Workshop

talend

- Talend Data Integration

Cloud Integration · Stitch Data Loader · Talend Data Integration · Talend Big Data Integration

Cloud API Services · Talend Data Catalog · Talend Data Quality · Talend Data Preparation

Talend Data Fabric

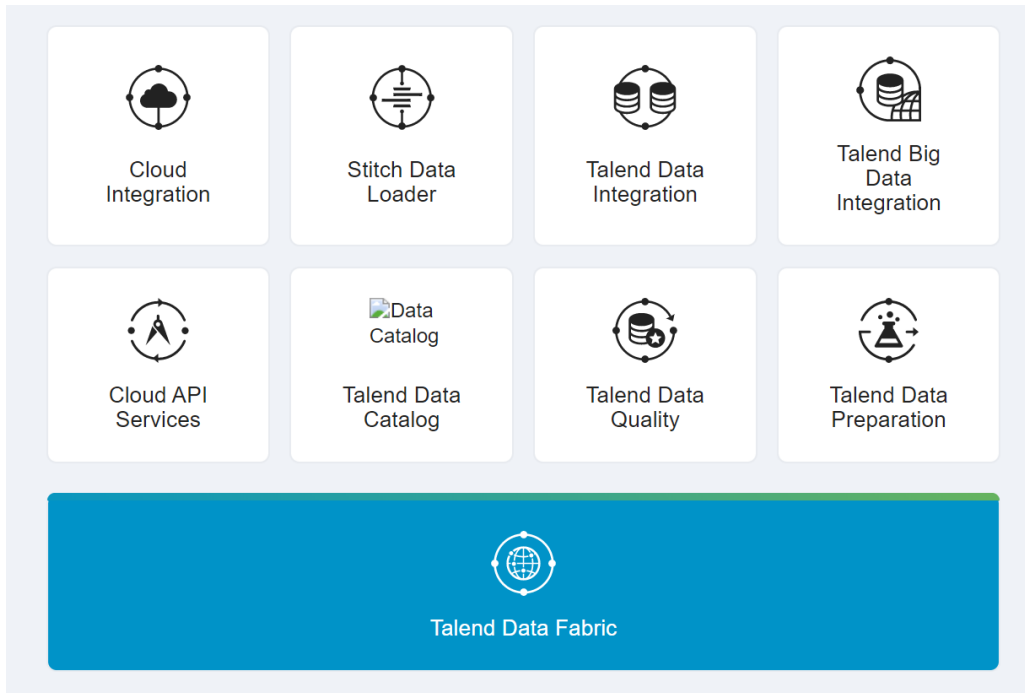ATHENA
IT SOLUTIONS

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

ATHENA
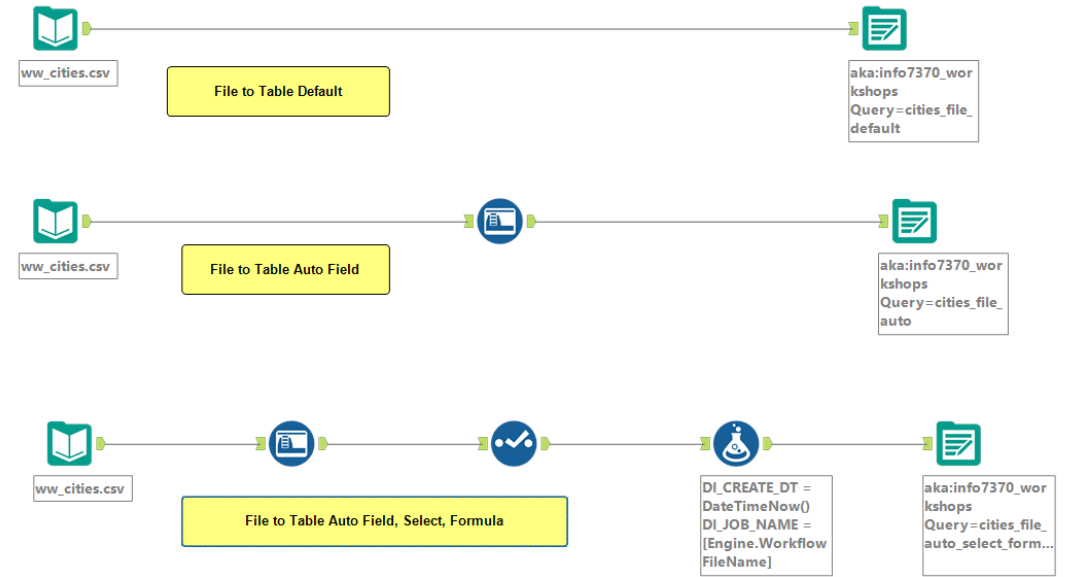IT SOLUTIONS

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

ATHENA
IT SOLUTIONS

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()

ATHENA
IT SOLUTIONS

Talend

Alteryx: cities_file_input.yxmd

ww_cities.csv → File to Table Default → aka:info7370_workshops Query=cities_file_default

ww_cities.csv → File to Table Auto Field → aka:info7370_workshops Query=cities_file_auto

ww_cities.csv → File to Table Auto Field, Select, Formula → DI_CREATE_DT = DateTimeNow() DI_JOB_NAME = [Engine.WorkflowFileName] → aka:info7370_workshops Query=cities_file_auto_select_form...

ATHENA
IT SOLUTIONS

Talend

Alteryx: cities_Excel_input.yxmd

Talend
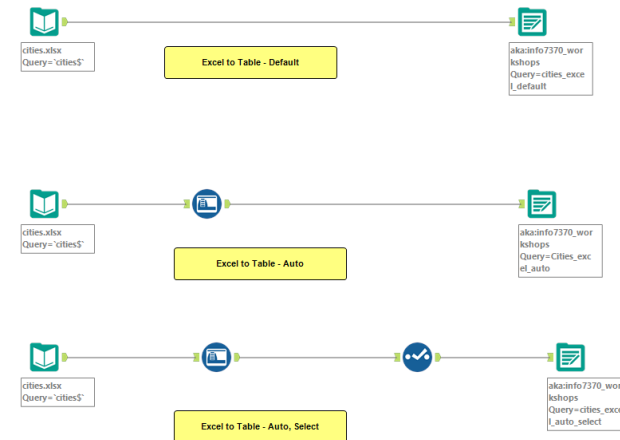
Alteryx: cities_Table_input.yxmd

Table to Table Transfer with Formulas

aka:info7370_input
Query=select
Cities.CityID,
Citi...

