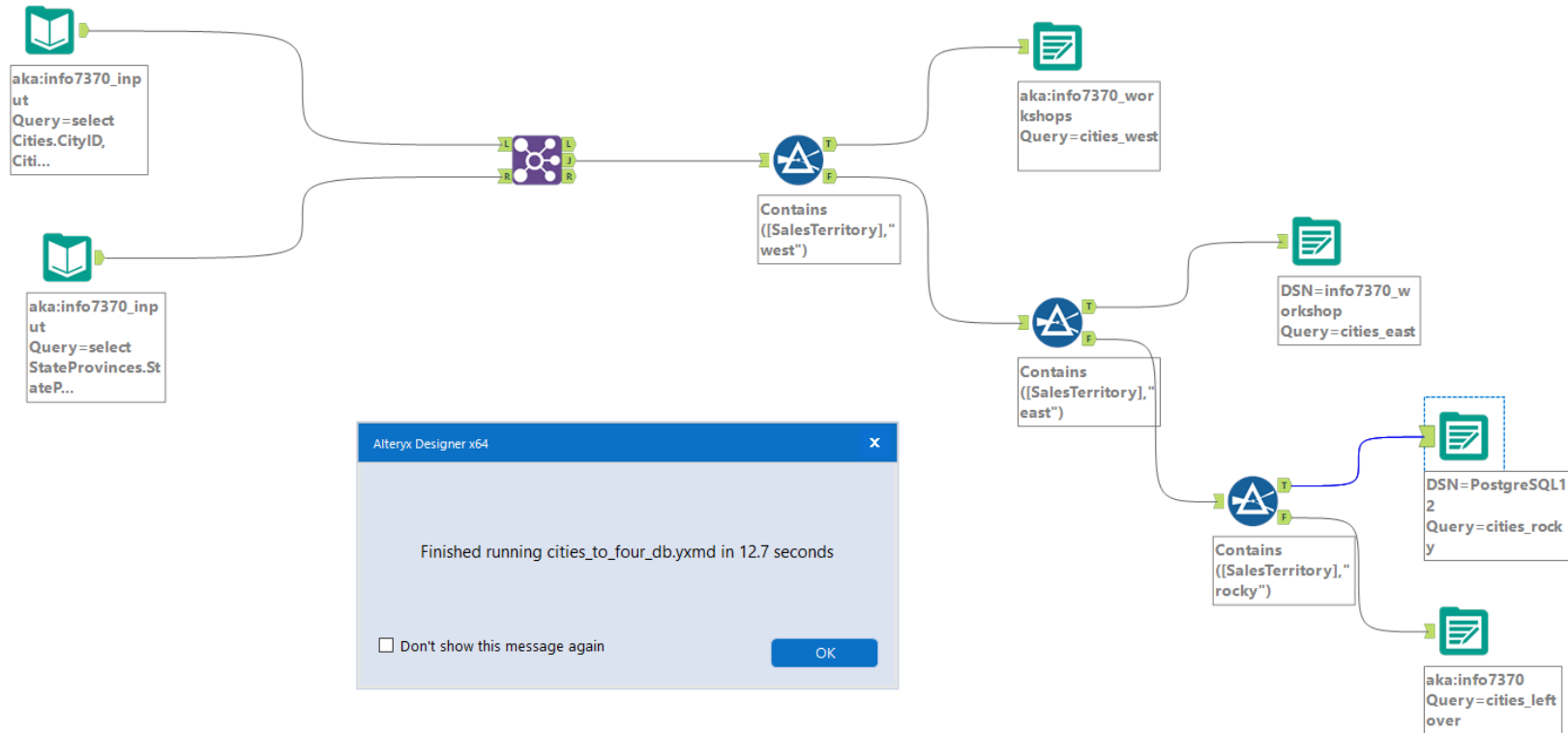
Alteryx: create_dim_geography_workshop.yxmd



Data Preparation

More examples

Alteryx: cities_to_four_db.yxmd



Output Data (11) - Input

3 of 8 Fields | Cell Viewer | 2,213 records displayed

Record	CityID	CityName	StateProvinceID	Right_StateProvinceID	StateProvinceCode	StateProvinceName	CountryID	SalesTerritory
1	110	Adams City	6	6	CO	Colorado	230	Rocky Mountain
2	184	Agate	6	6	CO	Colorado	230	Rocky Mountain
3	214	Aguilar	6	6	CO	Colorado	230	Rocky Mountain

