



**Anàlisi de la mobilitat dels estudiants en el  
programa ERASMUS.  
Construcció i explotació d'un Magatzem de dades.**

**Bartomeu Alemany Far**  
Grau en Enginyeria Informàtica

**Bartomeu Antich Luque**

12/01/2016



Aquesta obra està subjecta a una llicència de [Reconeixement-NoComercial-SenseObraDerivada 3.0 Espanya de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

## FITXA DEL TREBALL FINAL

<b>Títol del treball:</b>	<b>Anàlisi de la mobilitat dels estudiants dins del programa ERASMUS. Construcció i explotació d'un Magatzem de dades.</b>
<b>Nom de l'autor:</b>	<i>Bartomeu Alemany Far</i>
<b>Nom del consultor:</b>	<i>Bartomeu Antich Luque</i>
<b>Data de lliurament (mm/aaaa):</b>	<i>Gen/2016</i>
<b>Àrea del Treball Final:</b>	<i>Magatzem de dades</i>
<b>Titulació:</b>	<i>Grau en Enginyeria Informàtica</i>
<b>Resum del Treball (màxim 250 paraules):</b>	
<p>La mobilitat de l'alumnat europeu ha esdevingut una peça clau en el marc docent del vell continent. Aquest treball ha pretès dotar d'una eina que faciliti l'estudi d'aquest fenomen.</p> <p>La resposta ha estat la construcció d'un magatzem de dades que permeti integrar la informació disponible procedent de diferents fonts d'informació dels sistemes operacionals de les institucions acadèmiques europees.</p> <p>Aquest magatzem serà la base per a la construcció del cub OLAP que serà la font de la qual es nodriran els informes prefixats i les consultes que els analistes necessitin definir.</p> <p>La implementació de la solució s'ha basat en programari lliure, tant pel que fa al programari de base (sistema operatiu i sistema gestor de base de dades) com el que s'utilitza pel disseny dels processos de càrrega i construcció del cub OLAP. També de programari lliure ha estat la solució utilitzada per a la construcció dels informes prefixats, així com la plataforma emprada per a la seva explotació.</p>	

**Abstract (in English, 250 words or less):**

Students' mobility has become essential in European education framework. This work has had the intention to produce a tool that make easy to study this phenomenon.

A data warehouse construction has been the answer. This product will permit to integrate and analyse data from different data sources. Sources from different operational information systems of different European academic institutions.

This data warehouse will be a base to construct an OLAP cube that will become source for fixed reports and queries that analysts need to define.

Solution implementation has based in free software. This is Server Operating System, RDBSM software, OLAP cube creation software and Business Analytics Platform (Pentaho BA) for reporting and analysis.

**Paraules clau (entre 4 i 8):**

DW, ETL, OLAP, Magatzem de dades, mobilitat, MDX RDBMS

## Índex

1. INTRODUCCIÓ .....	1
1.1. Context i justificació .....	1
1.2. Objectius del Treball .....	1
1.3. Enfocament i mètode seguit .....	2
1.4. Planificació del Treball .....	2
1.4.1. Fases .....	2
1.4.2. Tasques .....	3
1.4.3. Diagrama de Gant .....	7
1.4.4. Fites .....	8
1.5. Riscos i contingències .....	8
1.5.1. Descripció del risc i del pla de contingència .....	8
1.5.2. Riscos materialitzats .....	9
2 ANÀLISI .....	9
2.1. Requeriments .....	9
2.1.1. Funcionals .....	9
2.1.2. No funcionals .....	10
2.2. Casos d'ús .....	11
2.2.1. Actors .....	11
2.2.2. Subsistemes .....	11
2.2.3. Descripció dels casos d'ús (CU) .....	11
2.3. Fonts de dades .....	17
2.3.1. Studentmobility_2012-2013_Datadictionary .....	17
2.3.2. ISOCountryCodes081507 .....	18
2.3.3. EUC_Consolidated_Table_2007_2013 .....	18
2.3.4. SCED97_Erasmus_subject_codes .....	18
2.3.5. SM_2011_2012 .....	19
2.3.6. SM_2012_2013 .....	20
2.3.7. NACE CODES .....	21
2.3.8. Resultat de l'anàlisi de les fonts de dades .....	21
3 DISSENY .....	22
3.1. Arquitectura .....	22
3.1.1. Programari .....	22
3.1.2. Maquinari .....	23
3.2. Disseny conceptual .....	24
3.2.1. Fet/fets .....	24
3.2.2. Granularitat .....	24
3.2.3. Mesures .....	24
3.2.4. Dimensions .....	25
3.2.5. Atributs de les dimensions .....	26
3.2.6. Cel·les .....	28
3.2.7. Restriccions .....	29
3.2.8. Viabilitat .....	29
3.3. Disseny tècnic .....	31
3.3.1. Model lògic .....	32
3.3.2. Model físic .....	33
3.4. Disseny processos d'ETL a alt nivell, tractament d'errades i automatitzacions .....	34
3.4.1. Extracció .....	35
3.4.2. Transformació .....	36
3.4.3. Càrrega .....	36
4 IMPLEMENTACIÓ .....	37
4.1. Creació del DW .....	37
4.2. Processos ETL .....	39
4.2.1. Carrega del DW .....	39
4.2.2. Càrrega de les dimensions .....	40

4.2.3	Tractament de la dimensió <i>Course</i> .....	40
4.2.4	Tractament de la dimensió <i>Age</i> .....	41
4.2.5	Tractament de la dimensió <i>Subject Area</i> .....	42
4.2.6	Tractament de dimensió <i>Nationality</i> .....	42
4.2.7	Tractament dimensió <i>Institution: Academic Institutions and Placements</i> .....	42
4.2.8	Tractament de la taula de fets <i>Mobility</i> .....	45
4.3	Cub OLAP .....	48
4.3.1	Configuració de la connexió.....	48
4.3.2	Creació de l'esquema i el cub OLAP .....	49
4.3.3	Assignació de la taula de fets .....	49
4.3.4	Creació de les dimensions externes.....	49
4.3.5	Definició de la <i>dimension usage</i> .....	50
4.3.6	Definició de les mesures .....	50
4.3.7	Creació d'una mesura calculada .....	51
4.3.8	Publicació.....	51
4.4	Informes .....	52
4.4.1	Top 10 d'institucions emissores/receptores .....	53
4.4.2	Distribució en % d'estudiants per nacionalitat dels estudiants .....	53
4.4.3	Distribució en % d'estudiants per àrea de coneixement.....	54
4.4.4	Evolució comparativa per curs.....	54
4.4.5	Edat mitjana dels estudiants per nacionalitat de la institució receptora/emissora .....	55
4.4.6	Tant per cent de beques per nacionalitat emissora/receptora .....	55
4.4.7	Percentatge d'estudiants en pràctiques per àrea de coneixement .....	55
4.4.8	Tant per cent d'estudiants en pràctiques per sector empresarial.....	56
4.4.9	Percentatge d'estudiants en pràctiques per mida d'empresa .....	56
4.4.10	Distribució segons la preparació en idiomes .....	57
5	CONCLUSIONS .....	58
6	GLOSSARI DE TERMES .....	59
7	ANNEXES.....	60
7.1	Annex I: Taules de l' <i>Staging Area</i> .....	60
7.2	Annex II: Scripts SQL DW .....	61
7.2.1	Annex II: Script SQL dw.....	61
7.2.2	Annex II: Script SQL SA .....	65
7.3	Annex III: Cub OLAP TFG_ErasmusMobility .....	68
7.4	Annex IV: Reports Saiku .....	72
8	BIBLIOGRAFIA.....	77

## Lista d'imatges

Imatge 1.Cas d'ús Anàlisi i informes	11
Imatge 2.Cas d'ús Administració DW i cubs OLAP	13
Imatge 3 Cas d'ús Gestió d'usuaris	16
Imatge 4 Arquitectura del programari	23
Imatge 5 Arquitectura del maquinari	23
Imatge 6 Relació mesures/elements del DW	29
Imatge 7 Model conceptual	31
Imatge 8 Model lògic	33
Imatge 9 Diagrama model físic	34
Imatge 10. Entrada a l'eina <i>Workbench</i> de MySQL	37
Imatge 11. Creació de la taula de fets	38
Imatge 12. Claus alines de la taula de fets	38

Imatge 13. Càrrega inicial del DW	39
Imatge 14. Càrrega anual del DW	40
Imatge 15. Càrrega de les dimensions al DW	40
Imatge 16. Càrrega dels cursos	40
Imatge 17. Data Grid	41
Imatge 18. Càrrega de les edats en el DW	41
Imatge 19. Generació dels anys	41
Imatge 20. Càrrega de l'àrea de coneixement en el DW	42
Imatge 21. Càrrega de la nacionalitat al DW	42
Imatge 22. Extracció de la nacionalitat	42
Imatge 23. Càrrega d' <i>Institution</i>	43
Imatge 24. Extracció d' <i>Academic Institution</i>	43
Imatge 25. Extracció de <i>Placement</i>	43
Imatge 26 33 T <i>Set Sector, countryCode and Size values</i>	44
Imatge 27. 34 T <i>Set SectorLong</i>	44
Imatge 28. 35 T <i>Set SizeLong</i>	44
Imatge 29. 38 T <i>Set CountryName</i> if null	45
Imatge 30. 37 <i>Set CountryName</i> from Country	45
Imatge 31. 39 T <i>Load Institution</i> into DW	45
Imatge 32. Exemple de càrrega font tipus 2011_2012.	46
Imatge 33. 46 <i>Job Prepare 2011 Like</i>	46
Imatge 34. 40 T <i>Extract SM_2011_2012 Like</i>	46
Imatge 35. 44 T <i>Set Grants</i> to 0 or 1	47
Imatge 36. <i>Insert Placement grain</i> into <i>stg_mobility2</i>	47
Imatge 37. <i>Load Placement</i> into DW from <i>stg_mobility2</i>	47
Imatge 38. <i>Lookup Placement and Academic Institution</i>	48
Imatge 39. Configuració connexió	48
Imatge 40. Creació de l'esquema i cub	49
Imatge 41. Creació de la taula de fets	49
Imatge 42. Creació la dimensió	49
Imatge 43. Creació del nivell	50
Imatge 44. Creació de la <i>dimension usage</i>	50
Imatge 45. Creació de la mesura <i>Students</i>	50
Imatge 46. Creació de la mesura <i>averageAge</i>	51
Imatge 47. Publicar cub/Importar anàlisi	51
Imatge 48. Cub OLAP	51
Imatge 49. Accés al sistema	52
Imatge 50. Accés als informes	52
Imatge 51. Informes disponibles	52
Imatge 52. Top 10 institucions emissores/receptores	53
Imatge 53. Distribució en % d'estudiants per nacionalitat	53
Imatge 54. Distribució en % d'estudiants per àrea de coneixement	54
Imatge 55. Evolució comparativa per curs	54
Imatge 56. Edat mitjana dels estudiants per nacionalitat receptora/emissora	55
Imatge 57. Tant per cent de beques per nacionalitat emissora/receptora	55

Imatge 58. Percentatge d'estudiants en pràctiques per àrea de coneixement	56
Imatge 59. Percentatge d'estudiants en pràctiques per sector empresarial	56
Imatge 60. Percentatge d'estudiants en pràctiques segons la mida de l'empresa	57
Imatge 61. Distribució segon la preparació en idiomes	57

## Llista de taules

Taula 1. Cronologia de les fases	3
Taula 2. Cronologia de tasques	5
Taula 3. Lliurables	8
Taula 4. Cas d'ús Executar informe	12
Taula 5. Cas d'ús Crear informe	12
Taula 6. Cas d'ús Exportar informe	12
Taula 7. Cas d'ús Editar informe	12
Taula 8. Cas d'ús Crear Anàlisi	13
Taula 9. Cas d'ús Eliminar Anàlisi	13
Taula 10. Cas d'ús Crear cub OLAP	14
Taula 11. Cas d'ús Eliminar cub OLAP	14
Taula 12. Cas d'ús Editar cub OLAP	14
Taula 13. Cas d'ús Carregar dades	15
Taula 14. Cas d'ús Editar procés ETL	15
Taula 15. Cas d'ús Crear procés ETL	15
Taula 16. Cas d'ús Eliminar procés ETL	16
Taula 17. Cas d'ús Crear usuari	16
Taula 18. Cas d'ús Eliminar usuari	17
Taula 19. Cas d'ús Modificar usuari	17
Taula 20. Estructura <i>ISOCountrycodes081507</i>	18
Taula 21. Estructura <i>EUC_Consolidated_Table_2007_2013</i>	18
Taula 22. Estructura dels codis <i>ISCDE97</i>	18
Taula 23. Estructura <i>SCED97_Erasmus_subject_codes</i>	18
Taula 24. Estructura <i>SM_2011_2012</i>	19
Taula 25. Estructura <i>SM_2012_2013</i>	20
Taula 26. Resultat de l'anàlisi	21
Taula 27. Arquitectura	22
Taula 28. Dimensions del DW	25
Taula 29. Curs	26
Taula 30. Edat	26
Taula 31. Nacionalitat	26
Taula 32. Gènere	26
Taula 33. Preparació en llengües	26
Taula 34. Nivell d'estudis	27



Taula 35. Tipus de mobilitat	27
Taula 36. Institució (emissora/receptora)	27
Taula 37. Institució acadèmica	27
Taula 38. Empresa on s'efectuen les pràctiques	27
Taula 39. Àrees de coneixement	28
Taula 40. Estimació aproximada de la mida del DW	30
Taula 41. Model relacional	32
Taula 42. Correspondència entre atributs	60
Taula 43. SM_2011_2012	60
Taula 44. SM_2012_2013	61
Taula 45. Taula per a la càrrega del grànul	61

# 1. Introducció

## 1.1. Context i justificació

La Unió Europea ofereix estadístiques i fonts de dades en modalitat "open data" sobre diferents organismes i institucions europees. En el marc educatiu, i dins de la Unió Europea, són d'especial interès les dades de mobilitat d'estudiants del programa Erasmus que són un indicador del grau d'integració de la societat europea i, en especial, de la joventut.

Fins ara, les dades dels sistemes operacionals són extretes a fitxers en format CSV i XLS, i, mitjançant fulls de càlcul i programes fets a mida creuades i analitzades. Ara bé, això comporta una gran despesa de temps i feina degut a la baixa integració de les dades que s'utilitzen i al seu alt volum.

Aquest projecte pretén dotar d'una eina que faciliti la integració de les dades de les diferents fonts en un únic repositori i que permeti un estudi en profunditat dels fluxos de l'alumnat dins del teixit educatiu europeu.

## 1.2. Objectius del Treball

Els Treballs Fi de Grau (d'ara en endavant, TFG) estan caracteritzats per una sèrie d'objectius comuns a tots ells, els quals es relacionen tot seguit:

- Posar en pràctica els coneixements i competències adquirits durant la titulació.
- Realitzar un projecte passant per totes les seves fases, seleccionant els procediments adequats per dur-ho a terme.
- Documentar i justificar el desenvolupament i el resultat del treball.
- Presentar i defensar el treball realitzat.
- Autoavaluar el treball d'acord amb uns criteris determinats.

En aquest TFG, en estar emmarcat dins l'Àrea de Magatzem de dades, es materialitzen en:

- Aprendre a planificar i executar un projecte des de l'inici fins el tancament.
- Entendre la importància dels magatzems de dades en la presa de decisions.
- Aprendre els conceptes fonamentals dels magatzems de dades quant a:
  - L'anàlisi dels requeriments (funcionals i no funcionals) del projecte.
  - Entendre la influència de les fonts de dades de què es disposa en el disseny final del magatzem de dades.
  - Dissenyar els esquemes (conceptual, lògic i físic) del magatzem.
  - Implementar l'esquema relacional del DW i del cub OLAP.
  - Aprendre a dissenyar els processos ETL (extracció, transformació i càrrega de les dades al magatzem) per a la càrrega inicial i la seva actualització, així com la

utilització d'eines específiques per a tal fi.

- Dissenyar informes i consultes MBX que ataquin el DW i que extreguin la informació necessària per a la presa de decisions.

L'objectiu específic d'aquest *TFG* és la creació i posterior actualització d'un magatzem de dades (*DW*, d'ara en endavant) que integri les fonts procedents de la Unió Europea i que permeti obtenir una sèrie d'informes prefixats i que alhora estigui obert ha inferir altra informació mitjançant eines de *BI* per part dels analistes.

Finalment, i no per això menys important, presentar aquesta Memòria i la Presentació virtual, on s'expliqui l'anàlisi, disseny i construcció del *DW*, així com dels informes sol·licitats.

### 1.3. Enfocament i mètode seguit

La falta d'una solució senzilla amb dades integrades provinents dels diferents sistemes operacionals, aconsellen la construcció d'un nou sistema d'informació que doni resposta a les necessitats d'anàlisi d'aquest fenomen de la mobilitat de l'alumnat.

El mètode seguit pel seu desenvolupament ha estat l'*SDLC* (*Systems Development Life Cycle*), conegut com a de desenvolupament en cascada, el qual es caracteritza perquè per avançar a la següent fase és necessari haver completat l'anterior.

La tria ha estat més per l'encotillament a què obliga el desenvolupament d'un treball de fi de grau que per la idoneïtat de la metodologia en si mateixa, la qual pressuposa que els requisits no variaran durant el desenvolupament, alhora que confia en què en el moment de la realització de cada fase es té tot el coneixement necessari per a la seva correcta execució.

### 1.4. Planificació del Treball

#### 1.4.1. Fases

##### Pla de treball i Anàlisi preliminar de requeriments

Documents a lliurar:

- Anàlisi preliminar de requeriments, que recollirà les necessitats funcionals del *TFG* amb la relació dels informes a presentar, un estudi preliminar de les fonts de dades disponibles i dels elements del *DW*.
- Pla de treball on es recollirà la planificació estimada del projecte, on es relacionaran les fases, tasques, fites, riscos i plans de contingències.

##### Anàlisi de requeriments i disseny conceptual i tècnic

Documents a lliurar:

- Anàlisi detallada de requeriments basat en l'anàlisi preliminar realitzada.
- Disseny conceptual i tècnic amb la descripció del model dimensional que donarà suport

a les necessitats dels usuaris, segons l'anàlisi realitzada i el disseny dels procediments d'extracció de dades a alt nivell (processos, pseudocodi, etc.).

### **Implementació**

Tasques a realitzar:

- Construcció del magatzem de dades (base de dades, càrregues, etc.)
- Configuració de l'eina d'exploració de dades.
- Construcció dels informes i anàlisi de la informació.

### **Memòria i Presentació virtual**

Presentació de la memòria del treball realitzat i de la presentació virtual on es resumeix d'una forma visual la realització del projecte.

En la taula següent es presenta la cronologia de cada una d'elles (taula 1).

**Taula 1 Cronologia de les fases**

Títol	Inici	Lliurament
<b>PAC1</b>		
Pla de treball i anàlisi preliminar	19/09/2015	01/10/2015
<b>PAC2</b>		
Anàlisi de requeriments i disseny conceptual i tècnic	02/10/2015	29/10/2015
<b>PAC3</b>		
Implementació	30/10/15	17/12/02
<b>PAC 4</b>		
Memòria	18/12/15	12/01/16
Presentació virtual		

#### **1.4.2 Tasques**

Tot seguit es passa a relacionar les tasques en què s'ha descompost el projecte i la seva descripció.

1. Lectura i comprensió de l'enunciat del TFG

Lectura acurada del cas plantejat i primera anàlisi.

2. Lectura inicial documentació i recerca

Lectura de la documentació inicial aportada i recerca d'informació al web.

3. Anàlisi preliminar

Anàlisi dels requeriments (enunciat), alhora que s'estudien les fonts de dades (estructura, informació i format de la informació que contenen). Ambdós menaran cap a la definició

dels elements del DW: fets, granularitat, dimensions, atribut, etc.

4. Pla del treball

Confecció, amb el coneixement que proporciona l'Anàlisi preliminar, de la planificació del TFG, on es definiran: actors, recursos, fases del projecte, tasques, riscos i plans de contingència.

5. Revisió dels documents a entregar.

Repàs i correcció de la redacció dels documents Anàlisi preliminar i Pla de treball.

6. Anàlisi de requeriments

Aprofundiment en l'anàlisi dels requeriments del client, les fonts de dades disponibles, els elements de DW resultants i la determinació dels informes definitius.

7. Disseny conceptual del DW

Determinació dels elements del DW a utilitzar: fets, granularitat, mesures, dimensions i els seus atributs (descriptors i jerarquies d'agregació) a utilitzar en el disseny del DW.

Determinació de l'esquema de DW més convenient pel projecte (ja sigui el d'estrella o el de flocc de neu). Per acabar estudiant la viabilitat i les restriccions.

8. Disseny tècnic

Determinació de l'esquema relacional (lògic) corresponent a l'esquema conceptual elegit i l'esquema relacional (físic) a implementar, així com dels índexs, claus alienes i restriccions necessàries.

9. Disseny dels procediments ETL (alt nivell)

Creació del codi d'alt nivell on es detallen els processos d'extracció, transformació i càrrega, tenint en compte la problemàtica específica de cada font.

10. Revisió documentació PAC 2

Correcció i refinament de la documentació a entregar.

11. Estudi documentació de les eines d'explotació

Estudi d'eines OLAP i ETL de la suite PENTAHO.

12. Construcció del DW

Implementació en el RDBSM MySQL del model físic dissenyat en la fase anterior, així com de les taules temporals de càrrega.

13. Càrrega de les dades al magatzem

Implementació dels procediments d'ETL prèviament dissenyats.

14. Construcció del model Multidimensional

Connexió a la base de dades relacional i creació del model multidimensional.

15. Construcció dels informes i anàlisi de la informació

Creació dels informes sol·licitats pels analistes.

16. Redacció i revisió de la Memòria

Redacció de la memòria del TFG.

17. Creació de la presentació

Cerca d'imatges i creació de la presentació del projecte.