DI_Create_Date =
DateTimeNow()
DI_Job_Name =
[Engine.Workflow
FileName]

aka:info7370_workshops
Query=cities_Table_Transfer

ATHENA
IT SOLUTIONS

Talend
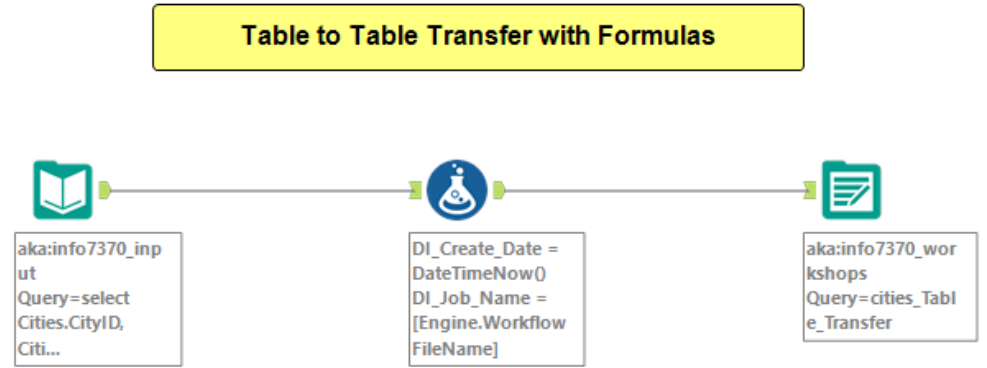
Alteryx: create_dim_geography_workshop.yxmd

ATHENA
IT SOLUTIONS
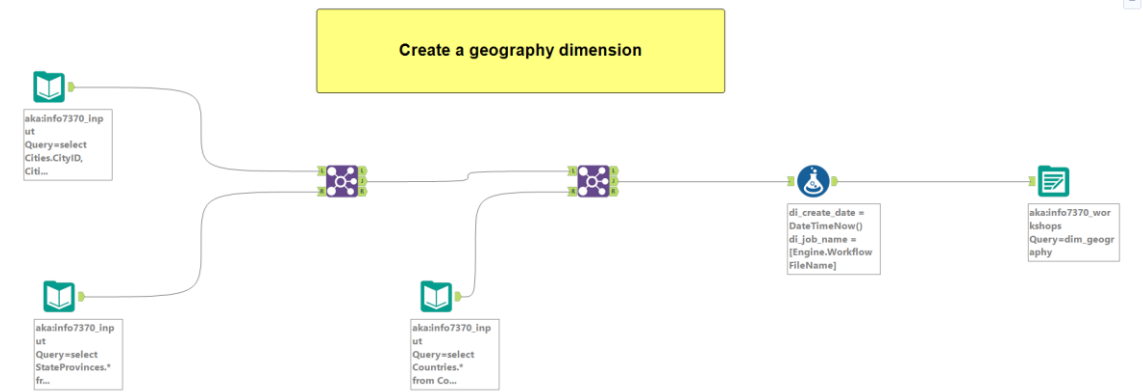
Data Preparation
More examples

ATHENA
IT SOLUTIONS

Alteryx: cities_to_four_db.yxmd

Copyright © 2020 Athena IT Solutions   All rights reserved.

