

# Ejercicios Talend

## MDA (EDEM). Curso 2021/2022

Lluna Sanz Montrull

23/11/2021

## Ejercicio 1

Leer un fichero CSV y escribirlo a fichero Json en la misma carpeta. Hacer lo mismo con MASTERS.csv. Leer de un json y escribir en un CSV. - Comentario: se producen incompatibilidades con las versiones de Java al pasar un fichero de .csv a .json.

## Talend

The screenshot displays the Talend Open Studio for Data Integration (8.0.1.20211109\_1610) interface. The main workspace shows a job design for 'Job ejercicio\_01 0.1'. The job consists of three components: 'tFileInputDelimited\_1' (reading from 'washington\_bike\_rentals'), 'tFileOutputJSON\_1' (writing to 'tFileOutputJSON\_1'), and 'tFileInputJSON\_1' (reading from 'master\_original'). The job is currently in the 'Designer' view. The 'Execution' tab is selected, showing the log of the job run. The log indicates that the job started at 18:48:22 on 22/11/2021, connected to the socket on port 3560, and ended at 18:49:22 on 22/11/2021 with an exit code of 0. The log also shows that the job was disconnected from the socket. The 'Basic Run' tab is selected, showing the job's configuration and the 'Run' button. The 'Advanced settings' tab is also visible, showing the job's target execution settings. The 'Memory Run' tab is also visible, showing the job's memory usage settings. The 'Outline' view on the left shows the job's structure, including the 'Job Designs' folder, 'Contexts', 'Codigo', 'SQL Templates', 'Metadata', 'Db Connections', 'File delimited', 'File positional', 'File regex', 'File XML', and 'File Fxxml'.

## JSON

### Washington bike rentals

```
{
  "data": [
    {
      "hum": "80,5833",
      "mnth": 1,
      "weekday": 6,
      "cnt": 985,
      "registered": 654,
      "holiday": 0,
      "instant": 1,
      "weathersit": 2,
      "workingday": 0,
      "dteday": "01\01\2011",
      "casual": 331,
      "atemp": "0,363625",
      "season": 1,
      "yr": 0,
      "temp_celsius": "14,110847",
      "windspeed_kh": "10,749882"
    },
    {
      "... "
    },
    {
      "hum": "57,75",
      "mnth": 12,
      "weekday": 1,
      "cnt": 2729,
      "registered": 2290,
      "holiday": 0,
      "instant": 731,
      "weathersit": 2,
      "workingday": 1,
      "dteday": "31\12\2012",
      "casual": 439,
      "atemp": "0,223487",
      "season": 1,
      "yr": 1,
      "temp_celsius": "8,849153",
      "windspeed_kh": "10,374682"
    }
  ]
}
```

731 Rows

## Masters

```
{
  "data": [
    {
      "id": 1,
      "Nom": "Latz"
    },
    {
      "id": 2,
      "Nom": "Janyx"
    },
    {
      "id": 3,
      "Nom": "Eabox"
    },
    {
      "id": 4,
      "Nom": "Izio"
    },
    {
      "id": 5,
      "Nom": "Browsebug"
    },
    {
      "id": 96,
      "Nom": "Realpoint"
    },
    {
      "id": 97,
      "Nom": "Skyble"
    },
    {
      "id": 98,
      "Nom": "Kimia"
    },
    {
      "id": 99,
      "Nom": "Skiba"
    },
    {
      "id": 100,
      "Nom": "Blognation"
    }
  ]
}
```

100 Rows

## Ejercicio 2

Debéis leer un fichero CSV(alumnos) y reemplazar: - China -> CH - France -> FR - Brazil -> BR - Cuba -> CU Debéis leer un fichero CSV(alumnos) y filtrar, de forma que solo me devuelva los alumnos que son de Portugal. Leer CSV ALUMNOS, reemplazar Brazil por BR, y filtrar solo los de BR.

## SQL

```
SELECT ID, NOM, COGNOM, REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(PAIS,
'Portugal', 'Portugal'), 'France', 'FR'), 'Cuba', 'CU'), 'China', 'CH'),
'Brazil', 'BR'), DIRECCION, EMAIL
FROM SQL_session.ALUMNOS
WHERE PAIS = "Portugal"
```

```
SELECT ID, NOM, COGNOM, REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(PAIS,
'Portugal', 'Portugal'), 'France', 'FR'), 'Cuba', 'CU'), 'China', 'CH'),
'Brazil', 'BR'), DIRECCION, EMAIL
FROM SQL_session.ALUMNOS
WHERE PAIS = "BR"
```

## Talend

Talend Open Studio for Data Integration (8.0.1.20211109\_1610) | Local\_Project (Conexión: Local)

Archivo Editar View Search Ventana Ayuda

Support 100%

Repositorio LOCAL: Local\_Project

- Job Designs
  - ejercicio\_02
  - ejercicio\_04
  - ejercicio\_05
  - ejercicio\_06
  - ejercicio\_07
  - ejercicio\_08
  - ejercicio\_08.0.1
- Contexts
- Código
- SQL Templates
- Metadata
  - Db Connections
    - File delimited
      - alu\_master 0.1
      - alumnos 0.1
      - masters 0.1
    - File positional
    - File regex
    - File XML
    - File Xerox

Designer | Code

Trabajo(Job ejercicio\_02 0.1) Contexts(ejercicio\_02) Component Run (Trabajo ejercicio\_02)

**tReplace\_1**

Basic settings

Advanced settings

Dynamic settings

View

Documentation

Esquema Built-in Edit schema Sync columns

Simple mode

Columna de ...	Search	Reemplazar ...	Whole w...	Case sen...	Glob exp...	Comentario
PAIS	"Brazil"	"BR"	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
PAIS	"France"	"FR"	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
PAIS	"Cuba"	"CU"	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
PAIS	"China"	"CH"	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
PAIS	"Portugal"	"Portugal"	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

152 rows in 0.01s

152 rows in 0.02s

5 rows in 0.02s

9 rows in 0.02s

450 rows in 0.02s

alumnos

tReplace\_1

tMap\_1

FileOutputDelimited\_1

FileOutputDelimited\_2

Buscar componente...