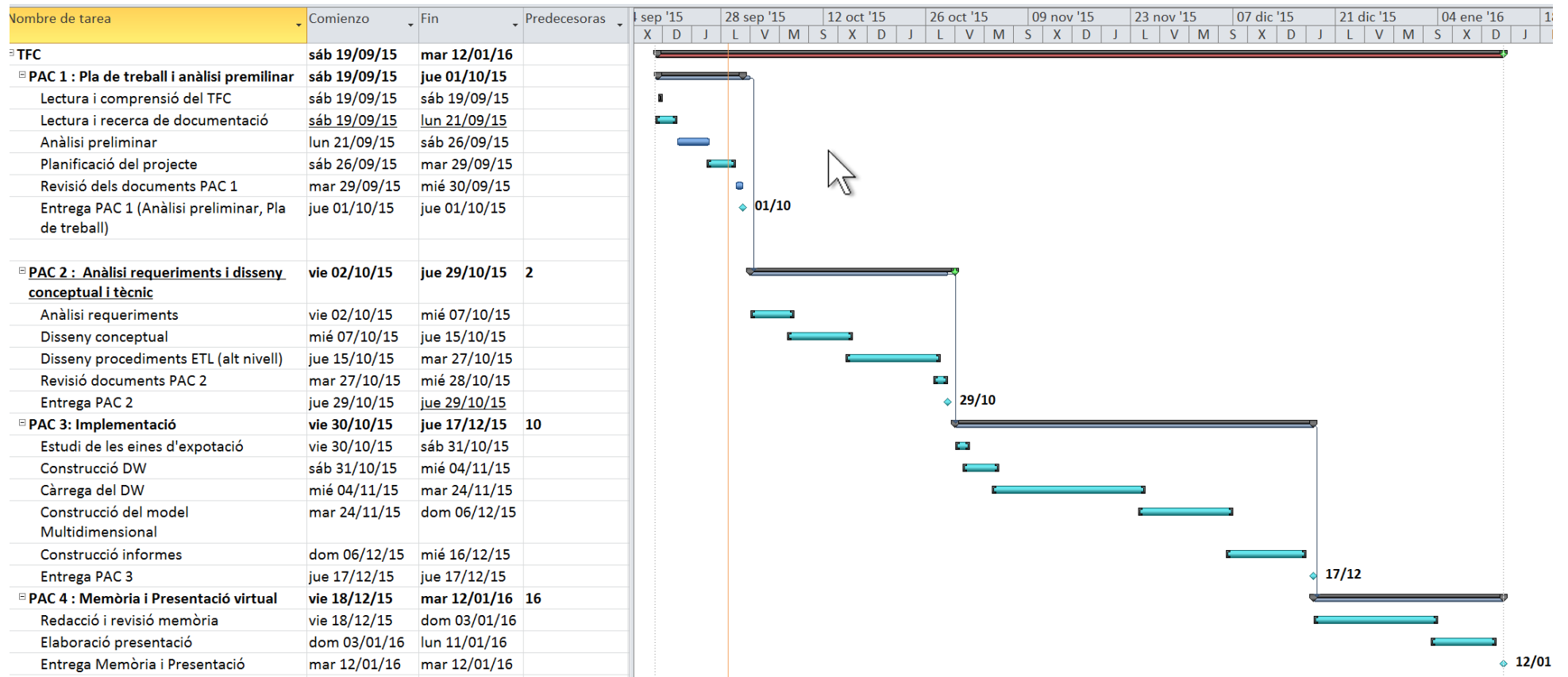
En la taula 2 es presenta la cronologia de les tasques enumerades anteriorment.

**Taula 2 Cronologia de tasques**

Nom de la tasca	Data inici	Data final	Hores
<b><u>Fase, Pla de treball i Anàlisi preliminar de requeriments</u></b>			
Lectura i comprensió de l'enunciat <i>TFG</i>	19/09/15	19/09/15	2
Lectura inicial documentació i recerca	19/09/15	21/09/15	8
Anàlisi preliminar	21/09/15	26/09/15	14
Pla de treball i diagrama de Gantt	26/09/15	29/09/15	20
Revisió dels documents a entregar	29/09/15	30/09/15	4
<b><u>Anàlisi de requeriments i disseny conceptual i tècnic</u></b>			
Anàlisi de requeriments	02/10/15	07/10/15	22
Disseny conceptual del <i>DW</i>	07/10/15	15/09/15	28
Disseny tècnic	15/09/15	22/09/15	36
Disseny dels procediments ETL (alt nivell)	22/09/15	27/09/15	40
Revisió documentació PAC 2	27/10/15	28/10/15	6
<b><u>Implementació</u></b>			
Estudi documentació de les eines d'exploració	30/10/15	31/10/15	10
Construcció del <i>DW</i>	31/10/15	04/11/15	16
Càrrega de les dades al magatzem	04/11/15	24/11/15	80
Construcció del model Multidimensional	24/11/15	06/12/15	44
Construcció dels informes	06/12/15	16/12/16	40
<b><u>Memòria i Presentació virtual</u></b>			
Redacció i revisió de la Memòria	18/12/15	03/01/16	50
Creació de la presentació	03/01/15	12/01/16	36
<b><u>Fase, Pla de treball i Anàlisi preliminar de requeriments</u></b>			
Lectura i comprensió de l'enunciat <i>TFG</i>	19/09/15	19/09/15	2
Lectura inicial documentació i recerca	19/09/15	21/09/15	8
Anàlisi preliminar	21/09/15	26/09/15	14
Pla de treball i diagrama de Gantt	26/09/15	29/09/15	20
Revisió dels documents a entregar	29/09/15	30/09/15	4

<b><u>Anàlisi de requeriments i disseny conceptual i tècnic</u></b>			
Anàlisi de requeriments	02/10/15	07/10/15	22
Disseny conceptual del <i>DW</i>	07/10/15	15/09/15	28
Disseny tècnic	15/09/15	22/09/15	36
Disseny dels procediments <i>ETL</i> (alt nivell)	22/09/15	27/09/15	40
Revisió documentació PAC 2	27/10/15	28/10/15	6
<b><u>Implementació</u></b>			
Estudi de la documentació de les eines d'explotació	30/10/15	31//10/15	10
Construcció del <i>DW</i>	31/10/15	04/11/15	16
Càrrega de les dades al magatzem	04/11/15	24/11/15	80
Construcció del model Multidimensional	24/11/15	06/12/15	44
Construcció dels informes	06/12/15	16/12/16	40
<b><u>Memòria i Presentació virtual</u></b>			
Redacció i revisió de la Memòria	18/12/15	03/01/16	50
Creació de la presentació	03/01/15	12/01/16	36

### 1.4.3 Diagrama de Gantt





#### 1.4.4 Fites

Les fites coincideixen en l'acabament de cada una de les fases i es corresponen amb l'entrega dels productes sol·licitats, ja siguin documents o elements de programari (veure relació en la taula 3).

**Taula 3. Lliurables**

Documents/Productes	Data lliurament
Pla de treball i Anàlisi preliminar	01/10/215
Anàlisi de requeriments i Disseny conceptual i tècnic	29/10/15
Creació de base de dades <i>DW</i> , model OLAP i informes sol·licitats	17/12/02
Memòria i Presentació	12/01/16

#### 1.5 Riscos i contingències

Incidències i riscos són inherents als projectes, i s'han de gestionar. Una mala gestió pot dur un projecte al fracàs (Ramón Rodrigues, Maríné Jove, 2011, pàg.50). Per tant, en la gestió d'un projecte és imprescindible haver-los detectat i fer els plans de contingència corresponents.

##### 1.5.1 Descripció del risc i del pla de contingència.

###### 1. Malaltia

###### Descripció

Malaltia menor: la malaltia més freqüent és el grip o el constipat. Una bona suposició seria una durada de dos a tres dies de feina.

Hospitalització: pròpia o d'un familiar en primer o segon grau.

###### Pla de contingència

En el primer cas, aquesta possible indisponibilitat seria fàcilment solucionat incrementant l'horari de dedicació, tant dels dies feiners com festius (veure apartat recursos). En tot cas, si coincidís amb dates properes a la presentació d'un lliurable, s'hauria de posar en coneixement del Consultor, i correspondria a ell determinar si la fita es pot o no posposar.

Pel que fa al segon cas, la solució dependria de la durada i de la proximitat de les fites. Però, en tot cas, s'hauria de comunicar al Consultor, i confiar en què hi hagués una solució. Si no n'hi hagués, s'hauria de procedir al tancament del projecte.

###### 2. Problemes tècnics amb l'equipament informàtic personal

###### Descripció

Avaria de l'ordinador personal principal i/o del programari de ofimàtica.

###### Pla de contingència

S'utilitzaria un portàtil de suport amb el programari d'ofimàtica idèntic al de treball i es faria ús de les còpies al núvol o al disc USB.

### **3. Problemes tècnics a la màquina virtual d'Amazon**

#### Descripció

Per motius aliens a l'executor del projecte i a la pròpia UOC, hi ha vegades que la màquina Amazon pot no ser accessible, o que el programari de base de dades i/o de la Suite Pentaho de BA es desconfiguri.

#### Pla de contingència

S'ha de transferir la gestió del risc a la UOC. La UOC hauria de proveir d'una solució per a la continuació del TFG, ja proveint uns altres recursos, ja modificant els terminis d'entrega dels lliurables.

### **4. Falta d'expertesa en el camp dels DW**

#### Descripció

El món dels magatzems de dades és nou per a l'alumne executor d'aquest TFG. Això té com a conseqüències un menor coneixement sobre el tema i una incertesa sobre les tasques que s'han de dur a terme i com s'han d'executar.

#### Pla de contingència

Dedicació en el començament de cada fase a l'estudi dels temes tècnics necessaris per a la seva execució.

#### **1.5.2 Riscos materialitzats**

Durant l'execució del projecte, de fet, se'n materialitzaren dos dels quatre riscos detectats: el número 2 i el 4. Mentre que el número 4 era una contingència que estava prevista que passaria, la número 2 no era així. Però en qualsevol cas, els plans de contingència es posaren en marxa i els riscos foren mitigats. El primer totalment, per contra el segon penso que només parcialment, ja que encara em queda camí per recórrer en el camp dels DW.

## **2 Anàlisi**

### **2.1 Requeriments**

El DW haurà d'oferir una sèrie de funcionalitats que donin resposta a les necessitats dels usuaris (requeriments funcionals). Però, a més, haurà d'oferir una sèrie de restriccions i qualitats en l'execució de les funcionalitats (requeriments no funcionals).

#### **2.1.1 Funcionals**

Els requeriments funcionals detectats són:

- Proporcionar els següents informes fixos:
  - Top 10 d'universitats més receptores, i més emissores d'estudiants Erasmus.
  - Distribució en % d'estudiants per nacionalitat.
  - Distribució en % d'estudiants per àrea de coneixement.
  - Evolució comparativa del nombre d'estudiants per curs.

- Edat mitjana d'estudiants per nacionalitat receptora i emissora.
- % d'estudiants amb beca d'estudis per nacionalitat receptora i emissora.
- % d'estudiants amb beca de pràctiques per nacionalitat receptora i emissora
- Distribució en % d'estudiants en pràctiques per sector empresarial.
- Distribució en % d'estudiants en pràctiques per mida d'empresa.
- Nombre d'estudiants en pràctiques per àrea de coneixement.
- Distribució en % d'estudiants segons la preparació en llengües.
- Oferir als analistes poder:
  - Executar els informes prefixats.
  - Elegir el format d'exportació del contingut dels informes.
  - Modificar els informes.
  - Navegar pels cubs OLAP i generar informes que els ataquin.
- Oferir als administradors poder:
  - Modificar l'estructura del DW.
  - Generar cubs OLAP.
  - Carregar les dades al DW.
  - Modificar els processos d'ETL.
  - Generar-ne de nous.
- A més:
  - El sistema demanarà la identificació als usuaris per entrar-hi.
  - El sistema extraurà les dades, les transformarà i les carregarà al DW.
  - Les dades incorrectes no seran tractades, donant lloc a un informe d'errors.

### 2.1.2 No funcionals

Els requeriments no funcionals detectats són:

- Es disposarà d' una interfície gràfica web per a l'accés al sistema d'anàlisi del negoci.
- La interfície haurà de ser amigable, autoexplicativa i de fàcil accés.
- El DW haurà de tenir l'estructura adequada al volum de dades a tractar.
- El sistema construït haurà de permetre absorbir la càrrega de treball durant cinc anys sense haver de canviar de plataforma ni tecnologia.
- El sistema haurà d'estar construït amb eines "open source", en concret: MySQL CE, MySQL Workbench CE, Modrian Workbench, Pentaho PDI i Suite Pentaho BA.
- El DW haurà d'estar hostatjat en una màquina virtual al núvol.
- Hi haurà mecanismes de seguretat per al control d'accés al sistema d'informació, ja sigui per a l'administració ja sigui per a l'anàlisi de la informació.
- S'haurà d'utilitzar una eina fàcilment programable i gràfica pel procés d'ETL.
- El sistema haurà de ser fàcilment depurable de dades errònies.
- En els processos d'ETL s'hauran de poder carregar fonts d'informació de diferents tipus de fitxers.

## 2.2 Casos d'ús

En aquest apartat es determinen els actors que interactuen amb el sistema a dissenyar, els casos d'ús en què intervenen i els subsistemes en els quals s'agrupen.

### 2.2.1 Actors

Els actors que s'han identificat en el sistema d'informació a dissenyar són:

- Administradors del DW, que s'encarregaran de l'administració dels elements del DW, dels informes i dels usuaris que hi tindran accés.
- Analistes, que seran els encarregats d'executar els informes i analitzar la informació.

### 2.2.2 Subsistemes

Els casos d'ús s'agruparan en els tres subsistemes següents:

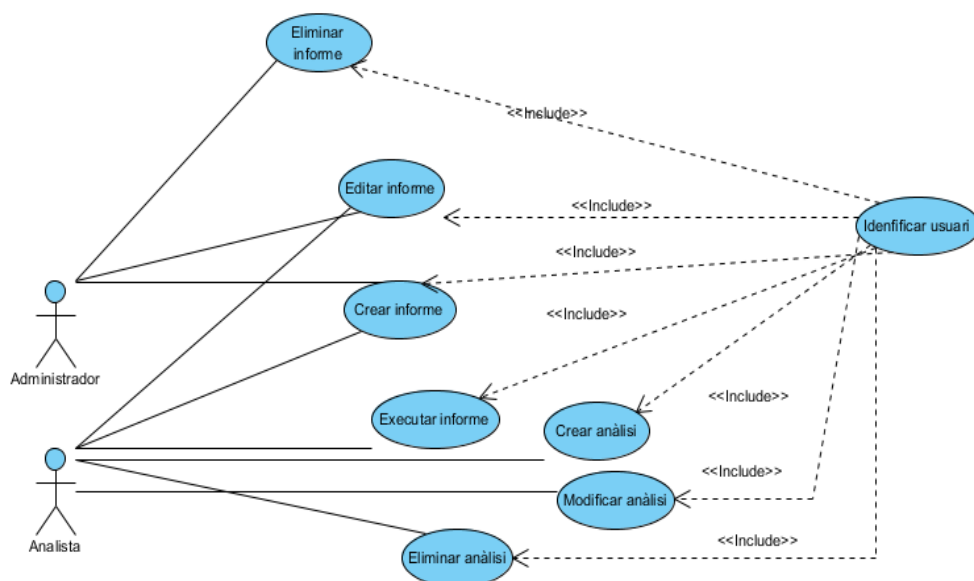
- Administració d'informes i anàlisi  
Gestió i accés als informes i a les anàlisis.
- Administració del DW i cubs OLAP  
Administració dels elements del DW i dels esquemes multidimensionals OLAP.
- Administració d'usuaris.  
Gestió dels usuaris del sistema.

### 2.2.3 Descripció dels casos d'ús (CU)

Per a cada cas d'ús es presentarà un gràfic i una taula explicativa.

#### Subsistema d'anàlisi i informes

Imatge 1 Cas d'ús Anàlisi i informes



**Taula 4 Cas d'ús Executar informe**

<b>Cas d'ús</b>	Executar informe.
<b>Descripció</b>	Presenta les dades generades per l'informe.
<b>CU relacionats</b>	Identificar usuari.
<b>Actors</b>	Analista.
<b>Precondicions</b>	<ol style="list-style-type: none"> <li>1. Usuari donat d'alta i amb permisos.</li> <li>2. Usuari identificat en el sistema.</li> </ol>
<b>Flux</b>	<ol style="list-style-type: none"> <li>1. L'actor elegeix l'informe a executar.</li> <li>2. El sistema mostra la informació que en resulta.</li> </ol>

**Taula 5 Cas d'ús Crear informe**

<b>Cas d'ús</b>	Crear informe.
<b>Descripció</b>	Creació d'un nou informe.
<b>CU relacionats</b>	Identificar usuari.
<b>Actors</b>	Administrador i Analista.
<b>Precondicions</b>	<ol style="list-style-type: none"> <li>1. Usuari donat d'alta i amb permisos.</li> <li>2. Usuari identificat en el sistema.</li> </ol>
<b>Flux</b>	<ol style="list-style-type: none"> <li>1 L'actor elegeix l'esquema.</li> <li>2 L'actor selecciona els camps i paràmetres.</li> <li>3 L'actor selecciona la visualització.</li> <li>4 El sistema valida les dades i l'emmagatzema.</li> </ol>

**Taula 6. Cas d'ús Exportar informe**

<b>Cas d'ús</b>	Exportar informe.
<b>Descripció</b>	Permet l'exportació del resultat de l'execució de l'informe.
<b>CU relacionats</b>	Identificar usuari i Executar informe.
<b>Actors</b>	Analista.
<b>Precondicions</b>	<ol style="list-style-type: none"> <li>1. Usuari donat d'alta i amb permisos.</li> <li>2. Usuari identificat en el sistema.</li> </ol>
<b>Flux</b>	<ol style="list-style-type: none"> <li>2 L'actor elegeix l'informe a consultar.</li> <li>3 El sistema mostra la informació resultant.</li> <li>4 L'actor exporta l'informe amb el format elegit.</li> </ol>

**Taula 7. Cas d'ús Editar informe**

<b>Cas d'ús</b>	Editar informe.
	Modificació d'un informe.
<b>CU relacionats</b>	Identificar usuari.
<b>Actors</b>	Administrador i Analista.
<b>Precondicions</b>	<ol style="list-style-type: none"> <li>1. Usuari donat d'alta i amb permisos.</li> <li>2. Usuari identificat en el sistema.</li> </ol>
<b>Flux</b>	<ol style="list-style-type: none"> <li>1 L'actor elegeix l'informe.</li> <li>2 L'actor selecciona els camps i paràmetres.</li> <li>3 L'actor selecciona la visualització</li> <li>4 El sistema valida les dades i l'emmagatzema</li> </ol>

**Taula 8. Cas d'ús Crear Anàlisi**

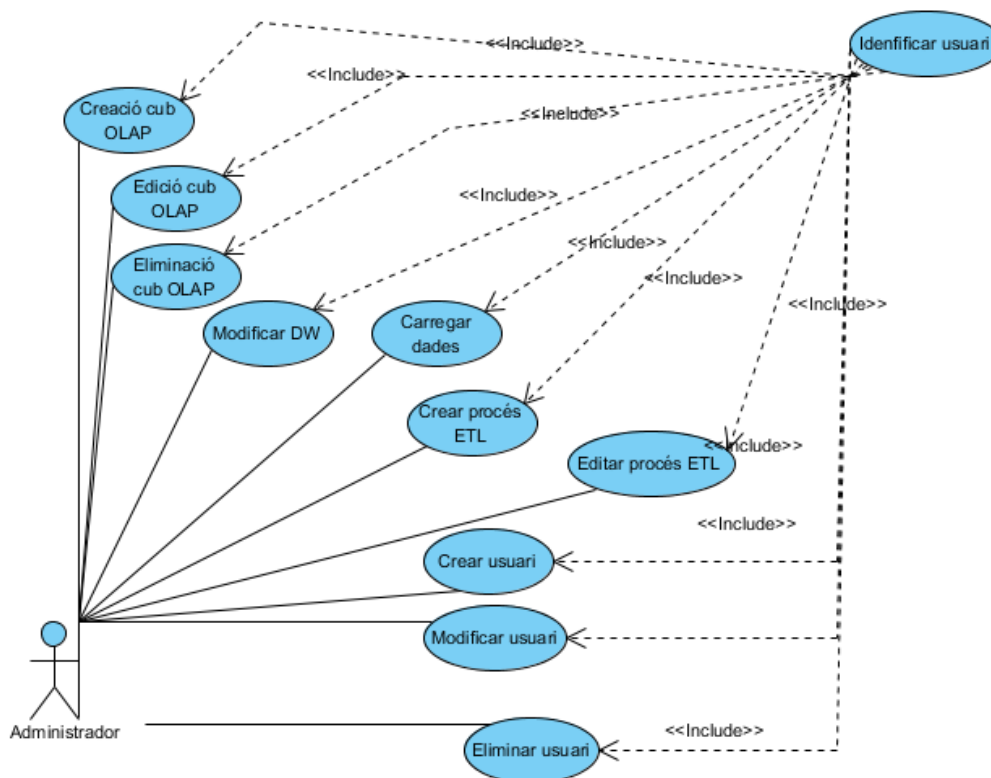
<b>Cas d'ús</b>	Crear Anàlisi.
<b>Descripció</b>	Permet l'edició d'una anàlisi.
<b>CU relacionats</b>	Identificar usuari.
<b>Actors</b>	Analista.
<b>Precondicions</b>	<ol style="list-style-type: none"> <li>1. Usuari donat d'alta i amb permisos.</li> <li>2. Usuari identificat en el sistema.</li> </ol>
<b>Flux</b>	<ol style="list-style-type: none"> <li>1. L'actor elegeix l'esquema.</li> <li>2. L'actor crea l'anàlisi.</li> <li>3. L'actor sol·licita l'emmagatzemament i el sistema executa.</li> </ol>

**Taula 9. Cas d'ús Eliminar Anàlisi**

<b>Cas d'ús</b>	Eliminar Anàlisi.
<b>Descripció</b>	Permet l'eliminació d'una anàlisi.
<b>CU relacionats</b>	Identificar usuari.
<b>Actors</b>	Analista.
<b>Precondicions</b>	<ol style="list-style-type: none"> <li>1. Usuari donat d'alta i amb permisos.</li> <li>2. Usuari identificat en el sistema.</li> </ol>
<b>Flux</b>	<ol style="list-style-type: none"> <li>1. L'actor elegeix l'anàlisi.</li> <li>2. El sistema mostra la informació resultant.</li> <li>3. L'actor confirma l'eliminació i el sistema executa.</li> </ol>

**Administració DW i cubs OLAP**

**Imatge 2 Cas d'ús Administració DW i cubs OLAP**



**Taula 10. Cas d'ús Crear cub OLAP**

<b>Cas d'ús:</b>	Crear cub OLAP.
<b>Descripció</b>	Creació d'un cub a partir de l'esquema relacional del DW.
<b>CU relacionats</b>	Identificar usuari.
<b>Actors</b>	Administrador.
<b>Precondicions</b>	<ol style="list-style-type: none"> <li>1. Usuari donat d'alta i amb permisos.</li> <li>2. Usuari identificat en el sistema.</li> </ol>
<b>Flux:</b>	<ol style="list-style-type: none"> <li>1 L'actor elegeix l'esquema relacional.</li> <li>2 L'actor crea el cub OLAP.</li> <li>3 El sistema demana confirmació.</li> <li>4 L'actor confirma/cancel·la, i el sistema executa.</li> </ol>

**Taula 11. Cas d'ús Eliminar cub OLAP**

<b>Cas d'ús:</b>	Eliminar cub <i>OLAP</i>
<b>Descripció</b>	<i>Elimina un cub OLAP.</i>
<b>CU relacionats</b>	Identificar usuari.
<b>Actors</b>	Administrador.
<b>Precondicions</b>	<ol style="list-style-type: none"> <li>1. Usuari donat d'alta i amb permisos.</li> <li>2. Usuari identificat en el sistema.</li> </ol>
<b>Flux:</b>	<ol style="list-style-type: none"> <li>1 L'actor elegeix el cub OLAP</li> <li>2 L'actor sol·licita la seva eliminació.</li> <li>3 El sistema demana confirmació.</li> <li>4 L'actor confirma/cancel·la i el sistema executa</li> </ol>