Data Integration

More

Installation & Setup

Talend

ATHENA
IT SOLUTIONS

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

ATHENA
IT SOLUTIONS

Installation

ATHENA
IT SOLUTIONS
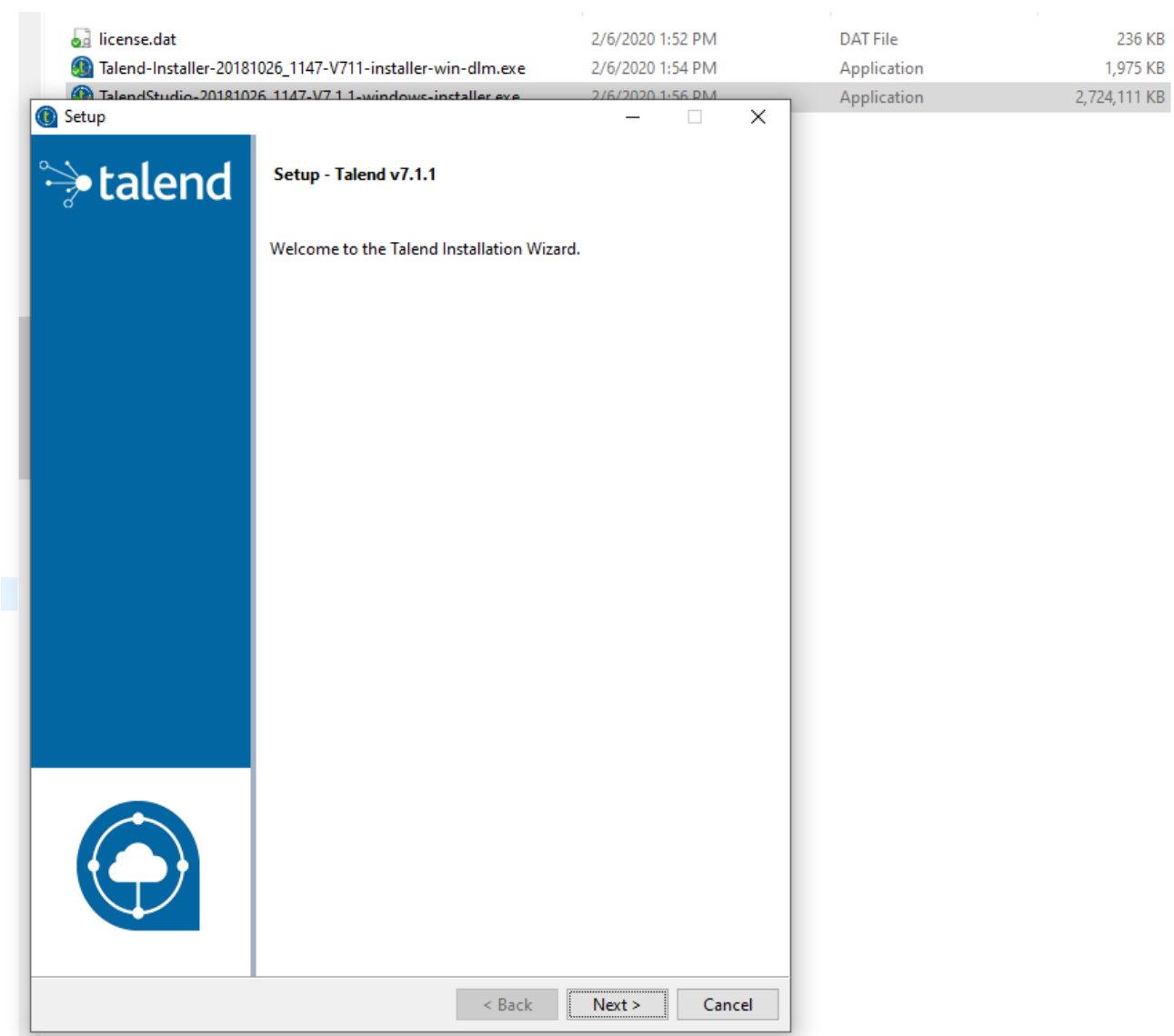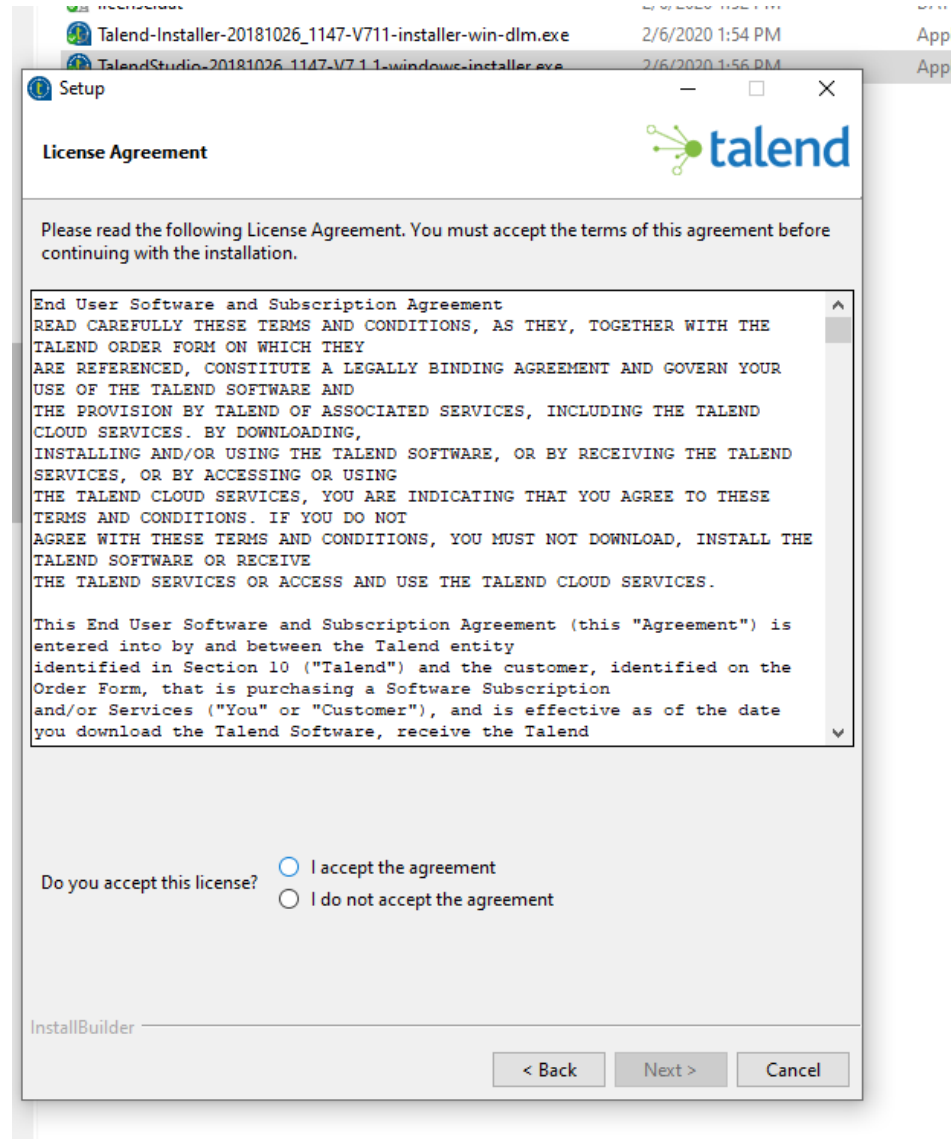
Installation
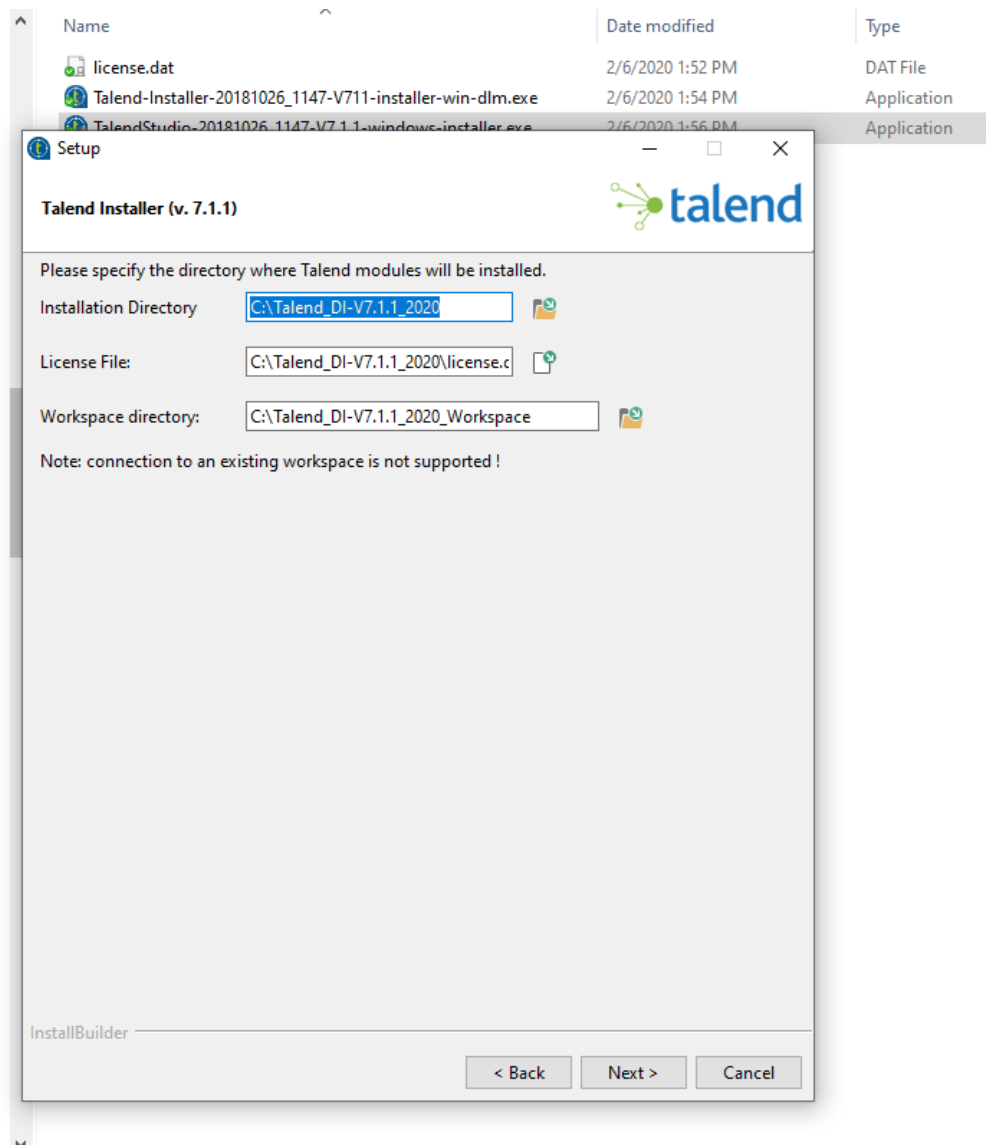
Installation