Favoritos

Utilizados recientemente

- tMap
- tFileOutputDelimited
- tReplace
- tFileInputDelimited
- tFileOutputJSON
- tBigQueryInput
- tAggregateRow
- tSortRow

Big Data

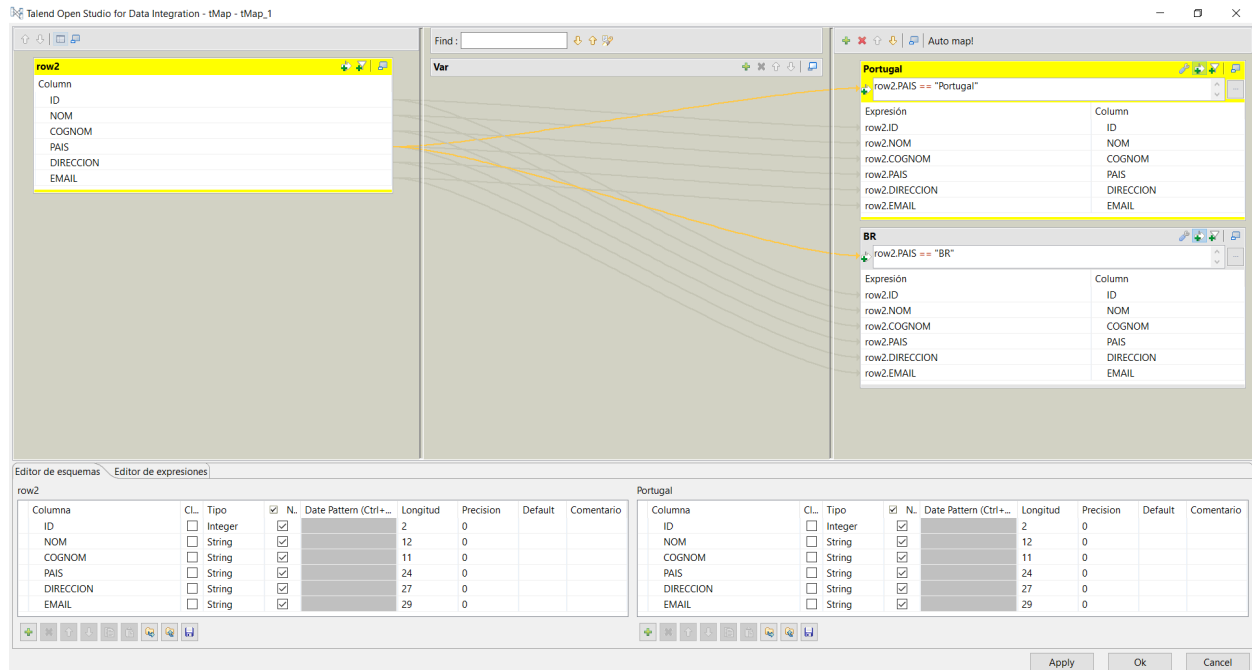
- Google BigQuery
- Google Storage
- Hive

Business Intelligence

Business

Cloud

- Amazon
- Azure Storage
- Azure
- Box
- Dropbox
- Google Drive
- Google Storage
- Google
- Marketo
- NetSuite
- Salesforce
- ServiceNow
- Snowflake
- tCloudStart
- tCloudStop



## CSV

### Alumnos de Portugal

	ID	NOM	COGNOM	PAIS	DIRECCION	EMAIL
0	1	Noëlla	Orchart	Portugal	54 Veith Parkway	rorchart0@huffingtonpost.com
1	15	Hélène	Brooke	Portugal	1068 Sherman Crossing	tbrookee@icio.us
2	45	Marylène	Duddy	Portugal	485 Delladonna Terrace	oduddy18@samsung.com
3	96	Bénédicte	Bampforth	Portugal	5 Blaine Road	cbampforth2n@imageshack.us
4	124	Yú	Keaves	Portugal	423 Spenser Alley	akeaves3f@chicagotribune.com

