

Cas pràctic: Magatzem de dades per l'anàlisi de la conciliació laboral i familiar o *work-life balance*

PR2 – Càrrega de dades

Nom: Marc Bracons Cucó

Aula 2

Índex de continguts

1. Càrrega de dades

2. Identificació dels processos ETL

- 2.1. Bloc IN (de les fonts a taules intermèdies)
- 2.2. Bloc TR (poblar les taules del magatzem)

3. Disseny dels processos ETL

- 3.1. Creació de taules intermèdies (*staging area*)
- 3.2. Creació del model multidimensional
 - 3.2.1. Dimensions
 - 3.2.2. Taules de fets
- 3.3. Creació del procés d'extracció, transformació i càrrega
 - 3.3.1. Taules de fets
 - 3.3.2. Connexió a la base de dades *SQL Server*
 - 3.3.3. Bloc IN
 - 3.3.4. Bloc TR

4. Implementació de treballs amb processos ETL

- 4.1. JOB_IN
- 4.2. JOB_TR_DIMS
- 4.3. JOB_TR_FACTS
- 4.4. JOB_CÀRREGA_DW

1. Càrrega de dades

En aquesta etapa del projecte hem de dur a terme la implementació i càrrega de dades del magatzem de dades. La implementació del model físic d'un magatzem de dades dependrà del tipus de gestor de base de dades que es vulgui utilitzar. Cada gestor de base de dades té el seu propi model de dades i regles, així com la seva sintaxi per a la creació dels objectes que componen la base de dades (taules, índexs, vistes, etc. en el cas dels relacionals), per la qual cosa és important tenir en compte aquestes diferències en implementar el model físic.

El primer pas serà descarregar les dades a l'entorn VDI facilitat per la UOC. Ho guardarem al disc local F, ja que aquest no es borra en tancar la sessió a la màquina virtual (figura 1).

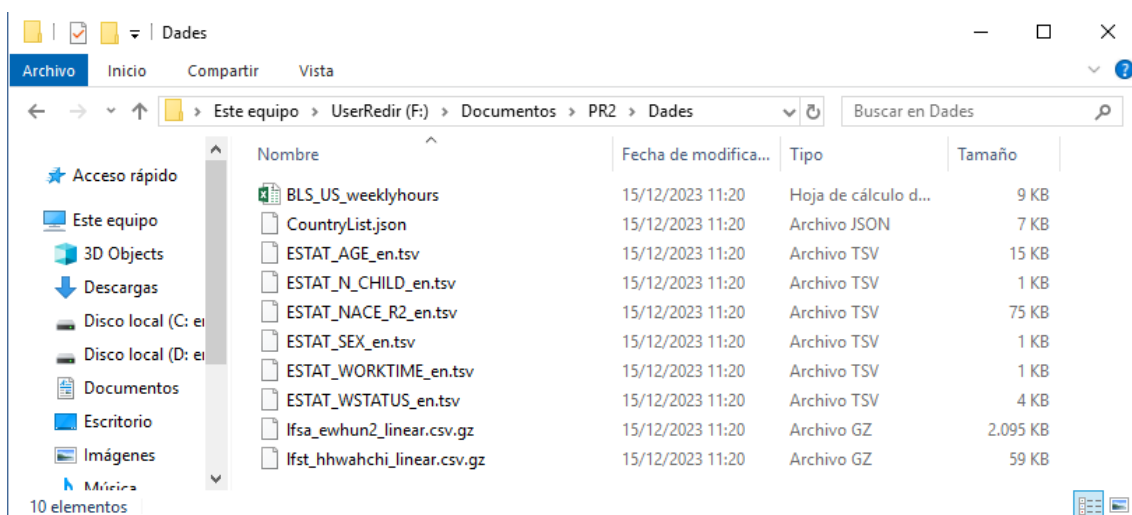


Figura 1. Dades al path 'F:\Documents\PR2\Dades'

2. Identificació dels processos ETL

A l'hora de dissenyar els processos de càrrega d'una base de dades analítica no hi ha una estratègia única. És habitual estructurar els processos ETL sobre la base de les entitats de dades que s'han d'actualitzar, ja que hi ha diferències conceptuais en l'actualització d'una dimensió respecte a la d'una taula de fets. La divisió del procés de càrrega inicial en diferents blocs d'actualització facilitarà el disseny d'un ordre d'execució i la gestió de les dependències. Cadascun d'aquests blocs d'actualització es dividirà en les corresponents etapes d'ETL.

S'identifiquen els dos blocs següents:

Bloc IN: processos de càrrega de les dades des de les fonts fins a les taules intermèdies en l'àrea de maniobres (STG). Aquests processos es distingeixen pel prefix IN_ en el nom.

- **Bloc TR:** processos de transformació per carregar les dades des de les taules intermèdies fins al magatzem, segons el model multidimensional dissenyat. Així doncs, els processos ETL de transformació per carregar les dimensions són diferents d'aquells que es porten a terme per carregar les taules de fets. Aquests processos es distingeixen amb el prefix TR_ en el nom.

2.1 Bloc IN (de les fonts a taules intermèdies)

Nom de l'ETL	Descripció	Orígens de les dades	Taula de destinació(stage)
IN_COUNTRY	Conté els noms dels països en ordre alfabètic, agrupats per zona i amb el seu codi ISO 3166 1 alpha 2 en format JSON. L'estructura del fitxer és la	CountryList.json	STG_COUNTRY
IN_PCT_EMPLOYEES_HOME	Percentatge de persones que treballen a casa per país, grup d'edat, nombre i edat de fills.	lfst_hhwahchi_linear.csv	STG_PCT_EMPLOYEES_HOME

IN_ HOURS_WORKED	Mitjana d'hores treballade s setmanal s a la UE per país, sexe, edat, categoria professio nal i activitat econòmic a.	lfsa_ewhun2_linear.c sv	STG_ HOURS_WORKED
IN_ HOURS_WORKED_ USA	Mitjana d'hores treballade s setmanal s als Estats Units. Dades obtingude s de US Departme nt of Labor.	BLS_US_weeklyhour s.xlsx	STG_ HOURS_WORKED_USA
IN_ ESTAT_AGE	Llista de trams d'edats segons classifica ció de l'Eurostat.	ESTAT_AGE_en.tsv	STG_ ESTAT_AGE
IN_ ESTAT_N_CHILD	Llista del nombre de fills segons classifica ció de l'Eurostat.	ESTAT_N_CHILD_e n.tsv	STG_ ESTAT_N_CHILD
IN_ ESTAT_SEX	Llista de valors relatius al sexe de les persones	ESTAT_SEX_en.tsv	STG_ ESTAT_SEX

	que treballen segons classificació de l'Eurostat.		
IN_ ESTAT_NACE_R2	Llista d'activitats econòmiques segons classificació de l'Eurostat.	ESTAT_NACE_R2_en.tsv	STG_ ESTAT_NACE_R2
IN_ ESTAT_WORKTIME	Llista de tipus d'horaris laborals de les persones que treballen segons classificació de l'Eurostat.	ESTAT_WORKTIME_en.tsv	STG_ ESTAT_WORKTIME
IN_ ESTAT_WSTATUS	Llista d'estats laborals segons classificació de l'Eurostat.	ESTAT_WSTATUS_en.tsv	STG_ ESTAT_WSTATUS

2.2 Bloc TR (de les taules intermèdies al magatzem)

El bloc TR de processos ETL per poblar el model multidimensional del magatzem té dues parts diferenciades. D'una banda, els processos de càrrega i transformació de les dimensions i, per un altre, els de les taules de fets. L'ordre d'execució és important perquè la càrrega de dades sigui la correcta. Les dimensions es carregaran primer i, després, les taules de fets perquè no hi hagi errors durant la càrrega.

D'una banda, alguns dels processos del bloc TR de càrrega i transformació de les dimensions són els següents:

Nom de l'ETL	Descripció	Taula d'origen	Taula de destinació (dimensions)
TR_DIM_COUNTRY	Càrrega de la informació relativa al país	STG_COUNTRY_LIST	DIM_COUNTRY
TR_DIM_SEX	Càrrega de la informació relativa al sexe de les persones	STG_ESTAT_SEX	DIM_SEX
TR_DIM_AGE	Càrrega de la informació relativa al tram d'edat	STG_ESTAT_AGE	DIM_AGE
TR_DIM_SECTOR	Càrrega de la informació relativa a l'activitat econòmica	STG_ESTAT_NACE_R2	DIM_SECTOR
TR_DIM_WORKTIME	Càrrega de la informació relativa al tipus d'horari laboral	STG_ESTAT_WORKTIME	DIM_WORKTIME
TR_DIM_WORKSTATUS	Càrrega de la informació relativa als estats laborals	STG_ESTAT_WSTATUS	DIM_WORKSTATUS
TR_DIM_N_CHILD	Càrrega de la informació relativa al nombre de fills	STG_ESTAT_N_CHILD	DIM_N_CHILD

Observem que els diversos processos de càrrega són similars entre si en termes, així doncs, el patró de càrrega seria el següent per al procés **TR_DIM_***:

1. Extracció de dades: Les dades s'extreuen de la taula d'origen TR_DIM_* tenint en compte el format del origen de les dades (json, csv, tsv o xlsx).
2. Transformació de les dades: Realització d'operacions de neteja i transformació de dades en funció de les necessitats de les dades originals.
3. Càrrega de dades: Les dades transformades són carregades a les corresponents taules **DIM_***.

D'altra banda, s'han d'identificar els processos de càrrega dels fets i les seves taules d'origen i destinació.

Nom de l'ETL	Descripció	Taula d'origen	Taula de destinació
TR_FAC_PCT_EMP PLOYEES_HOME	Càrrega i transformació de la taula de fets de 'FACT_PCT_EMP LOYEES_HOME'	STG_PCT_EMP LOYEES_HOME	FACT_PCT_EMP LOYEES_HOME
TR_FACT_HOURS _WORKED	Càrrega i transformació de la taula de fets de 'FACT_HOURS_ WORKED'	STG_HOURS_W ORKED STG_HOURS_W ORKED_USA	FACT_HOURS_ WORKED

3. Disseny dels processos ETL

En aquest apartat s'han de dissenyar els processos de càrrega identificats en el punt anterior amb l'eina de disseny proporcionada. En aquest cas és *Pentaho Data Integration* (PDI).

3.1 Creació de taules intermèdies (staging area)

En el nostre cas, la implementació la farem amb Microsoft SQL Server, sistema gestor de bases de dades relacional integrat a l'entorn VDI facilitat per la UOC. Utilitzarem un fitxer facilitat en l'enunciat de l'assignatura anomenat '*DLL_STG+DW.csv*'.

Aquest fitxer conté tots els scripts de creació de les taules de l'*staging area* (STG) i del magatzem de dades. Per a tal de usar-lo, iniciarem sessió a Microsoft SQL Server i farem una *nova query* sobre el nostre *source* '**SOURCE_braconsm**' (figura 1). A continuació hi posarem els script del fitxer '*DLL_STG+DW.csv*' i amb la tecla F5 farem un *run* sobre la *query* (figura 2) i si tot ha anat bé, veurem un missatge al terminal informant sobre la comanda finalitzada correctament (figura 3). Per assegurar que el procés s'ha dut a terme correctament podem mirar les taules creades a la carpeta *Tables* de **SOURCE_braconsm** (figura 4).