Data Integration

More

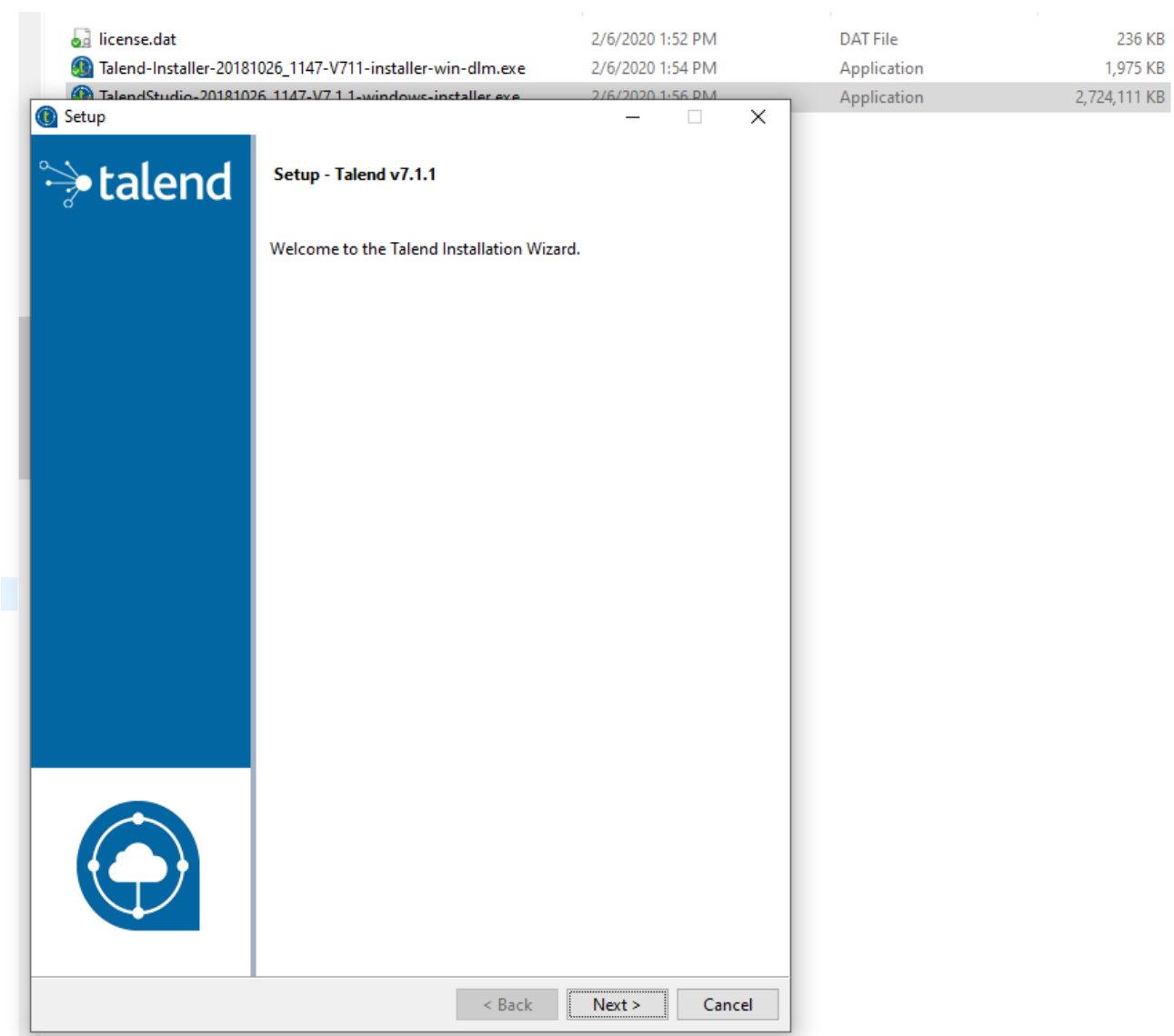
Installation & Setup

Talend

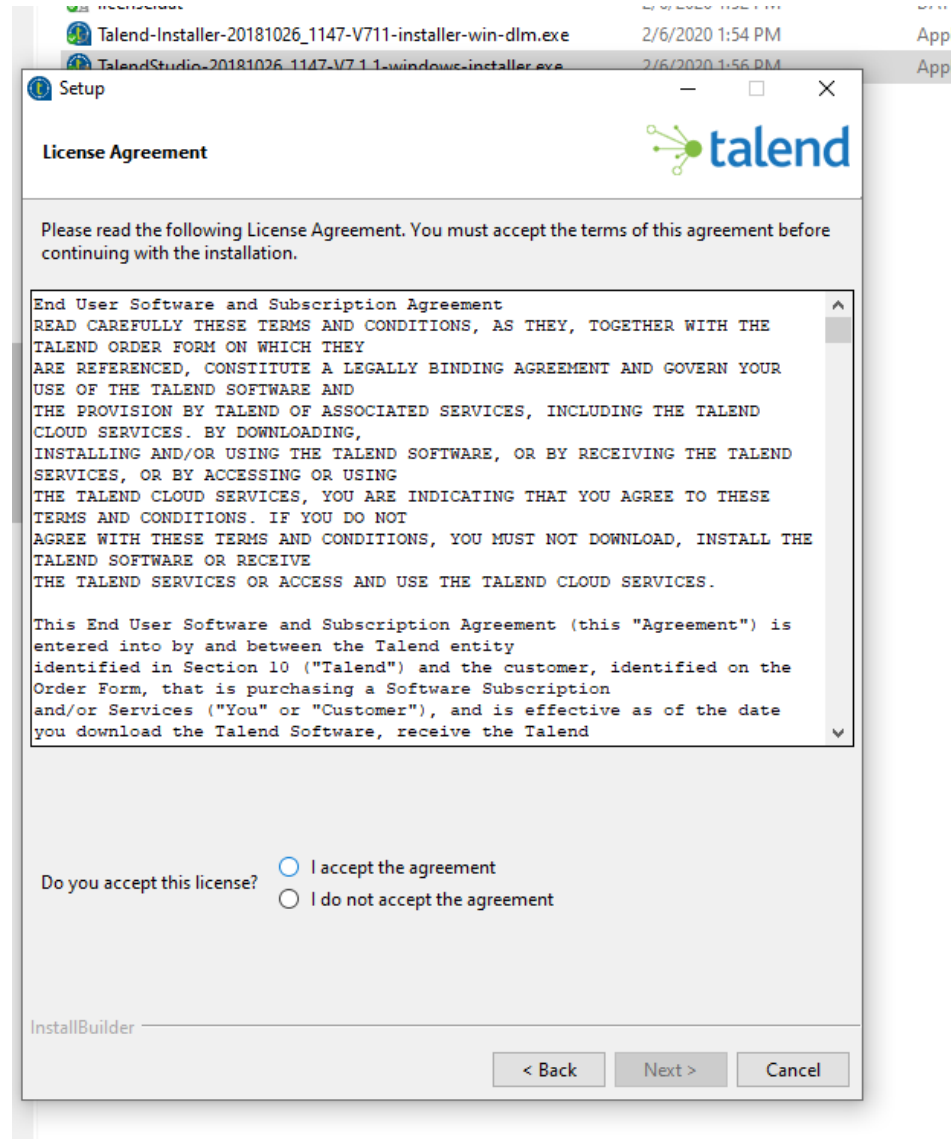
Talend Installation & Setup

- Installing software - assumption
- Modify Studio configuration
 - C:\Talend_DI-V7.1.1_2020 - Primary\studio\Talend-Studio-win-x86_64.ini
- Starting up Talend & installing required 3rd party
- Setup Show view: modules
- Ensuring jdbc & jars setup for databases
- Import archive file: workshop_prep.zip
 - Sample jobs
 - Sample metadata db connections
- Run jobs to validate above setup