**Taula 12. Cas d'ús Editar cub OLAP**

<b>Cas d'ús:</b>	Editar cub <i>OLAP</i> .
<b>Descripció</b>	<i>Modificació dels elements del cub OLAP.</i>
<b>CU relacionats</b>	Identificar usuari.
<b>Actors</b>	Administrador.
<b>Precondicions</b>	<ol style="list-style-type: none"> <li>1 Usuari donat d'alta i amb permisos.</li> <li>2 Usuari identificat en el sistema.</li> </ol>
<b>Flux:</b>	<ol style="list-style-type: none"> <li>1 L'actor elegeix el cub OLAP.</li> <li>2 L'actor edita el cub.</li> <li>3 El sistema demana confirmació.</li> <li>4 L'actor confirma/cancel·la i el sistema executa.</li> </ol>

**Taula 13. Cas d'ús Carregar dades**

<b>Cas d'ús</b>	Carregar dades.
<b>Descripció</b>	Execució procés/processos d'ETL del DW.
<b>CU relacionats</b>	Identificar usuari.
<b>Actors</b>	Administrador.
<b>Precondicions</b>	<ol style="list-style-type: none"> <li>1 Usuari donat d'alta i amb permisos.</li> <li>2 Usuari identificat en el sistema.</li> </ol>
<b>Flux</b>	<ol style="list-style-type: none"> <li>1 L'actor elegeix el treball a executar.</li> <li>2 El sistema demana confirmació.</li> <li>3 El sistema executa/cancel·la el treball sol·licitat.</li> </ol>

**Taula 14. Cas d'ús Editar procés ETL**

<b>Cas d'ús</b>	Editar procés ETL
<b>Descripció</b>	<i>Modificació d'un procés ETL.</i>
<b>CU relacionats</b>	Identificar usuari.
<b>Actors</b>	Administrador.
<b>Precondicions</b>	<ol style="list-style-type: none"> <li>1 Usuari donat d'alta i amb permisos.</li> <li>2 Usuari identificat en el sistema.</li> </ol>
<b>Flux</b>	<ol style="list-style-type: none"> <li>1 L'actor elegeix un treball o transformació.</li> <li>2 L'actor el modifica.</li> <li>3 El sistema demana confirmació.</li> <li>4 El sistema emmagatzema/cancel·la la modificació.</li> </ol>

**Taula 15. Cas d'ús Crear procés ETL**

<b>Cas d'ús</b>	Eliminar procés ETL.
<b>Descripció</b>	Eliminació d'un procés i/o transformació ETL.
<b>CU relacionats</b>	Identificar usuari.
<b>Actors</b>	Administrador.
<b>Precondicions</b>	<ol style="list-style-type: none"> <li>1. Usuari donat d'alta i amb permisos.</li> <li>2. Usuari identificat en el sistema.</li> </ol>
<b>Flux</b>	<ol style="list-style-type: none"> <li>1 L'actor elegeix un treball o transformació.</li> <li>2 L'actor demana l'eliminació.</li> <li>3 El sistema sol·licita la confirmació.</li> <li>4 L'actor confirma/cancel·la i el sistema executa.</li> </ol>

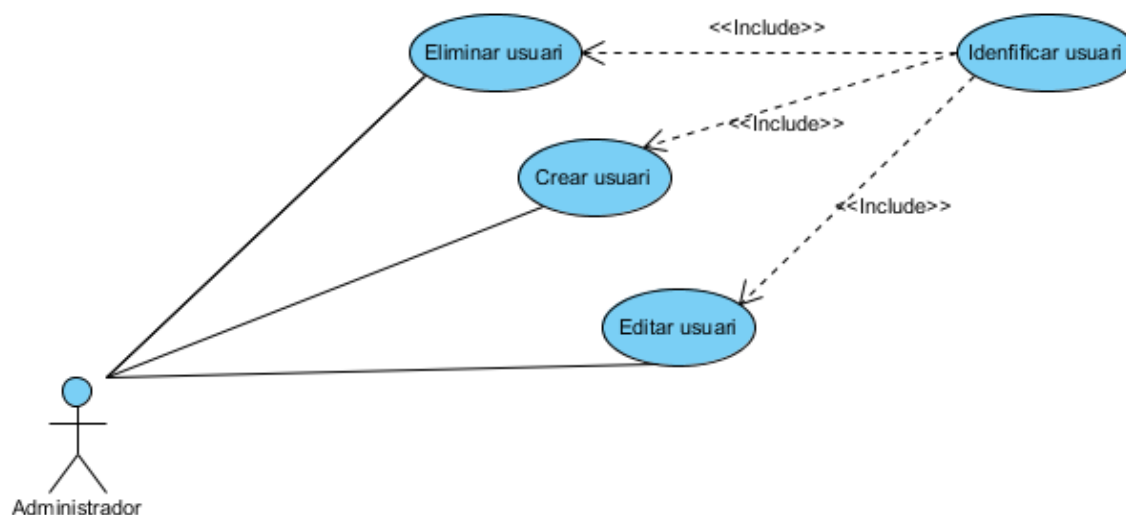


**Taula 16. Cas d'ús Eliminar procés ETL**

<b>Cas d'ús</b>	Eliminar procés ETL.
<b>Descripció</b>	Eliminació d'un procés i/o transformació ETL.
<b>CU relacionats</b>	Identificar usuari.
	Administrador.
<b>Precondicions</b>	1. Usuari donat d'alta i amb permisos. 2. Usuari identificat en el sistema.
<b>Flux</b>	1 L'actor elegeix un treball o transformació. 2 L'actor demana l'eliminació. 3 El sistema sol·licita la confirmació. 4 L'actor confirma/cancel·la i el sistema executa.

### Gestió d'usuaris

**Imatge 3 Cas d'ús Gestió d'usuaris**



**Taula 17. Cas d'ús Crear usuari**

<b>Cas d'ús</b>	Crear usuari.
<b>Descripció</b>	Alta d'usuari en el sistema de BA.
<b>CU relacionats</b>	Identificar usuari.
<b>Actors:</b>	Administrador.
<b>Precondicions</b>	1 Usuari donat d'alta i amb permisos. 2 Usuari identificat en el sistema.
<b>Flux</b>	1. L'actor crea l'usuari. 2. El sistema valida les dades i el dona d'alta.

**Taula 18. Cas d'ús Eliminar usuari**

<b>Cas d'ús</b>	Eliminar usuari.
<b>Descripció</b>	Eliminació d'un usuari del sistema de BA.
<b>CU relacionats</b>	Identificar usuari.
<b>Actors:</b>	Administrador.
<b>Precondicions</b>	1 Usuari donat d'alta i amb permisos. 2 Usuari identificat en el sistema.
<b>Flux</b>	1 L'actor elegeix l'usuari. 2 El sistema valida les dades i l'elimina.

**Taula 19 Cas d'ús Modificar usuari**

<b>Cas d'ús</b>	Modificar usuari.
<b>Descripció</b>	Edició dels permisos d'usuari en el sistema de BA.
<b>CU relacionats</b>	Identificar usuari.
<b>Actors:</b>	Administrador.
<b>Precondicions</b>	1 Usuari donat d'alta i amb permisos. 2 Usuari identificat en el sistema.
<b>Flux</b>	1 L'actor selecciona l'usuari. 2 L'actor modifica els permisos de l'usuari. 3 El sistema valida les dades i les enregistra.

## 2.3 Fonts de dades

S'ha de fer palès que l'anàlisi de les fonts de dades es fonamentarà en l'anàlisi preliminar feta. A més, d'acord amb les decisions que foren preses pel que fa als atributs dels elements del DW, aquells que no foren considerats aptes per a tal fi i que en el subsegüent estudi amb més profunditat no hagin estat tinguts en compte, seran obviats.

S'ha d'aclarir que tots els fitxers proporcionats des del sistema operacionals tenen una estructura en què la primera fila conté els noms dels camps que constitueixen els registres.

### 2.3.1 Studentmobility\_2012-2013\_Datadictionary

Aquesta no és una font de dades pròpiament dita, per tant, no es relacionarà el seu contingut. Però, sí que serà una font cabdal alhora de determinar l'estructura i tipologia de les dades esperades en els fitxers de mobilitat i, en menor grau, en els mestres de països, institucions i àrees de coneixement per referències

### 2.3.2 ISOCountryCodes081507

- Descripció: llistats de codis i atributs ISO normalitzats de les nacionalitats.
- Format: full de càlcul XLS.
- Nombre de registres: 252.

**Taula 20. estructura ISOCountrycodes081507**

Nom	Tipus	Domini	Observacions rellevants
Code	String	Alfabètic (ISO Country code)	4 caràcters
Country	String	Alfanumèric (text)	Màx. 48 caràcters.

### 2.3.3 EUC\_Consolidated\_Table\_2007\_2013

- Descripció: llistat consolidat de codis i atributs d'institucions educatives.
- Format: full de càlcul XLS.
- Nombre de registres: 4.918.

**Taula 21 Estructura EUC\_Consolidated\_Table\_2007\_2013.**

Nom	Tipus	Domini	Observacions rellevants
Country	String	ISO Country Code	2 caràcters
Charter type code	String	[EUC  EUCP  EUCX]	Màx. 4 caràcters
Organization Name	String	Alfabètic (text)	117 caràcters
Street	String	Alfabètic (text)	
PostCode	String	Alfanumèric	
City	String	Alfanumèric (text)	

### 2.3.4 SCED97\_Erasmus\_subject\_codes

Hi ha dues pestanyes, la primera conté els codis ISCED97:

- Descripció: llistat de codis i atributs d'àrees ISCED97.
- Format: full de càlcul XLS.
- Registres: 146.

**Taula 22. Estructura dels codis ISCED97.**

Nom	Tipus	Domini	Observacions rellevants
Code	String	Numèric	3 dígits
Subject Area	String	Text lliure	Màx. 60

La segona pestanya conté els codis Erasmus:

- Descripció: llistat de codis i atributs d'àrees de coneixement.
- Format: full de càlcul XLS.
- Registres: 140.

**Taula 23. Estructura SCED97\_Eramus\_subject\_codes.**

Nom	Tipus	Domini	Observacions rellevants
Full Code	String	Alfanumèric	4 caràcters amb punt
Subject Area	String	Alfanumèric	Màx. 60
Equivalent ISCD97	String	Numèric	3 dígits
ISCD97 Description	String	Alfanumèric	3 dígits

### 2.3.5 SM\_2011\_2012

- Descripció: mobilitat d'estudiants Erasmus curs 2011-2012.
- Procedència: mobilitat distintes institucions.
- Format: CSV separat per “;”.
- Número de registres: 252.827.

**Taula 24. Estructura SM\_2011\_2012.**

Nom Cam	Tipus	Domini	Obs. rellevants
HOMEINSTITUTION	String	Erasmus code,	13 car.
COUNTRYCODEOFHOMEINSTITUTION	String	ISO Country Code	
AGE	Numèric	2 dígit	
GENDRE	String	[M F]	
NATIONALITY	String	SO Country Code	
SUBJECTAREA	Numèric	ISCED97Erasmus	
LEVELSTUDY	String	[1   2   3   S]	
YEARS PRIOR	Numèric	Sencer [0 ..14]	
MOBILITYTYPE	String	[S   C   P]	
HOSTINSTITUTION	String	Erasmus Code	
COUNTRYCODEOFHOSTINSTITUTION	String	ISO Country codes	
PLACEMENTENTERPRISE	String	Text lliure	Màx. 208 car.
COUNTRYOFPLACEMENT	String	ISO Country Codes	
ENTERPRISESIZE	String	[L   M   S]	
TYPEPLACEMENTSECTOR	String	[?   NACE sector codes]	Màx. 1 car.
LENGTHSTUDYPERIOD	Numèric	. [0.00 –13.50]	Fins a 2 dec.
LENGTHPLACEMENT	Numèric	[0.00 –13.50]	Fins a 2 dec.
SHORTDURATION	String	[?   T]	
STUDYSTARTDATE	Data	format mm-yy	
PLACEMENTSTARTDATE	Data	format mm-yy	
CONSORTIUMAGREEMENTNUMBE	String		
ECTSCREDITSSTUDY	Numèric	Sencer [0..90]	
ECTSCREDITSPLACEMENT	Numèric	Sencer [0..90]	
TOTALECTSCREDITS	Numèric	Sencer [0..90]	
SNSUPPLEMENT	Numèric	Sencer [0 ..1600]	
TAUGHTHOSTLANG	String	[Y   N]	
LANGUAGETAUGHT	String	ISO Languages Code XX	
LINGPREPARATION	Alfabètic	EC  HS   HM  NN]	
STUDYGRANT	String	Fins a 2 decimals	Pot ser sencer
PLACEMENTGRANT	String	Fins a 2 decimals	Pot ser sencer
PREVIOUSPARTICIPATION	String	[N   S]	
QUALIFICATIONATHOST	String	[D   J   N]	

### 2.3.6 SM\_2012\_2013

- Descripció: mobilitat d'estudiants Erasmus curs 2012-2013
- Procedència: mobilitat de distintes institucions.
- Format: CSV separat per “;”.
- Número de registres: 267.547.

**Taula 25 Estructura SM\_2012\_201**

<b>Nom Cam</b>	<b>Tipus</b>	<b>Descripció</b>	<b>Obs. rellevants</b>
STUDENT_ID	String	Alfanumèric	Fins a 20 car.
ID_MOBILITI_CDE,	String	Alfanumèric	15 car.
HOME_INSTITUTION_CDE	String	Erasmus code	13 car.
HOME_INSTITUTION_CTRY_CDE	String	ISO Country code	
ESTUDENT_AGE_VALUE	Numèric	Sencer [19..63]	
ESTUDENT_GENDER_CDE	String	[F   M]	
ESTUDENT_NATIONALITY_CDE	String	ISO Country Code  XX	
STUDENT_SUBJECT_AREA_VALUE	Numèric	Equivalent ISCED97	
STUDENT_STUDY_LEVEL_CDE	String	[1   2   3   S]	
NUMB_YRS_HIGHER_EDUCAT_VALUE	Numèric	Sencer (3 d )	
MOBILITY_TYPE_CDE	String	[P   S]	
HOST_INSTITUTION_CDE	String	['???'   Erasmus Code]	
HOST_INSTITUTION_COUNTRY_CDE	String	ISO Country code or '???'	
PLACEMENT_ENTERPRISE_VALUE	String	? Unknow ?   freetext	Màx. 208 car.
PLACEMENT_ENTERPRISE_CTRY_CDE	String	? Unknow ?'   freetext)	
PLACEMENT_ENTERPRISE_SIZE_CDE	String	[?   L   M   S]	
TYPE_PLACEMENT_SECTOR_VALUE	String	[?   NACE sector codes] r.	
LENGTH_STUDY_PERIOD_VALUE,	Numèric	Fins a 2 dec. [0.00 –13.50]	
LENGTH_PLACEMENT_VALUE	Numèric	Fins a 2 dec. [0.00 –13.50]	
SHORT_DURATION_CDE	String	[?   T]	
STUDY_START_DATE	Data	[mm-yy o 00/0000]	(*)
PLACEMENT_START_DATE	Data	[mm-yy o 00/0000]	
CONSORTIUM_AGREEMENT_NUMBER	String	1Alfabètic	18 car.
ECTS_CREDITS_STUDY_AMT	Numèric	Sencer [0..99]	
ECTS_CREDITS_PLACEMENT_AMT	Numèric	Sencer [0..99]	
TOTAL_ECTS_CREDITS_AMT,	Numèric	Sencer [0..99]	
SPECIAL_NEEDS_SUPPLEMENT_VALUE	Numèric	Fins a 2 dec. [0.00 –13.50]	
TAUGHT_HOST_LANGUAGE_CDE	String	[Y   N]	
LANGUAGE_TAUGHT_CDE	String	ISO Code languages	
LINGUISTIC_PREPARATION_CDE	String	[EC   HM   HS   NN]	
ESTUDY_GRANT_AMT	Numèric	Fins a 2 decimals	
PLACEMENT_GRANT_AMT	Numèric	Fins a 2 decimals	
PREVIOUS_PARTICIPATION_CDE	String	[N   P   S]	
QUALIFICATION_AT_HOST_CDE	String	[D   J   N]	

### 2.3.7 NACE CODES

- Descripció: Codis dels sectors empresarials segons la UE.
- Procedència: Internet (veure bibliografia).
- Format: PDF.

### 2.3.8 Resultat de l'anàlisi de les fonts de dades

El fitxer *ISCED97\_Erasmus\_subject\_codes* presenta dues pestanyes que es corresponen amb l'entitat *SubjectArea* amb dues codificacions: codificació Erasmus i codificació ISCED97. En no haver-hi en la segona tots els codis Erasmus, es presenta el problema de què no és possible fer-ne la unió de les dues i que en resulti una dimensió amb valors en tots els camps: sempre quedaran registres amb valors nuls als atributs *Full Code* i *Subject Area*. En ser una possible font de problemes serà millor centrar-se en la codificació ISCED97, que, per altra banda, és la que presenten les fonts de dades de mobilitat.

En qualsevol cas, els punts més conflictius es presenten en els fitxers de mobilitat, punts que s'hauran de tenir en compte a l'hora de muntar el magatzem de dades i que es relacionen a continuació en la taula 26.

**Taula 26. Resultat de l'anàlisi.**

Incidència	Acció correctiva/Conseqüències
Camps dels registres amb nom diferent	Homogeneïtzar la nomenclatura.
Diferent nombre de camps en els fitxers de mobilitat	Pot afectar el gràdul.
Només hi ha "id". d'alumne en una de les fonts	Afectarà a la granularitat del fet.
Els camps que fan referència al país de les institucions receptores presenten valors de 4 caràcters que són combinació de dos codis de país permesos (casos: BEFR i BENL)	A tenir en compte en els processos d'ETL, Només es consideraran els dos primers.
Una de les fonts (SM_2012_2013), en lloc de deixar el camp buit d'acord a l' <a href="#">annex II</a> , els emplena amb valors tals com '?', '???' o '? Unknow ?'	A tenir en compte en els processos d'ETL. S'hauran de posar a nul.
Els valors del camp subject àrea és el codi <i>Erasmus Short Code</i>	Afectarà al procés d'ETL.
Camps de dates (SM_2011_2012) amb dates incorrectes, o quan hauria de ser buit presenta els valors '00/0000'. A més, s'haurà de tenir en compte el control de la data de començament, que haurà de ser >= que el rang inferior de la data del curs. També hi haurà que tenir en compte la llengua per defecte dels sistema alhora de carregar les dades.	A tenir en compte en els processos d'ETL.
Els camps de les beques d'estudi o de pràctiques poden tenir valors sencers o decimals	S'ha de tenir en compte en els processos d'ETL.
Hi ha fitxers en què no hi ha total de beca.	El camp total serà calculat.
El camp <i>Placement</i> és text lliure. A més, per una mateixa instància es pot presentar en valors de sector i mida d'empresa diferents.	Presentarà problemes a l'hora de fer les insercions i actualitzacions .

## 3 Disseny

### 3.1 Arquitectura

En el projecte de construcció del DW s'han utilitzat dos entorns: el de producció i el de desenvolupament, ambdós virtualitzats, i amb el mateix programari de base (sistema operatiu i sistema de gestió de bases de dades relacionals) i el mateix programari BA (Pentaho Business BA) i OLAP (motor OLAP i eines d'informes, anàlisi i disseny). Per tant, els administradors i els analistes hi poden accedir utilitzant la mateixa interfície sigui quin sigui l'entorn considerat (a la taula 27, es presenten les característiques del programari i maquinari utilitzat en el projecte).

**Taula 27. Arquitectura.**

	<b>Producció</b>	<b>Desenvolupament</b>
<b>Maquinari</b>	MV Amazon (2 nuclis 3.7 GB RAM)	MVWare (1 nuclis 2 GB RAM)
<b>Programari</b>	SO Ubuntu 14 LTS	SO Ubuntu 14 LTS
	SGBDR MySQL 5.5.4 CE	SGBDR MySQL 6.0 CE
	MySQL Workbench 6.0 CE	MySQL Workbench 6.3 CE
	Pentaho BA 5.4.0.1-130 (Haiku, CDE)	Pentaho BA 5.4.0.1-130 (Haiku i CDE)
	PDI 5.4.0.1-130 (Kettle)	PDI 5.4.0.1-130 (Kettle)
	Modrian Schema Workbench 3.10.0.1-130	Modrian Schema Workbench 3.10.0.1-130

#### 3.1.1 Programari

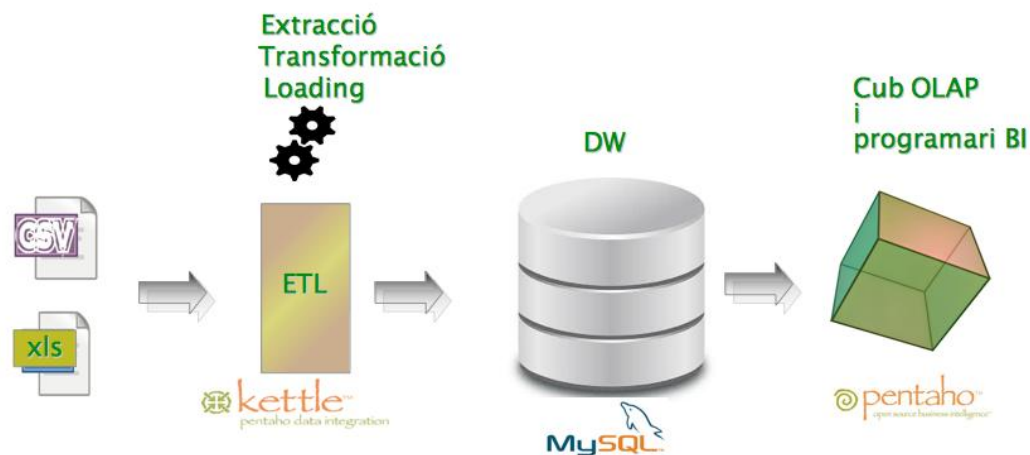
En primer lloc, es disposa de les fonts de dades que es presenten en formats CSV o XLS. Aquestes contenen les dades que poblaran el DW, un cop hagin passat pel programari d'integració, el qual es conegut com d'ETL (*Extraction, Transformation and Loading*). Un cop feta la càrrega, ja pot ser utilitzat pel programari OLAP on es defineixen els cubs a utilitzar i es poden especificar les consultes MDX. Esquemàticament, l'engranatge de tot el programari del DW és (veure imatge 4):

1. Extracció de les fonts externes procedents dels sistemes operacionals cap a l'àrea de maniobra o *Staging Area* (SA, d'ara en endavant).
2. Transformació de les dades i càrrega al DW.
3. Definició dels cubs OLAP.
4. Definició de les consultes MBX necessàries.
5. Generació dels informes fixos a utilitzar.
6. Extracció d'informació.