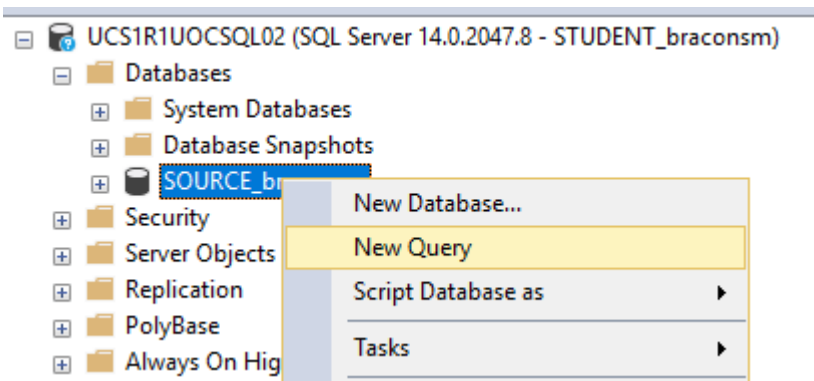


Figura 1. New Query

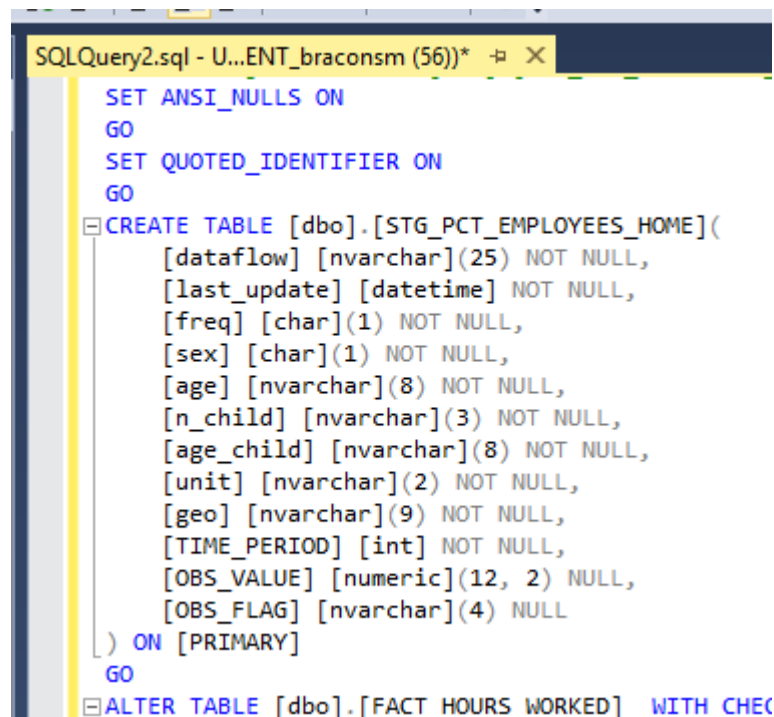


Figura 2. Script creació taules

```
SQLQuery2.sql - U...ENT_braconsm (56))* X
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
CREATE TABLE [dbo].[STG_PCT_EMPLOYEES_HOME](
    [dataflow] [nvarchar](25) NOT NULL,
    [last_update] [datetime] NOT NULL,
    [freq] [char](1) NOT NULL,
    [sex] [char](1) NOT NULL,
    [age] [nvarchar](8) NOT NULL,
    [n_child] [nvarchar](3) NOT NULL,
    [age_child] [nvarchar](8) NOT NULL,
    [unit] [nvarchar](2) NOT NULL,
    [geo] [nvarchar](9) NOT NULL,
    [TIME_PERIOD] [int] NOT NULL,
    [OBS_VALUE] [numeric](12, 2) NULL,
    [OBS_FLAG] [nvarchar](4) NULL
) ON [PRIMARY]
GO
ALTER TABLE [dbo].[FACT_HOURS_WORKED] WITH CHECK ADD FOREIGN KEY([fk_age])
REFERENCES [dbo].[DIM_AGE] ([pk_age])
GO
ALTER TABLE [dbo].[FACT_HOURS_WORKED] WITH CHECK ADD FOREIGN KEY([fk_country])
REFERENCES [dbo].[DIM_COUNTRY] ([pk_country])
GO
ALTER TABLE [dbo].[FACT_HOURS_WORKED] WITH CHECK ADD FOREIGN KEY([fk_sex])
REFERENCES [dbo].[DIM_SEX] ([pk_sex])
GO
```

100 %

Messages

Commands completed successfully.

Figura 3. Missatge de Query finalitzada correctament

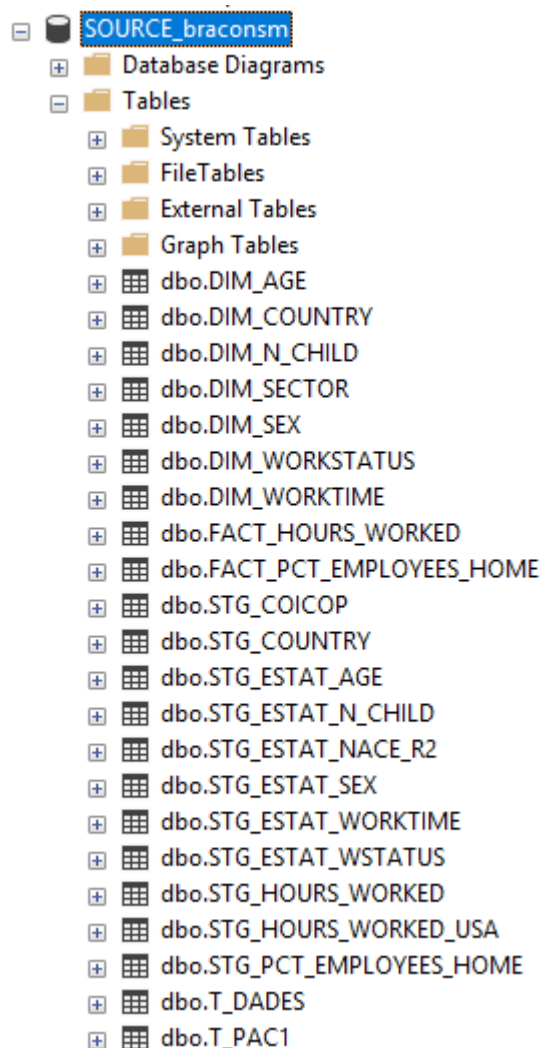


Figura 4. Taules creades a **SOURCE_braconsm**

3.1.1 Bloc IN

Primer de tot definirem el directori en el que estan les dades, per això anirem a ‘*Set Environment Variables*’ de Pentaho Data Integration i ho omplirem de la següent manera (figura 5):

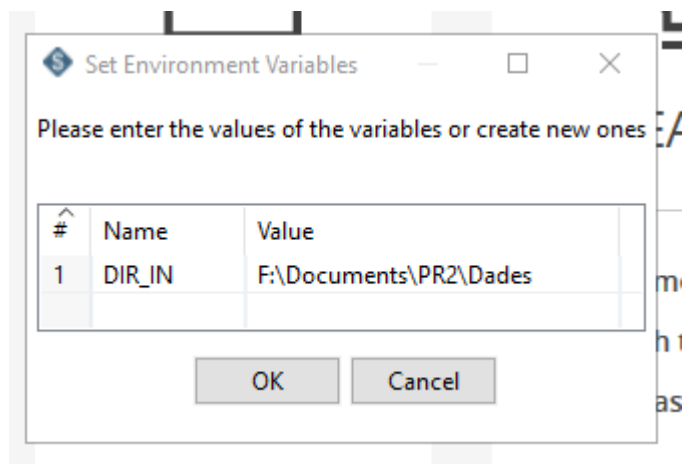


Figura 5. ‘*Set Environment Variables*’

Transformació d'IN_ESTAT_AGE

A continuació es descriu el desenvolupament de la transformació d'IN_ESTAT_AGE mitjançant Spoon. L'objectiu és carregar les dades originals de ESTAT_AGE_en.tsv, en la taula STG_ESTAT_AGE de l'àrea intermèdia (STG). La taula intermèdia ha estat creada en el punt anterior en la base de dades analítica mitjançant el script facilitat.

La transformació d'IN_ESTAT_AGE conté les etapes següents:

- Lectura del fitxer .tsv.
- Normalització de text a majúscules.
- Reordenació de files en ordre alfabètic.
- Selecció de valors.
- Càrrega a la taula intermèdia STG_ESTAT_AGE.

A continuació es detallen les diferents etapes que s'implementaran per poder fer la càrrega de dades.

En primer lloc, es crea una nova transformació i s'hi afegeix un ‘Text file input’. A continuació seleccionem el fitxer i indiquem la ruta amb *DIR_IN* (figura 6). En la pestanya ‘Content’ hi posem una tabulació com a Separador i marquem ‘Header’ indicant que la capçalera té 1 línia (figura 7). Ara volem obtenir els camps, per a això, anirem a la pestanya ‘Fields’ i farem click a ‘Get Fields’ (figura 8) i finalment podrem veure que la càrrega s’ha dut a terme correctament al donar click en ‘Preview rows’ (figura 9).

Per a normalitzar les dades farem servir 'String operations', seleccionarem les dades amb el botó 'Get Fields' i donarem el valor 'upper' a la columna 'Lower/Upper' (figura 10). El següent pas consisteix en ordenar les files, per això farem servir 'Sort rows' on, després de fer click en 'Get Fields', posarem 'Ascending' amb valor 'Y' (figura 11). Abans del pas final, hem de seleccionar els valors desitjats, això ho fem amb 'Select Values' (figura 12). Finalment, es carregaran les dades en la taula intermèdia de l'*stage* utilitzant el pas Table Output (figura 15). Per a aquest pas, serà necessari especificar la connexió de la base de dades, que es podrà generar dins del mateix component prement el botó New i indicant que es vol connectar a Microsoft SQL Server i utilitzar l'accés natiu amb les meves credencials (figura 13), fent posteriorment un test per veure que la connexió és correcta (figura 14). Si tot ha sortit correctament, hauríem de veure un missatge de finalització de la transformació al terminal (figura 16) i podríem veure les dades a la taula de Microsoft SQL Server (figura 17). La transformació completa es pot veure en la figura 18.

#	File/Directory	Wildcard (RegExp)	Exclude wildcard	Required	Include subfolders
1	\${DIR_IN}\ESTAT_AGE_en.tsv			N	N

Figura 6

Input fields

Filetype	CSV
Separator	
Enclosure	
Escape	
Header	<input checked="" type="checkbox"/> Number of header lines 1
Footer	<input type="checkbox"/> Number of footer lines 1
Wrapped lines?	<input type="checkbox"/> Number of times wrapped 1
Paged layout (printout)?	<input type="checkbox"/> Number of lines per page 80
	Document header lines 0

Figura 7

Text file input													
Step name: Text file input													
File	Content	Error Handling	Filters	Fields	Additional output fields								
#	Name	Type	Format	Position	Length	Precision	Currency	Decimal	Group	Null if	Default	Trim type	Repeat
1	TOTAL	String			9		\$.	,	-		none	N
2	Total	String			32		\$.	,	-		none	N

Figura 8



Examine preview data

Rows of step: Text file input (653 rows)

#	TOTAL	Total_1
1	LFD	Late foetal death
2	LFD1	Late foetal death (group 1)
3	LFD2	Late foetal death (group 2)
4	MN0	Zero minutes
5	MN1-9	From 1 to 9 minutes
6	MN1-14	From 1 to 14 minutes
7	MN1-59	From 1 to 59 minutes
8	MN1-74	From 1 to 74 minutes
9	MN1-149	From 1 to 149 minutes
10	MN_GE1	1 minute or over
11	MN10-14	From 10 to 14 minutes
12	MN15-19	From 15 to 19 minutes
13	MN15-29	From 15 to 29 minutes
14	MN20-29	From 20 to 29 minutes
15	MN30-44	From 30 to 44 minutes
16	MN30-59	From 30 to 59 minutes
17	MN_GE30	30 minutes or over
18	MN45-59	From 45 to 59 minutes
19	MN60-149	From 60 to 149 minutes
20	MN60-179	From 60 to 179 minutes
21	MN_GE60	60 minutes or over
22	MN75-149	From 75 to 149 minutes
23	MN150-299	From 150 to 299 minutes
24	MN_GE150	150 minutes or over
25	MN180-299	From 180 to 299 minutes
26	MN_GE300	300 minutes or over
27	H_VARY	Hours vary
28	H0	Zero hours
29	H_LT1	Less than 1 hour
30	H1-9	From 1 to 9 hours
31	H1-19	From 1 to 19 hours
32	H1-29	From 1 to 29 hours
33	H GE1	1 hour or over

Figura 9

String operations

Step name: String operations

The fields to process:

#	In stream field	Out stream field	Trim type	Lower/Upper	Padding	Pad char	Pad Length	InitCap	Escape	Digits	Remove Special character
1	TOTAL		none	upper	none			N	None	none	none
2	Total_1		none	upper	none			N	None	none	none

Help OK Get fields Cancel

Figura 10

Sort rows

Step name: Sort rows

Sort directory: %%java.io.tmpdir%% Browse...

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

Compress TMP Files? ☐

Only pass unique rows? (verifies keys only) ☐

Fields :

#	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1	TOTAL	Y	N	N	0	N
2	Total_1	Y	N	N	0	N

Help OK Cancel Get Fields

Figura 11

Select values

Step name
Select values

Select & Alter
Remove
Meta-data

Fields :

#	Fieldname	Rename to	Length	Precision
1	TOTAL	code		
2	Total_1	description		

Get fields to select

Edit Mapping

Include unspecified fields, ordered by name ☐

? Help
OK
Cancel

Figura 12

Connection name:
cn_stage

Connection type:

Ingres
Ingres VectorWise
Intersystems Cache
KingbaseES
LucidDB
MS Access
MS SQL Server
MS SQL Server (Native)
MariaDB
MaxDB (SAP DB)
MonetDB
MySQL
Native Mondrian
Neoview
Netezza
Oracle
Oracle RDB
Palo MOLAP Server
Pentaho Data Services

Access:
Native (JDBC)
ODBC
JNDI

Settings
Host Name:
UCS1R1UOCSQL02
Database Name:
SOURCE_braconsm
Instance Name:
Port Number:
1433
Username:
STUDENT_braconsm
Password:
Use Integrated Security
Use ... to Separate Schema and Table

Test
Feature List
Explore

OK
Cancel

Figura 13



Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2023/12/16 17:31:42 - Spoon - Running transformation using the Kettle execution engine

2023/12/16 17:31:42 - Spoon - Transformation opened.

2023/12/16 17:31:42 - Spoon - Launching transformation [IN_ESTAT_AGE]...

2023/12/16 17:31:42 - Spoon - Started the transformation execution.