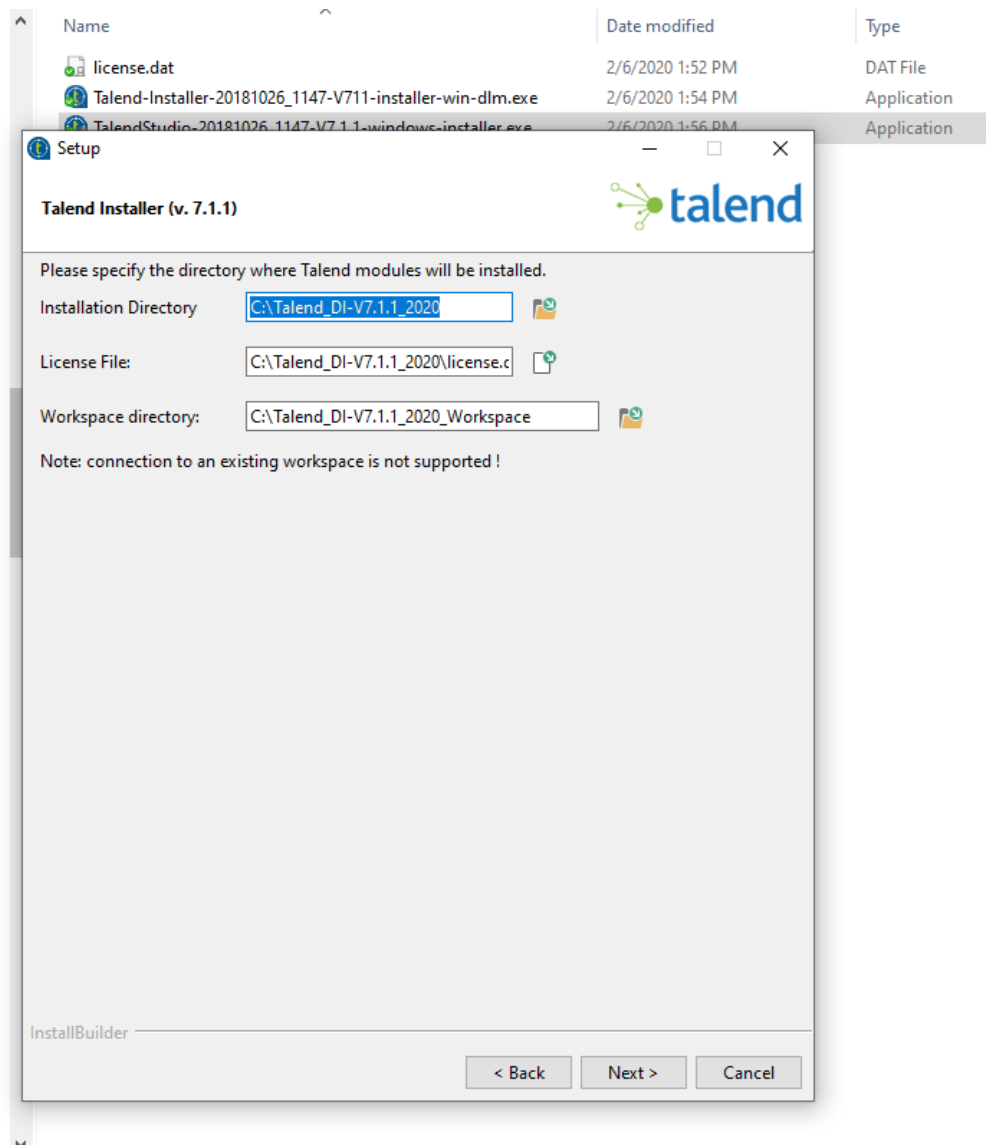
Installation



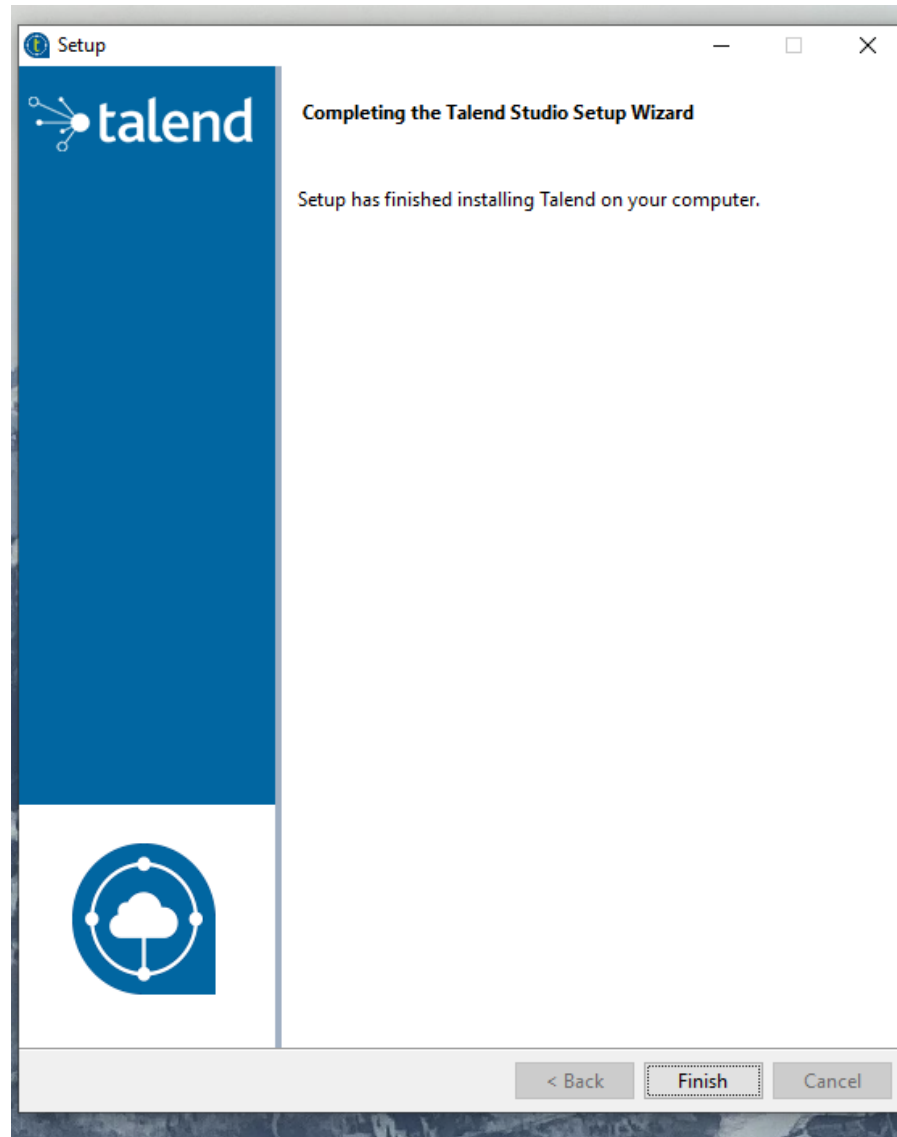
Installation



Installation



Installation




Change Studio Maximum Memory

C:\Talend_DI-V7.1.1_2020 - Primary\studio\Talend-Studio-win-x86_64.ini

Change this to:

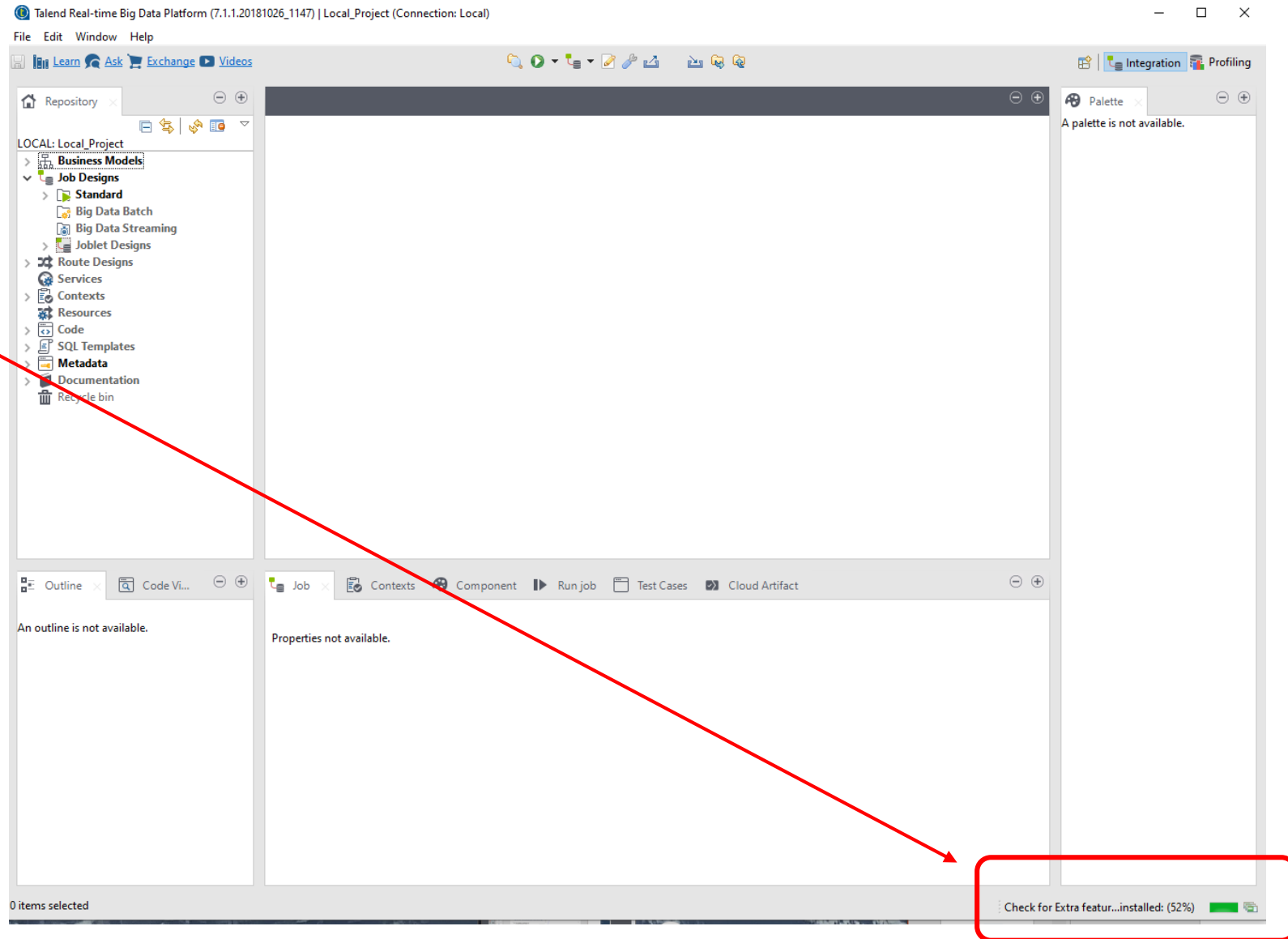
notebook	-Xmx
8G	2G
16G	4G
32G	8G



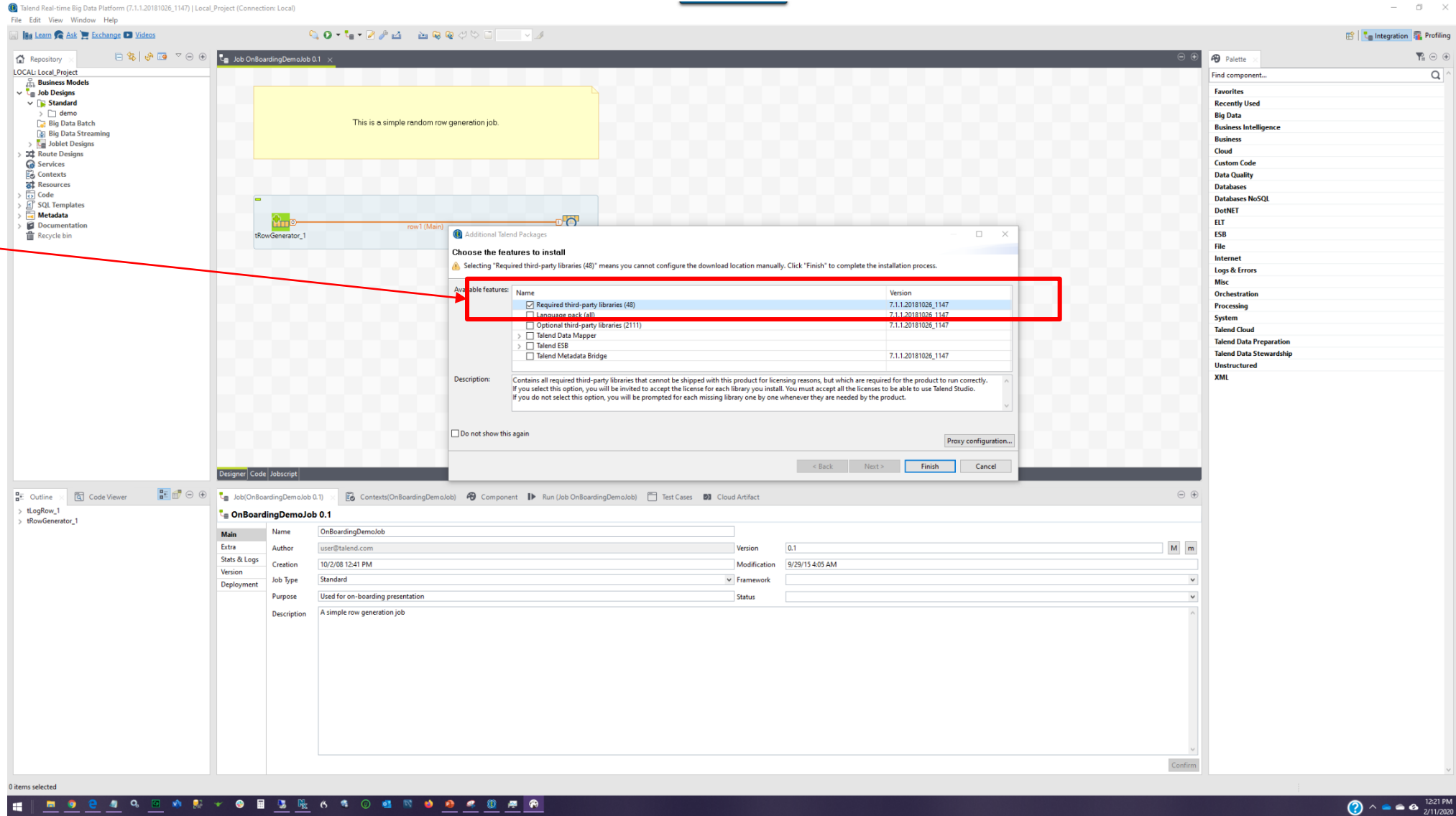
```
Talend-Studio-win-x86_64 - Copy.ini - Notepad
File Edit Format View Help
-vm
C:\Talend_DI-V7.1.1_2020\jre1.8.0_131\bin\server\jvm.dll
-vmargs
-Xms512m
-Xmx1536m
-Dfile.encoding=UTF-8
-Dosgi.requiredJavaVersion=1.8
-XX:+UseG1GC
-XX:+UseStringDeduplication
```