Installation
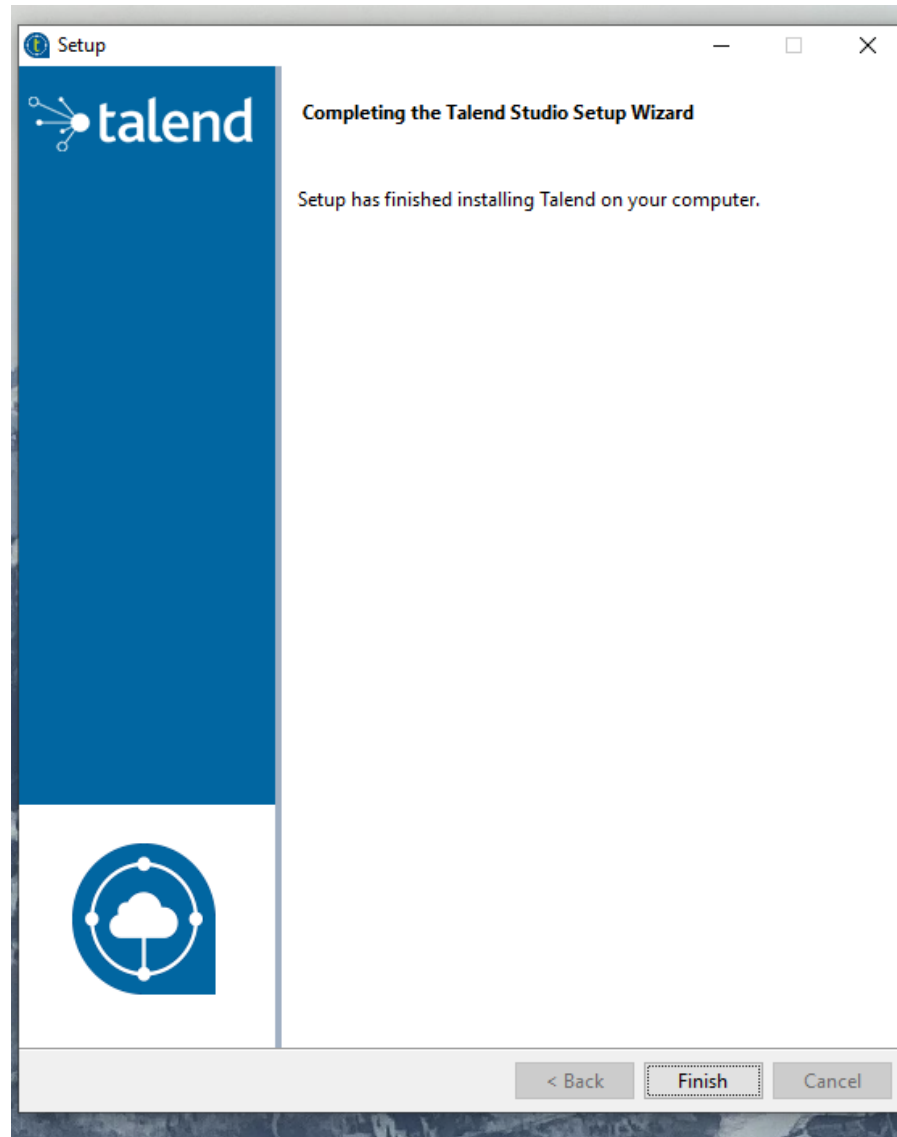
Change Studio Maximum Memory