El programari utilitzat, *Suite Pentaho de BA*, ha estat elegit per ser capaç de donar solució als requeriments funcionals i no funcionals i, en concret, per:

1. Capacitat d'oferir programari d'ETL de fàcil utilització.
2. Generació de cubs OLAP utilitzant com a font un DW relacional independent del RDBMS utilitzat.
3. Generació dels cubs mitjançant una interfície gràfica amigable.
4. Capacitat d'oferir la utilització, creació i modificació d'informes mitjançant consultes MDX.
5. Capacitat d'anàlisi a la carta pels analistes i extracció d'informes al vol.
6. Accés als elements del DW utilitzant una interfície gràfica web amb accés restringit per rol i fàcilment gestionable per l'administrador.

**Imatge 4 Arquitectura del programari**

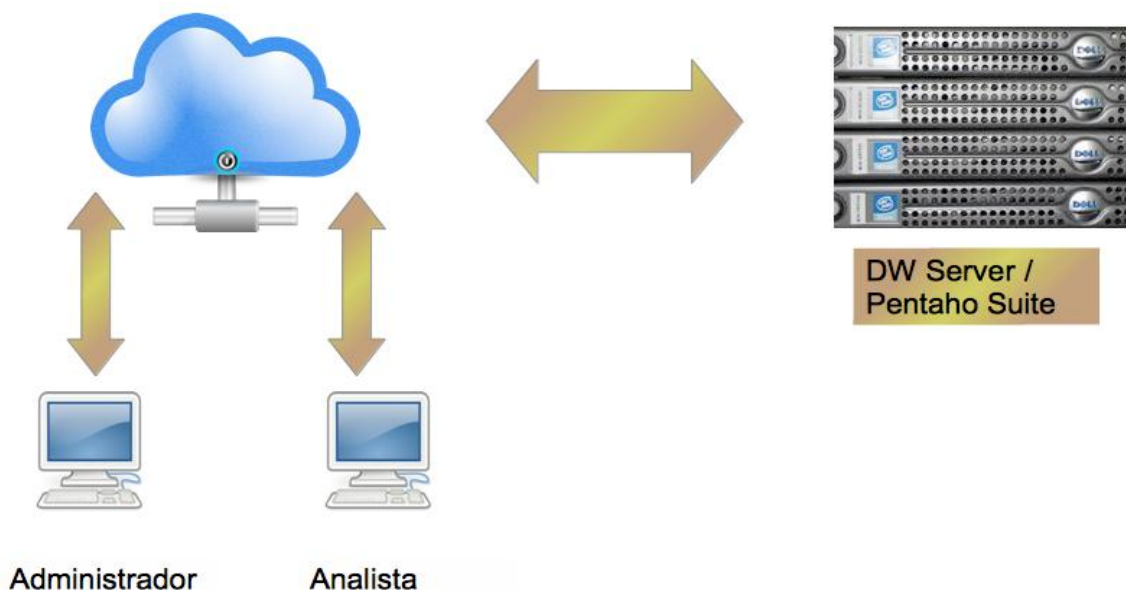


### 3.1.2 Maquinari

El maquinari que dona suport al programari utilitzat per a la implementació de DW i a la suite de BA és una màquina virtualitzada en el núvol (requeriment no funcional).

Tant els administradors del sistema de DW com els analistes accedeixen a les funcionalitats, descrites en l'apartat anterior i en la resta del document, a través de connexions de xarxa utilitzant la xarxa pública (imatge 5).

**Imatge 5 Arquitectura del maquinari**





## 3.2 Disseny conceptual

Per determinar el model conceptual de dades del DW s'hauran d'especificar:

- El fet o fets.
- La seva granularitat.
- Les mesures que es necessiten.
- Les dimensions que intervenen i els seus atributs.
- La cel·la del grànul utilitzat.
- Les restriccions que s'han d'aplicar.
- La viabilitat.

### 3.2.1 Fet/fets

El fet és la mobilitat dels estudiants en del programa europeu ERASMUS (Mobility), independent de la tipologia de les institucions de destinació (*Institution/Placement*) dels estudiants.

### 3.2.2 Granularitat

Quant major sigui el nivell de detall del grànul, de més informació es disposarà i més precises seran les mesures dels fets. Ara bé, també augmentarà exponencialment la mida del DW. Per contra, quan menys detall presenti el grànul elegit, més informació es perdrà i, en conseqüència, menor serà la precisió de determinades mesures. En una paraula, en escollir el nivell de detall del grànul sempre s'ha de ponderar la precisió que es necessita i l'espai de què es disposa.

La granularitat no s'ha pogut arribar a considerar a nivell atòmic, per falta de dades identificatives dels estudiants en tots i cada un dels fitxers de moviments. El fet *Mobility* ha abocat a plantejar un nivell d'agregació superior en què una mobilitat qualsevol serà descrita per les dimensions següents:

- Gènere (*Gender*).
- Edat (*Age*).
- Preparació lingüística (*LingualPreparation*).
- Nivell d'estudis (*StudyLevel*).
- Nacionalitat (*Nacionality*).
- Àrea de coneixement (*SubjectArea*).
- Tipus de mobilitat (*MobilityType*).
- Institució emissora (*HomeInstitution*).
- Institució receptora (*HostInstitution/Placement*).
- Curs, que donarà el marc temporal (*Course*).

### 3.2.3 Mesures

Són els atributs que conformen un fet, generalment són numèrics, com en el cas objecte d'anàlisi. Aquestes mesures, indicadors que es tindran en compte per poder cobrir els requeriments funcionals, seran:

- Students (nombre d'estudiants, suma).
- %\_Students (% d'estudiants, suma dels estudiants en tant per cent).
- ageAverage (edat mitjana dels estudiants respecte del total).
- %\_StudentsWithGrant (% d'estudiants amb beques).

És important aclarir que la mesura "Students" fa referència **no als estudiants diferents que participen en el programa Erasmus, si no al nombre de mobilitats que es produeixen independentment dels estudiants particulars que hi participen**, ja que hi poden haver dues o més mobilitats del mateix estudiant, però es consideraran mobilitats diferents. De fet, el fet mobilitat individual associat a un estudiant en particular o a un altre no interessa al programa, el que sí que ho fa **és el fet en si**.

Ara bé, aquestes mesures seran les que es definiran en l'eina OLAP. En la implementació relacional del DW s'hauran d'emmagatzemar les següents mesures (que contindran els valor a nivell de grànul):

- *students* (nombre d'estudiants).
- *ageAvgContribution* (edat del estudiants multiplicat pel nombre d'estudiants del mateix grànul).
- *withStudyGrant* (número d'estudiants amb beca d'estudi del grup).
- *withPlacementGrant* (número d'estudiants amb beca de pràctiques del grup).
- *withTotalGrant* (suma de les dues quantitats anteriors).

### 3.2.4 Dimensions

Les dimensions ofereixen punts de vista discrets dels fets des dels quals analitzar-los. S'infereixen de l'enunciat del TFG, de l'anàlisi de les fonts de dades, dels informes prefixats i de la granularitat especificada anteriorment. Les dimensions identificades es relacionen a la taula 28.

**Taula 28. Dimensions del DW**

Dimensió	Rol	Descriptor / Nivells de jerarquia
Course		Nivells de jerarquia: *all/course.
Institution Academic Placement	HomeInstitution	Nivells de jerarquia *all/institution/nacionalitat.
	HostInstitution	Nivells de jerarquia d'agregació: Per ambdues tipologies: *all/institution/nacionalitat. Pel cas de l'empresa de pràctiques: * all/institution/sector. * all/institution/size.
SubjectArea		Nivells de jerarquia: *all/subjectArea.
MobilityType		Nivell de jerarquia. *all/mobilityType.
StudyLevel		Nivells de jerarquia: *all/studyLevel.
Nationality		Nivells de jerarquia: *all/nationality.
Gender		Nivells de jerarquia d'agregació: *all/gender.
Age		Nivells de jerarquia d'agregació: *all/age.
LingPreparation		Nivells de jerarquia d'agregació: *all/lingPreparation.

### 3.2.5 Atributs de les dimensions

Els atributs elegits estan descrits en el fitxer (*Studentmobility\_2012-2013\_Datadictionary*) analitzat en l'apart **Font de dades** d'aquest document. Ara bé, se n'han afegit d'altres, en segons quines dimensions, per facilitar als analistes l'ús en la selecció. En les taules següents s'han resumit: la dimensió, l'atribut o atributs de la dimensió i la modalitat d'atribut de què es tracte (descriptor, nivell de jerarquia).

**Taula 29. Curs**

DIMENSIÓ			
Course	Atributs	Tipus	Modalitat
	course	Alfanumèric (ex.: 2011_2012)	Descriptor

**Taula 30 Edat**

DIMENSIÓ			
Age	Atributs	Tipus	Modalitat
	Age	Sencer (2 dígit)	Descriptor

**Taula 31. Nacionalitat**

DIMENSIÓ			
Nationality	Atributs	Tipus	Classe
	countryCode	Alfabètic (ISO Country Code)	Descriptor
	countryName	Text lliure	Descriptor

**Taula 32. Gènere**

DIMENSIÓ			
Gender	Atributs	Tipus	Classe
	genderShort	Alfabètic [M F]	Descriptor
	genderLarge	Alfabètic cadena [Male Female]	Descriptor

**Taula 33. Preparació en llengües**

DIMENSIÓ			
LingPreparation	Atributs	Tipus	Classe
	lingPreparationShort	[EC HS HM NN]	Descriptor
	lingPreparationLarge	[EILC   Host  Home   None]	Descriptor

**Taula 34. Nivell d'estudis**

<b>DIMENSIÓ</b>			
<b>StudyLevel</b>	<b>Atributs</b>	<b>Tipus</b>	<b>Classe</b>
	<i>studyLevelShort</i>		[1   2   3   S]
	<i>studyLevelLong</i>	[First Cycle  Second Cycle Third Cycle Short Cycle]	<i>Descriptor</i>

**Taula 35. Tipus de mobilitat**

<b>DIMENSIÓ</b>			
<b>MobilityType</b>	<b>Atributs</b>	<b>Tipus</b>	<b>Classe</b>
		<i>mobilityTypeCode</i>	[S C P]
	<i>mobiltyTypeDescription</i>	[Mobility for study Combined S&P Mobility for placement]	<i>Descriptor</i>

**Taula 36. Institució (emissora/receptora)**

<b>DIMENSIÓ</b>				
<b>Institution</b>	<b>Atributs</b>	<b>Tipus</b>	<b>Classe</b>	
	<i>institutionName</i>		Alfanumèric	<i>Descriptor</i>
	<i>countryCode</i>		Alfabetic (ISO Country Code)	<i>Nivell de jerarquia</i>

**Taula 37. Institució acadèmica**

<b>DIMENSIÓ</b>			
<b>Academic (Institution)</b>	<b>Atributs</b>	<b>Tipus</b>	<b>Classe</b>
		<i>erasmusCode</i>	Afanumèric (Erasmus Code)

**Taula 38. Empresa on s'efectuen les pràctiques**

<b>DIMENSIÓ</b>			
<b>Placement (Institution)</b>	<b>Atributs</b>	<b>Tipus</b>	<b>Classe</b>
	<i>sizeShort</i>	<i>Alfabètic [L M S]</i>	<i>Nivell de jerarquia</i>
	<i>sizeLong</i>	<i>Alfabètic [Large Medium Small]</i>	<i>Nivell de jerarquia</i>
	<i>sectorShort</i>	<i>Alfabètic..( NACE Sector Code)</i>	<i>Nivell de jerarquia</i>
	<i>sectorLong</i>	<i>text lliure descriptiu</i>	<i>Nivell de jerarquia</i>

**Taula 39. Àrees de coneixement**

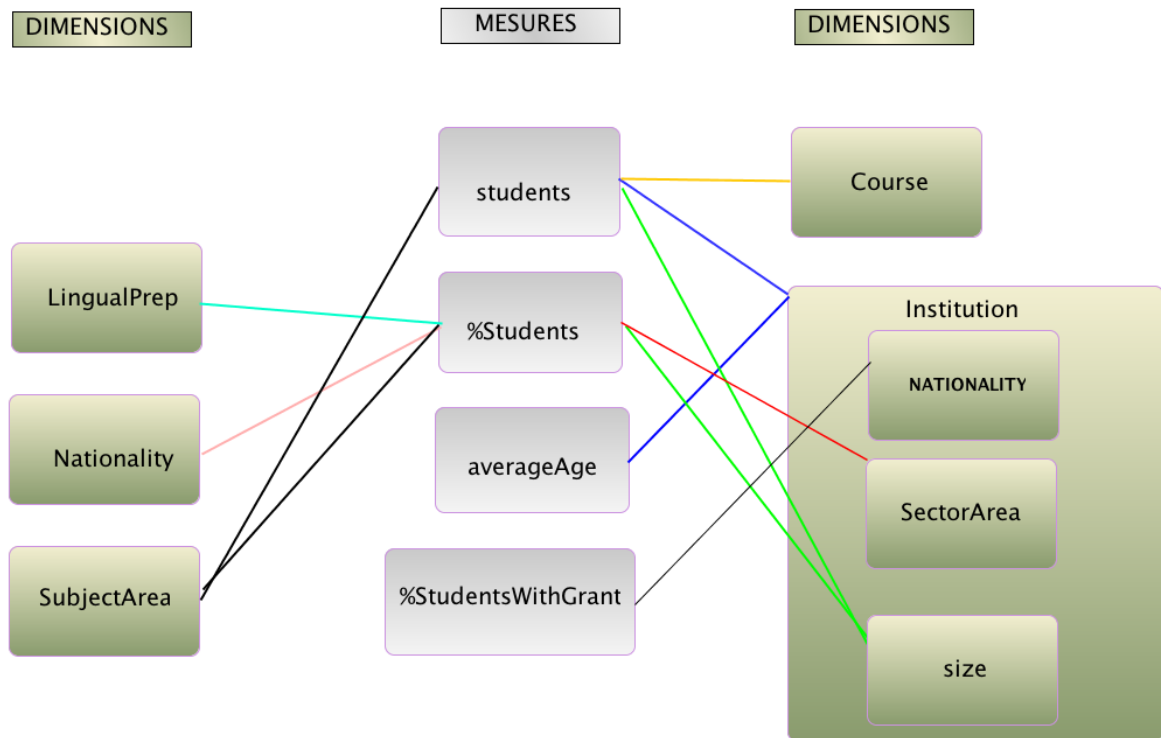
<b>DIMENSIÓ</b>			
<b>SubjectArea</b>	<b>Atributs</b>	<b>Tipus</b>	<b>Classe</b>
	ISCD97Code	String (3 dígit)	Descriptor
	ISCD97Description	Text lliure	Descriptor
	erasmusShortCode	String (1 dígit)	Nivell jerarquia

### 3.2.6 Cel·les

És el moment d'enllaçar la granularitat amb les mesures escollides, ja que els indicadors seleccionats entren en joc amb un gràdul específic, i no afecten a totes les dimensions. Tot seguit es passa a relacionar indicador i dimensió, i, si s'escau, nivell de jerarquia afectat (veure Imatge 6):

- Students per institució (receptora/emissora).
- Students per curs.
- Students per àrea de coneixement i per sector empresarial.
- %Students per nacionalitat.
- %Students per àrea de coneixement.
- %Students per sector empresarial.
- %Students per mida d'empresa.
- %students per preparació en llengües.
- averageAge per institució (receptora/emissora).
- %\_StudentsWithGrant per nacionalitat de la institució (receptora/emissora).

Imatge 6 Relació mesures/elements del DW (Dimensions i nivell de jerarquia)<sup>1</sup>



### 3.2.7 Restriccions

Pel que fa a la dimensió *Institution* i als nivells de jerarquia de la subclasse *Placement*, s'haurà de tenir en compte que tant *Size* com *SectorArea* són disjunctes i que, per tant, per considerar totes les seves instàncies (*all*) s'haurà de fer l'agregació a nivell de *Placement*.

### 3.2.8 Viabilitat

Per determinar la viabilitat de la solució s'haurà de fer una estimació de l'espai ocupat pel DW. Amb aquest objectiu es farà un càlcul anual, suposant les següents premisses:

- Els fitxers mestres són poc volàtils i, en conseqüència, no tindran augments significatius en nombre de registres.
- Els fitxers de mobilitat seran un per curs.
- El nombre total de registres del fitxers de mobilitats:
  - Seran de l'ordre del 255.000 registres, però en agrupar-se per grànuls no atòmics (veure l'apartat **Fet**) hi pot haver una disminució d'uns 25.000 .Tot plegat, dona uns 220.000 registres per fitxer.

Els càlculs es faran considerant que els atributs:

<sup>1</sup> En la figura, la dimensió *Institution* actua amb dos papers, *HostInstitution* i *HomeInstitution*, i aquesta darrera, a més,

- Alfanumèrics s'emmagatzemaran en tipus *VARCHAR* que ocupa el nombre de caràcters més un.
- Els numèrics menors de 255 ho faran en *UNSIGNED TINYINT* que ocupa un byte.
- Els numèrics menors de 65.535 ho faran en *UNSIGNED SMALLINT* que ocupa 2 bytes.
- Els numèrics menors de 16.777.215 en *UNSIGNED MEDIUMINT* que ocupa 3 bytes.
- Els numèrics menors de 4.294.967.295 en *UNSIGNED INT*, que ocupa 4 bytes.
- Els numèrics decimals ho faran en *decimal(9,2)* que ocupa 4 bytes i *decimal(10,2)*, que n'ocupa 5.

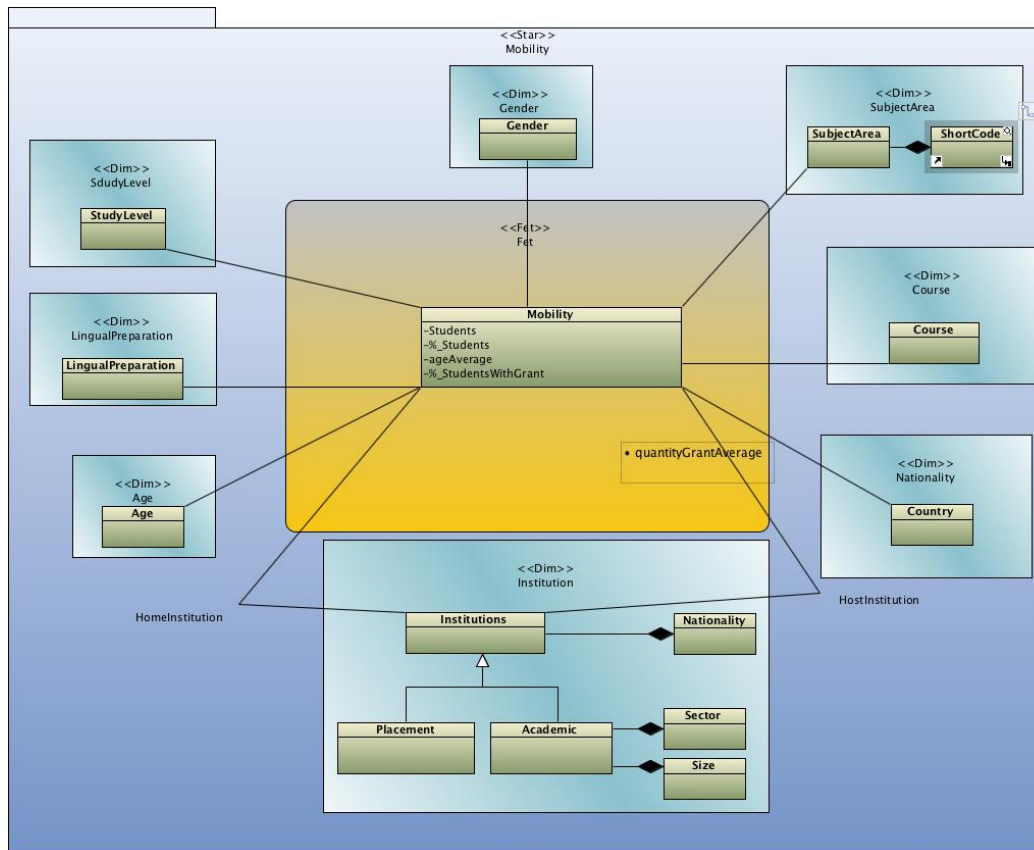
Els resultats de l'anàlisi es resumeixen en la taula 40, on s'observa que el DW ocuparà el primer any uns 23 MB amb una ràtio de creixement de l'ordre dels 8 MB anuals, amb la qual cosa **es pot concloure la seva viabilitat**.

**Taula 40. Estimació aproximada de la mida del DW.**

<i>Taula</i>	<i>Registres</i>	<i>Bytes/registre</i>	<i>Bytes/taula</i>	<i>MB/taula</i>
<i>Mobility</i>	220.000	38	8.360.000	7,97
<i>Course</i>	1	10	10	-
<i>Country</i>	253	58	14674	-
<i>Institution</i>	4,918	389	1.913.102	1,82
<i>Placement</i>	~70.000	212	14.840.000	14,15
<i>SubjectArea</i>	146	80	11.680	0,01
<i>MobilityType</i>	3	29	87	-
<i>StudyLevel</i>	4	19	76	-
<i>Age</i>	89	2	178	-
<i>Gender</i>	2	14	28	-
<i>MobilityType</i>	3	32	96	-
<i>LinguisticPrep</i>	4	14	56	-
			<b>TOTAL aproximat</b>	<b>22.95</b>

El model conceptual que en resulta, presentat en la imatge 7, defineix el DW des del punt de vista conceptual, el més proper al negoci i, per tant, a l'usuari.

## Imatge 7 Model conceptual<sup>2</sup>



### 3.3 Disseny tècnic

Un cop realitzat el disseny multidimensional, el disseny conceptual s'ha de traspasar al disseny lògic (model relacional) i al físic. D'una banda, s'hauran de passar les entitats a relacions (taules) i, de l'altra, donar-les una solució física sobre el RDBMS elegit.

Pel model relacional, cada dimensió donarà lloc a una taula (relació) i el mateix succeirà amb el fet. Ara bé, la dimensió *Institution* presenta herència i haurà de ser tractada aplicant-hi una de les tres solucions següents:

- Una sola taula: amb els atributs tant de la superclasse (*Institution*) com els de les subclasses (*Acadèmic* i *Placement*).
- Dues taules: una per a cada subclasse que continguin cada una els atributs propis i els de la superclasse *Institution*.
- Tres taules diferents: una per a la superclasse i una per a cada subclasse.

Pel que fa a l'esquema del DW, una altra decisió a prendre serà considerar quina de les configuracions, la d'estrella o la de floc de neu, és la més adequada pel nostre cas. Totes les consideracions presentades s'han de ponderar per fer un disseny que doni lloc al DW que, a més d'eficaç, sigui eficient.