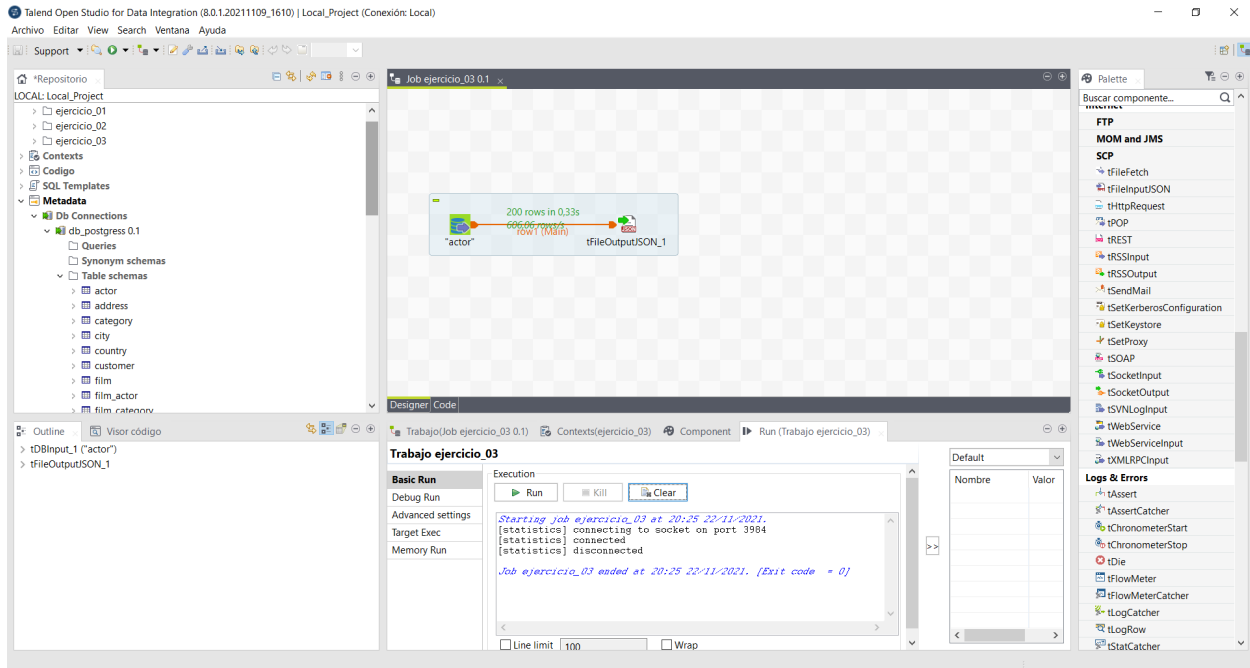
### Alumnos de Brasil

	ID	NOM	COGNOM	PAIS	DIRECCION	EMAIL
0	13	Séverine	Heathcoat	BR	1602 Welch Way	jheathcoatc@wisc.edu
1	22	Maëlann	Clavering	BR	29120 David Road	bclaveringl@china.com.cn
2	44	Pélagie	Arlow	BR	0 Briar Crest Pass	harlow17@netlog.com
3	74	Méng	Greenlies	BR	2 Derek Pass	rgreenlies21@storify.com
4	94	Joséphine	Treuge	BR	6543 Jackson Parkway	btreuge21@state.tx.us

### Ejercicio 3

Leer tabla de actores y volcarlo a fichero JSON. - Instrucciones

## Talend



## JSON

```
{
  "data": [
    {
      "last_update": "26-05-2013",
      "last_name": "Guinness",
      "actor_id": 1,
      "first_name": "Penelope"
    },
    {
      "last_update": "26-05-2013",
      "last_name": "Wahlberg",
      "actor_id": 2,
      "first_name": "Nick"
    },
    {
      "last_update": "26-05-2013",
      "last_name": "Chase",
      "actor_id": 3,
      "first_name": "Ed"
    },
    {
      "last_update": "26-05-2013",
      "last_name": "Davis",

```

```

        "actor_id": 4,
        "first_name": "Jennifer"
    },
    {
        "last_update": "26-05-2013",
        "last_name": "Lollobrigida",
        "actor_id": 5,
        "first_name": "Johnny"
    },
    {
        "...",
    },
    {
        "last_update": "26-05-2013",
        "last_name": "Walken",
        "actor_id": 196,
        "first_name": "Bela"
    },
    {
        "last_update": "26-05-2013",
        "last_name": "West",
        "actor_id": 197,
        "first_name": "Reese"
    },
    {
        "last_update": "26-05-2013",
        "last_name": "Keitel",
        "actor_id": 198,
        "first_name": "Mary"
    },
    {
        "last_update": "26-05-2013",
        "last_name": "Fawcett",
        "actor_id": 199,
        "first_name": "Julia"
    },
    {
        "last_update": "26-05-2013",
        "last_name": "Temple",
        "actor_id": 200,
        "first_name": "Thora"
    }
]
}