2023/12/16 17:31:42 - IN_ESTAT_AGE - Dispatching started for transformation [IN_ESTAT_AGE]

2023/12/16 17:31:42 - Insert into SQL Server.0 - Connected to database [cn_stage] (commit=1000)

2023/12/16 17:31:42 - Text file input.0 - Opening file: file:///F:/Documents/PR2/Dades/ESTAT_AGE_en.tsv

2023/12/16 17:31:42 - Text file input.0 - Finished processing (I=654, O=0, R=0, W=653, U=1, E=0)

2023/12/16 17:31:42 - String operations.0 - Finished processing (I=0, O=0, R=653, W=653, U=0, E=0)

2023/12/16 17:31:42 - Sort rows.0 - Finished processing (I=0, O=0, R=653, W=653, U=0, E=0)

2023/12/16 17:31:42 - Select values.0 - Finished processing (I=0, O=0, R=653, W=653, U=0, E=0)

2023/12/16 17:31:42 - Insert into SQL Server.0 - Finished processing (I=0, O=653, R=653, W=653, U=0, E=0)

2023/12/16 17:31:42 - Spoon - The transformation has finished!!

Figura 16

100 %

Results Messages

	code	description
1	ADULT	ADULT
2	AVG	AVERAGE
3	CSA-Y12	FROM MINIMUM COMPULSORY SCHOOL AGE TO 12 YEARS
4	D0	ZERO DAYS
5	D0-27	FROM 0 TO 27 DAYS
6	D0-3	FROM 0 TO 3 DAYS
7	D1-3	FROM 1 TO 3 DAYS
8	D1-6	FROM 1 TO 6 DAYS
9	D1-7	FROM 1 TO 7 DAYS
10	D1-90	FROM 1 TO 90 DAYS
11	D10-19	FROM 10 TO 19 DAYS
12	D14-20	FROM 14 TO 20 DAYS
13	D14-29	FROM 14 TO 29 DAYS
14	D181-360	FROM 181 TO 360 DAYS
15	D2-3	FROM 2 TO 3 DAYS
16	D21-29	FROM 21 TO 29 DAYS
17	D21-M1	FROM 21 DAYS TO 1 MONTH
18	D28-364	FROM 28 DAYS TO LESS THAN 1 YEAR
19	D4-5	FROM 4 TO 5 DAYS
20	D4-6	FROM 4 TO 6 DAYS

Query executed successfully.

Figura 17

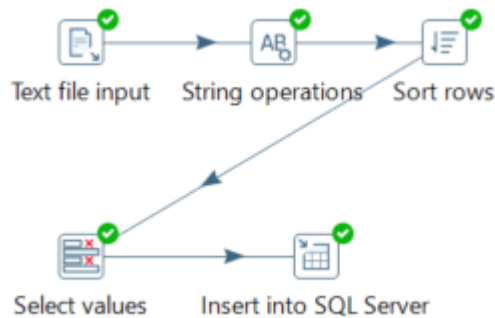


Figura 18

Transformació d'IN_ESTAT_N_CHILD

Fent exactament els mateixos passos que al punt anterior carreguem les dades del fitxer 'ESTAT_N_CHILD_en.tsv' (figura 21). Només hem canviat el nom del fitxer al input i el target table a 'STG_ESTAT_N_CHILD'. La càrrega ha sigut exitosa (figura 19) i podem veure les dades al servidor (figura 20).

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2023/12/16 17:36:53 - Spoon - Running transformation using the Kettle execution engine

2023/12/16 17:36:53 - Spoon - Transformation opened.

2023/12/16 17:36:53 - Spoon - Launching transformation [IN_ESTAT_N_CHILD]...

2023/12/16 17:36:53 - Spoon - Started the transformation execution.

2023/12/16 17:36:54 - IN_ESTAT_N_CHILD - Dispatching started for transformation [IN_ESTAT_N_CHILD]

2023/12/16 17:36:54 - Insert into SQL Server.0 - Connected to database [cn_stage] (commit=1000)

2023/12/16 17:36:54 - Text file input.0 - Opening file: file:///F:/Documents/PR2/Dades/ESTAT_N_CHILD_en.tsv

2023/12/16 17:36:54 - Text file input.0 - Finished processing (I=20, O=0, R=0, W=19, U=1, E=0)

2023/12/16 17:36:54 - String operations.0 - Finished processing (I=0, O=0, R=19, W=19, U=0, E=0)

2023/12/16 17:36:54 - Sort rows.0 - Finished processing (I=0, O=0, R=19, W=19, U=0, E=0)

2023/12/16 17:36:54 - Select values.0 - Finished processing (I=0, O=0, R=19, W=19, U=0, E=0)

2023/12/16 17:36:54 - Insert into SQL Server.0 - Finished processing (I=0, O=19, R=19, W=19, U=0, E=0)

2023/12/16 17:36:54 - Spoon - The transformation has finished!!

Figura 19

Results		Messages
	code	description
1	0	NO CHILDREN
2	1	1 CHILD
3	10	10 CHILDREN
4	2	2 CHILDREN
5	3	3 CHILDREN
6	4	4 CHILDREN
7	5	5 CHILDREN
8	6	6 CHILDREN
9	7	7 CHILDREN
10	8	8 CHILDREN
11	9	9 CHILDREN
12	GE1	1 CHILD OR MORE
13	GE3	3 CHILDREN OR MORE
14	GE4	4 CHILDREN OR MORE
15	GE5	5 CHILDREN OR MORE
16	GE6	6 CHILDREN OR MORE
17	GE7	7 CHILDREN OR MORE
18	NRP	NO RESPONSE
19	UNK	UNKNOWN

Query exe... UCS1R1UOCSQL02 (14.

Figura 20

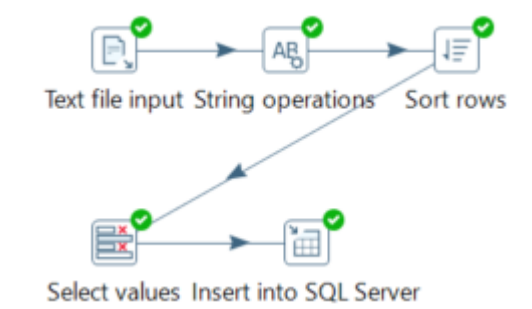


Figura 21

Transformació d'IN_ESTAT_SEX

Fent exactament els mateixos passos que als punts anteriors carreguem les dades del fitxer 'ESTAT_SEX_en.tsv'. Només hem canviat el nom del fitxer al input i el target table a 'STG_SEX' i a 'Select Values' el Fieldname 'T' és el code i 'Total' és description. La càrrega ha sigut exitosa (figura 22) i podem veure les dades al servidor (figura 23).

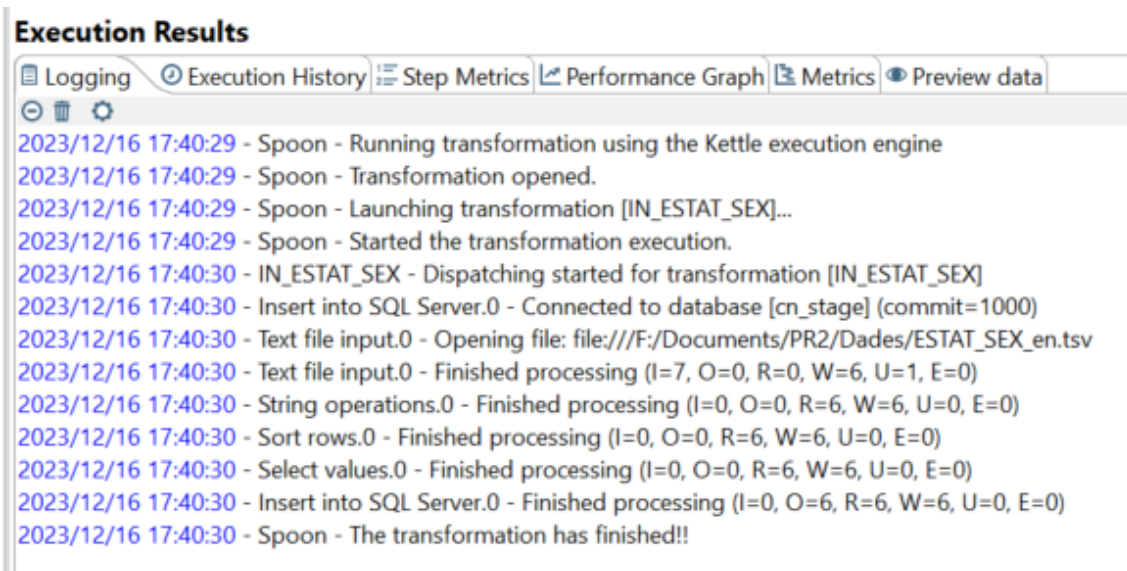


Figura 22

Results		
	code	description
1	DIFF	ABSOLUTE DIFFERENCE BETWEEN MALES AND FEMALES
2	F	FEMALES
3	M	MALES
4	NAP	NOT APPLICABLE
5	NRP	NO RESPONSE
6	UNK	UNKNOWN

Figura 23

Transformació d'IN_ESTAT_NACE_R2

Fent exactament els mateixos passos que als punts anteriors carreguem les dades del fitxer 'ESTAT_NACE_R2_en.tsv'. Només hem canviat el nom del fitxer al input i el target table a 'ESTAT_NACE_R2' i a 'Select Values' el Fieldname 'TOTAL' és el code i 'Total__all_NACE_activities' és description. La càrrega ha sigut exitosa i podem veure les dades al servidor (figura 24).

Results		Messages
	code	description
1	A	AGRICULTURE, FORESTRY AND FISHING
2	A-S	ALL NACE ACTIVITIES (EXCEPT ACTIVITIES OF HOUSEHOLDS A...
3	A-T	ALL NACE ACTIVITIES (EXCEPT EXTRA-TERRITORIAL ORGANIS...
4	A01	CROP AND ANIMAL PRODUCTION, HUNTING AND RELATED SE...
5	A011	GROWING OF NON-PERENNIAL CROPS
6	A0111	GROWING OF CEREALS (EXCEPT RICE), LEGUMINOUS CROPS ...
7	A0112	GROWING OF RICE
8	A0113	GROWING OF VEGETABLES AND MELONS, ROOTS AND TUBERS
9	A0114	GROWING OF SUGAR CANE
10	A0115	GROWING OF TOBACCO
11	A0116	GROWING OF FIBRE CROPS
12	A0119	GROWING OF OTHER NON-PERENNIAL CROPS
13	A012	GROWING OF PERENNIAL CROPS
14	A0121	GROWING OF GRAPES
15	A0122	GROWING OF TROPICAL AND SUBTROPICAL FRUITS
16	A0123	GROWING OF CITRUS FRUITS
17	A0124	GROWING OF POME FRUITS AND STONE FRUITS
18	A0125	GROWING OF OTHER TREE AND BUSH FRUITS AND NUTS
19	A0126	GROWING OF OLEAGINOUS FRUITS
20	A0127	GROWING OF BEVERAGE CROPS
<div> ✓ Query e... UCS1R1UOCSQL02 (14.0 RTM) STUDENT_braconsm (66) SOU </div>		

Figura 24

Transformació d'IN_ESTAT_WORKTIME

Fent exactament els mateixos passos que als punts anteriors carreguem les dades del fitxer 'ESTAT_WORKTIME_en.tsv'. Només hem canviat el nom del fitxer al input i el target table a 'ESTAT_WORKTIME' i a 'Select Values' el Fieldname 'TOTAL' és el code i 'Total_1' és description. La càrrega ha sigut exitosa i podem veure les dades al servidor (figura 25).