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

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
```

ATHENA
IT SOLUTIONS

Talend - Setup

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

ATHENA
IT SOLUTIONS
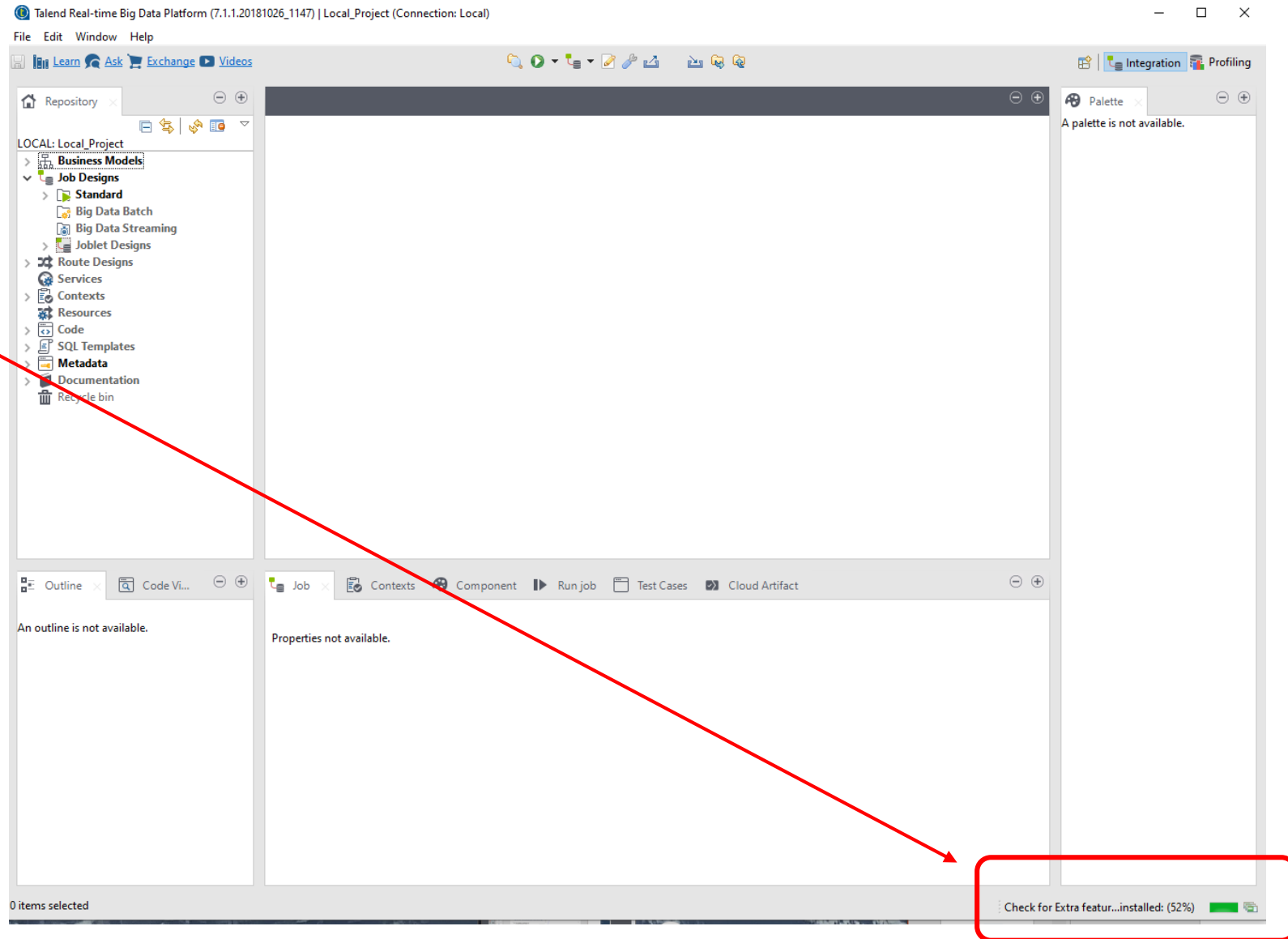
Select Required third party libraries

ATHENA
IT SOLUTIONS
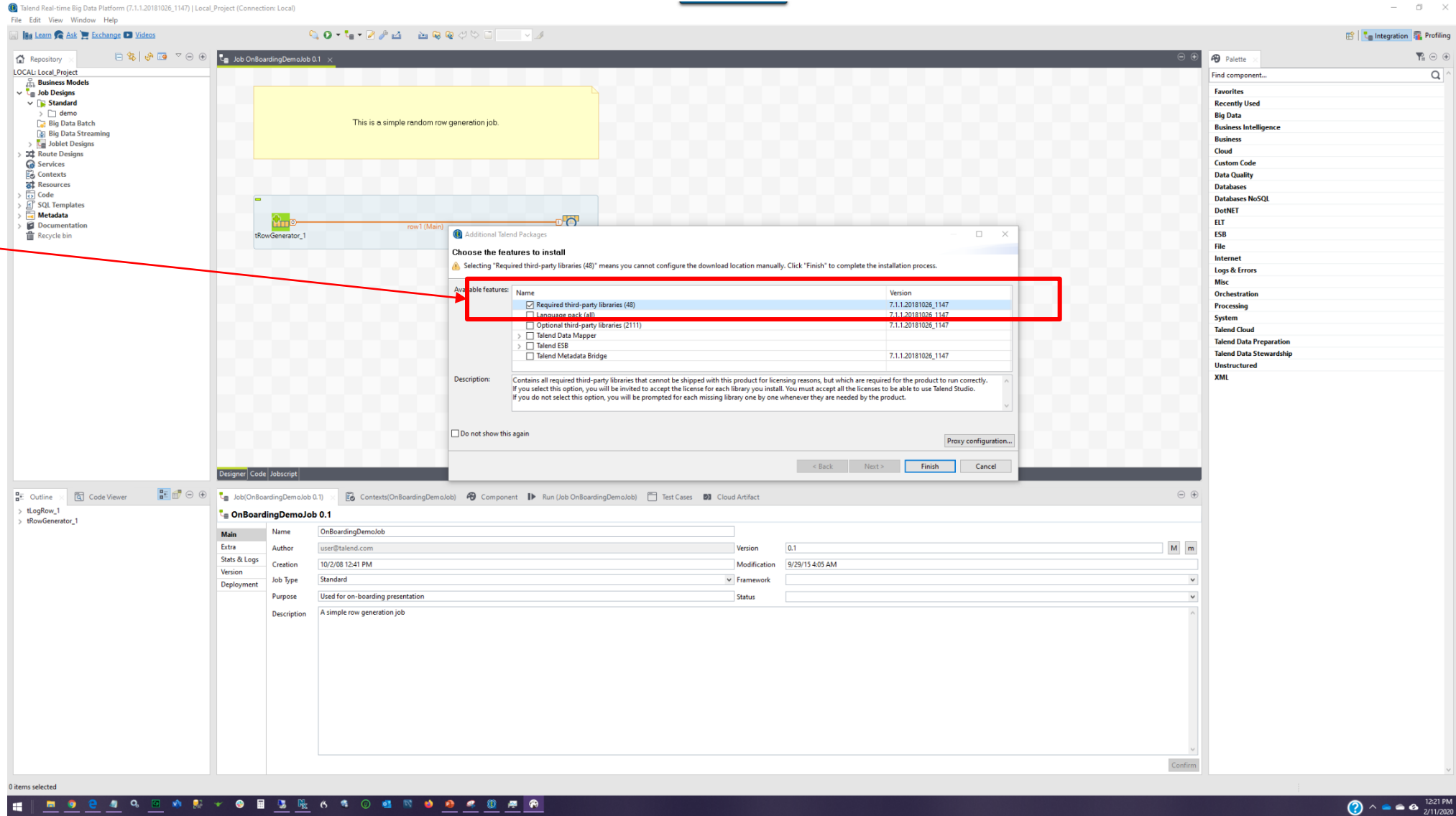
"I will accept…" & then Accept All button

ATHENA
IT SOLUTIONS
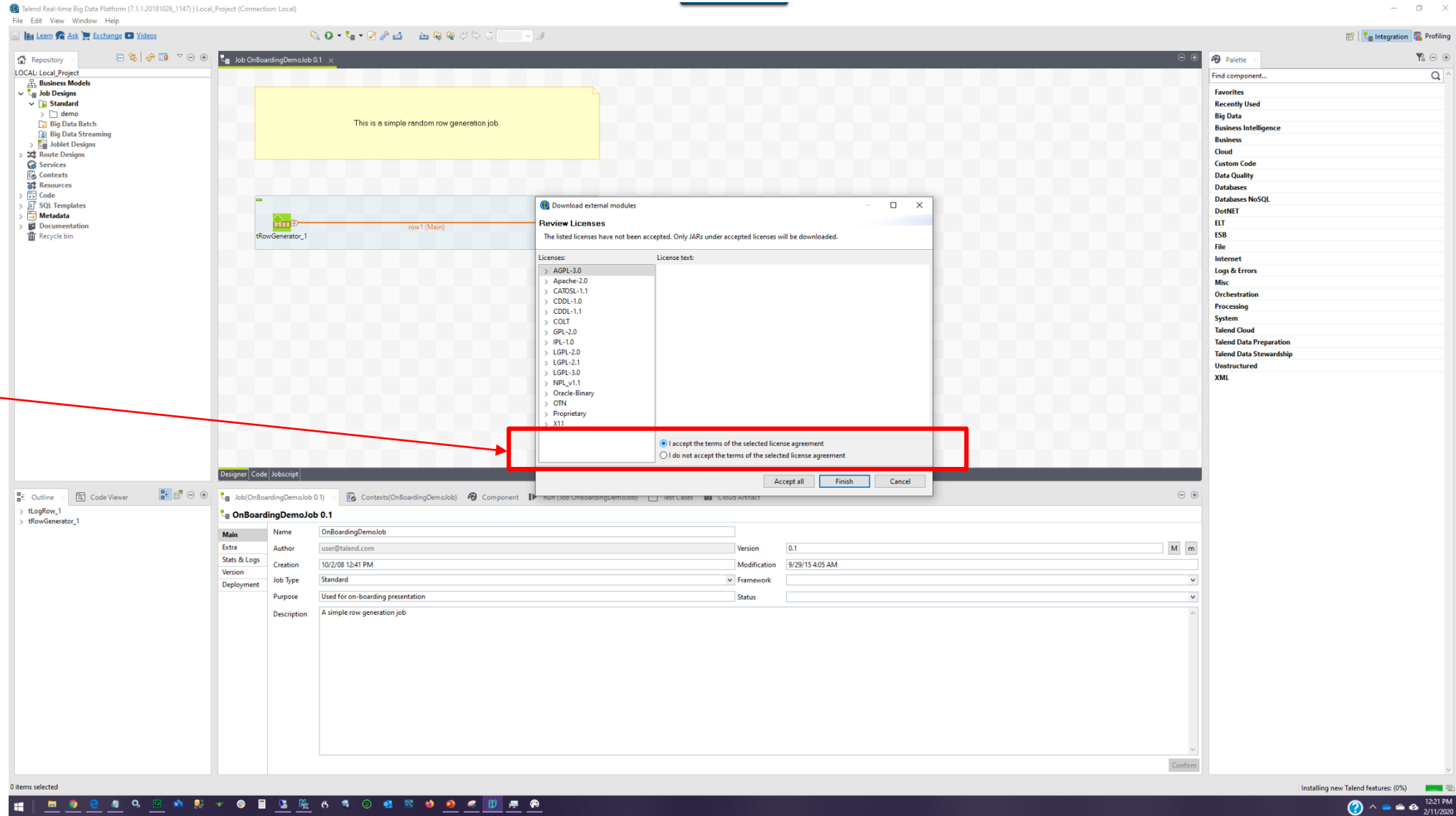