```

200 rows

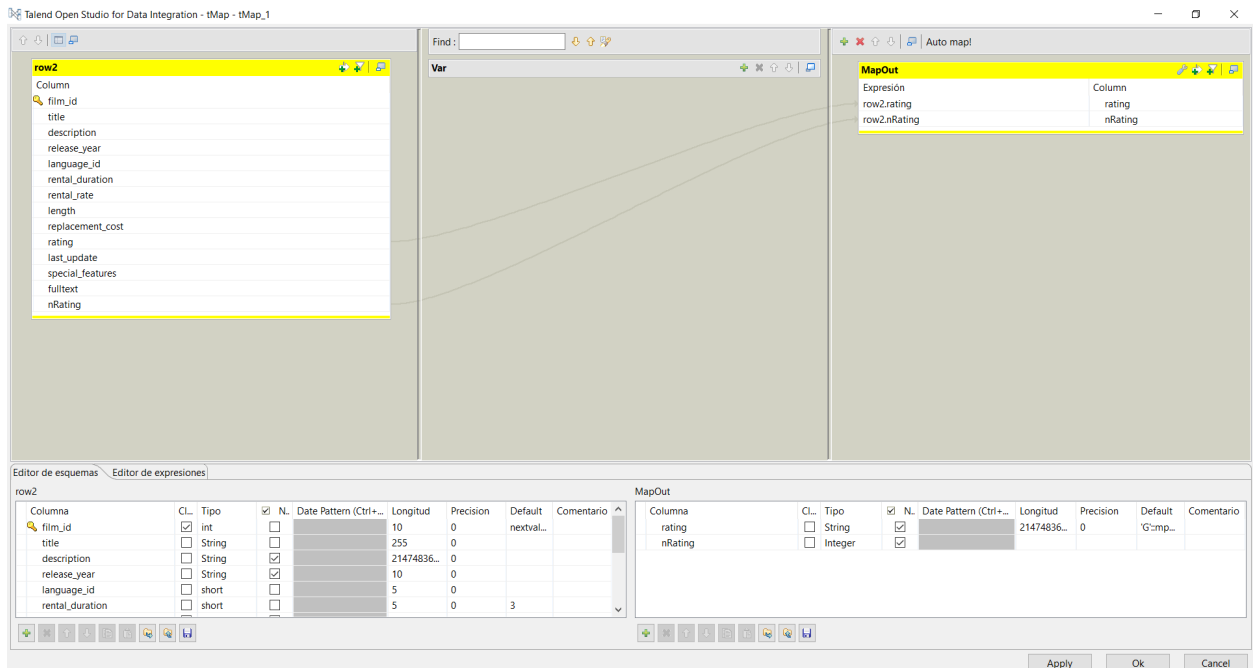
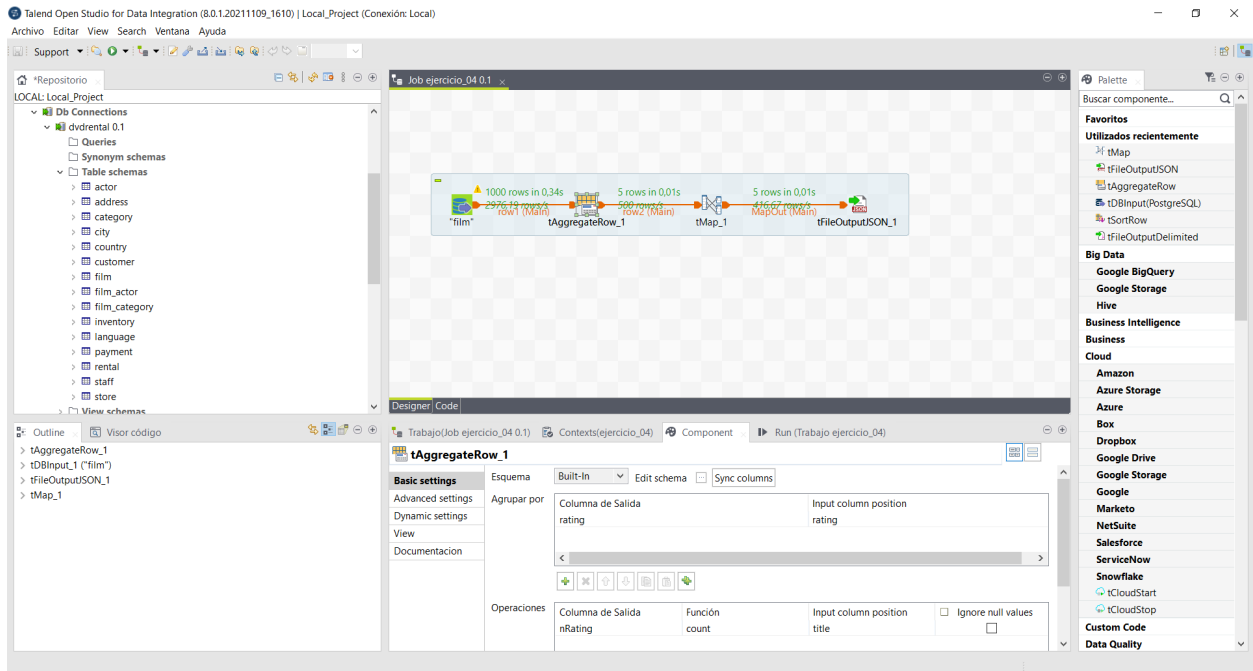
## Ejercicio 4

Agregar las películas por rating y mostrar un count, volcar a json el resultado

## SQL

```
SELECT rating, count(title) FROM film GROUP BY rating
```

## Talend





## JSON

```
{
  "count_rating": [
    {
      "nRating": 195,
      "rating": "R"
    },
    {
      "nRating": 194,
      "rating": "PG"
    },
    {
      "nRating": 178,
      "rating": "G"
    },
    {
      "nRating": 210,
      "rating": "NC-17"
    },
    {
      "nRating": 223,
      "rating": "PG-13"
    }
  ]
}
```

5 Rows

## Ejercicio 5

Realizar un Join entre Actor / Film / Film\_Actor y volcar a json un fichero con estos campos:

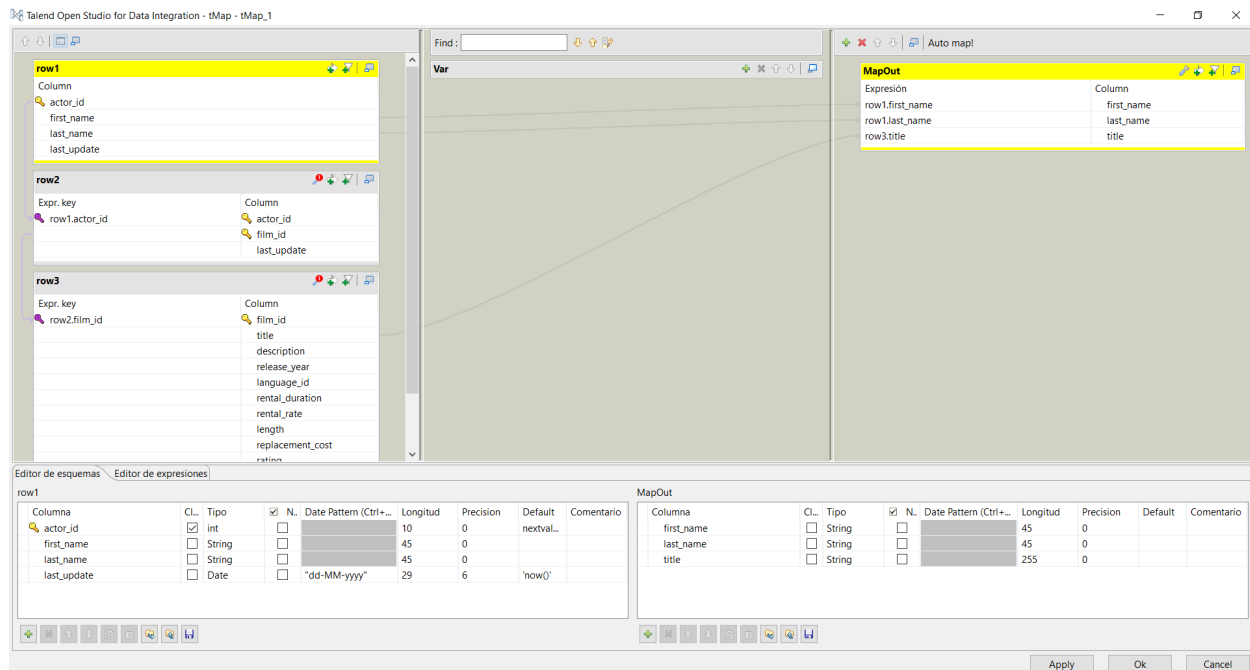
- Nombre
- Apellido
- Película

## SQL

```
SELECT actor.first_name, actor.last_name, film.title
FROM actor
JOIN film_actor ON actor.actor_id=film_actor.actor_id
JOIN film ON film_actor.film_id=film.film_id
```

## Talend

The screenshot displays the Talend Open Studio for Data Integration (8.0.1.20211109\_1610) interface. The main workspace shows a job configuration for 'Job ejercicio\_05 0.1'. The job consists of three input components: 'actor' (200 rows in 0.03s), 'film\_actor' (5462 rows in 0.03s), and 'film' (72851.76 rows in 0.05s). These are connected to a 'tMap\_1' component, which is configured with a 'tMapOut (Main)' output. The output is connected to a 'tFileOutputJSON\_1' component, which outputs to a file named 'output.json'. The left sidebar shows the 'LOCAL: Local\_Project' repository with a tree view of table schemas including actor, address, category, city, country, customer, film, film\_actor, film\_category, inventory, language, payment, rental, staff, store, and view schemas. The right sidebar shows the 'Palettes' with various components like tCloudStart, tCloudStop, Custom Code, Data Quality, Databases, DotNET, ELT, Combined SQL, Conexiones, Connections, Map, SQLTemplate, ESB, Fichero, Entrada, Management, NamedPipe, Salida, tAdvancedFileOutputXML, tExternalSortOutput, tFileOutputARFF, tFileOutputDelimited, tFileOutputExcel, tFileOutputJSON, tFileOutputLDIF, tFileOutputMSDelimited, tFileOutputMSPositional, tFileOutputMSXML, tFileOutputPositional, tFileOutputProperties, tFileOutputRaw, and tFileOutputXML. The bottom status bar indicates 'Run (Trabajo ejercicio\_05)'.



## JSON

```
{
  "data": [
    {
      "last_name": "Guinness",
      "title": "Wizard Coldblooded",
      "first_name": "Penelope"
    },
    {
      "last_name": "Wahlberg",
      "title": "Wardrobe Phantom",
      "first_name": "Nick"
    },
    {
      "last_name": "Chase",
      "title": "Young Language",
      "first_name": "Ed"
    },
    {
      "last_name": "Davis",
      "title": "Unforgiven Zoolander",
      "first_name": "Jennifer"
    },
    {
      "last_name": "Lollobrigida",
      "title": "Sunrise League",
      "first_name": "Johnny"
    },
    {

```

```

    "...",
  },
  {
    "last_name": "Walken",
    "title": "Whisperer Giant",
    "first_name": "Bela"
  },
  {
    "last_name": "West",
    "title": "Yentl Idaho",
    "first_name": "Reese"
  },
  {
    "last_name": "Keitel",
    "title": "Youth Kick",
    "first_name": "Mary"
  },
  {
    "last_name": "Fawcett",
    "title": "Wait Cider",
    "first_name": "Julia"
  },
  {
    "last_name": "Temple",
    "title": "Wrong Behavior",
    "first_name": "Thora"
  }
]
}

```

200 Rows

## Ejercicio 6

Cargar en un csv la cantidad de dinero Gastada por usuario, nombre y apellido

### SQL

```
SELECT customer.first_name, customer.last_name, sum(payment.amount) FROM customer JOIN payment ON customer
```

### Talend

Talend Open Studio for Data Integration (8.0.1.20211109\_1610) | Local\_Project (Conexión: Local)

Archivo Editar View Search Ventana Ayuda

Support 100%

Repositorio LOCAL: Local\_Project

- Job Designs
  - ejercicio\_04
  - ejercicio\_05
  - ejercicio\_06
  - ejercicio\_06 0.1**
  - ejercicio\_07
- Contexts
- Código
- SQL Templates
- Metadata
  - Db Connections
    - dvdrental 0.1
      - Queries
      - Synonym schemas
      - Table schemas
        - actor
        - address
        - category
        - city
        - country
        - customer

Job ejercicio\_06 0.1

Diagrama de flujo de datos:

- row1 (Main) -> tMap\_1 -> MapOut (Main) -> row3 (Main) -> tSortRow\_1 -> row4 (Main) -> tFileOutputDelimited\_1
- row2 (lookupt) -> tMap\_1
- customer -> tMap\_1

Componentes:

- tAggregateRow\_1
- tDBInput\_1 ("customer")
- tDBInput\_2 ("payment")
- tFileOutputDelimited\_1
- tMap\_1
- tSortRow\_1

Paleta:

- Favoritos
- Utilizados recientemente
  - tMap
  - tAggregateRow
  - tSortRow
  - tFileOutputJSON
  - tFileOutputDelimited
- Big Data
  - Google BigQuery
  - Google Storage
  - Hive
- Business Intelligence
- Business
- Cloud
  - Amazon
  - Azure Storage
  - Azure
  - Box
  - Dropbox
  - Google Drive
  - Google Storage
  - Google
  - Marketo
  - NetSuite
  - Salesforce
  - ServiceNow
  - Snowflake
  - tCloudStart
  - tCloudStop
- Custom Code
- Data Quality
- Databases

Trabajo(Job ejercicio\_06 0.1) Contexts(ejercicio\_06) Component Run (Trabajo ejercicio\_06)

**tSortRow\_1**

Esquema: Built-In Edit schema Sync columns

Basic settings

Criterio(s)

Columna del esquema	Orden numérico o alfabético?	Orden Ascendente o Descendente?
amount	numérico	desc

Advanced settings

Dynamic settings

View

Documentation

Talend Open Studio for Data Integration (8.0.1.20211109\_1610) | Local\_Project (Conexión: Local)

Archivo Editar View Search Ventana Ayuda

Support 100%

Repositorio LOCAL: Local\_Project

- Job Designs
  - ejercicio\_04
  - ejercicio\_05
  - ejercicio\_06
  - ejercicio\_06 0.1**
  - ejercicio\_07
- Contexts
- Código
- SQL Templates
- Metadata
  - Db Connections
    - dvdrental 0.1
      - Queries
      - Synonym schemas
      - Table schemas
        - actor
        - address
        - category
        - city
        - country
        - customer

Job ejercicio\_06 0.1

Diagrama de flujo de datos:

- row1 (Main) -> tMap\_1 -> MapOut (Main) -> row3 (Main) -> tSortRow\_1 -> row4 (Main) -> tFileOutputDelimited\_1
- row2 (lookupt) -> tMap\_1
- customer -> tMap\_1

Componentes:

- tAggregateRow\_1
- tDBInput\_1 ("customer")
- tDBInput\_2 ("payment")
- tFileOutputDelimited\_1
- tMap\_1
- tSortRow\_1

Paleta:

- Favoritos
- Utilizados recientemente
  - tMap
  - tAggregateRow
  - tSortRow
  - tFileOutputJSON
  - tFileOutputDelimited
- Big Data
  - Google BigQuery
  - Google Storage
  - Hive
- Business Intelligence
- Business
- Cloud
  - Amazon
  - Azure Storage
  - Azure
  - Box
  - Dropbox
  - Google Drive
  - Google Storage
  - Google
  - Marketo
  - NetSuite
  - Salesforce
  - ServiceNow
  - Snowflake
  - tCloudStart
  - tCloudStop
- Custom Code
- Data Quality
- Databases

Trabajo(Job ejercicio\_06 0.1) Contexts(ejercicio\_06) Component Run (Trabajo ejercicio\_06)

**tAggregateRow\_1**

Esquema: Built-In Edit schema Sync columns

Basic settings

Agrupar por

Columna de Salida	Input column position
first_name	first_name
last_name	last_name

Advanced settings

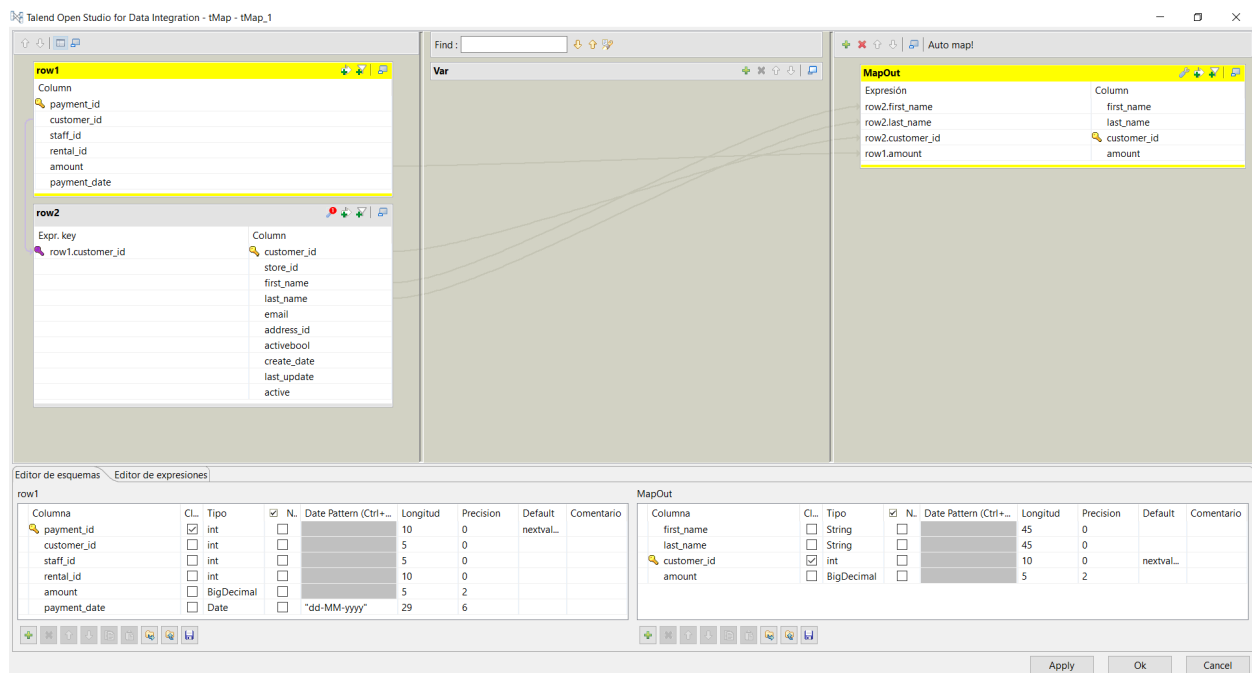
Dynamic settings

View

Documentation

Operaciones

Columna de Salida	Función	Input column position	Ignore null values
amount	Suma	amount	<input type="checkbox"/>