Results		Messages
	code	description
1	ATYP	ATYPICAL WORKING TIME
2	AVG_FTE	YEARLY AVERAGE IN FULL-TIME EQUIVALENTS
3	DECR	DECREASE WORKING HOURS
4	FT	FULL-TIME
5	H0	ZERO HOURS
6	H1-19	FROM 1 TO 19 HOURS
7	H1-9	FROM 1 TO 9 HOURS
8	H10-19	FROM 10 TO 19 HOURS
9	H20-29	FROM 20 TO 29 HOURS
10	H20-31	FROM 20 TO 31 HOURS
11	H30-39	FROM 30 TO 39 HOURS
12	H40-49	FROM 40 TO 49 HOURS
13	H_GE32	32 HOURS OR OVER
14	H_GE50	50 HOURS OR OVER
15	INCR	INCREASE WORKING HOURS
16	NAP	NOT APPLICABLE
17	NCHG	NO CHANGES TO WORKING HOURS
18	NORM	NORMAL WORKING TIME
19	NRP	NO RESPONSE
20	P00	ZERO PERCENT OF A FULL TIME

Query exe...
UCS1R1UOCSQL02 (14.0 RTM)
STUDENT_braco

Figura 25

Transformació d'IN_ESTAT_WSTATUS

Fent exactament els mateixos passos que als punts anteriors carreguem les dades del fitxer 'ESTAT_WSTATUS_en.tsv'. Només hem canviat el nom del fitxer al input i el target table a 'ESTAT_WSTATUS' i a 'Select Values' el Fieldname 'TOTAL' és el code i 'Total_1 és description. La càrrega ha sigut exitosa i podem veure les dades al servidor (figura 26).

	code	description
1	ACT	PERSONS IN THE LABOUR FORCE (FORMER NAME: ACTIVE PE...
2	ACT_INAC	PERSONS IN- AND OUTSIDE THE LABOUR FORCE (FORMER NA...
3	ACT_UNK	PERSONS IN THE LABOUR FORCE (FORMER NAME: ACTIVE PO...
4	B_AGE_MIN	PERSONS BELOW THE NATIONAL MINIMUM AGE FOR ECONOM...
5	CFAM	CONTRIBUTING FAMILY WORKERS
6	CFAM_COOP	CONTRIBUTING FAMILY WORKERS AND MEMBERS OF PRODU...
7	CLIENT	CLIENTS
8	COOP	MEMBERS OF PRODUCERS' COOPERATIVES
9	COWN	CO-OWNER
10	EDUC	STUDENTS
11	EMP	EMPLOYED PERSONS
12	EMP_FT	EMPLOYED PERSONS WORKING FULL-TIME
13	EMP_OTH	OTHER EMPLOYED PERSONS
14	EMP_PT	EMPLOYED PERSONS WORKING PART-TIME
15	EMP_TEMP	EMPLOYED PERSONS WITH TEMPORARY CONTRACT
16	EMP_UNK	EMPLOYED PERSONS - UNKNOWN
17	EMP_W	PERSONS EMPLOYED OR PREVIOUSLY EMPLOYED
18	FLF_D	FARM LABOUR FORCE, DIRECTLY EMPLOYED BY THE FARM O...
19	FLF_D_NR	FARM LABOUR FORCE, DIRECTLY EMPLOYED BY THE FARM O...
20	FLF_D_P	FARM LABOUR FORCE, DIRECTLY EMPLOYED BY THE FARM O...

Query exe... | UCS1R1UOCSQL02 (14.0 RTM) | STUDENT_braconsm (62) | SOURCE_b...

Figura 26

Transformació d'IN_COUNTRY

En aquest cas el fitxer d'origen és un json així que farem servir 'JSON input' on hi especificarem el directori al fitxer i a la pestanya 'Fields' al fer click en 'Select Fields' seleccionarem les caixetes de 'name:', 'code:' i 'region:' (figura 27). La resta del job és igual que en els punts anteriors (figura 28). Podem veure que la càrrega ha sigut exitosa en el servidor (figura 30).

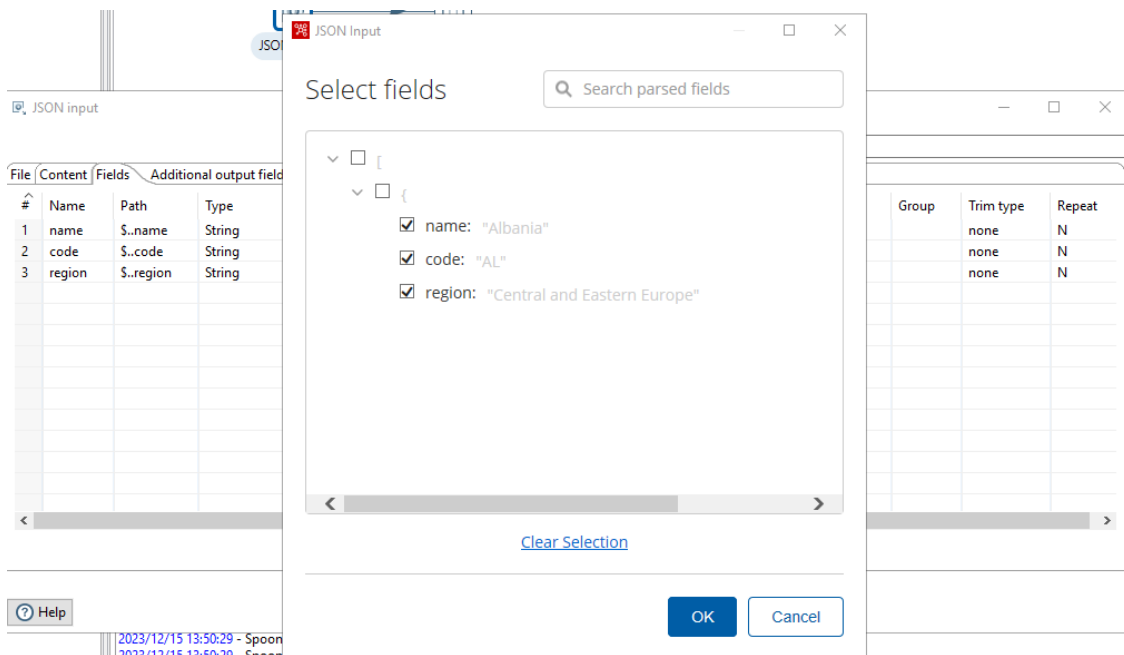


Figura 27

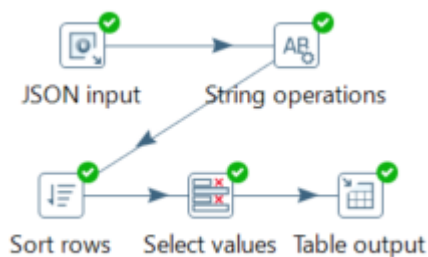


Figura 28

Results Messages			
	code	name	region
1	AL	ALBANIA	CENTRAL AND EASTERN EUROPE
2	AD	ANDORRA	WESTERN EUROPE
3	AR	ARGENTINA	SOUTH AMERICA
4	AM	ARMENIA	CENTRAL AND EASTERN EUROPE
5	AT	AUSTRIA	WESTERN EUROPE
6	AZ	AZERBAIJAN	CENTRAL AND EASTERN EUROPE
7	BY	BELARUS	CENTRAL AND EASTERN EUROPE
8	BE	BELGIUM	WESTERN EUROPE
9	BZ	BELIZE	CENTRAL AMERICA
10	BM	BERMUDA	NORTH AMERICA
11	BO	BOLIVIA	SOUTH AMERICA
12	BA	BOSNIA AND HERZEGOVINA	CENTRAL AND EASTERN EUROPE
13	BR	BRAZIL	SOUTH AMERICA
14	BG	BULGARIA	CENTRAL AND EASTERN EUROPE
15	CA	CANADA	NORTH AMERICA
16	CL	CHILE	SOUTH AMERICA
17	CO	COLOMBIA	SOUTH AMERICA
18	CR	COSTA RICA	CENTRAL AMERICA
19	HR	CROATIA	CENTRAL AND EASTERN EUROPE
20	CU	CUBA	NORTH AMERICA

✓ Query executed successfully.

Figura 30

Transformació d'IN_PCT_EMPLOYEES_HOME

En aquest cas el fitxer d'origen és un csv.gz així que hem fet servir 'GZIP CSV input' (figura 31). AL fer click en 'Get Fields' hem obtingut tots els atributs i el seu tipus. Ho hem deixat tot igual menys el de n_child que per defecte s'ha detectat de tipus 'int' però és de tipus 'string', ja que té valors com GE3. Al igual que en els punts anteriors hem aplicat 'String operations' i 'Sort rows'. Com en aquest cas hi ha valors <null> a OBS_VALUE i OBS_FLAG, hem fet servir 'If field value is null' per donar 'None' als de OBS_FLAG i hem eliminat els de OBS_VALUE amb 'Sort rows', ja que no ens serveixen (figura 32). A continuació, em fet servir 'Select Values' per canviar el nom dels 'Fieldname' quan ha sigut necessari (figura 33) i finalment ho hem carregat al servidor (figura 34).

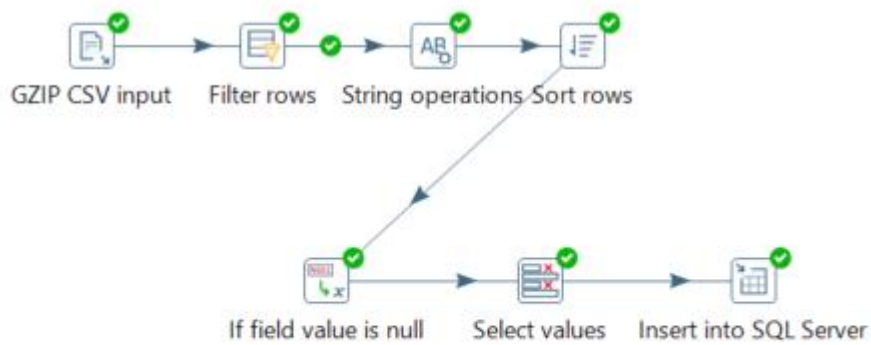


Figura 31

Filter rows

Step name

Filter rows

Send 'true' data to step:

String operations

Send 'false' data to step:

The condition:

NOT

OBS_VALUE

IS NULL

-

Help

OK

Cancel

If field value is null

Step name

If field value is null

Replace Null for all fields

Replace by value

Set empty string?

☐

Mask (Date)

Select fields

☒

Select value type

☐

Value types

#	Type	Replace by value	Conversion mask (Date)	Set empty string?

Fields

#	Field	Replace by value	Conversion mask (Date)	Set empty string?
1	DATAFLOW			N
2	LAST UPDATE			N
3	freq			N
4	sex			N
5	age			N
6	n_child			N
7	agechild			N
8	unit			N
9	geo			N
1.	TIME_PERIOD			N
1.	OBS_VALUE			N
1.	OBS_FLAG	None		N

Help

OK

Get Fields

Cancel

Figura 32

Select values

Step name

Select & Alter

Remove

Meta-data

Fields:

#	Fieldname	Rename to	Length	Precision
1	DATAFLOW	dataflow		
2	LAST UPDATE	last_update		
3	freq			
4	sex			
5	age			
6	n_child			
7	agechild	age_child		
8	unit			
9	geo			
10	TIME_PERIOD			
11	OBS_VALUE			
12	OBS_FLAG			

Get fields to select

Edit Mapping

☐ Include unspecified fields, ordered by name

Help

OK

Cancel

Figura 33

Results		Messages										
	dataflow	last_update	freq	sex	age	n_child	age_child	unit	geo	TIME_PERIOD	OBS_VALUE	OBS_FLAG
1	ESTAT:LFST_HHWAHCHI(1.0)	2023-02-15 23:00:00.000	A	F	Y18-24	0	NAP	PC	AT	2017	9.70	None
2	ESTAT:LFST_HHWAHCHI(1.0)	2023-02-15 23:00:00.000	A	F	Y18-24	0	NAP	PC	AT	2018	10.00	None
3	ESTAT:LFST_HHWAHCHI(1.0)	2023-02-15 23:00:00.000	A	F	Y18-24	0	NAP	PC	AT	2019	9.00	None
4	ESTAT:LFST_HHWAHCHI(1.0)	2023-02-15 23:00:00.000	A	F	Y18-24	0	NAP	PC	AT	2020	19.30	None
5	ESTAT:LFST_HHWAHCHI(1.0)	2023-02-15 23:00:00.000	A	F	Y18-24	0	NAP	PC	AT	2021	21.90	B
6	ESTAT:LFST_HHWAHCHI(1.0)	2023-02-15 23:00:00.000	A	F	Y18-24	0	NAP	PC	BE	2017	13.70	B
7	ESTAT:LFST_HHWAHCHI(1.0)	2023-02-15 23:00:00.000	A	F	Y18-24	0	NAP	PC	BE	2018	15.10	None
8	ESTAT:LFST_HHWAHCHI(1.0)	2023-02-15 23:00:00.000	A	F	Y18-24	0	NAP	PC	BE	2019	14.20	None
9	ESTAT:LFST_HHWAHCHI(1.0)	2023-02-15 23:00:00.000	A	F	Y18-24	0	NAP	PC	BE	2020	18.20	U
10	ESTAT:LFST_HHWAHCHI(1.0)	2023-02-15 23:00:00.000	A	F	Y18-24	0	NAP	PC	BE	2021	20.00	B
11	ESTAT:LFST_HHWAHCHI(1.0)	2023-02-15 23:00:00.000	A	F	Y18-24	0	NAP	PC	CY	2021	7.80	BU
12	ESTAT:LFST_HHWAHCHI(1.0)	2023-02-15 23:00:00.000	A	F	Y18-24	0	NAP	PC	CZ	2017	4.70	U
13	ESTAT:LFST_HHWAHCHI(1.0)	2023-02-15 23:00:00.000	A	F	Y18-24	0	NAP	PC	CZ	2018	2.80	U
14	ESTAT:LFST_HHWAHCHI(1.0)	2023-02-15 23:00:00.000	A	F	Y18-24	0	NAP	PC	CZ	2019	3.70	U

Figura 34

Transformació d'IN_HOURS_WORKED

Al igual que en el punt anterior el fitxer és un csv.gz. Hem seguit els mateixos passos i em carregat les dades al servidor correctament (figura 35).