De l'anàlisi, i tenint en compte les recomanacions que es recullen a la bibliografia, el disseny elegit presentarà les següents característiques:

<sup>2</sup> En el diagrama de l'esquema conceptual no es mostren les mesures a nivell de taula (estudiants a nivell de grànul, ni els imports de les beques a nivell de grànul)



- S'utilitzarà la configuració en estrella, d'aquesta manera primarem l'eficiència en la resposta sobre la d'estalvi d'espai del DW
- El disseny es representarà utilitzant una sola taula. L'elecció té en compte que la taula de fets, *Mobility*, es relaciona amb *Institution* directament, i amb dos rols (*HomeInstitution* i *HostInstitution*), per la qual cosa primarà la velocitat d'accés sobre la d'estalvi d'espai.
  - També es considerarà en la taula de dimensió *Institution* els atributs de nivell jeràrquic:
    - countryCode i countryName.
    - sectorShort i sectorLarge.
    - sizeShort i sizeLong.
  - Els atributs que siguin propis d'una subclasse tindran en les ocurrencies de l'altre un valor fix per evitar l'aparició de valors nuls.

A nivell físic, es canviaran les claus primàries naturals per claus subrogades per a totes i cada una de les taules, siguin de dimensió o de fets.

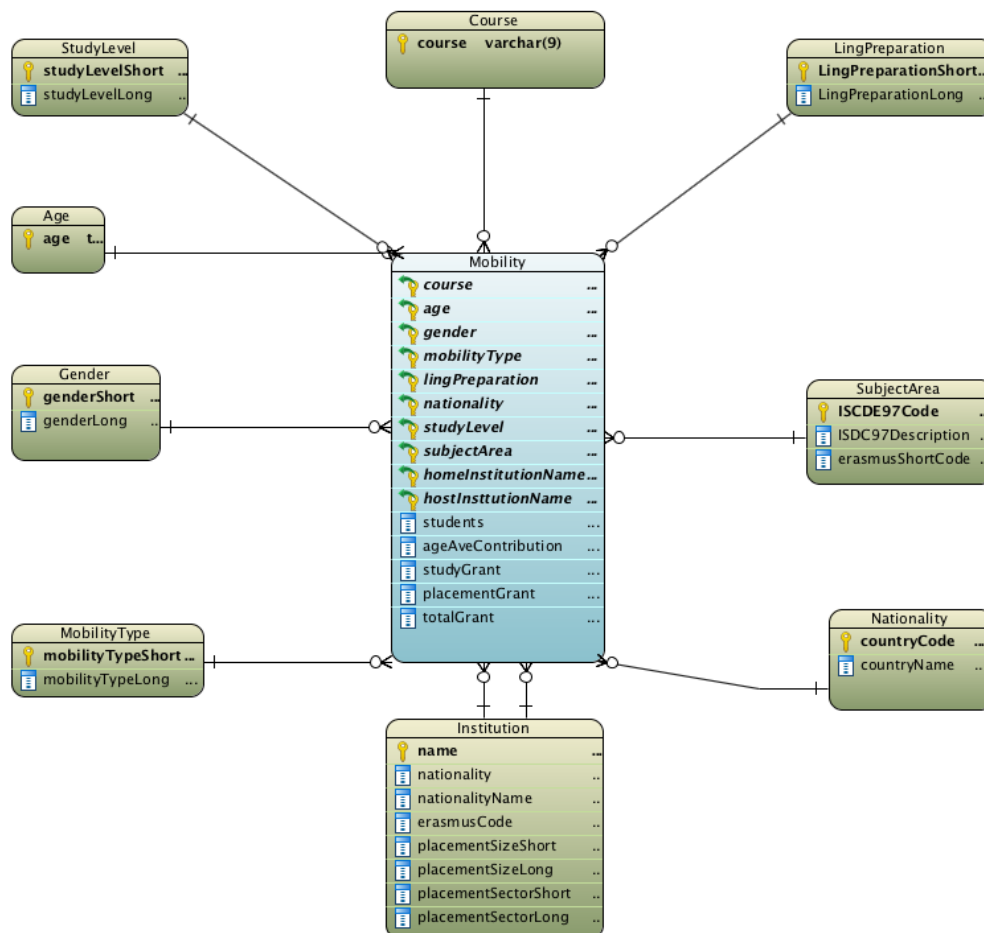
### 3.3.1 Model lògic

Primerament, es relacionaran les taules que en resulten, i, tenint en compte la simplicitat de l'esquema en estrella, s'utilitzarà la següent notació: ***negreta, cursiva i subratllat*** per la clau primària, *cursiva* per les alternatives i *cursiva i subratllat* per les alienes. El resultat es presenta a la taula 41, on les relacions donen lloc a l'esquema conceptual representat en la imatge 8:

Taula 41 Model relacional.

Relació	Atributs	Claus estrangeres	
		Claus	Taula referenciada
Course	<b><u>course</u></b>		
Age	<b><u>age</u></b>		
Gender	<b><u>genderShort</u></b> , genderLarge		
Nationality	<b><u>countryCode</u></b> , countryName		
StudyLevel	<b><u>studyLevelShort</u></b> , studyLevelLarge		
MobilityType	<b><u>mobilityTypeShort</u></b> , mobilityTypeLong		
LingPreparation	<b><u>lingPreparationShort</u></b> , lingPreparationLong		
SubjectArea	<b><u>ISCED97Code</u></b> , ISCED97Description, erasmusShortCode		
Institution	<b><u>name</u></b> , nationality, erasmusCode, placementSizeShort, placementSizeLong, sectorShort, sectorLarge		
Mobility	<b><u>course</u></b> , <b><u>age</u></b> , <b><u>gender</u></b> , <b><u>studyLevel</u></b> , <b><u>nationality</u></b> , <b><u>mobilityType</u></b> , <b><u>lingPreparation</u></b> , <b><u>subjectArea</u></b> , <b><u>hostInstitution</u></b> , <b><u>homeInstitution</u></b> , students, ageAvgContribution, studyGrant, placementGrant, totalGrant	<i>course</i> <i>age</i> <i>mobilityType</i> <i>gender</i> <i>lingPreparation</i> <i>studyLevel</i> <i>subjectArea</i> <i>nationality</i> <i>hostInstitution</i> <i>homeInstitution</i>	Course Age MobilityType Gender LingPreparation StudyLevel SubjectArea Nationality Instiution Institution

Figura 8 Diagrama lògic.



### 3.3.2 Model físic

Ara que ja es disposa del model lògic del DW, resta obtenir el model físic. Per fer-ho s'hauran de considerar aspectes com:

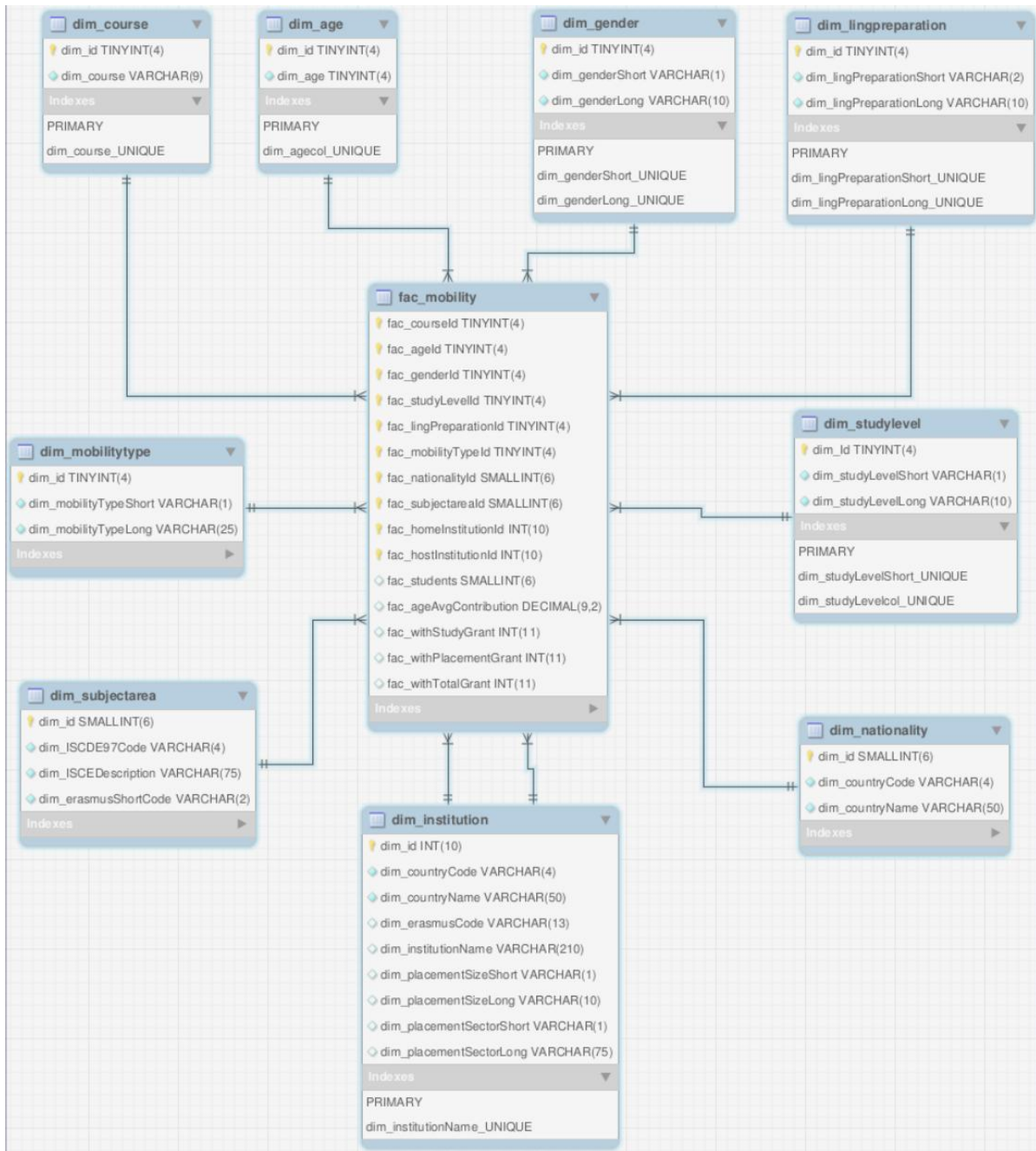
- L'RDBMS utilitzat.
- Els tipus de clau primària a utilitzar: la pròpia del model lògic o una de subrogada.
- La determinació del tipus dels atributs i de les claus utilitzades.
- La determinació de si els atributs poden ser nuls o no.
- La definició, si escau, de claus úniques per a les claus alternatives.

Pel nostre cas, s'ha decidit utilitzar claus subrogades en la construcció del DW. Això permet, d'una banda, que els índexs generats ocupin menys espai, i de l'altra, que la taula de fet en utilitzar com a clau primària la clau composta corresponent ho faci amb les subrogades que tenen una mida molt inferior a la del model lògic.

S'ha de pensar que només amb **la no utilització del nom de la institució com a clau aliena** es produeix un estalvi d'uns 210 bytes per registre, que tenint en compte els 220.000 registres anuals que tindrà per terme mitjà, suposarà un estalvi d'uns 47 MB anuals.

Les taules implementades amb els seus atributs, índexs i altres restriccions es veuen representades en l'esquema presentat a la imatge 9.

Imatge 9. Diagrama model físic



### 3.4 Disseny processos d'ETL a alt nivell, tractament d'errades i automatitzacions

L'eina que s'utilitzarà per dur a terme el procés serà la eina *Pentaho PDI* (coneguda com a *Spoon*), la qual permet dissenyar els fluxos de tasques d'una manera gràfica i requereix una mínima programació. En *Pentaho PDI* cada tasca (*job*), que es pot agrupar en tasques més grosses, està constituïda per un conjunt de transformacions individuals (*transformation*).

En el disseny de l'ETL del DW del projecte, s'aprofitaran aquestes funcionalitats per dissenyar un procés dividit en dues grans branques: ETL de fitxers mestres i ETL de fitxers de mobilitat; cada una amb una o més subtasques per font de dades a importar. Ara bé, no s'haurà que executar cada subgrup com un tot, ja que es podrà elegir quina executar, i d'aquesta manera es podrà incorporar al DW la font que es requereixi

sense necessitat d'activar les transformacions que afecten les altres fonts de dades. En qualsevol cas, es crearan dos treballs de càrrega totalment automatitzada del DW, un per a la càrrega inicial i l'altre per a les actualitzacions anuals.

El procés ETL de fitxers mestres actuarà sobre les fonts de dades de: nacionalitats (*ISOCountryCodes081507*), àrees de coneixement (*ISCED97 Erasmus subject codes*) i institucions acadèmiques (*EUC\_Consolidated\_Table\_2007\_2013*). Per contra, el de mobilitats dels alumnes atacarà els fitxers de mobilitat tipus (*SM\_2011\_2012* i *SM\_2012\_2013*), que es diferenciïn, principalment (a part de la nomenclatura dels camps) per incloure l'identificador de l'alumne o no.

El procés d'ETL utilitzarà un magatzem de dades intermedi, o de maniobra (SA) en el qual es durà a terme l'extracció i transformació de les fonts de dades. En aquesta àrea es tindran taules que es correspondran amb els fitxers mestres i amb els de fets per encabir els registres a importar.

També s'ha de mencionar que les taules de dimensions que s'han afegit (*stg\_course*, *stg\_gender*, *stg\_mobilityType*, *stg\_lingPreparation*, *stg\_age*, *stg\_size* i *stg\_sector*) es crearan amb un Script SQL que executarà l'administrador directament. *Per tant*, la seva creació en el DW serà també utilitzant un *script*. En el cas de què aquestes taules tinguessin alguna variació posterior, s'actualitzaran amb un *Script SQL* que executarà l'administrador -en l'Annex 1 es presenta la correspondència entre els fitxers de fonts de dades i les taules de l'SA-.

### 3.4.1 Extracció

Tot seguit es descriurà a alt nivell les transformacions que s'utilitzaran per fer l'extracció de cada font dels fitxers corresponents: Per norma general s'aplicarà a tota extracció el següent algorisme:

1. Buidar la taula de l'SA relacionada.
2. Seleccionar els camps d'origen que s'han de carregar a la taula.
3. Si un registre és en blanc no es tractarà.
4. Si les mides o el tipus no es corresponen amb els dels camps corresponents a la taula receptora no s'extraurà i es notificarà, si així escau<sup>3</sup>.
5. S'emplenarà la taula amb els registres que hagin passat els punt 3 i 4.

Pels fitxers mestres, hi haurà tres extraccions (subtasques): Nacionalitats, Institucions i Àrees de coneixement, que es caracteritzaran per:

- Nacionalitats: només necessita fer l'extracció seguint les passes generals.
- Àrees de coneixement: fer l'extracció de les dades de la pestanya ISCED97 codes, seguint les passes generals.
- Institucions:
  - Extracció de les institucions acadèmiques del fitxer mestre seguint les passes generals.

---

<sup>3</sup> Les mides són les que s'han elegit per a la creació del DW en el disseny físic: En tot cas, l'administrador haurà de prendre la decisió de si modificar l'atribut o no.

- Extracció de les empreses de pràctiques (*Placement*), aquesta es farà des dels fitxers de mobilitat<sup>4</sup>, en aquest cas:
  - S'elegiran els registres amb nom d'empresa no buit.
  - No es farà l'actualització dels registres existents, només la inserció dels encara no donats d'alta, d'aquesta manera s'evitarà actualitzar els atributs de sector i mida d'empresa i nacionalitat. Aquest fet, permetrà que les empreses amb nom, nacionalitat i sector diferents es considerin empreses diferents.

Pel que fa als fitxers de mobilitat, hi haurà dues subtasques, una per a cada tipus de fitxer. Només s'extrauran els registre que compleixin amb els criteris generals, deixant la validació per data per a la següent fase, la de transformació.

### 3.4.2 Transformació

En aquesta fase ja es disposa de les taules de l'SA carregades amb les dades extretes de les fonts, i és el moment de realitzar les transformacions i validacions pertinents.

#### 1. *stg\_country*

No hi ha a fer cap passa addicional.

#### 2. *stg\_subjectArea*

El mateix succeeix amb aquesta taula.

#### 3. *stg\_institution*.

S'emplenaran els atributs:

- Es truncarà el codi de la nacionalitat als dos primers caràcters (veure apartat **Anàlisi de les Fonts de dades**)
- Es truncarà el nom de les institucions i empreses a 210 caràcters.
- *stg\_countryName* amb el *stg\_countryName* procedent de la taula *stg\_country*.
- *stg\_sizeLong* amb el corresponent de la taula *stg\_Size*.
- *stg\_sectorLong* amb el corresponent de la taula *stg\_Sector*

#### 4. *stg\_mobility1*

- a) Es seleccionaran tots els dels fitxers de mobilitat, siguin quines siguin les seves dates de començament dels estudis o de les pràctiques.
- b) Es truncarà el nom de les empreses de pràctiques a la mateixa longitud que en la càrrega. A més es durant a terme les mateixes transformacions i neteja sobre ell.

### 3.4.3 Càrrega

Un cop dutes a terme les fases d'extracció i transformació de les dades s'ha de procedir a fer la càrrega del magatzem. Es trobarà, un altre cop, dues casuístiques: càrrega de les taules de dimensions i càrrega de la taula de fets.

#### a. Càrrega de les taules de dimensions

---

<sup>4</sup> En haver dues estructures de fitxers de fonts de mobilitat hi haurà d'haver dues transformacions diferents, i s'haurà d'elegir quina de les dues du a terme.

Les taules de dimensions s'actualitzaran amb les procedents de l'SA. La política a seguir serà la següent;

- a) Si un registre de la taula de l'SA no existeix, se'l dona d'alta
  - b) Si resulta que les dades no claus han variat es bescanvien per les procedents del registre de la taula de l'SA.
  - c) Si es presenta qualque error no es carrega el registre i s'indica l'error.
- b. Càrrega de la taula de fets

Es procedirà a emplenar la taula `stg_mobility2` amb els valors dels agregats del grànul. Per fer-ho es procedirà a fer insercions des d'una sentència SQL que agruparà els fets segon el grànul i calcularà: la suma dels imports de les seves beques (`stg_wsithStudyGrant`, `stg_withPlacementGrant`, `stg_withTotalGrant`), contarà el nombre d'estudiants que el formen i el valor de la seva contribució a l'edat mitjana (`stg_ageAvgContribution`).

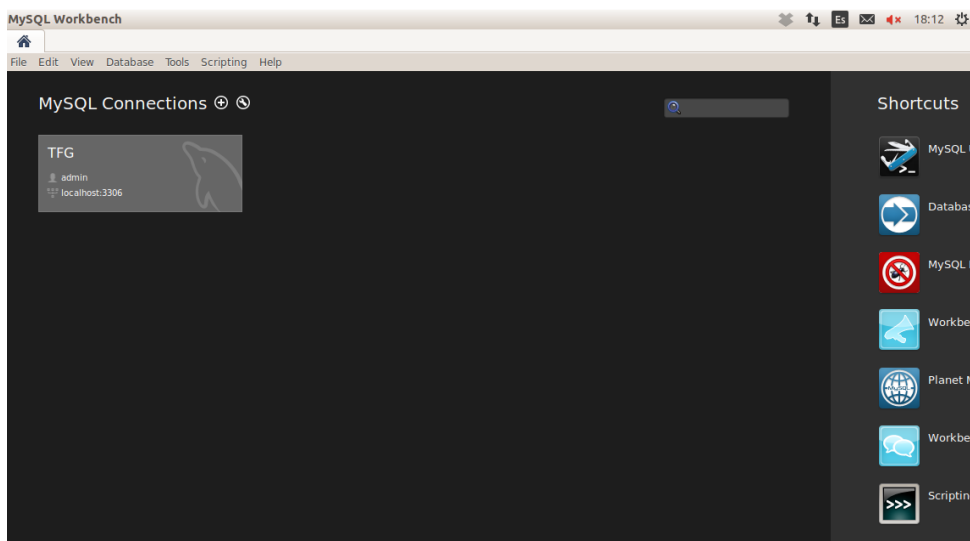
S'ha de fer notar que els valors de les dimensions del grànul en la taula de maniobra que no existeixin en les dimensions del DW s'ignoraran. Essent conscients que tal mesura provocarà una baixa en el rendiment de estudiants comptabilitzats.

## 4 Implementació

### 4.1 Creació del DW

El *data warehouse* s'ha construït utilitzant el sistema gestor de bases de dades *MySQL*, en la versió i el maquinari especificats en els requeriments no funcionals del sistema a construir. Per a la seva construcció, s'ha utilitzat la utilitat *MySQL Workbench* (imatge 10 on es mostra la connexió utilitzada, *TFG*, amb l'usuari *admin*), eina que permet connectar-se al gestor de BD i construir els esquemes, taules i relacions amb una interfície gràfica amigable.

#### Imatge 10. Entrada a l'eina *Workbench* de *MySQL*



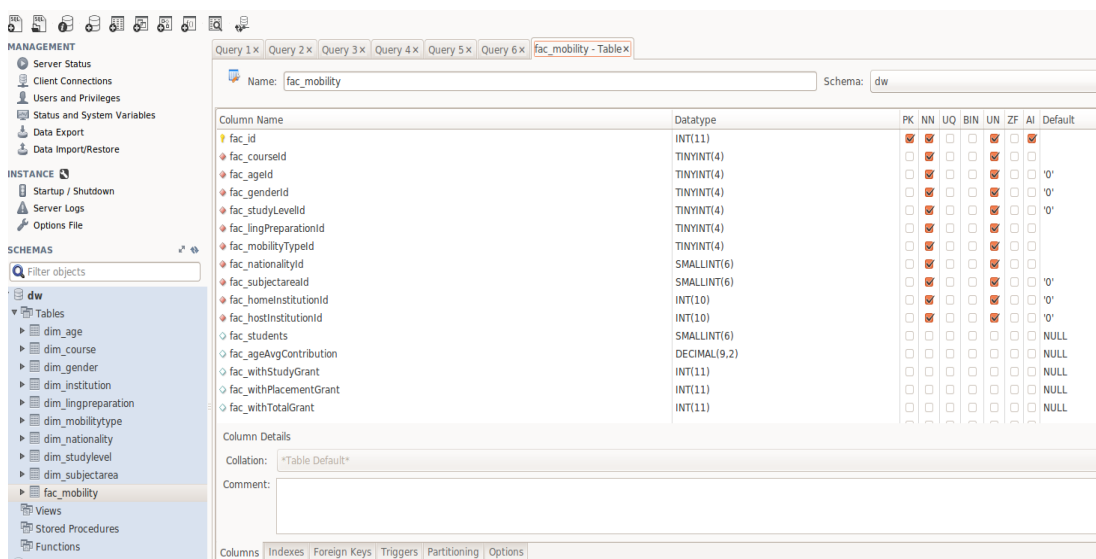
S'han construït dos esquemes anomenats **dw** i **stg**. El primer és el magatzem de dades pròpiament dit, mentre que el segons és la zona de maniobra utilitzada pels processos d'ETL. En l'esquema **dw**, la taula

de fets utilitza restriccions, en canvi a la zona de maniobra s'han obviat per poder treballar amb més llibertat i eficiència.