## CSV

	first_name	last_name	amount
0	Eleanor	Hunt	211.55
1	Karl	Seal	208.58
2	Marion	Snyder	194.61
3	Rhonda	Kennedy	191.62
4	Clara	Shaw	189.6
...	...	...	...
594	Tiffany	Jordan	49.88
595	Anthony	Schwab	47.85
596	Caroline	Bowman	37.87
597	Leona	Obrien	32.9
598	Brian	Wyman	27.93

599 rows × 3 columns

## Ejercicio 7

Cargar en un csv el numero de veces que se ha alquilado cada película.

### SQL

```
SELECT title, count(rental.rental_id)
FROM film
JOIN inventory ON film.film_id=inventory.film_id
JOIN rental ON inventory.inventory_id=rental.inventory_id
GROUP BY title
```

### Talend

Talend Open Studio for Data Integration (8.0.1.20211109\_1610) | Local\_Project (Conexión: Local)

Archivo Editar View Search Ventana Ayuda

Support 100%

Repositorio LOCAL: Local\_Project

- Job Designs
  - ejercicio\_04
  - ejercicio\_05
  - ejercicio\_06
  - ejercicio\_06.0.1
  - ejercicio\_07
    - ejercicio\_07.0.1
- Contexts
- Código
- SQL Templates
- Metadata
  - Db Connections
    - divdental 0.1
      - Queries
      - Synonym schemas
      - Table schemas
        - actor
        - address
        - category
        - city
        - country

Outline

- tAggregateRow\_1
- tDBInput\_1 ("film")
- tDBInput\_2 ("rental")
- tDBInput\_3 ("inventory")
- tFileOutputDelimited\_1
- tMap\_1

Diagrama de flujo:

El diagrama muestra un flujo de datos que comienza con tres fuentes de datos: "rental", "film" e "inventory". "rental" se conecta a "tMap\_1" (row1 (Main)). "film" se conecta a "tMap\_1" (row2 (Lookup)). "inventory" se conecta a "tMap\_1" (row3 (Lookup)). El flujo continúa a "MapOut0 (Main)", luego a "tAggregateRow\_1" (row4 (Main)), y finalmente a "tFileOutputDelimited\_1".

Configuración de tAggregateRow\_1:

Esquema: Built-in Edit schema Sync columns

Basic settings

Advanced settings

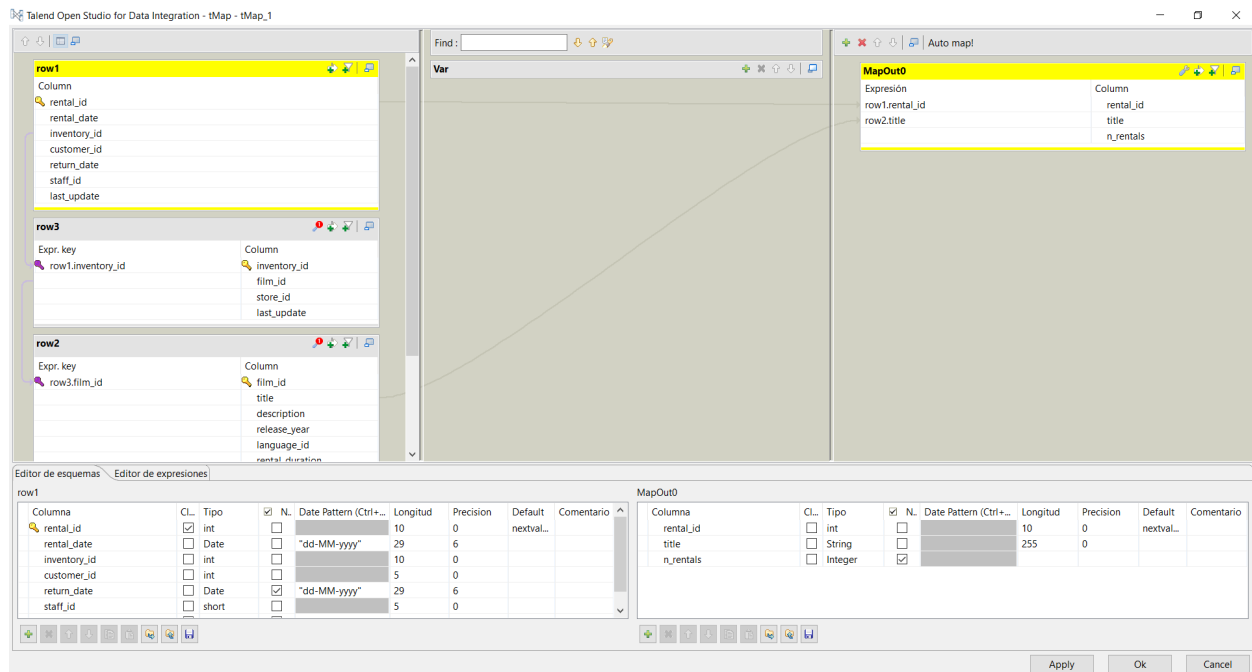
Dynamic settings

View

Documentation

Operaciones

Columna de Salida	Función	Input column position	Ignore null values
n_rentals	count	rental_id	<input type="checkbox"/>