Talend - Setup

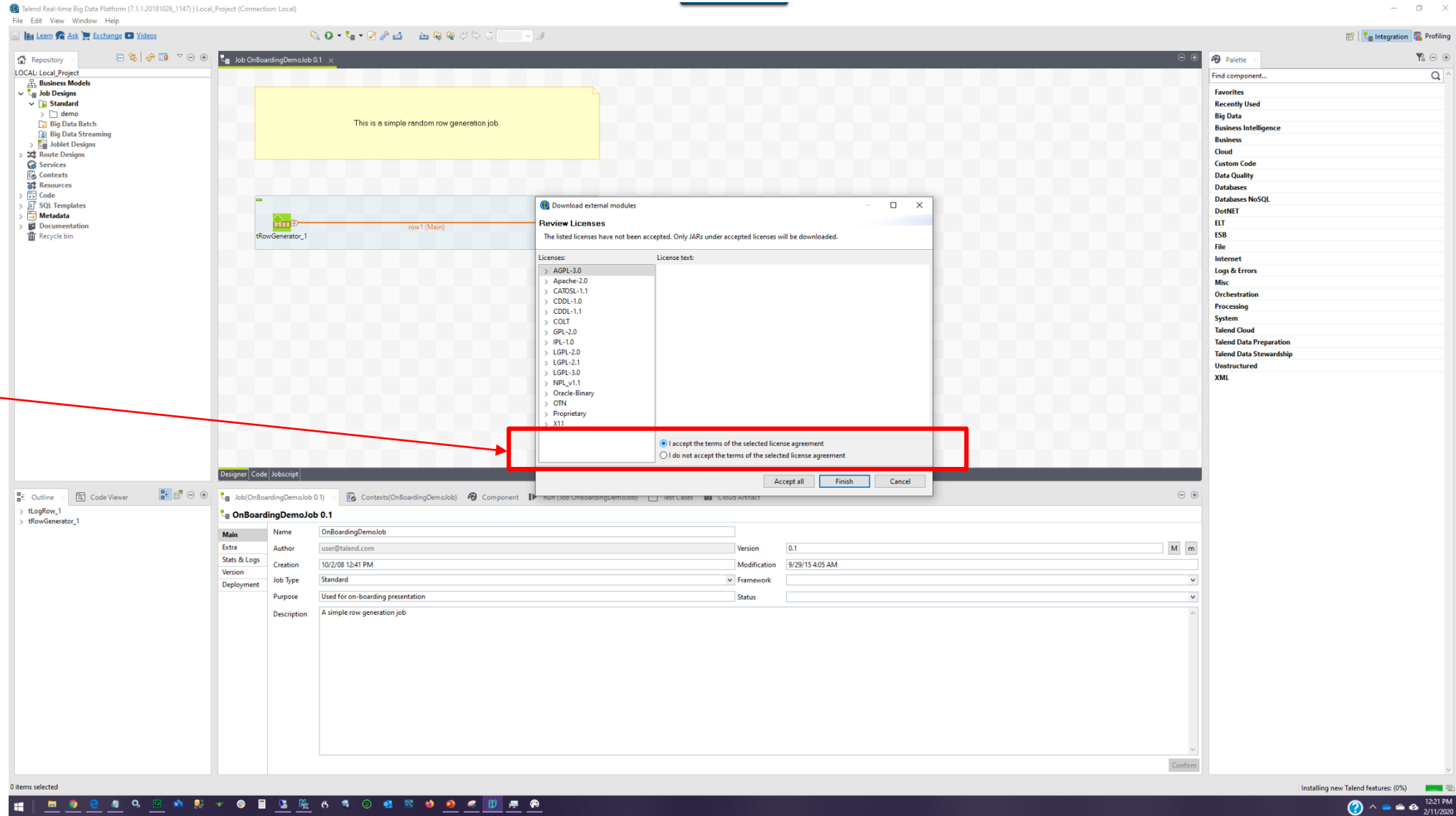
When 1st start
Talend need to
let this finish
before doing
anything