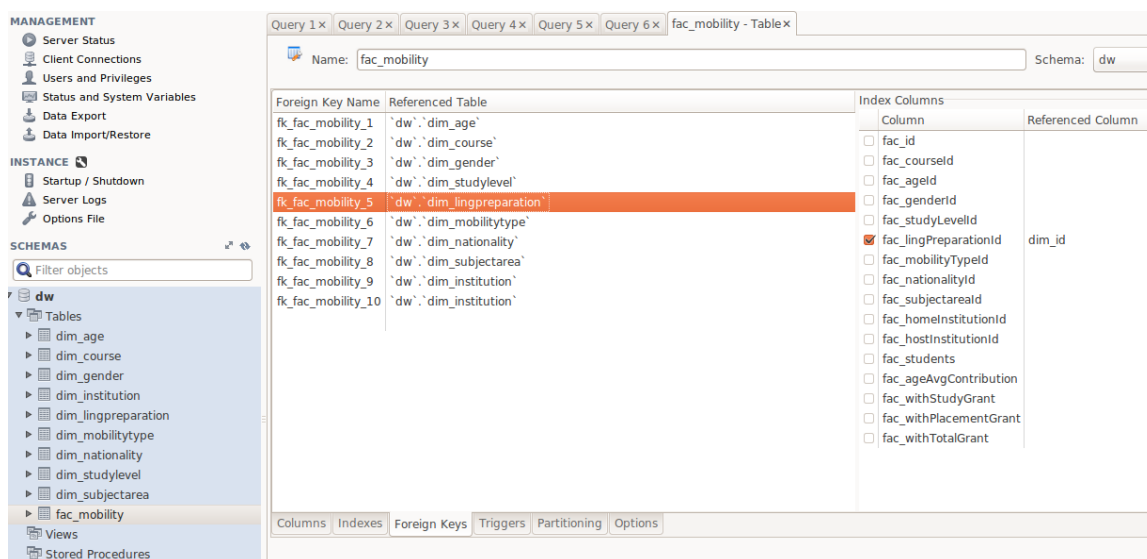
El treball amb l'eina gràfica ha donat lloc a les taules que es detallen en els annexos I (*Scripts SQL DW*) i II (*Scripts SQL SA*), d'acord amb les característiques descrites en la fase de disseny.

En la imatge 11 es mostra, en la creació de la taula de fets, els atributs, mentre que en la imatge 12, les claus alienes.

### Imatge 11. Creació de la taula de fets



### Imatge 12. Claus alienes de la taula de fets



## 4.2 Processos ETL

En la implementació dels processos d'ETL s'han seguit les pautes de disseny especificades en l'apartat **Disseny a alt nivell dels processos d'ETL**, i s'ha utilitzat una estructura de directoris i de codificació, que tot seguit es detalla:

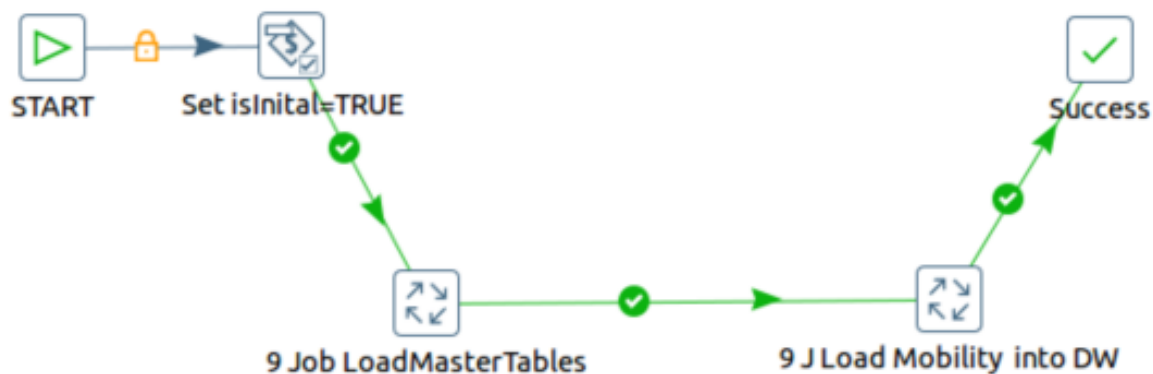
- Les fonts de dades romanen en el directori '/home/user/dades'.
- Tots els processos comencen per un identificador de fins a 1 dígit (0..9, A...Z).
- Darrera de l'identificador hi ha una **T** per a les transformacions (*Transformation*) i una **J** per als treballs (*Job*).
- En els treballs es generen fitxers de log en text pla que permeten seguir la traça dels treballs engegats i es dipositen en el directori '/home/user/dades/logFiles'.
- El repositori de les transformacions utilitzades està ubicat en el directori '/home/dades/Kettelrepository'.
- En els apartats següents, es relacionen els treballs utilitzats en l'ordre emprat per a la càrrega del DW, mostrant i detallant les transformacions involucrades més rellevants.

### 4.2.1 Carrega del DW

Per a la càrrega del DW s'han creat dos treballs que automatitzen el procés: el de càrrega inicial i el d'actualització anual. En el primer s'han de carregar els fets de mobilitat de dos cursos: 2011-2012 i 2012-2013, mentre que en el segon només s'haurà de carregar el de l'any en curs (veure imatges 13 i 14)

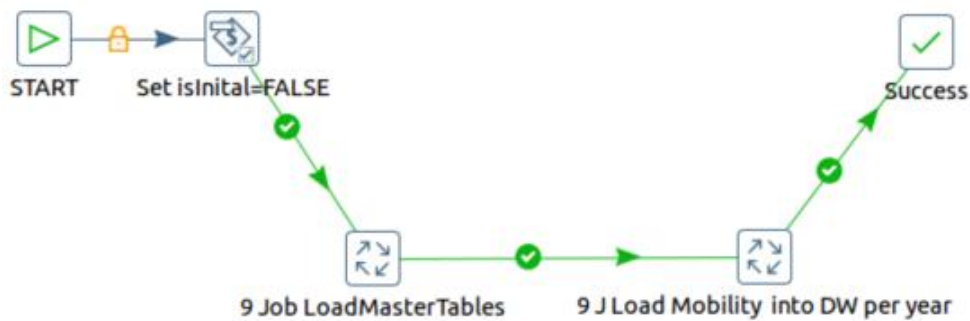
Per dur-ho a terme s'ha creat una variable global booleana, '*isInitial*', que permet discriminar quan es tracta d'un cas o de l'altre. A més, els noms dels fitxers de mobilitat hauran de ser del tipus 1SM\_XXXX\_YYYY pels d'estructura similar al curs 2011\_2012 i 2SM\_XXXX\_YYYY pels que la tinguin similar a la del 2012\_2013.

Imatge 13. Càrrega inicial del DW





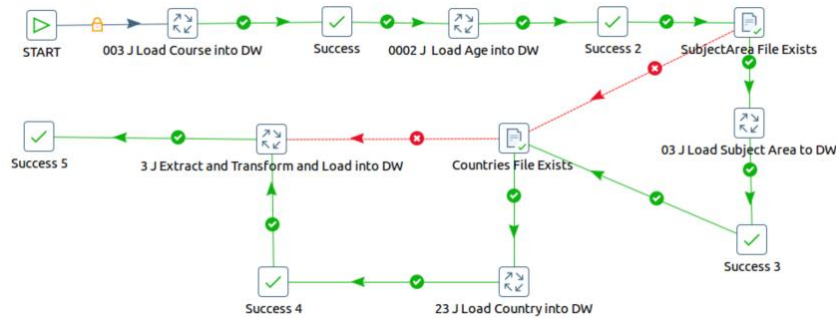
**Imatge 14. Càrrega anual del DW.**



### 4.2.2 Càrrega de les dimensions

En la imatge 15 es mostra la càrrega de les taules mestres de les dimensions.

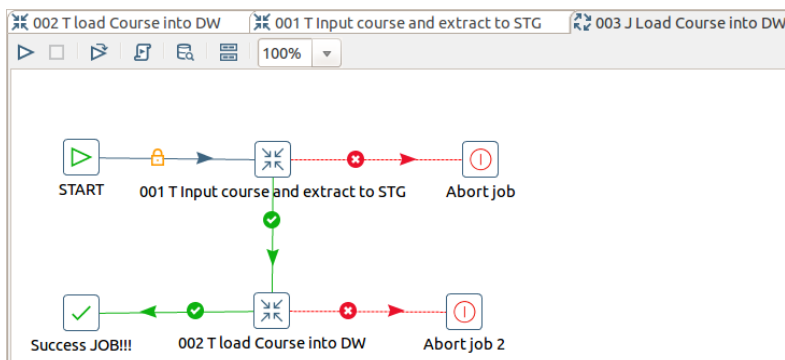
**Imatge 15. Càrrega de les dimensions al DW.**



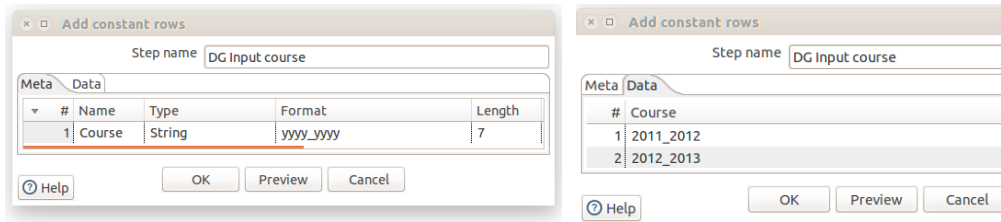
### 4.2.3 Tractament de la dimensió Course

Els cursos es carreguen utilitzant el treball “003 J Load Course into DW” que consta de dues transformacions (imatge 16). La primera, mitjançant una graella de dades (*Data Grid*), introdueix els codis dels cursos a carregar (imatges 17) i emplena la taula *stg\_course* -s’ha de fer notar que s’ha canviat la càrrega per *script* d’SQL per l’emplenament del *data grid*-. Mentre que la segona traspassa les dades de l’SA al DW (taula *dim\_course*).

**Imatge 16. Càrrega dels cursos**



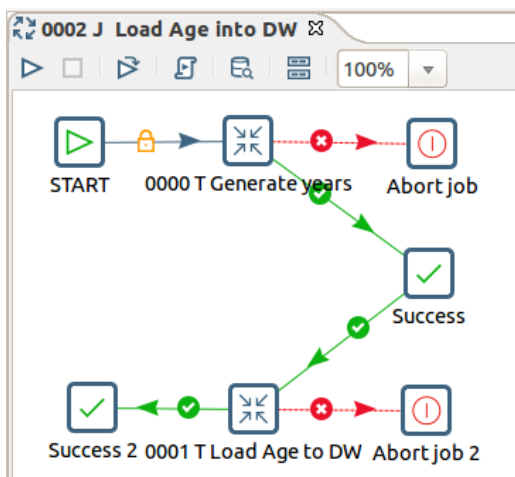
## Imatge 17 Data Grid



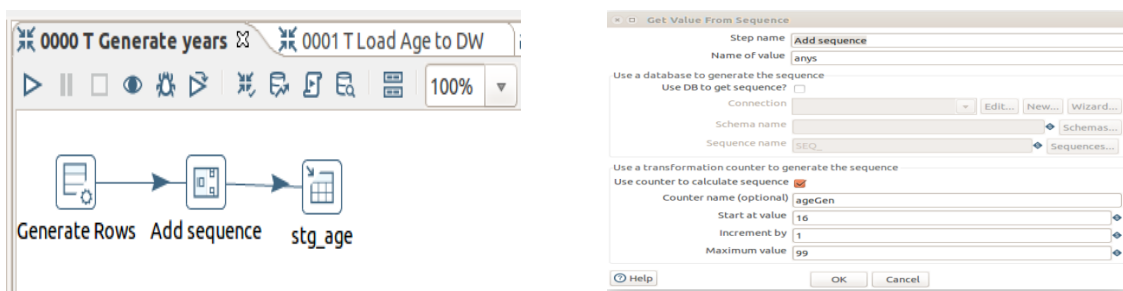
## 4.2.4 Tractament de la dimensió Age

Es carreguen les dades amb el treball "005 J Load Age into DW" (imatge 18). La passa més destacada és la transformació que genera els anys, la qual utilitza un generador seqüencial (imatge 19)

### Imatge 18. Càrrega de les edats en el DW



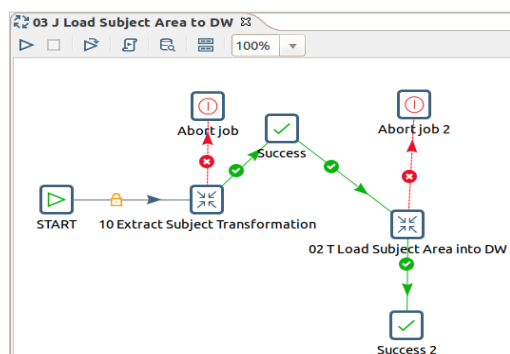
### Imatge 19. 000T Generació dels anys



#### 4.2.5 Tractament de la dimensió *Subject Area*

El treball encarregat de la càrrega és el “03 J Load Subject Area into DW” (imatge 20), el qual extrau les àrees de coneixement cap a l’SA amb una transformació que valida els seus atributs. Posteriorment, una altra transformació incorpora les dades de la taula *stg\_subjectArea* cap a la de dimensions, *dim\_subjectArea*, en el DW.

Imatge 20. Càrrega de l’àrea de coneixements en el DW.



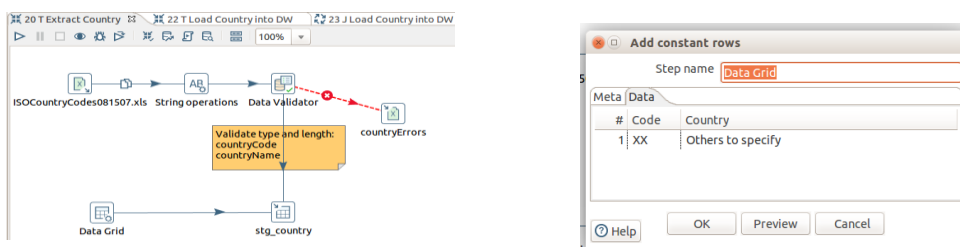
#### 4.2.6 Tractament de dimensió *Nationality*

Cas similar a l’anterior, on el treball “23 J Load Country into DW” és l’encarregat de coordinar el procés (imatge 21). La passa més destacada és la utilització d’un *Data Grid* per afegir el codi XX als països no especificats (imatge 22).

Imatge 21. Càrrega de la nacionalitat al DW



Imatge 22. Extracció de la nacionalitat

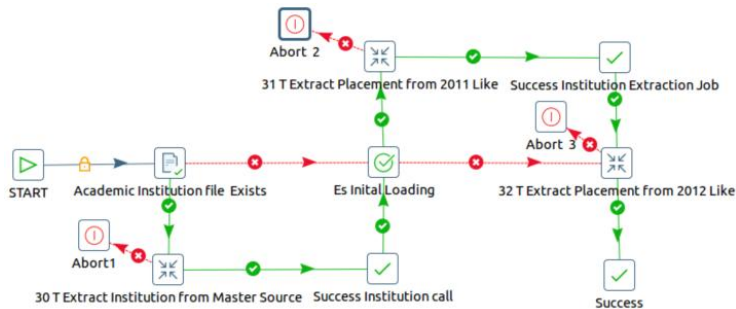


#### 4.2.7 Tractament de la dimensió *Institution: Academic Institutions and Placements*

El tractament de les fonts d’institucions és el més complex dels que tracten les taules de dimensions. Per una banda, es tracten les institucions acadèmiques (font d’institucions consolidades) i, per una altra, les

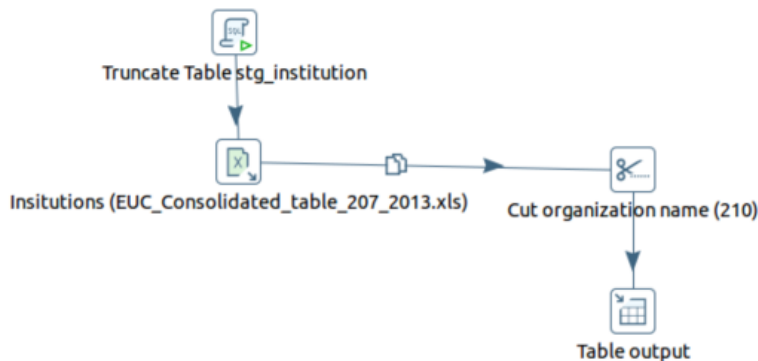
empreses que col·laboren oferint places en pràctiques (fitxers de moviments). El treball encarregat és el "3 J Extract and Transform and Load into DW" (imatge 23) on s'encadenen els treballs d'extracció, transformació i càrrega. La passa més important està en la càrrega de la dimensió *Institution* la qual ha de carregar de fonts de diferent tipologia -s'observa que els treballs es llancen dependent de l'existència o no del fitxer corresponent -.

**Imatge 23. Càrrega d'*Institution***



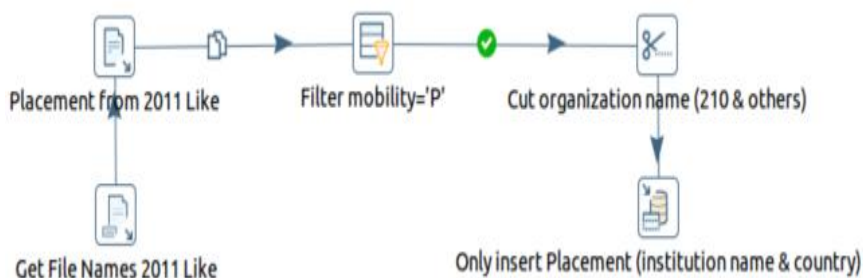
Les transformacions de les extraccions es mostren a les imatges 24 i 25.

**Imatge 24. Extracció d'*Academic Institution***



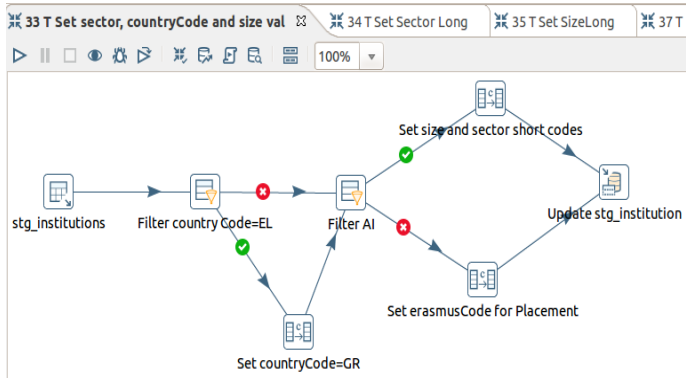
En la imatge 25 es mostra com exemple el cas d'empreses de pràctiques de tipologia de fitxer de fonts de dades 2011\_2012 (el de 2012\_2013 és idèntic):

**Imatge 25. Extracció de *Placement***

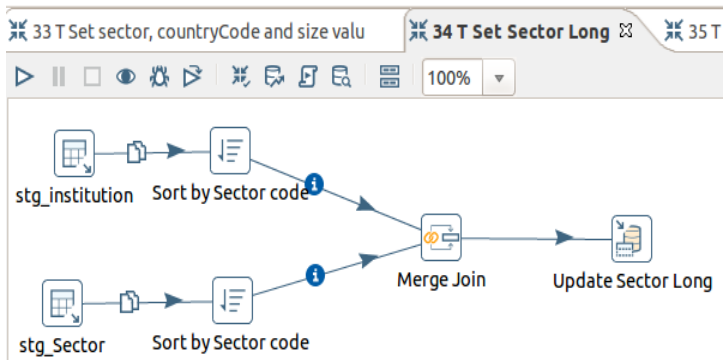


A més, es disposa de cinc processos de transformació on s'obtenen els descriptors dels codis de nació, sector empresarial i mida de l'empresa (imatges 26, 27, 28, 29 i 30).

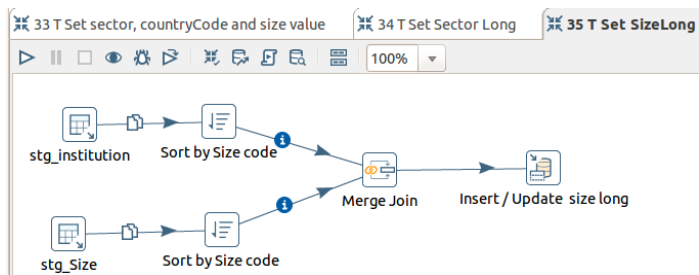
**Imatge 26. 33 T Set sector, countryCode and size values**



**Imatge 27. 34 T Set SectorLong**



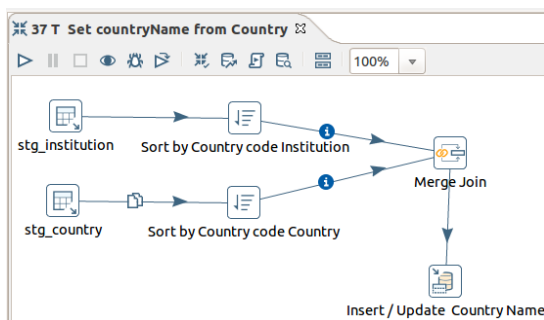
**Imatge 28. 35 T Set sizeLong**



Imatge 29. 38 T Set countryName if null

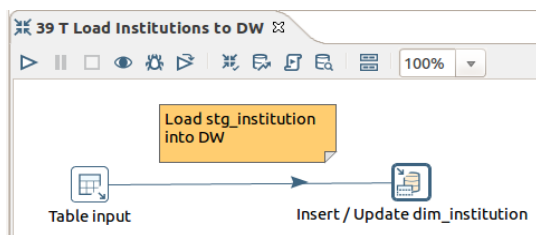


Imatge 30. 37 T Set countryName from Country



Les càrregues al DW es fan utilitzant dues transformacions, on mitjançant una sentència SQL es discrimina pel codi ERASMUS, que té el valor “*EC PLACEMENT*” per a les empreses de pràctiques, i qualsevol altre per a les institucions acadèmiques (a la imatge 31 es mostra el cas d’institucions acadèmiques).