## CSV

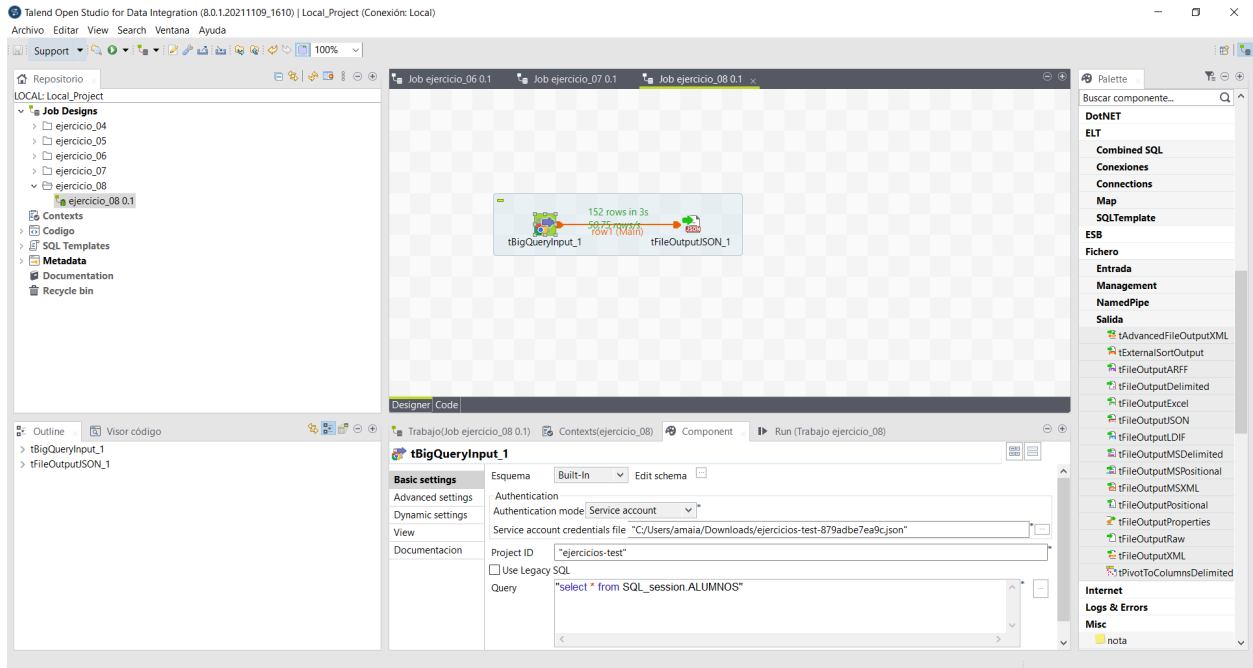
	title	rentals
0	Enough Raging	18
1	Microcosmos Paradise	13
2	Harry Idaho	30
3	Operation Operation	27
4	Gilbert Pelican	9
...	...	...
953	Vanishing Rocky	18
954	Ice Crossing	24
955	Bonnie Holocaust	16
956	Heaven Freedom	18
957	Saturday Lambs	28

958 rows  $\times$  2 columns



# Ejercicio 8

## Import from BigQuery



## JSON

```
{
  "data": [
    {
      "DIRECCION": "16340 La Follette Place",
      "ID": 9,
      "EMAIL": "epolleye8@businessinsider.com",
      "NOM": "Lén",
      "COGNOM": "Polleye",
      "PAIS": "Cuba"
    },
    {
      "DIRECCION": "90940 Roxbury Road",
      "ID": 121,
      "EMAIL": "drugge3c@ox.ac.uk",
      "NOM": "Märta",
      "COGNOM": "Rugge",
      "PAIS": "Cuba"
    },
    {
      "DIRECCION": "3224 McGuire Hill",
      "ID": 55,
      "EMAIL": "alukasen1i@ox.ac.uk",
      "NOM": "Marie-hélène",
      "COGNOM": "Lukasen",
      "PAIS": "Peru"
    }
  ]
}
```

```

},
{
  "DIRECCION": "983 Cascade Trail",
  "ID": 89,
  "EMAIL": "cmoffet2g@biglobe.ne.jp",
  "NOM": "Amélie",
  "COGNOM": "Moffet",
  "PAIS": "Peru"
},
{
  "DIRECCION": "5 New Castle Alley",
  "ID": 3,
  "EMAIL": "mcatherick2@yale.edu",
  "NOM": "Eliès",
  "COGNOM": "Catherick",
  "PAIS": "China"
},
{
  "... "
},
{
  "DIRECCION": "7 Manitowish Hill",
  "ID": 148,
  "EMAIL": "ttaggart43@hhs.gov",
  "NOM": "Dà",
  "COGNOM": "Taggart",
  "PAIS": "Czech Republic"
},
{
  "DIRECCION": "73 Green Ridge Hill",
  "ID": 114,
  "EMAIL": "vgoudie35@constantcontact.com",
  "NOM": "Miléna",
  "COGNOM": "Goudie",
  "PAIS": "United Kingdom"
},
{
  "DIRECCION": "0 Miller Road",
  "ID": 95,
  "EMAIL": "gcommucci2m@buzzfeed.com",
  "NOM": "Thérèse",
  "COGNOM": "Commucci",
  "PAIS": "Papua New Guinea"
},
{
  "DIRECCION": "56 Londonderry Road",
  "ID": 135,
  "EMAIL": "astert3q@aboutads.info",
  "NOM": "Cécilia",
  "COGNOM": "Stert",
  "PAIS": "Bosnia and Herzegovina"
},
{

```

```
"DIRECCION": "5 Arizona Crossing",  
"ID": 32,  
"EMAIL": "eglisenanv@technorati.com",  
"NOM": "Véronique",  
"COGNOM": "Glisenan",  
"PAIS": "Central African Republic"  
}  
]  
}
```

152 Rows