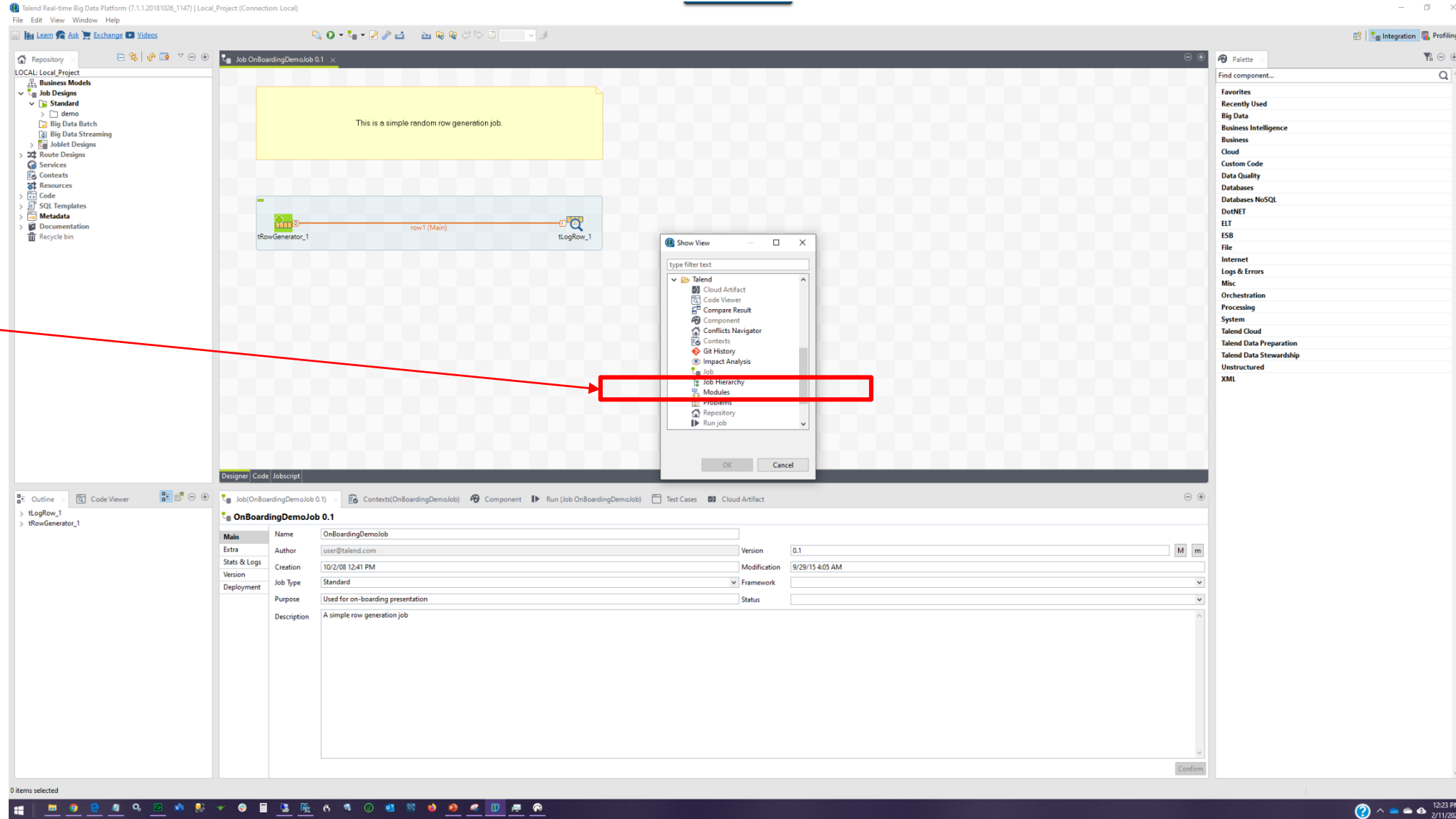
Select Required third party libraries



“I will accept...” & then Accept All button

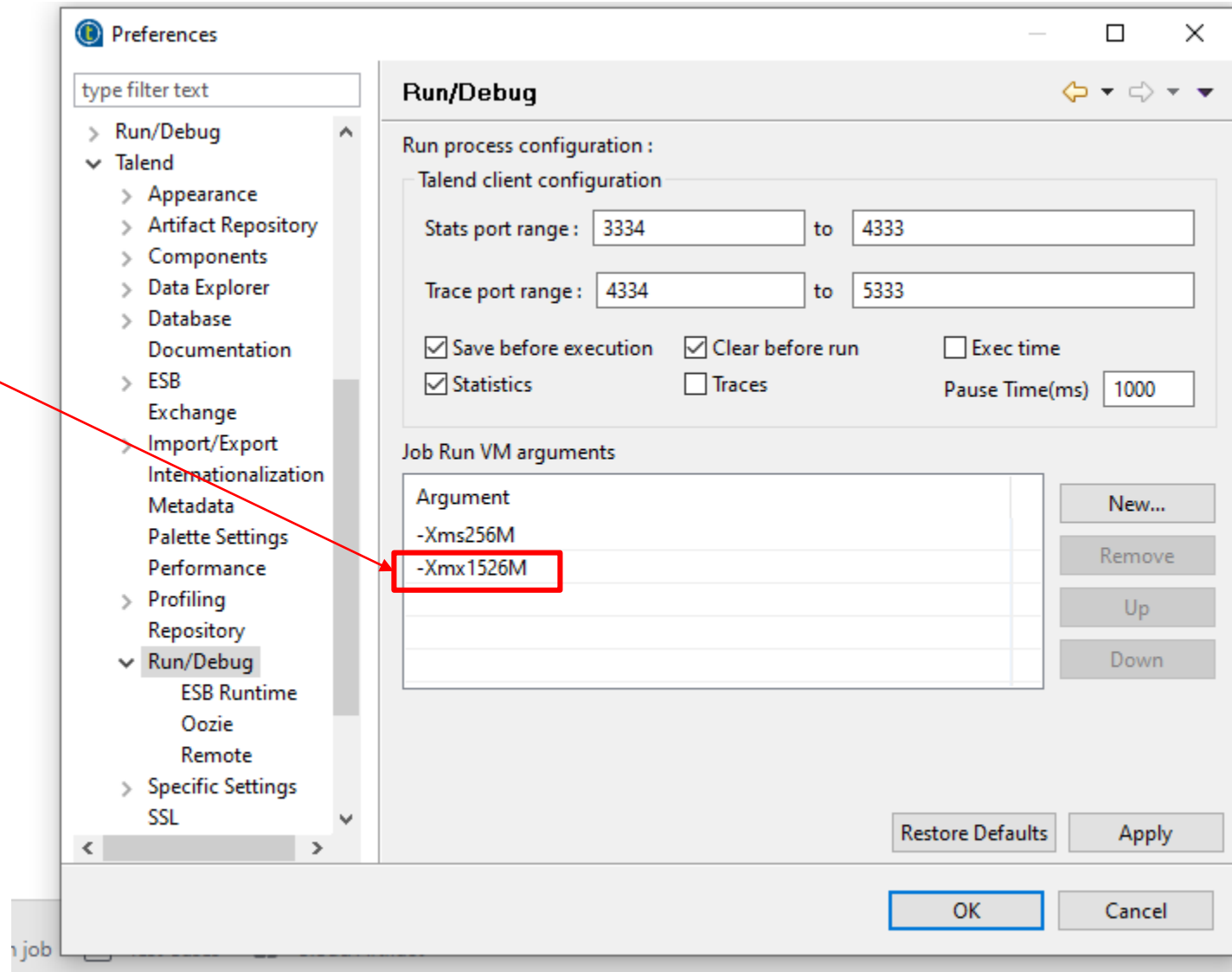


Windows, Show Views, Select Talend-Modules

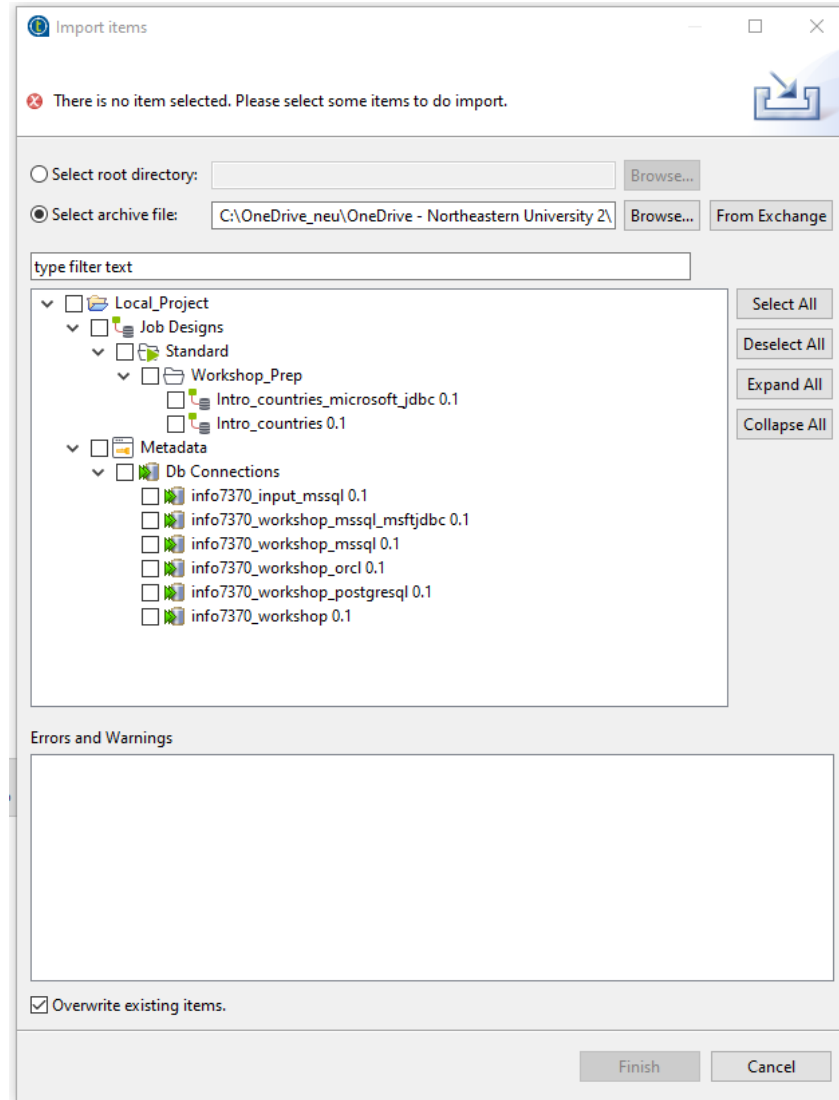


Change Run/Debug Maximum Memory

Change this to:
notebook -Xmx
8G 3G
16G 8G
32G 12G



Import Archive File



- Sample archive file
 - Jobs to test connections to our workshop databases
 - DB Connections

Run sample jobs

Talend Real-time Big Data Platform (7.1.1.20181026_1147) | Local_Project (Connection: Local)

File Edit View Window Help

Learn Ask Exchange Videos

Repository LOCAL: Local_Project

- Business Models
- Job Designs
 - Standard
 - demo
 - Workshop_01
 - Workshop_Preparation
 - Intro_countries_microsoft_jdbc 0.1
 - Intro_countries 0.1
 - Big Data Batch
 - Big Data Streaming
 - Joblet Designs
 - Route Designs
 - Services
 - Contexts
 - Resources
 - Code
 - SQL Templates
 - Metadata
 - Db Connections
 - info7370_input_mssql 0.1
 - info7370_workshop_input_msfjdbc 0.1
 - info7370_workshop_mssql_msfjdbc 0.1
 - info7370_workshop_mssql 0.1
 - info7370_workshop_orcl 0.1
 - info7370_workshop_postgresql_port5432 0.1
 - info7370_workshop_postgresql 0.1
 - info7370_workshop 0.1
 - REST API Definitions
 - SAP Connections
 - File delimited
 - File positional
 - File regex
 - File xml
 - File Excel