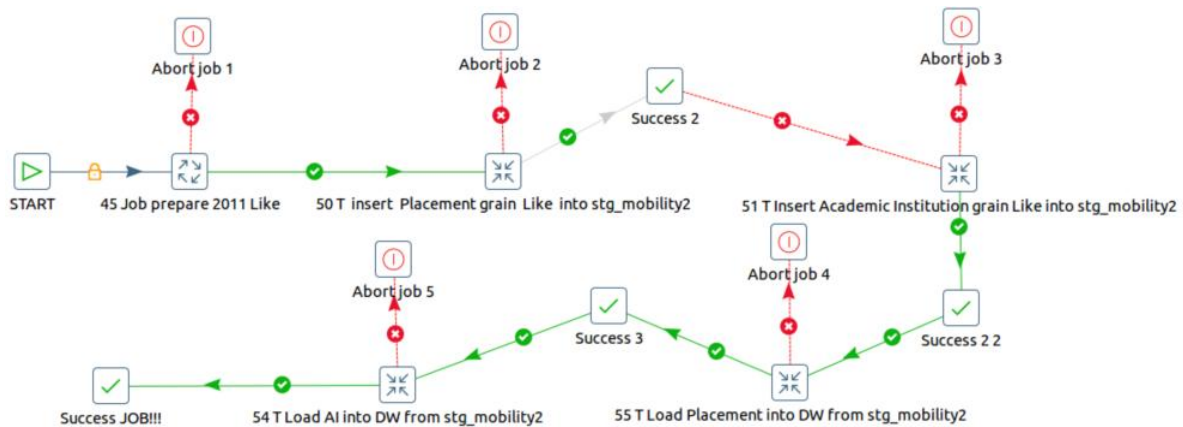
Imatge 31. 39 T Load institution into DW



#### 4.2.8 Tractament de la taula de fets *Mobility*

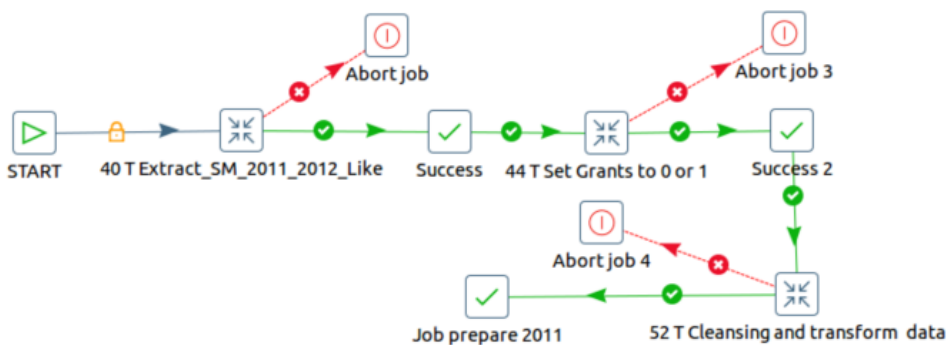
Els fitxers de mobilitat seran els que proporcionaran les dades per a la càrrega de la taula de fets. Cada curs (que es correspon amb un fitxer) es tracta per separat. Així es farà l’extracció, transformació i càrrega de cada un d’ells individualment, automatitzant-ho tot en un treball (un per a cada tipus de fitxer). Tot seguit, es presenten les passes més importats prenent com a exemple els cursos tipus 2011\_2012, on el treball que coordina l’extracció, transformació i càrrega es presenta en la imatge 32.

**Imatge 32. Exemple de càrrega de mobilitat tipus 2011\_2012**

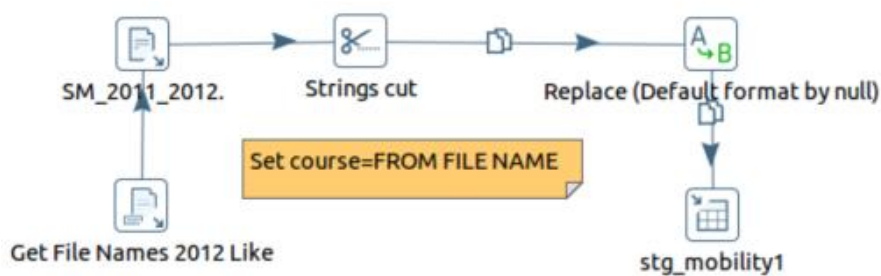


En primer lloc, s'executa el treball de preparació on es fa l'extracció dels moviments i la neteja i transformacions pertinents (imatge 33), per posteriorment agrupar pel grànul i fer la càrrega al DW (procés que es veurà més endavant). Les transformacions utilitzades es presenten a les imatges 34, 35, 36, 37, 38 i 39.

**Imatge 33. 46 Job Prepare 2011 Like**

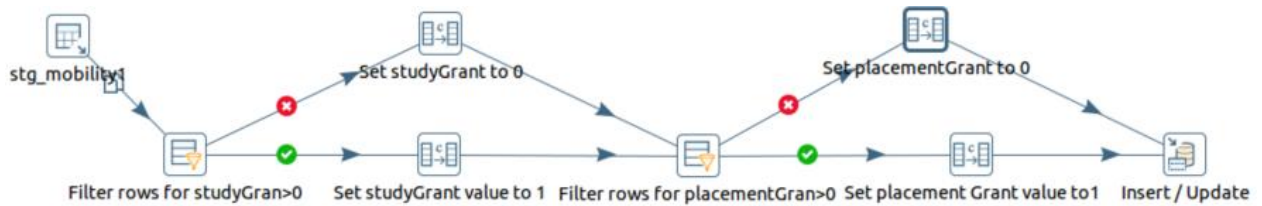


**Imatge 34. 40 T Extract SM\_2011\_2012 Like**



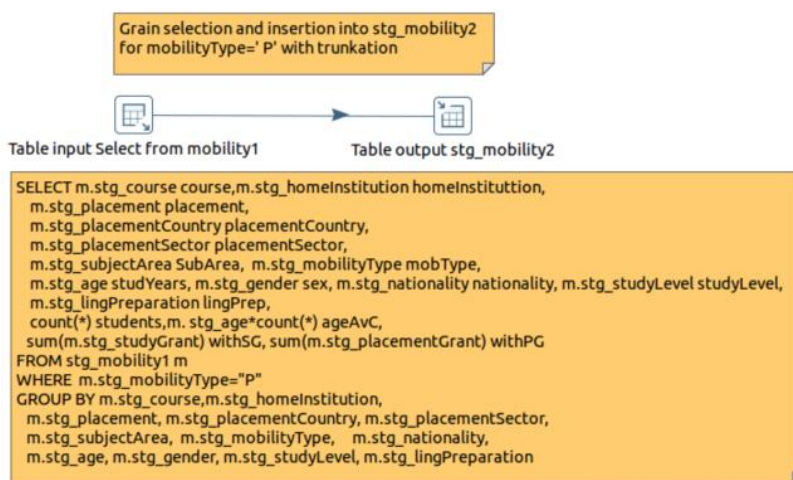
S'ha de fer notar que no es fan validacions de les dates de les beques, ja que es considera que tot registre d'un fitxer, en ser anual, és correcte.

**Imatge 35. 44 T Set grants to 0 or 1**



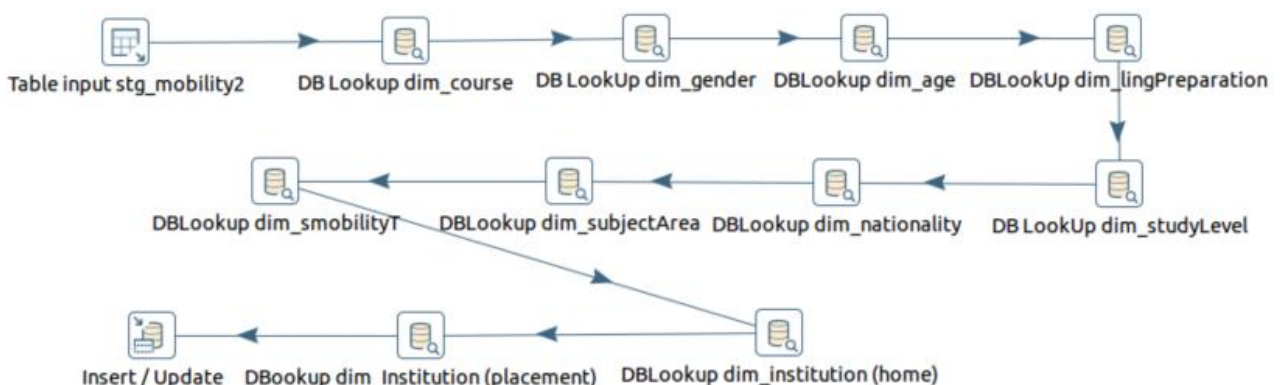
Per acabar, un cop preparades les dades en la taula de l'SA *stg\_mobility1*, només queda agrupar-les segon el grànul (imatge 36) amb les mesures calculades i traspassar-les al DW (imatge 37).

**Imatge 36. Insert Placement grain into stg\_mobility2**



S'ha d'aclarir que s'han separat els tractaments de les mobilitats d'estudi i de pràctiques per la manera diferent en què s'ha de fer el càlcul del grànul, ja que si bé a Placement s'agrupa per empresa (*Placement*), país de l'empresa i sector, en el procés que tracta les institucions acadèmiques només es fa per codi Erasmus (*HostInstitution*).

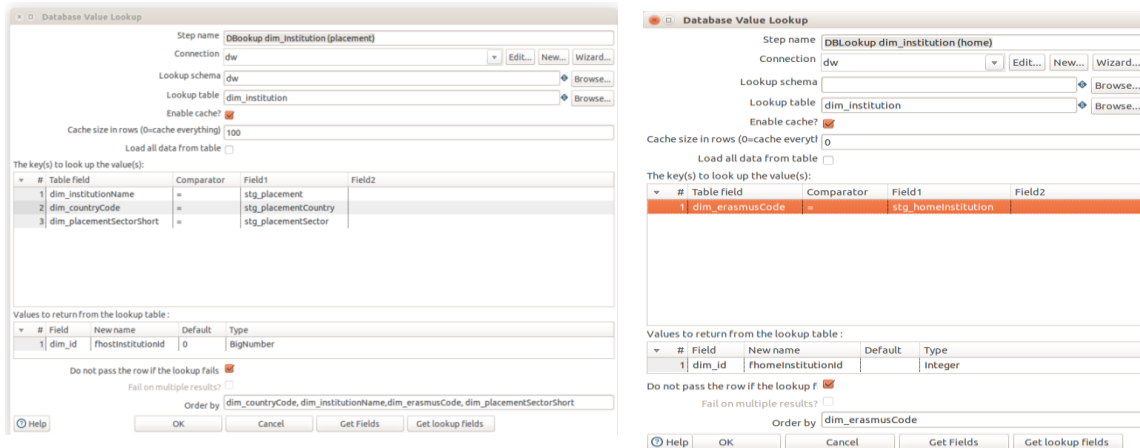
**Imatge 37. Load Placement into DW from stg\_mobility2**



Es mostren com exemple de *Data Base Lookup* el d'institucions i el d'empreses (imatge 38).



### Imatge 38. Lookup Placement and Academic Institution



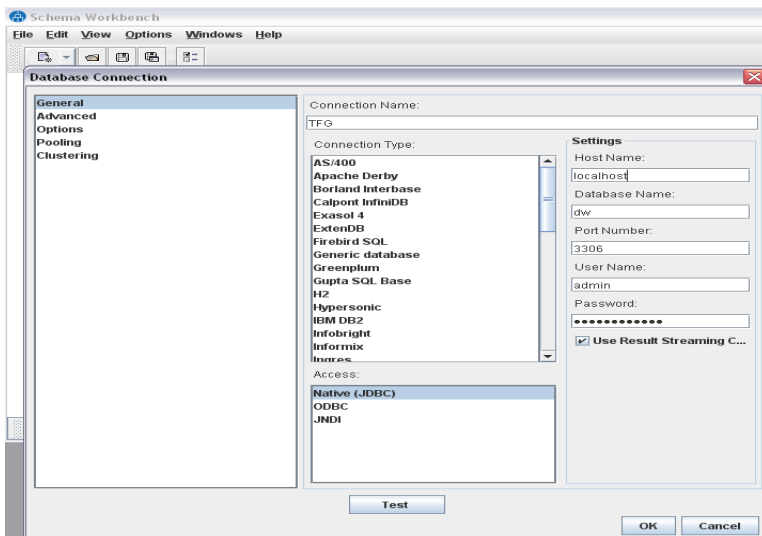
## 4.3 Cub OLAP

Per construir el cub OLAP s'ha utilitzat l'eina *Schema Workbench* de Pentaho en la versió indicada en l'apartat de requeriments no funcionals del document de disseny. Aquesta eina és un editor gràfic que crea un fitxer en format XML que és la representació del cub OLAP: De cubs en poden haver amb dimensions internes o externes, en aquest treball s'ha preferit la utilització d'un esquema amb el cub OLAP que utilitza dimensions externes perquè en aquestos les dimensions és poden compartir amb altres cubs que es puguin desenvolupar en un futur. Per acabar aquesta petita introducció, dir que el fitxer XML s'ha desat en el directori '/home/user/dades/Cube' (fitxer TFG\_EM.mondriant.xml). Les passes més importants s'expliquen tot seguit.

### 4.3.1 Configuració de la connexió

El primer que s'ha de fer és configurar una connexió a l'esquema que emmagatzema el DW des de l'opció *Connections* del menú *Options* (imatge 39).

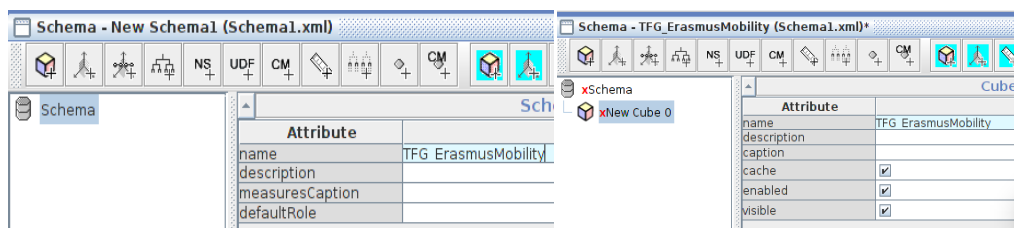
### Imatge 39. Configuració connexió



### 4.3.2 Creació de l'esquema i el cub OLAP

Ara es crea un nou esquema des de *File->New->Schema* i el cub corresponent (imatge 40).

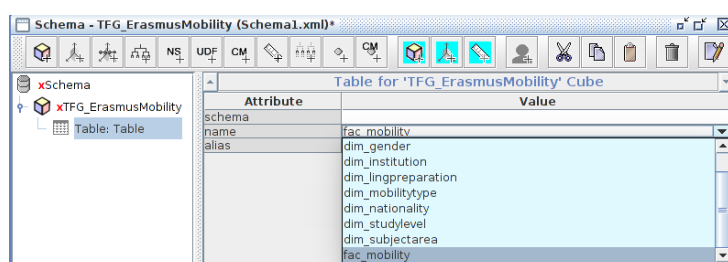
Imatge 40. Creació esquema i cub.



### 4.3.3 Assignació de la taula de fets

La propera passa és l'assignació de la taula de fets: es pica amb el botó secundari sobre el cub creat i es selecciona *Table*, per a tot seguit elegir la taula (imatge 41).

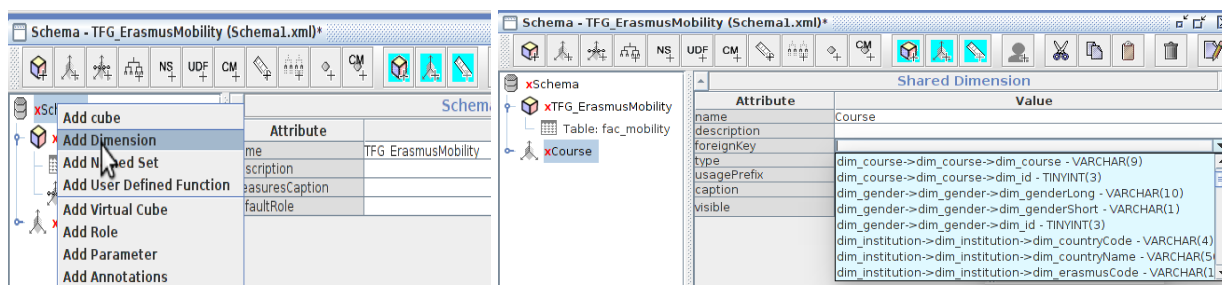
Imatge 41. Creació de la taula de fets



### 4.3.4 Creació de les dimensions externes

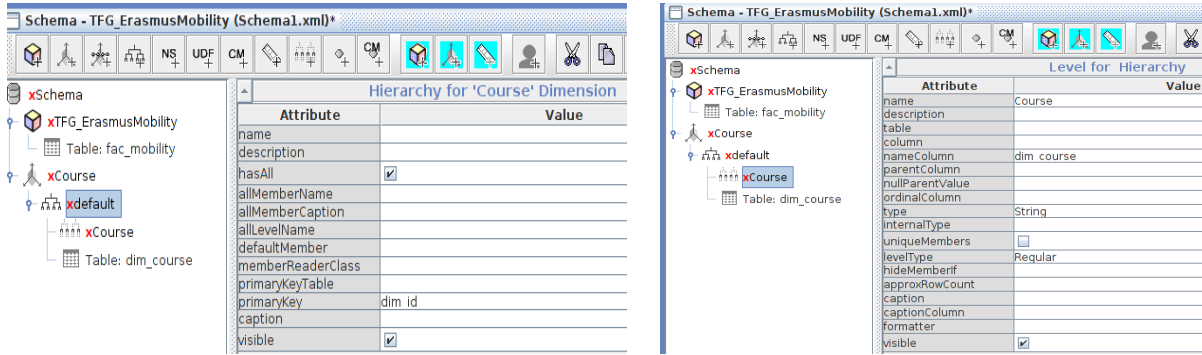
És el moment de crear les dimensions externes (des de l'esquema) amb una jerarquia d'un nivell per a les dimensions: *Course*, *Age*, *Gender*, *Nationality*, *StudyLevel*, *LingPreparation*, *MobilityType* i *SubjectArea* (imatges 42 i 43)<sup>5</sup>.

Imatge 42. Creació de la dimensió



<sup>5</sup> Les dimensions *HomeInstitution* i *HostInstitution* són uns casos especials, ja que la primera té una jerarquia amb dos nivells, mentre que la segona en té tres; però, tenint en compte que el procés és similar, no es detalla en aquest document.

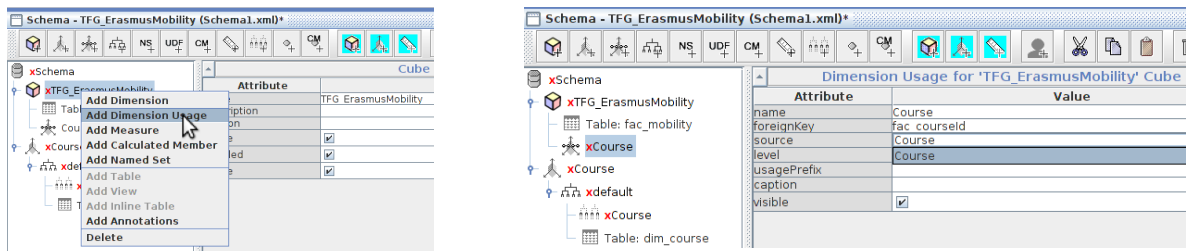
### Imatge 43. Creació del nivell



#### 4.3.5 Definició de la *dimension usage*

Per a cada dimensió externa es definirà una *dimension usage* en el cub per lligar-la a la taula de fets, amb la clau aliena (*foreignKey*), i a la dimensió externa, amb el camp *source* (imatge 44).

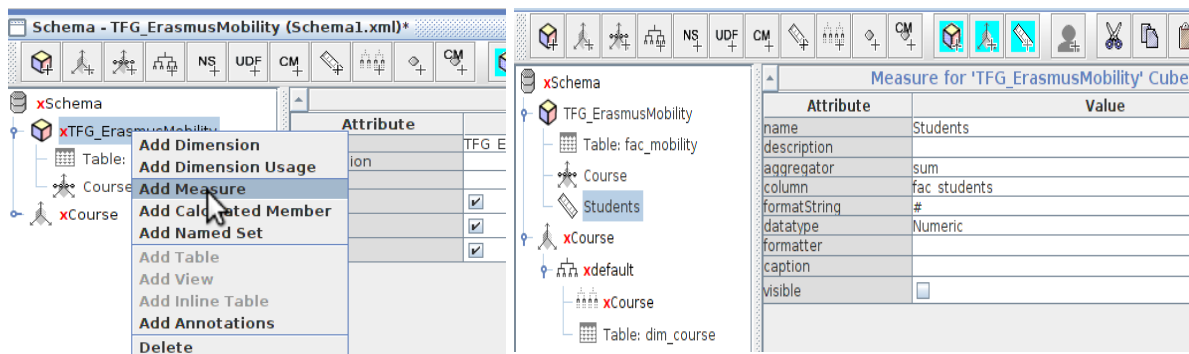
### Imatge 44. Creació de la *dimension usage*



#### 4.3.6 Definició de les mesures

Resta només definir les mesures. Amb tal fi, es mostra com es fa per a la mesura *Student* (imatge 45), on s'especifica l'operació d'agregació que es vol fer servir (*agregator*).

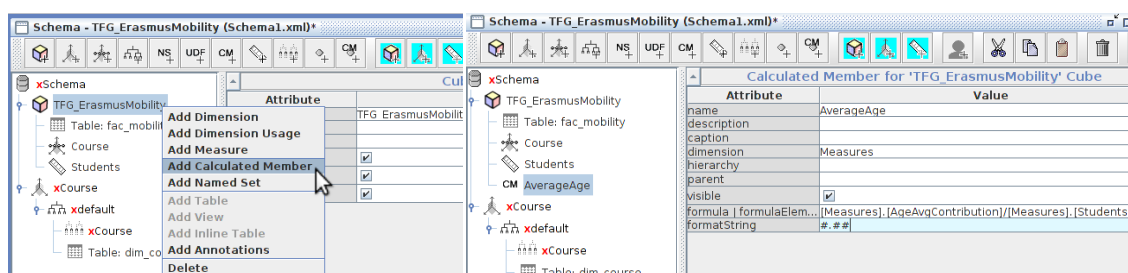
### Imatge 45. Creació de la mesura *Students*



### 4.3.7 Creació d'una mesura calculada

Per acabar, es mostrarà la creació d'un membre calculat, en concret *AverageAge* (imatge 46).