	dataflow	last_update	freq	nace_r2	wstatus	worktime	age	sex	unit	geo	TIME_PERIOD	OBS_VALUE	OBS_FLAG
1	ESTAT:LFSA_EWHUN2(1.0)	15/02/23 23:00:00	A	A	CFAM	FT	Y15-24	F	HR	EU27_2020	2017	35.10	None
2	ESTAT:LFSA_EWHUN2(1.0)	15/02/23 23:00:00	A	A	CFAM	FT	Y15-24	F	HR	EU27_2020	2018	34.00	None
3	ESTAT:LFSA_EWHUN2(1.0)	15/02/23 23:00:00	A	A	CFAM	FT	Y15-24	F	HR	EU27_2020	2019	37.50	None
4	ESTAT:LFSA_EWHUN2(1.0)	15/02/23 23:00:00	A	A	CFAM	FT	Y15-24	F	HR	EU27_2020	2020	35.10	U
5	ESTAT:LFSA_EWHUN2(1.0)	15/02/23 23:00:00	A	A	CFAM	FT	Y15-24	F	HR	EU27_2020	2021	37.50	BU
6	ESTAT:LFSA_EWHUN2(1.0)	15/02/23 23:00:00	A	A	CFAM	FT	Y15-24	F	HR	FR	2017	19.30	U
7	ESTAT:LFSA_EWHUN2(1.0)	15/02/23 23:00:00	A	A	CFAM	FT	Y15-24	F	HR	MK	2017	45.40	None
8	ESTAT:LFSA_EWHUN2(1.0)	15/02/23 23:00:00	A	A	CFAM	FT	Y15-24	F	HR	MK	2018	45.30	None
9	ESTAT:LFSA_EWHUN2(1.0)	15/02/23 23:00:00	A	A	CFAM	FT	Y15-24	F	HR	MK	2019	43.80	U
10	ESTAT:LFSA_EWHUN2(1.0)	15/02/23 23:00:00	A	A	CFAM	FT	Y15-24	F	HR	MK	2020	40.20	U
11	ESTAT:LFSA_EWHUN2(1.0)	15/02/23 23:00:00	A	A	CFAM	FT	Y15-24	F	HR	PL	2017	45.60	U
12	ESTAT:LFSA_EWHUN2(1.0)	15/02/23 23:00:00	A	A	CFAM	FT	Y15-24	F	HR	RO	2017	32.60	None
13	ESTAT:LFSA_EWHUN2(1.0)	15/02/23 23:00:00	A	A	CFAM	FT	Y15-24	F	HR	RO	2018	31.70	None
14	ESTAT:LFSA_EWHUN2(1.0)	15/02/23 23:00:00	A	A	CFAM	FT	Y15-24	F	HR	RO	2019	34.80	None

Figura 35

Transformació d'IN_HOURS_WORKED_USA

En aquest cas el fitxer d'origen de les dades és un 'xlsx' així que hem fet servir 'Microsoft Excel Input'. A la pestanya 'Sheets' hem seleccionat el Sheet name 'CES0500000002', start row 11 i start column 0. A field hem obtingut les columnes 'Year', 'Period', 'Label' i 'Observation Value'. Hem normalitzat fent usant 'String operations' i a continuació hem fet servir 'Select Values' per a ajustar els noms als de la taula del servidor i finalment hem fet la càrrega de tota la transformació (figures 36 i 37). Observem que en aquesta taula hi ha informació repetida, ja que 'year' i 'period' ja tenen la informació de 'label'. Com la taula que s'ha donat a l'enunciat té per omplir els tres camps, no hem borrrat la columna label.

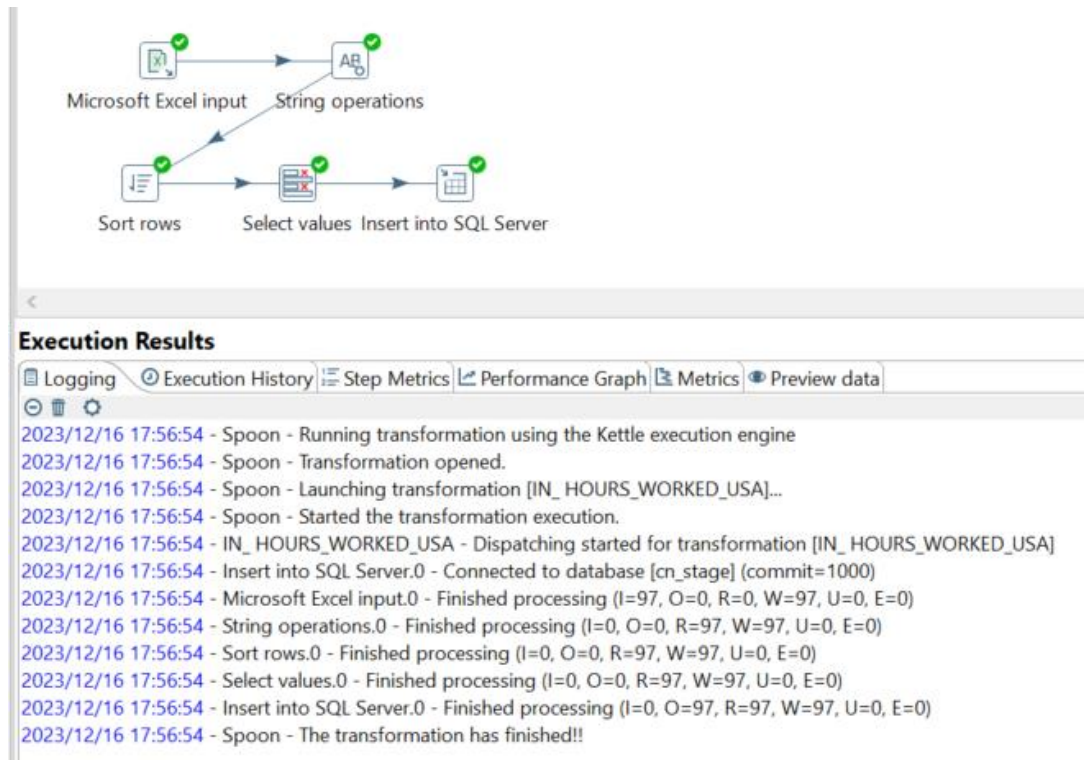


Figura 36

Results		Messages		
	year	period	label	observation_value
1	2015	M01	2015 JAN	34.50
2	2015	M02	2015 FEB	34.60
3	2015	M03	2015 MAR	34.50
4	2015	M04	2015 APR	34.50
5	2015	M05	2015 MAY	34.50
6	2015	M06	2015 JUN	34.50
7	2015	M07	2015 JUL	34.50
8	2015	M08	2015 AUG	34.60
9	2015	M09	2015 SEP	34.50
10	2015	M10	2015 OCT	34.50
11	2015	M11	2015 NOV	34.50
12	2015	M12	2015 DEC	34.50
13	2016	M01	2016 JAN	34.60
14	2016	M02	2016 FEB	34.40
15	2016	M03	2016 MAR	34.40
16	2016	M04	2016 APR	34.40
17	2016	M05	2016 MAY	34.40
18	2016	M06	2016 JUN	34.40
19	2016	M07	2016 JUL	34.40
20	2016	M08	2016 AUG	34.20

✓ Query executed successfully.

Figura 37

3.1.2 Bloc TR_DIM

Aquest bloc conté les transformacions per a la càrrega inicial de les dimensions al magatzem des de les taules intermèdies «IN_» de l' staging area. Es tindrà en compte que en una càrrega inicial poden executar-se les transformacions de càrrega de dimensions les vegades que siguin necessàries.

Transformació TR_DIM_COUNTRY

Per a fer aquesta transformació recuperarem les dades de la taula STG_COUNTRY mitjançant 'Table Input' on farem la següent crida SQL:

```
SELECT [code]
      , [name]
      , [region]
FROM [SOURCE_braconsm].[dbo].[STG_COUNTRY]
```

I per a la creació de pk_country farem servir el bloc 'Add sequence' on posarem un valor incremental que comenci en 1 (figura 38)

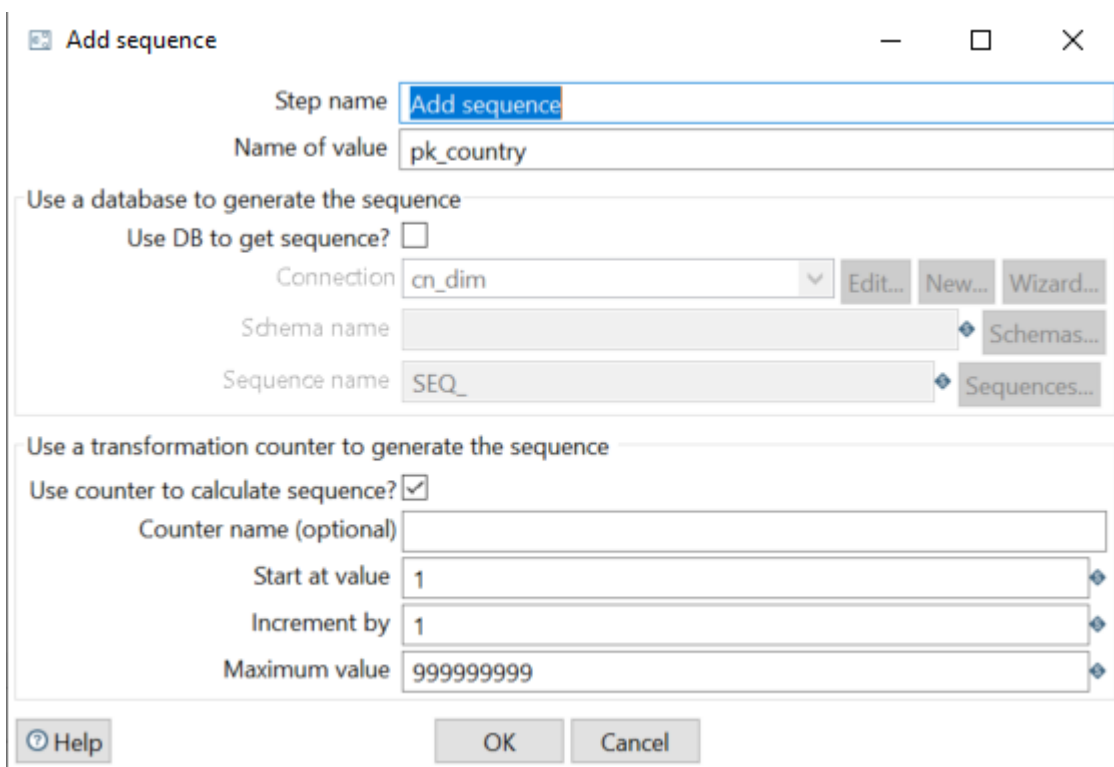


Figura 38

Al igual que hem fet anteriorment farem ús de 'Select Values' per a seleccionar les Dades que volem importar a la taula DIM_COUNTRY i canviar el nom de la columna 'name' a 'country'. Finalment, amb 'Table output' i una connexió a la nostre base de dades, carregarem les dades (figura 39). La transformació completa es pot veure a la figura 40.

SQLQuery12.sql -...ENT_bracons (62) X

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP (1000) [pk_country]
      , [code]
      , [country]
      , [region]
FROM [SOURCE_bracons]. [dbo]. [DIM_COUNTRY]

```

100 %

Results Messages

	pk_country	code	country	region
1	1	AL	ALBANIA	CENTRAL AND EASTERN EUROPE
2	2	AD	ANDORRA	WESTERN EUROPE
3	3	AR	ARGENTINA	SOUTH AMERICA
4	4	AM	ARMENIA	CENTRAL AND EASTERN EUROPE
5	5	AT	AUSTRIA	WESTERN EUROPE
6	6	AZ	AZERBAIJAN	CENTRAL AND EASTERN EUROPE
7	7	BY	BELARUS	CENTRAL AND EASTERN EUROPE
8	8	BE	BELGIUM	WESTERN EUROPE
9	9	BZ	BELIZE	CENTRAL AMERICA
10	10	BM	BERMUDA	NORTH AMERICA
11	11	BO	BOLIVIA	SOUTH AMERICA

Figura 39

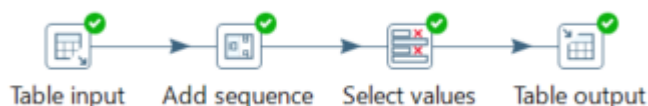


Figura 40

Transformació TR_DIM_SEX

Al igual que en el cas anterior, farem una crida SQL per a recuperar les dades,

```
SELECT [code], [description]
FROM [SOURCE_braconsms].[dbo].[STG_ESTAT_SEX];
```

I podem veure que la càrrega s'ha fet correctament al servidor en la figura 41.

SQLQuery14.sql -...ENT_braconsms (62)

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP (1000) [pk_sex]
, [code]
, [sex]
FROM [SOURCE_braconsms].[dbo].[DIM_SEX]
  
```

100 %

Results Messages

	pk_sex	code	sex
1	1	DIFF	ABSOLUTE DIFFERENCE BETWEEN MALES AND FEMALES
2	2	F	FEMALES
3	3	M	MALES
4	4	NAP	NOT APPLICABLE
5	5	NRP	NO RESPONSE
6	6	UNK	UNKNOWN