 - File Idif
 - File Json
 - LDAP
 - Azure Storage
 - Data Stewardship
 - Google Drive

Job Intro_countries_microsoft_jdbc 0.1 Job Intro_countries 0.1

Simple copy data to different databases

190 rows in 0.75s
253.33 rows/s
countries (Main)

tMap_2

190 rows in 3.94s
48.26 rows/s
put_mssql (Main order:1)

190 rows in 2.35s
80.89 rows/s
put_orcl (Main order:2)

190 rows in 1.42s
134.28 rows/s
put_post1 (Main order:3)

190 rows in 1.18s
160.61 rows/s
put_mysql (Main order:4)

"Countries_Copy_Limited"

"country_copy"

"country_copy"

Copy data to Microsoft SQL Server using Microsoft jdbc

"Countries"

get_countries (Main)

tMap_1

mssql (Main)

"country_copy"

Designer Code Jobscrip

Outline Code Viewer

- tDBInput_1 ("Countries")
- tDBInput_2 ("Countries")
- tDBOutput_1 ("country_copy")
- tDBOutput_3 ("country_copy")
- tDBOutput_4 ("Countries_Copy_Limited")
- tDBOutput_5 ("country_copy")
- tDBOutput_6 ("country_copy")
- tMap_1
- tMap_2

Job(Intro_countries 0.1) Contexts(Intro_countries) Component Run (Job Intro_countries) Test Cases Cloud Artifact

Job Intro_countries

Basic Run Debug Run Advanced settings Target Exec Memory Run

Execution

Run Kill Clear

Starting job Intro_countries at 16:25 12/02/2020.

[statistics] connecting to socket on port 3995
[statistics] connected
[statistics] disconnected

Job Intro_countries ended at 16:25 12/02/2020. [exit code=0]