Windows, Show Views, Select Talend-Modules

ATHENA
IT SOLUTIONS
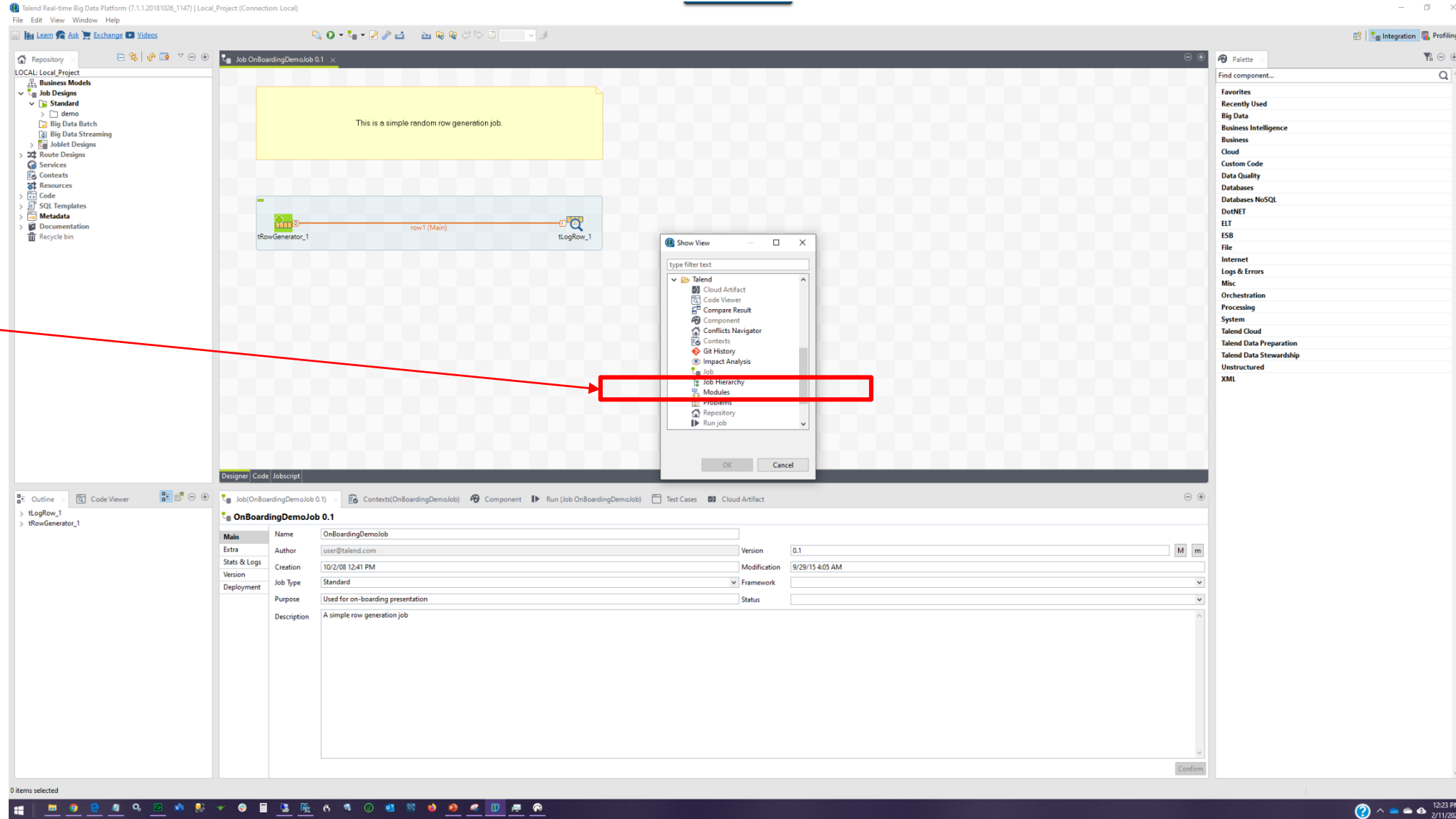
Change Run/Debug Maximum Memory

Change this to:

notebook	-Xmx
8G	3G
16G	8G
32G	12G

ATHENA
IT SOLUTIONS
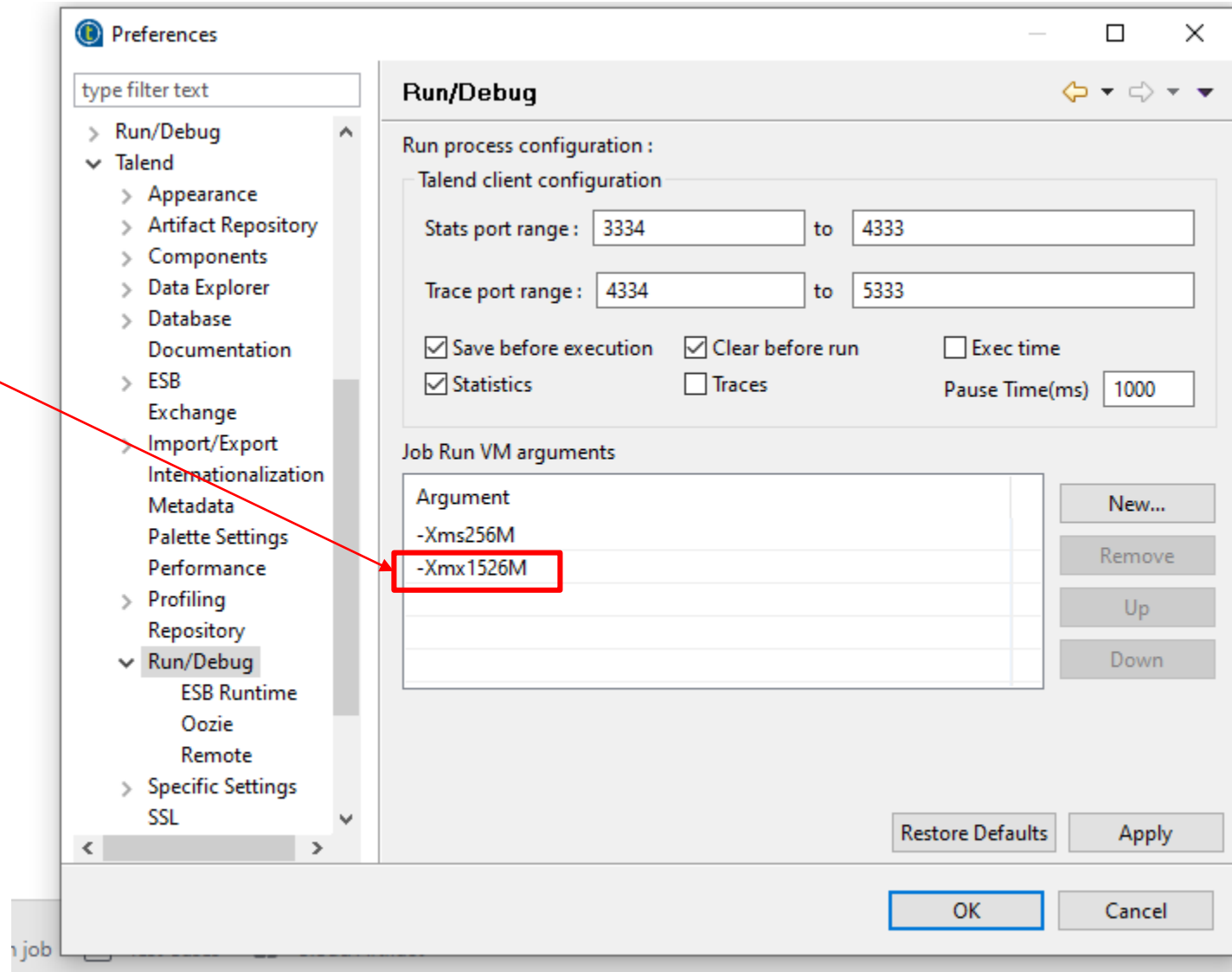
Import Archive File