Figura 41

Transformació TR_DIM_AGE

Al igual que en el cas anterior, farem una crida SQL per a recuperar les dades i em fet la càrrega a la base de dades (figura 42)

SQLQuery23.sql -...ENT_braconsms (63)

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP (1000) [pk_age]
, [code]
, [age]
FROM [SOURCE_braconsms].[dbo].[DIM_AGE]
  
```

100 %

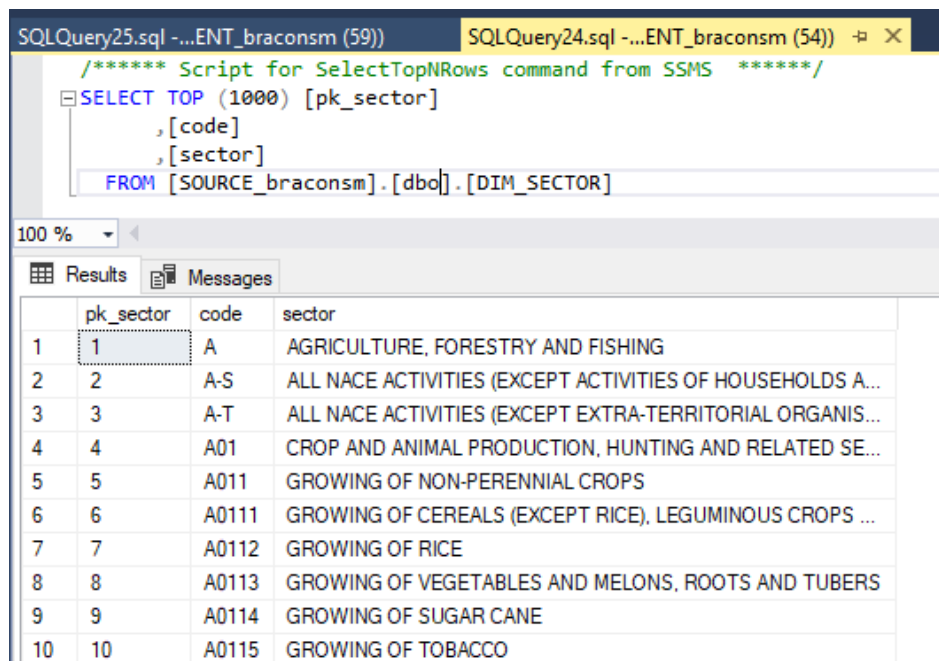
Results Messages

	pk_age	code	age
1	1	ADULT	ADULT
2	2	AVG	AVERAGE
3	3	CSA-Y12	FROM MINIMUM COMPULSORY SCHOOL AGE TO 12 YEARS
4	4	D0	ZERO DAYS
5	5	D0-27	FROM 0 TO 27 DAYS
6	6	D0-3	FROM 0 TO 3 DAYS
7	7	D1-3	FROM 1 TO 3 DAYS
8	8	D1-6	FROM 1 TO 6 DAYS
9	9	D1-7	FROM 1 TO 7 DAYS
10	10	D1-90	FROM 1 TO 90 DAYS
11	11	D10-19	FROM 10 TO 19 DAYS
12	12	D14-20	FROM 14 TO 20 DAYS

Figura 42

Transformació TR_DIM_AGE

Al igual que en el cas anterior, farem una crida SQL per a recuperar les dades i em fet la càrrega a la base de dades (figura 43)



SQLQuery25.sql -...ENT_bracons (59)) SQLQuery24.sql -...ENT_bracons (54))

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP (1000) [pk_sector]
, [code]
, [sector]
FROM [SOURCE_bracons]. [dbo]. [DIM_SECTOR]

```

100 %

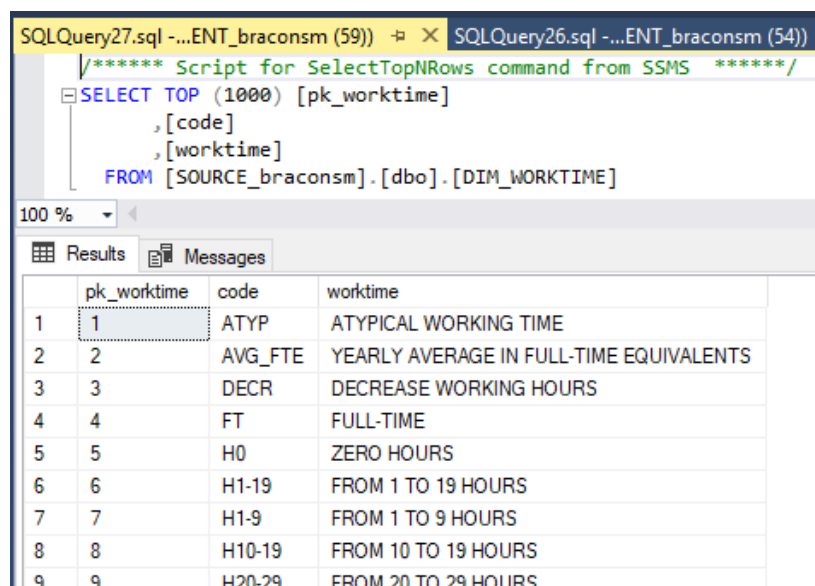
Results Messages

	pk_sector	code	sector
1	1	A	AGRICULTURE, FORESTRY AND FISHING
2	2	A-S	ALL NACE ACTIVITIES (EXCEPT ACTIVITIES OF HOUSEHOLDS A...
3	3	A-T	ALL NACE ACTIVITIES (EXCEPT EXTRA-TERRITORIAL ORGANIS...
4	4	A01	CROP AND ANIMAL PRODUCTION, HUNTING AND RELATED SE...
5	5	A011	GROWING OF NON-PERENNIAL CROPS
6	6	A0111	GROWING OF CEREALS (EXCEPT RICE), LEGUMINOUS CROPS ...
7	7	A0112	GROWING OF RICE
8	8	A0113	GROWING OF VEGETABLES AND MELONS, ROOTS AND TUBERS
9	9	A0114	GROWING OF SUGAR CANE
10	10	A0115	GROWING OF TOBACCO

Figura 43

Transformació TR_DIM_WORKTIME

Al igual que en el cas anterior, farem una crida SQL per a recuperar les dades i em fet la càrrega a la base de dades (figura 44)



SQLQuery27.sql -...ENT_bracons (59)) SQLQuery26.sql -...ENT_bracons (54))

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP (1000) [pk_worktime]
, [code]
, [worktime]
FROM [SOURCE_bracons]. [dbo]. [DIM_WORKTIME]

```

100 %

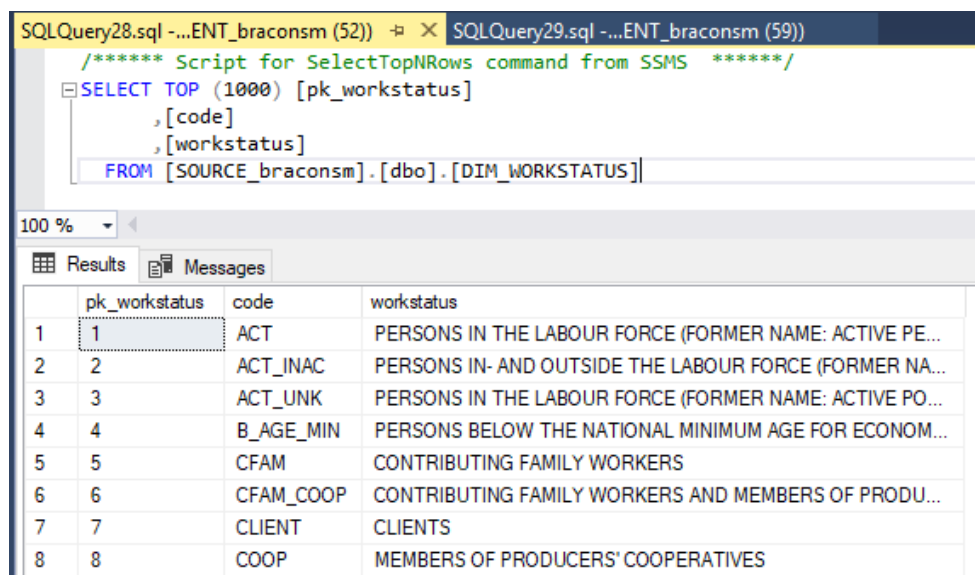
Results Messages

	pk_worktime	code	worktime
1	1	ATYP	ATYPICAL WORKING TIME
2	2	AVG_FTE	YEARLY AVERAGE IN FULL-TIME EQUIVALENTS
3	3	DECR	DECREASE WORKING HOURS
4	4	FT	FULL-TIME
5	5	H0	ZERO HOURS
6	6	H1-19	FROM 1 TO 19 HOURS
7	7	H1-9	FROM 1 TO 9 HOURS
8	8	H10-19	FROM 10 TO 19 HOURS
9	9	H20-29	FROM 20 TO 29 HOURS

Figura 44

Transformació TR_DIM_WORKSTATUS

Al igual que en el cas anterior, farem una crida SQL per a recuperar les dades i em fet la càrrega a la base de dades (figura 45)



```

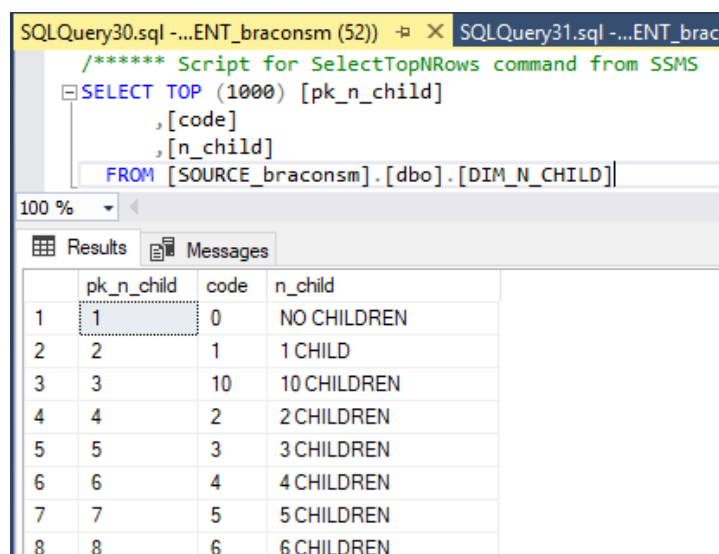
/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP (1000) [pk_workstatus]
, [code]
, [workstatus]
FROM [SOURCE_bracons]. [dbo]. [DIM_WORKSTATUS]
  
```

	pk_workstatus	code	workstatus
1	1	ACT	PERSONS IN THE LABOUR FORCE (FORMER NAME: ACTIVE PE...
2	2	ACT_INAC	PERSONS IN- AND OUTSIDE THE LABOUR FORCE (FORMER NA...
3	3	ACT_UNK	PERSONS IN THE LABOUR FORCE (FORMER NAME: ACTIVE PO...
4	4	B_AGE_MIN	PERSONS BELOW THE NATIONAL MINIMUM AGE FOR ECONOM...
5	5	CFAM	CONTRIBUTING FAMILY WORKERS
6	6	CFAM_COOP	CONTRIBUTING FAMILY WORKERS AND MEMBERS OF PRODU...
7	7	CLIENT	CLIENTS
8	8	COOP	MEMBERS OF PRODUCERS' COOPERATIVES

Figura 45

Transformació TR_DIM_N_CHILD

Al igual que en el cas anterior, farem una crida SQL per a recuperar les dades i em fet la càrrega a la base de dades (figura 46)



```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP (1000) [pk_n_child]
, [code]
, [n_child]
FROM [SOURCE_bracons]. [dbo]. [DIM_N_CHILD]
  
```

	pk_n_child	code	n_child
1	1	0	NO CHILDREN
2	2	1	1 CHILD
3	3	10	10 CHILDREN
4	4	2	2 CHILDREN
5	5	3	3 CHILDREN
6	6	4	4 CHILDREN
7	7	5	5 CHILDREN
8	8	6	6 CHILDREN

Figura 46

3.1.2 Bloc TR_FACT

Aquest bloc conté les transformacions per a la càrrega inicial de les taules de fet al magatzem des de les taules intermèdies «STG_» de l'staging area.

Transformació TR_FAC_PCT_EMPLOYEES_HOME

Per omplir aquesta taula de fets volem crear relacions entre les taules de dimensions i la taula 'STG_PCT_EMPLOYEES_HOME', per fer-ho farem servir la següent sentència de SQL:

```
SELECT
    s.[TIME_PERIOD] as [year],
    s.[unit],
    (SELECT pk_country FROM [SOURCE_bracons].[dbo].[DIM_COUNTRY] WHERE [code]
= s.[geo]) AS fk_country,
    (SELECT pk_age FROM [SOURCE_bracons].[dbo].[DIM_AGE] WHERE [code] =
s.[age]) AS fk_age,
    (SELECT pk_sex FROM [SOURCE_bracons].[dbo].[DIM_SEX] WHERE [code] =
s.[sex]) AS fk_sex,
    (SELECT pk_n_child FROM [SOURCE_bracons].[dbo].[DIM_N_CHILD] WHERE [code]
= s.[n_child]) AS fk_n_child,
    (SELECT pk_age FROM [SOURCE_bracons].[dbo].[DIM_AGE] WHERE [code] =
s.[age_child]) AS fk_age_child,
    s.[OBS_VALUE]
FROM
    [SOURCE_bracons].[dbo].[STG_PCT_EMPLOYEES_HOME] s;
```

a més, estem agafant canviant el nom de la columna TIME_PERIOD a year. A continuació filtre els possibles valors nuls de fk_country, ja que volem un model on no es perdin registres alhora d'analitzar les dades per una dimensió. Al igual que en les dimensions, afegirem la clau principal amb 'Add sequence' i carregarem la taula amb 'Table output' (figura 47)



Figura 47

Podem veure que s'ha carregat correctament a la base de dades a la figura 48

SQLQuery52.sql -...ENT_bracons (63)) SQLQuery51.sql -...ENT_bracons (57) + X

/****** Script for SelectTopNRows command from SSMS *****/

```

SELECT TOP (1000) [pk_id]
, [year]
, [unit]
, [fk_country]
, [fk_age]
, [fk_sex]
, [fk_n_child]
, [fk_age_child]
, [OBS_VALUE]
FROM [SOURCE_bracons].[dbo].[FACT_PCT_EMPLOYEES_HOME]

```

100 %

Results Messages

	pk_id	year	unit	fk_country	fk_age	fk_sex	fk_n_child	fk_age_child	OBS_VALUE
1	1	2017	PC	8	605	3	1	276	32.20
2	2	2018	PC	8	605	3	1	276	30.60
3	3	2019	PC	8	605	3	1	276	31.10
4	4	2020	PC	8	605	3	1	276	39.70
5	5	2021	PC	8	605	3	1	276	43.10
6	6	2017	PC	5	605	3	1	276	29.50
7	7	2018	PC	5	605	3	1	276	31.60
8	8	2019	PC	5	605	3	1	276	31.10
9	9	2020	PC	5	605	3	1	276	36.50

Figura 48

Transformació TR_FACT_HOURS_WORKED

Per a aquesta transformació hem hagut de tractar per una banda 'STG_HOURS_WORKED_USA' i per l'altre 'STG_WORKED_HOURS'. Després de tractar cada una de les taules les hem ajuntat i finalment n'hem fet la càrrega a la base de dades (figura 49)

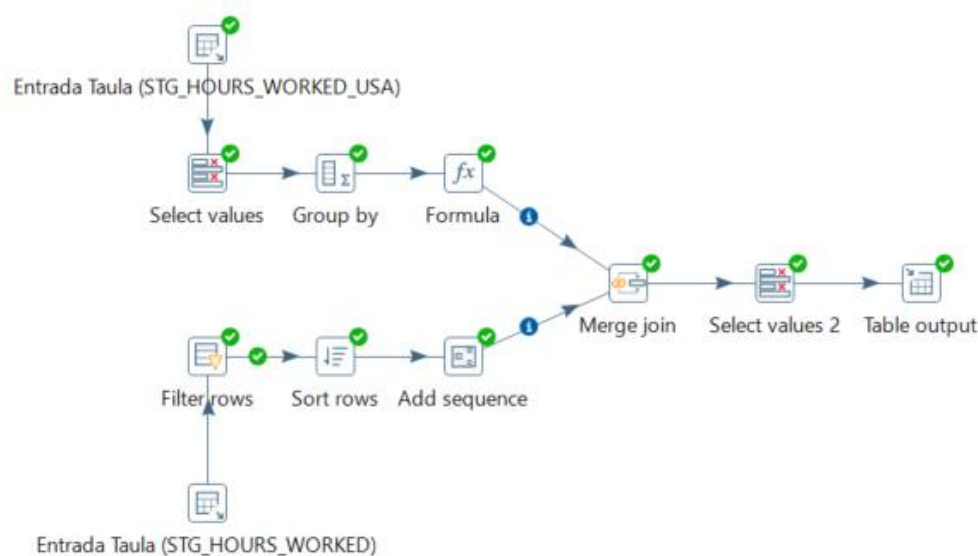


Figura 49

- STG_HOURS_WORKED_USA: S'ha carregat amb la següent sentència de SQL

```
SELECT [year]
      ,[period]
      ,[label]
      ,[observation_value]
FROM [SOURCE_braconsm].[dbo].[STG_HOURS_WORKED_USA]
```

A continuació s'han eliminat les columnes que no són necessàries mitjançant 'Select Values', quedant només amb 'year' i 'observation_value'. Posteriorment hem fet un 'Group by' ajuntant per 'year', guardant els resultats a una columna auxiliar 'obs_value_tmp' (figura 50).

Group by

Step name: Group by

Include all rows?: ☐

Temporary files directory: %%java.io.tmpdir%% [Browse...](#)

TMP-file prefix: grp

Add line number, restart in each group: ☐

Line number field name:

Always give back a result row: ☐

The fields that make up the group:

#	Group field
1	year

[Get Fields](#)

Aggregates :

#	Name	Subject	Type	Value
1	obs_value_tmp	observation_value	Sum	

[Get lookup fields](#)

[Help](#) [OK](#) [Cancel](#)

Figura 50

Al haver fet la suma, tenim les hores treballades de tots els mesos sumades així que per treure un promig a l'any s'ha de dividir el valor entre dotze, això ho hem fet amb 'Formula', guardant el resultat a la columna 'OBS_VALUE_USA' (figura 51).

[illegible]

Figura 51

- STG WORKED HOURS: S'ha carregat amb la següent sentència de SQL

```
SELECT
    s.[TIME_PERIOD] as [year],
    s.[unit],
    (SELECT pk_country FROM [SOURCE_braconsm].[dbo].[DIM_COUNTRY] WHERE [code]
= s.[geo]) AS fk_country,
    (SELECT pk_age FROM [SOURCE_braconsm].[dbo].[DIM_AGE] WHERE [code] =
s.[age]) AS fk_age,
    (SELECT pk_sex FROM [SOURCE_braconsm].[dbo].[DIM_SEX] WHERE [code] =
s.[sex]) AS fk_sex,
    (SELECT pk_sector FROM [SOURCE_braconsm].[dbo].[DIM_SECTOR] WHERE [code] =
s.[nace_r2]) AS fk_sector,
    (SELECT pk_worktime FROM [SOURCE_braconsm].[dbo].[DIM_WORKTIME] WHERE
[code] = s.[worktime]) AS fk_worktime,
    (SELECT pk_workstatus FROM [SOURCE_braconsm].[dbo].[DIM_WORKSTATUS] WHERE
[code] = s.[wstatus]) AS fk_workstatus,
    s.[OBS_VALUE]
FROM
    [SOURCE_braconsm].[dbo].[STG_HOURS_WORKED] s;
```

D'aquesta manera s'han relacionat les taules de dimensions amb les pk necessàries. A continuació s'ha fet servir 'Filters rows' per eliminar els valors nuls de fk_country, al igual que en el cas anterior. S'han ordenat les files alfabèticament amb 'Sort rows' i s'hi ha afegit una columna 'pk_id' que s'ha omplet de forma incremental amb 'Add sequence'.

- Unió de les taules: Amb 'Merge join' s'han unit les taules amb la columna clau 'year'. A més s'ha definit com a First Step la taula de STG_WORKED_HOURS (amb les transformacions esmentades) i com a Second Step STG_WORKED_HOURS_USA, així que el Join Type ha sigut del tipus LEFT OUTER per tal que la fusió fos correcte (figura 52).

Merge join

Step name: Merge join

First Step: Add sequence

Second Step: Formula

Join Type: LEFT OUTER

#	Key field
1	year

#	Key field
1	year

Get key fields (for both tables)

Help OK Cancel

Figura 52

Després hem fet un Select vàlues per eliminar les columnes que no ens interessaven i poder fer la càrrega a la base de dades amb 'Table output' (figura 53).

SQLQuery53.sql -...ENT_bracons (57) X

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP (1000) [pk_id]
, [year]
, [unit]
, [fk_country]
, [fk_age]
, [fk_sex]
, [fk_sector]
, [fk_worktime]
, [fk_workstatus]
, [OBS_VALUE]
, [OBS_VALUE_USA]
FROM [SOURCE_bracons].[dbo].[FACT_HOURS_WORKED]

```

100 %

Results Messages

	pk_id	year	unit	fk_country	fk_age	fk_sex	fk_sector	fk_worktime	fk_workstatus	OBS_VALUE	OBS_VALUE_USA
1	1	2017	HR	5	363	2	135	4	11	39.30	34.40
2	2	2017	HR	5	363	2	135	4	39	39.30	34.40
3	3	2017	HR	5	363	2	135	4	53	39.30	34.40
4	4	2017	HR	5	363	2	661	4	11	39.00	34.40
5	5	2017	HR	5	363	2	661	4	39	39.00	34.40
6	6	2017	HR	5	363	2	661	4	53	39.00	34.40
7	7	2017	HR	5	363	2	661	26	11	16.80	34.40
8	8	2017	HR	5	363	2	661	26	39	17.00	34.40
9	9	2017	HR	5	363	2	661	26	53	17.00	34.40
10	10	2017	HR	5	363	2	812	4	11	40.30	34.40
11	11	2017	HR	5	363	2	812	4	39	40.30	34.40
12	12	2017	HR	5	363	2	812	4	53	40.30	34.40
13	13	2017	HR	5	363	2	879	4	11	41.30	34.40

Figura 53

4. Implementació de treballs amb processos d'ETL

Tenint en compte els blocs de processos implementats:

- Bloc «IN_»: processos d'ETL de transformació i càrrega a l'àrea intermèdia.
- Bloc «TR_DIM»: processos d'ETL de transformació i càrrega de dimensions.
- Bloc «TR_FACT»: processos d'ETL de transformació i càrrega de fets.

En aquest apartat dissenyarem els treballs que permetran l'execució seqüencial de tots els processos d'ETL inclosos en cada bloc definit.

4.1 JOB IN

El treball «JOB_IN» processa totes les transformacions del bloc «IN_» per a la càrrega de dades des de les fonts de dades proporcionades a l'àrea intermèdia (staging area). El disseny el podem veure en la figura 54.

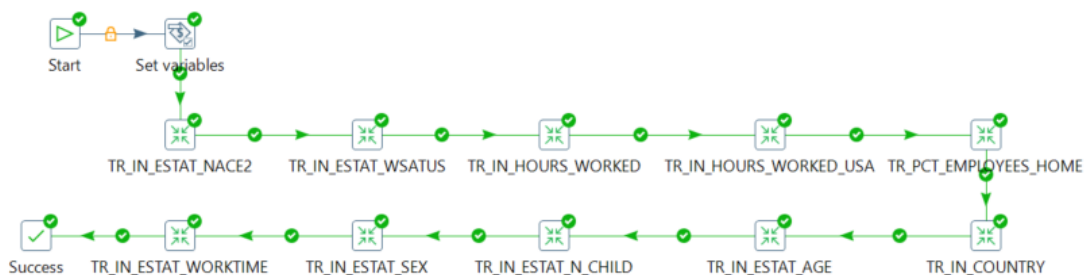


Figura 54

Els passos inclosos en el treball «JOB_IN» són:

- Inici del treball.
- Configuració de les variables d'entorn.
- Execució de les transformacions «IN_» de càrrega de l'staging area.
- Finalització del treball.

El resultat de l'execució de la transformació completa es pot veure en la figura 55

Execution Results

<div> Logging History Job metrics Metrics </div>						
Job / Job Entry	Comment	Result	Reason	Filename	Nr	Log date
▼ JOB_IN						
Job: JOB_IN	Start of job execution		start			2023/12/17 ...
Start	Start of job execution		start			2023/12/17 ...
Start	Job execution finished	Success			0	2023/12/17 ...
Set variables	Start of job execution		Followed uncondition...			2023/12/17 ...
Set variables	Job execution finished	Success			0	2023/12/17 ...
TR_IN_ESTAT_NACE	Start of job execution		Followed link after suc...	F:/Documents/PR2/ST...		2023/12/17 ...
TR_IN_ESTAT_NACE	Job execution finished	Success		F:/Documents/PR2/ST...	2	2023/12/17 ...
TR_IN_ESTAT_WSA	Start of job execution		Followed link after suc...	F:/Documents/PR2/ST...		2023/12/17 ...
TR_IN_ESTAT_WSA	Job execution finished	Success		F:/Documents/PR2/ST...	3	2023/12/17 ...
TR_IN_HOURS_WC	Start of job execution		Followed link after suc...	F:/Documents/PR2/ST...		2023/12/17 ...
TR_IN_HOURS_WC	Job execution finished	Success		F:/Documents/PR2/ST...	4	2023/12/17 ...
TR_IN_HOURS_WC	Start of job execution		Followed link after suc...	F:/Documents/PR2/ST...		2023/12/17 ...
TR_IN_HOURS_WC	Job execution finished	Success		F:/Documents/PR2/ST...	5	2023/12/17 ...
TR_PCT_EMPLOYEE	Start of job execution		Followed link after suc...	F:/Documents/PR2/ST...		2023/12/17 ...
TR_PCT_EMPLOYEE	Job execution finished	Success		F:/Documents/PR2/ST...	6	2023/12/17 ...
TR_IN_COUNTRY	Start of job execution		Followed link after suc...	F:/Documents/PR2/ST...		2023/12/17 ...
TR_IN_COUNTRY	Job execution finished	Success		F:/Documents/PR2/ST...	7	2023/12/17 ...
TR_IN_ESTAT_AGE	Start of job execution		Followed link after suc...	F:/Documents/PR2/ST...		2023/12/17 ...
TR_IN_ESTAT_AGE	Job execution finished	Success		F:/Documents/PR2/ST...	8	2023/12/17 ...
TR_IN_ESTAT_N_CH	Start of job execution		Followed link after suc...	F:/Documents/PR2/ST...		2023/12/17 ...
TR_IN_ESTAT_N_CH	Job execution finished	Success		F:/Documents/PR2/ST...	9	2023/12/17 ...
TR_IN_ESTAT_SEX	Start of job execution		Followed link after suc...	F:/Documents/PR2/ST...		2023/12/17 ...
TR_IN_ESTAT_SEX	Job execution finished	Success		F:/Documents/PR2/ST...	10	2023/12/17 ...
TR_IN_ESTAT_WOR	Start of job execution		Followed link after suc...	F:/Documents/PR2/ST...		2023/12/17 ...
TR_IN_ESTAT_WOR	Job execution finished	Success		F:/Documents/PR2/ST...	11	2023/12/17 ...
Success	Start of job execution		Followed link after suc...			2023/12/17 ...
Success	Job execution finished	Success			11	2023/12/17 ...
Job: JOB_IN	Job execution finished	Success	finished		11	2023/12/17 ...

Figura 55

S'observa el processament amb èxit de tots els passos del «JOB_IN» corresponent a l'execució de totes les transformacions que estan incloses en el treball.

4.1 JOB_TR_DIMS

El treball «JOB_TR_DIMS» processa totes les transformacions del bloc «TR_DIMS» per a la càrrega de dades, des de les taules intermèdies a les taules de dimensions del magatzem.

El disseny complet del treball «JOB_TR_DIMS» és el següent (figura 56):

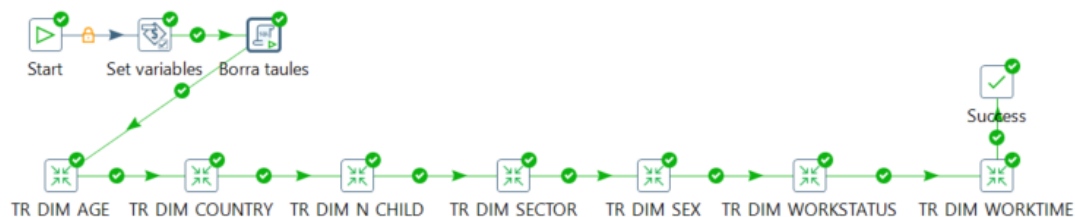


Figura 56

Els passos inclosos en el treball «JOB_TR_DIMS» són:

- Inici del treball.
- Càrrega de variables d'entorn (camí d'origens de dades i connexions).
- Esborrament de totes les taules. Això permet la recàrrega inicial en cas que sigui necessari. Encara que és important respectar l'ordre d'esborrament segons les relacions definides entre taules, atès que en aquest cas no hi ha relacions entre taules de dimensió, l'ordre no hi influirà.
- Execució seqüencial de totes les transformacions «TR_DIM» (extracció, transformació i càrrega de dimensions).
- Finalització del treball.

Per borrar les taules s'ha fet servir la següent sentència de SQL:

```

DELETE FROM FACT_HOURS_WORKED
DELETE FROM FACT_PCT_EMPLOYEES_HOME
DELETE FROM DIM_AGE
DELETE FROM DIM_COUNTRY
DELETE FROM DIM_N_CHILD
DELETE FROM DIM_SECTOR
DELETE FROM DIM_SEX
DELETE FROM DIM_WORKSTATUS
DELETE FROM DIM_WORKTIME
  
```

El resultat de l'execució de la transformació completa és el següent (figura 57):

Execution Results						
<div> Logging History Job metrics Metrics </div>						
Job / Job Entry	Comment	Result	Reason	Filename	Nr	Log date
▼ JOB_TR_DIMS						
Job: JOB_TR_DIMS	Start of job execution		start			2023/12/17 ...
Start	Start of job execution		start			2023/12/17 ...
Start	Job execution finished	Success			0	2023/12/17 ...
Set variables	Start of job execution		Followed uncondition...			2023/12/17 ...
Set variables	Job execution finished	Success			0	2023/12/17 ...
Borra taules	Start of job execution		Followed link after suc...			2023/12/17 ...
Borra taules	Job execution finished	Success			0	2023/12/17 ...
TR_DIM_AGE	Start of job execution		Followed link after suc...	F:/Documents/PR2/DI...		2023/12/17 ...
TR_DIM_AGE	Job execution finished	Success		F:/Documents/PR2/DI...	3	2023/12/17 ...
TR_DIM_COUNTRY	Start of job execution		Followed link after suc...	F:/Documents/PR2/DI...		2023/12/17 ...
TR_DIM_COUNTRY	Job execution finished	Success		F:/Documents/PR2/DI...	4	2023/12/17 ...
TR_DIM_N_CHILD	Start of job execution		Followed link after suc...	F:/Documents/PR2/DI...		2023/12/17 ...
TR_DIM_N_CHILD	Job execution finished	Success		F:/Documents/PR2/DI...	5	2023/12/17 ...
TR_DIM_SECTOR	Start of job execution		Followed link after suc...	F:/Documents/PR2/DI...		2023/12/17 ...
TR_DIM_SECTOR	Job execution finished	Success		F:/Documents/PR2/DI...	6	2023/12/17 ...
TR_DIM_SEX	Start of job execution		Followed link after suc...	F:/Documents/PR2/DI...		2023/12/17 ...
TR_DIM_SEX	Job execution finished	Success		F:/Documents/PR2/DI...	7	2023/12/17 ...
TR_DIM_WORKSTA	Start of job execution		Followed link after suc...	F:/Documents/PR2/DI...		2023/12/17 ...
TR_DIM_WORKSTA	Job execution finished	Success		F:/Documents/PR2/DI...	8	2023/12/17 ...
TR_DIM_WORKTIM	Start of job execution		Followed link after suc...	F:/Documents/PR2/DI...		2023/12/17 ...
TR_DIM_WORKTIM	Job execution finished	Success		F:/Documents/PR2/DI...	9	2023/12/17 ...
Success	Start of job execution		Followed link after suc...			2023/12/17 ...
Success	Job execution finished	Success			9	2023/12/17 ...
Job: JOB_TR_DIMS	Job execution finished	Success	finished		9	2023/12/17 ...

Figura 57

4.2 JOB_TR_FACTS

El treball «JOB_TR_FACTS» processa totes les transformacions del bloc «TR_FACT» per la càrrega de dades des de les taules intermèdies a les taules de fets del magatzem.

El disseny complet del treball «JOB_TR_FACTS» és el següent (figura 58):

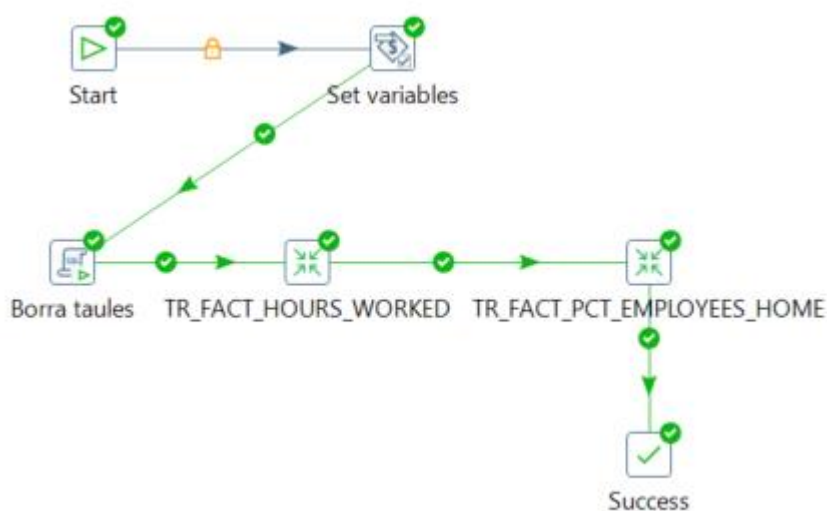


Figura 58

Els passos inclosos en el treball «JOB_TR_FACTS» són:

- Inici del treball.
- Càrrega de variables d'entorn.
- Eliminació de taules.
- Execució de les transformacions «TR_FACT».
- Finalització del treball

Al igual que en el cas anterior hem hagut de borrar taules en cas de necessitat. Això ho hem fet amb la sentència SQL següent:

```
DELETE FROM FACT_HOURS_WORKED
DELETE FROM FACT_PCT_EMPLOYEES_HOME
```

El resultat de l'execució de la transformació completa és el següent (figura 59)

Execution Results

Logging History Job metrics Metrics						
Job / Job Entry	Comment	Result	Reason	Filename	Nr	Log date
▼ JOB_TR_FACTS						
Job: JOB_TR_FACTS	Start of job execution		start			2023/12/17 ...
Start	Start of job execution		start			2023/12/17 ...
Start	Job execution finished	Success			0	2023/12/17 ...
Set variables	Start of job execution		Followed uncondition...			2023/12/17 ...
Set variables	Job execution finished	Success			0	2023/12/17 ...
Borra taules	Start of job execution		Followed link after suc...			2023/12/17 ...
Borra taules	Job execution finished	Success			0	2023/12/17 ...
TR_FACT_HOURS_1	Start of job execution		Followed link after suc...	F:/Documents/PR2/FA...		2023/12/17 ...
TR_FACT_HOURS_1	Job execution finished	Success		F:/Documents/PR2/FA...	3	2023/12/17 ...
TR_FACT_PCT_EMP	Start of job execution		Followed link after suc...	F:/Documents/PR2/FA...		2023/12/17 ...
TR_FACT_PCT_EMP	Job execution finished	Success		F:/Documents/PR2/FA...	4	2023/12/17 ...
Success	Start of job execution		Followed link after suc...			2023/12/17 ...
Success	Job execution finished	Success			4	2023/12/17 ...
Job: JOB_TR_FACTS	Job execution finished	Success	finished		4	2023/12/17 ...

Figura 59

4.3 JOB_CARGA_DW

El treball «JOB_CÀRREGA_DW» orquestra tots els treballs anteriors en un únic procés.

El disseny complet del treball «JOB_CARGA_DW» és el següent (figura 60)



Figura 60

Els passos inclosos en el treball «JOB_CARGA_DW» són:

- Inici del treball.
- Execució orquestrada dels treballs de càrrega de totes les transformacions («JOB_IN», «JOB_TR_DIMS», «JOB_TR_FACTS»).
- Finalització del treball.

El resultat de l'execució de la transformació completa és el següent (figura 61):

Execution Results						
Logging History Job metrics Metrics						
Job / Job Entry	Comment	Result	Reason	Filename	Nr	Log date
▼ JOB_CARGA_DW						
Job: JOB_CARGA_C	Start of job execution	start				2023/12/17 ...
Start	Start of job execution	start				2023/12/17 ...
Start	Job execution finished	Success			0	2023/12/17 ...
JOB_IN	Start of job execution		Followed uncondition...	F:/Documents/PR2/JO...		2023/12/17 ...
> Job: JOB_IN						
JOB_IN	Job execution finished	Success		F:/Documents/PR2/JO...	1	2023/12/17 ...
JOB_TR_DIMS	Start of job execution		Followed link after suc...	F:/Documents/PR2/JO...		2023/12/17 ...
> Job: JOB_TR_DIMS						
JOB_TR_DIMS	Job execution finished	Success		F:/Documents/PR2/JO...	2	2023/12/17 ...
JOB_TR_FACTS	Start of job execution		Followed link after suc...	F:/Documents/PR2/JO...		2023/12/17 ...
> Job: JOB_TR_FACTS						
JOB_TR_FACTS	Job execution finished	Success		F:/Documents/PR2/JO...	3	2023/12/17 ...
Success	Start of job execution		Followed link after suc...			2023/12/17 ...
Success	Job execution finished	Success			3	2023/12/17 ...
Job: JOB_CARGA_C	Job execution finished	Success	finished		3	2023/12/17 ...

Figura 61

S'observa el processament amb èxit de tots els passos del «JOB_ CARGA_DW», corresponent a l'execució de totes les transformacions que estan incloses en el treball.

El temps total de la càrrega inicial del data warehouse és d'aproximadament un 23 segons.

2023/12/17 01:36:01 - Spoon - Starting job...

2023/12/17 01:36:24 - Spoon - Job has ended.