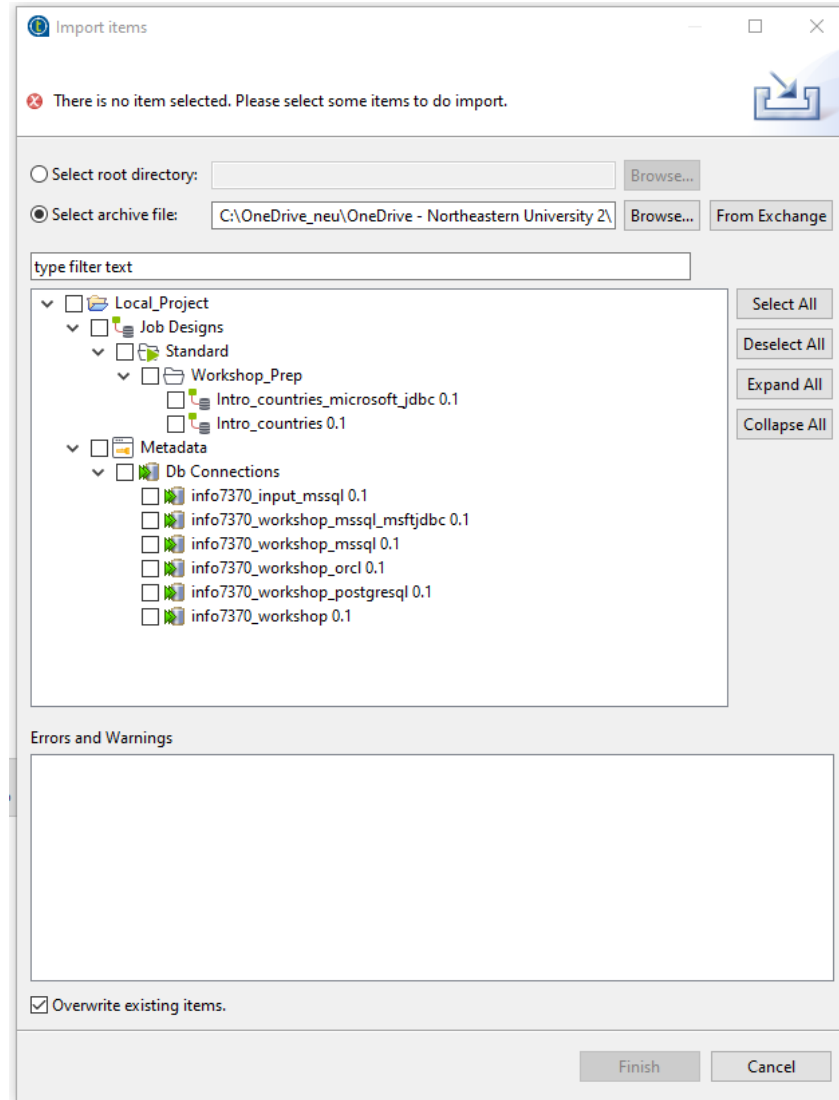
- Sample archive file
  - Jobs to test connections to our workshop databases
  - DB Connections

ATHENA
IT SOLUTIONS

Run sample jobs

Copyright © 2020 Athena IT Solutions   All rights reserved.