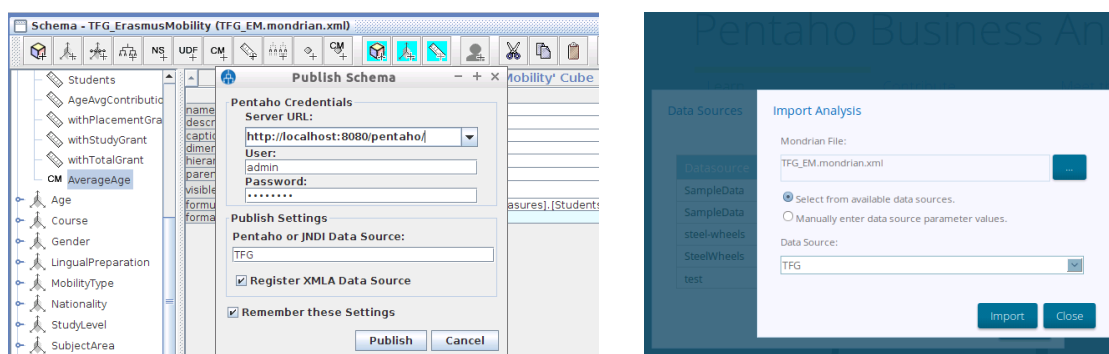
**Imatge 46. Creació de la mesura *AverageAge*.**



### 4.3.8 Publicació

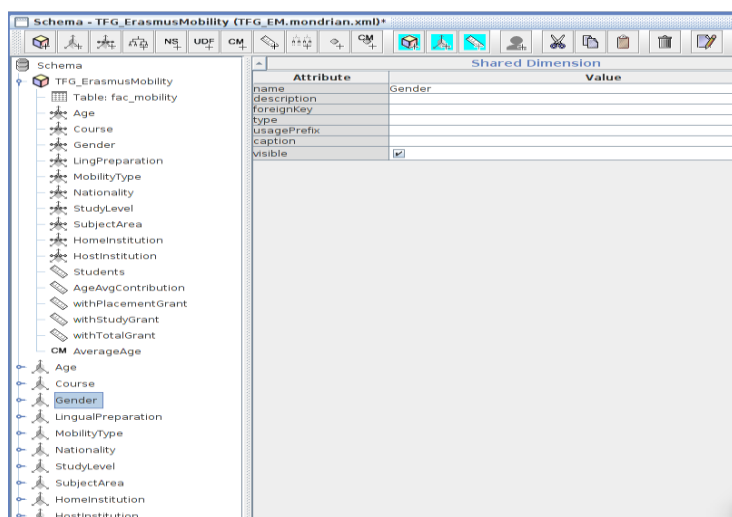
Configura el cub, amb la mateixa eina, es publica en el servidor o s'importa des d'aquest (imatge 47).

**Imatge 47. Publicar cub/Importar anàlisi**



El resultat és el cub mostrat a la imatge 48 (l'esquema resultant en format XML és el de l'annex III).

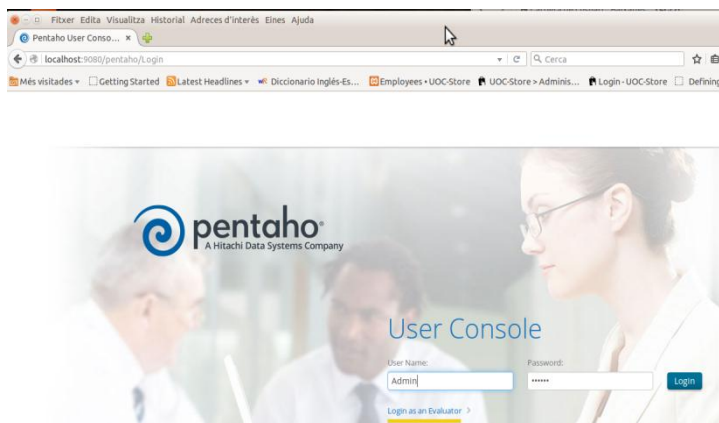
**Imatge 48. Cub OLAP**



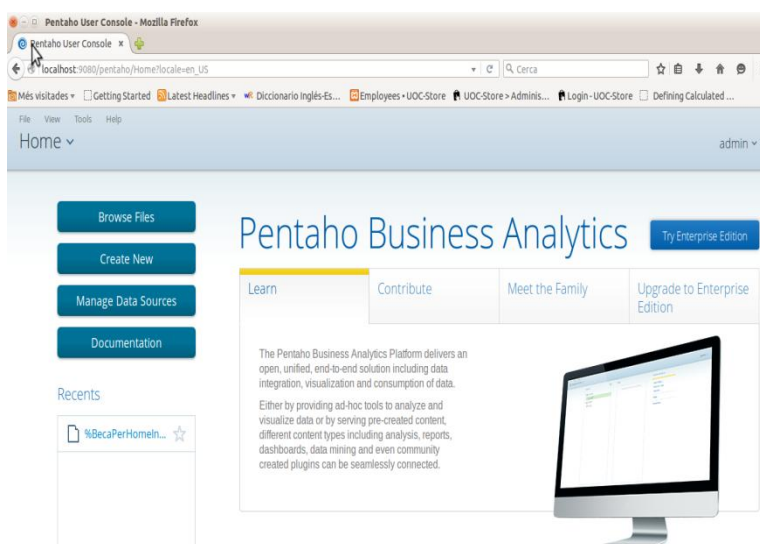
## 4.4 Informes

Per accedir als informes, els usuaris, mitjançant el navegador, han d'introduir el seu nom d'usuari i la contrasenya i navegar fins a la seva àrea personal, on hi trobaran els informes als quals tenen accés, picant sobre l'opció *Browse Files* (imatges 49, 50 i 51).

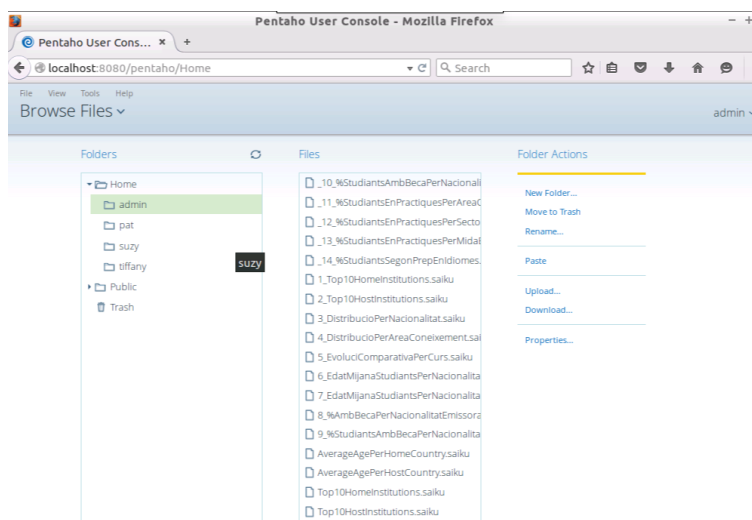
**Imatge 49. Accés al sistema**



**Imatge 50. Accés als informes**



**Imatge 51. Informes disponibles**



Tot seguit es presenten els informes que s'han dissenyat per acomplir els requeriments funcionals, mostrant la imatge del resultat obtingut en ser executats<sup>6</sup>.

#### 4.4.1 Top 10 d'institucions emissores/receptores

Els informes de la imatge 52 mostren les 10 primeres universitats per nombre d'estudiants que han participat en el programa Erasmus: un des del punt de vista de la recepció i l'altra del d'emissió d'estudiants.

**Imatge 52. Top 10 institucions emissores/receptores.**

Home Institution Name	Home Country Code	All Courses		
		Students No.	2011_2012	2012_2013
UNIVERSITÀ DI BOLOGNA - ALMA MATER STUDIORUM	IT	3.144	1.524	1.620
UNIVERSIDAD DE SEVILLA	ES	2.612	1.260	1.352
UNIVERSITAT DE VALENCIA (ESTUDI GENERAL) UVEG	ES	2.537	1.256	1.281
UNIVERSITAT POLITÈCNICA DE VALENCIA	ES	2.486	1.299	1.187
UNIWERSYTET WARSZAWSKI	PL	2.283	1.186	1.097
UNIVERZA V LJUBLJANI	SI	2.234	1.079	1.155
UNIVERZITA KARLOVA V PRAZE	CZ	2.044	1.038	1.006
UNIVERSITÄT WIEN	AT	2.021	972	1.049
MASARYKOVA UNIVERZITA	CZ	2.021	952	1.069
UNIVERSIDAD POLITÈCNICA DE MADRID (UPM)	ES	1.934	977	957

Host Institution Name	Host Country Code	Students No.		
		All Courses		
		2011_2012	2012_2013	2013_2014
UNIVERSITÀ DI BOLOGNA - ALMA MATER STUDIORUM	IT	2.924	1.471	1.453
UNIVERSITAT DE VALENCIA (ESTUDI GENERAL) UVEG	ES	2.738	1.329	1.409
UNIVERSIDAD DE SEVILLA	ES	2.630	1.339	1.291
UNIVERSITAT POLITÈCNICA DE VALENCIA	ES	2.461	1.298	1.163
UNIVERZITA KARLOVA V PRAZE	CZ	2.184	1.013	1.171
UNIVERZA V LJUBLJANI	SI	1.895	905	990
UNIVERSITAT DE BARCELONA	ES	1.785	889	896
FREIE UNIVERSITÄT BERLIN	DE	1.777	885	892
UNIVERSIDAD DE SALAMANCA	ES	1.715	899	816
KUNGLIGA TEKNISKA HOGSKOLAN	SE	1.667	830	837

#### 4.4.2 Distribució en % d'estudiants per nacionalitat dels estudiants

L'informe que mostra la imatge 53 permet als analistes confrontar les mobilitats produïdes en el programa Erasmus des del punt de vista de la nacionalitat dels estudiants participants.

**Imatge 53. Distribució en % d'estudiants per nacionalitat**

Name	Students %		
	All Courses		
		2011_2012	2012_2013
France	14.33%	14.33%	14.34%
Germany	14.13%	14.31%	13.96%
Spain	12.94%	13.32%	12.59%
Poland	6.77%	6.87%	6.69%
Turkey	5.86%	5.46%	6.23%
United Kingdom	4.85%	4.88%	4.81%
Italy	4.50%	4.47%	4.53%
Netherlands	3.52%	3.47%	3.58%
Czech Republic	2.82%	2.87%	2.77%
Portugal	2.76%	2.73%	2.79%
Belgium	2.51%	2.48%	2.53%

<sup>6</sup> El codi de les consultes es trobarà a l'annex IV Reports Saiku.

#### 4.4.3 Distribució en % d'estudiants per àrea de coneixement

En la imatge 54 s'observa la distribució de les mobilitats en el programa des de l'òptica de les àrees de coneixement.

Imatge 54. Distribució en % d'estudiants per àrea de coneixement

Subject Area Description	Students %		
	All Courses		
	2011_2012	2012_2013	2013_2014
Business and administration	8.79%	9.61%	8.03%
Foreign languages	8.73%	8.99%	8.49%
Business and administration (broad programmes)	7.92%	7.53%	8.29%
Law	5.02%	5.10%	4.95%
Engineering and engineering trades	3.98%	4.06%	3.91%
Economics	3.60%	3.77%	3.44%
Political science and civics	3.17%	3.27%	3.09%
Humanities	2.60%	2.24%	2.94%
Architecture and town planning	2.16%	2.31%	2.02%
Management and administration	2.16%	2.08%	2.22%
Medicine	2.11%	2.12%	2.09%
Computer science	1.74%	1.87%	1.62%

#### 4.4.4 Evolució comparativa per curs

La imatge 55 mostra l'informe de l'evolució de la mobilitat desglossada per curs i gènere.

Imatge 55. Evolució comparativa per curs

		Students No.	All Courses	
			2011_2012	2012_2013
(All)	Gender	Students No.	Students No.	Students No.
All Genders		415,954	200,727	215,227
Female		253,148	122,049	131,099
Male		162,806	78,678	84,128

#### 4.4.5 Edat mitjana dels estudiants per nacionalitat de la institució receptora/emissora

En la imatge 56, es presenta la mobilitat estudiantil desglossada per cursos des de l'òptica de la nacionalitat de les institucions emissores i receptores, respectivament.

**Imatge 56. Edat mitjana dels estudiants per nacionalitat receptora/emissora**

The image shows two screenshots of the Saiku Community Edition interface. The left screenshot, titled 'Editing:6\_EdatMijanaStudiansPerNacionalitatReceptora.saiku', displays a table of average ages for host countries. The right screenshot, titled 'Editing:7\_EdatMijanaStudiansPerNacionalitatEmissora.saiku', displays a table of average ages for home countries. Both tables show data for the years 2011\_2012 and 2012\_2013 across various countries.

Host Country Name	Average Age All Courses	
	2011_2012	2012_2013
Macedonia	24.00	24.00
Liechtenstein	23.29	23.38
Switzerland	23.20	23.20
Iceland	23.10	23.17
Croatia	22.86	22.89
Austria	22.85	22.86
Estonia	22.83	22.77
Turkey	22.83	22.81
Slovak Republic	22.71	22.73
Portugal	22.71	22.71
Norway	22.69	22.66
Sweden	22.64	22.63

Home Country Name	Average Age All Courses	
	2011_2012	2012_2013
Iceland	25.37	25.54
Liechtenstein	25.32	24.97
Denmark	23.93	23.89
Sweden	23.91	23.95
Norway	23.70	23.71
Finland	23.67	23.65
Germany	23.42	23.45
Switzerland	23.37	23.39
Austria	23.35	23.34
Czech Republic	23.28	23.29

#### 4.4.6 Tant per cent de beques per nacionalitat emissora/receptora

La imatge 57 mostra com els analistes poden confrontar la distribució dels estudiants amb beca des del punt de vista de la nacionalitat de les institucions emissores i receptores participants.

**Imatge 57. Tant per cent de beques per nacionalitat emissora/receptora**

The image shows two screenshots of the Saiku Community Edition interface. The left screenshot, titled 'Editing:8\_PercentageDeBequesPerNacionalitatE', displays a table of scholarship percentages for home countries. The right screenshot, titled 'Editing:9\_PercentageDeBequesPerNacion:', displays a table of scholarship percentages for host countries. Both tables show data for the years 2011\_2012 and 2012\_2013 across various countries.

Home Country Name	Students % All Courses	
	2011_2012	2012_2013
Germany	14.49%	14.73%
Spain	13.47%	13.88%
France	13.11%	13.09%
Poland	6.77%	6.78%
United Kingdom	5.99%	6.03%
Turkey	5.83%	5.46%
Netherlands	4.73%	4.72%
Italy	4.29%	4.25%
Czech Republic	3.33%	3.37%
Belgium	2.63%	2.58%

Host Country Name	Students % All Courses	
	2011_2012	2012_2013
Spain	13.38%	13.72%
Germany	11.70%	11.56%
France	11.67%	12.01%
United Kingdom	10.99%	11.01%
Italy	4.61%	4.75%
Netherlands	4.37%	4.45%
Poland	4.16%	3.91%
Sweden	4.06%	4.16%
Portugal	3.75%	3.77%
Belgium	3.04%	3.02%

#### 4.4.7 Percentatge d'estudiants en pràctiques per àrea de coneixement

La imatge 58 mostra l'informe que utilitzaran els analistes per estudiar les tendències de la mobilitat que les pràctiques produeixen atenint a les diverses àrees de coneixement



## Imatge 58. Percentatge d'estudiants en pràctiques per àrea de coneixement

Editing:A\_PercentatgeEnPractiquesPerAreaConeixment.saiku

You are using Saiku Community Edition, please consider upgrading to [Saiku Enterprise](#), or entering a

Subject Area Description	Students %		
	All Courses		
	2011_2012	2012_2013	
Foreign languages	9.48%	9.25%	9.70%
Business and administration	7.39%	8.21%	6.55%
Business and administration (broad programmes)	6.63%	6.06%	7.20%
Engineering and engineering trades	4.56%	4.90%	4.21%
Management and administration	3.98%	4.12%	3.84%
Travel, tourism and leisure	3.29%	3.34%	3.24%
Medicine	2.73%	2.55%	2.92%
Biology and biochemistry	2.61%	2.60%	2.61%
Economics	2.56%	2.66%	2.46%
Psychology	2.40%	2.43%	2.37%

### 4.4.8 Tant per cent d'estudiants en pràctiques per sector empresarial

Un cas interessant és el que es mostra a la imatge 59, on el focus d'estudi és la distribució de les mobilitats dels estudiants en pràctiques per sector empresarial.

## Imatge 59. Percentatge d'estudiants en pràctiques per sector empresarial

Editing:B\_PercentatgeEnPractiquesPerSectorEmpresarial.saiku

You are using Saiku Community Edition, please consider upgrading to [Saiku Enterprise](#), or entering a [sponsorship agreement](#)

Placement Sector	Students %		
	All Courses		
	2011_2012	2012_2013	
Private households with employed persons	17.36%	16.52%	18.22%
Education	14.24%	14.10%	14.37%
Extra-territorial organizations and bodies	9.33%	9.61%	9.05%
Other services activities	8.52%	8.29%	8.77%
Transport, storage and communication	8.43%	8.37%	8.51%
Mining and quarrying SubSection CA : Mining and quarrying of energy products	7.54%	7.51%	7.58%
Arts, entertainment and recreation	6.47%	6.51%	6.43%
Financial intermediation	6.25%	6.53%	5.96%
Health and social work	5.17%	5.43%	4.91%
Construction	2.95%	3.16%	2.74%
Wholesale and retail trade: repair of motor vehicles, motorcycles and personal tractors	2.62%	2.47%	2.77%

### 4.4.9 Percentatge d'estudiants en pràctiques per mida d'empresa

Un altre cas interessant és el que es mostra a la imatge 60, on el focus d'estudi és la distribució de les mobilitats dels estudiants en pràctiques per mida empresarial.

**Imatge 60. Percentatge d'estudiants en pràctiques per mida empresa**

C\_PercentatgeEnPractiquesPerMidaEmpresa.saiku

You are using Saiku Community Edition, please consider upgrading to [Saiku Enterprise](#)

Placement Size	Students %		
	All Courses		
		2011_2012	2012_2013
Large	25.37%	25.43%	25.30%
Medium	31.39%	31.87%	30.89%
Small	43.24%	42.69%	43.81%

#### 4.4.10 Distribució segons la preparació en idiomes

Per acabar, la imatge 61 presenta la relació que hi ha entre la preparació en idiomes segons el sexe i l'evolució per curs.

**Imatge 61. Distribució segons la preparació en idiomes**

Editing:D\_PercentatgeSegonPreparacioEnIdiomes.saiku

You are using Saiku Community Edition, please consider upgrading to [Saiku Enterprise](#), or entering a [sponsorship agreement with us](#) to support development.

LingualPreparation	All Courses									
	Students %	All Genders			2011_2012			2012_2013		
		All Genders		All Genders		All Genders		All Genders		
		Female	Male	Female	Male	Female	Male	Female	Male	
EIHC	2.34%	2.37%	2.28%	2.34%	2.36%	2.29%	2.34%	2.38%	2.27%	
Home	17.63%	17.25%	18.21%	18.13%	17.79%	18.65%	17.16%	16.76%	17.80%	
Host	11.19%	11.28%	11.05%	11.90%	11.99%	11.75%	10.53%	10.62%	10.39%	
None	68.85%	69.09%	68.46%	67.64%	67.85%	67.31%	69.97%	70.24%	69.55%	

## 5 Conclusions

S'han analitzat amb profunditat els requeriments dels usuaris, que s'han presentat en requeriments funcionals i no funcionals i s'han descrit els casos d'ús del sistema a crear. També s'ha aprofundit en el contingut i estructura de les fonts de dades. I a partir del que es vol i es necessita (requeriments) i del que es disposa (font de dades) s'ha proposat un model conceptual del magatzem, a partir del qual s'ha extret el model lògic, model que s'ha implementat, model físic, en un RDBMS concret (MySQL)

Un cop fet el disseny, s'ha procedit a la descripció a alt nivell dels processos d'ETL. Processos que a grans trets s'han estructurat en dos treballs: un per a les fonts de dades mestres i un altre per a les de mobilitat. Els treballs, al seu torn, s'han estructurat en transformacions. Tant els treballs com les transformacions poden ser executades individualment, i, així, l'administrador podrà, en tot moment, decidir si executar el procés que englobi totes les fonts o elegir la font a incorporar al magatzem. Quedaria per a una posterior millora del sistema el tractament dels valors del grànul no existents en les dimensions del DW.

Com ja s'apuntava en els treballs previs (**Pla de treball i Anàlisi preliminar**), el disseny conceptual del DW ha estat la part que més esforç ha requerit i més complicacions ha presentat. Això ha estat degut a la necessitat de més coneixements específics de l'àrea de magatzem de dades que, fins ara, eren aliens a l'alumne.

En definitiva, s'ha construït un sistema d'informació que compleix amb les peticions del client i amb les premisses de l'anàlisi, s'ha gestionat el projecte de manera eficient, acabant-lo en el termini especificat i amb l'obtenció del producte desitjat.

## 6 Glossari de termes

**Actors**, papers que els usuaris juguen en interaccionar amb un cas d'ús.

**BA**, *Business Analítics* o anàlisi del negoci, veure BI.

**BI**, *Business Intelligence* o intel·ligència del negoci, conjunt de sistemes i processos encarregats d'extreure informació de les dades.

**Casos d'ús**, estandardització de la descripció d'un requeriment funcional que es correspon amb una acció de l'usuari en el sistema dissenyat.

**Cub OLAP**, terme que es refereix a una estructura multidimensional de dades.

**Dimensió**, punt de vista des del que s'analitzen les mesures de la taula de fets.

**DW**, veure magatzem de dades.

**Estrella**, estructura d'informació que descriu les relacions entre les dimensions i els fets.

**Floc de neu**, estructura d'informació que modifica l'esquema en estrella introduint una major normalització.

**Fonts de dades**, dades procedents dels sistemes operacionals i que es presenten en diferents formats (fitxers plans, pdf, xml, etc.) que podran ser utilitzades per omplir els magatzem de dades.

**Magatzem de dades**, base de dades corporativa que s'utilitza per a la presa de decisions, i que es nodreix de les dades del sistemes operacionals.

**MDX**, acrònim de *Multidimensional Expression*, és un llenguatge declaratiu que és una extensió del SQL y que permet accedir i gestionar objectes multidimensionals.

**Mesura**, atribut de la taula de fets que dona a lloc als indicadors pel BI.

**ETL**, conjunt de processos que formen part del DW i que són els encarregats d'extreure les dades, transformar-les i carregar-les al seu esquema.

**OLAP**, acrònim de l'anglès *Online Analytical Processing*, solució que permet accedir en línia a una gran quantitat de dades agregades.

**Open data**, o dades obertes, és el conjunt de dades posades a disposició del públic i que es poden tornar a reutilitzar sense cap tipus de restricció.

**RDBMS**, o SGBD és el sistema relacional de gestió de bases de dades.

**Requeriments**, necessitats dels usuaris a les que el sistema dissenyat ha de donar resposta.

**SA**, *Staging Area*

**Stagin Area**, àrea de maniobra on es realitzen els processos d'ETL.

**Subsistemes**, part en què es divideix el sistema dissenyat per una major manejabilitat en el seu disseny.

## 7 Annexes

### 7.1 Annex I: Taules de l'*Staging Area*

Taula 42. Correspondència entre atributs

Font de dades		Taula de l' Staging Area	
Fitxer	Atribut	Atribut	Taula
ISOCountryCodes081507	Country	stg_countryCode	stg_country
	Name	stg_countryName	
ISCED97 Erasmus subject codes	Full code	stg_fullCode	stg_subjectArea
	ISCED97Code	stg_iscd97Code	
	ISCED97Description	stg_isced97description	
	erasmusShortCode	stg_erasmusShortCode	
EUC_Consolidated_Table_2007_2013	Country	stg_countryCode	stg_institution
	Organization Name	stg_institutionName	
	Erasmus code	stg_erasmusCode	
		stg_sizeShort	
		stg_sizeLong	
		stg_sectorShort	
		stg_sectorLong	

Taula 43. SM\_2011\_2012

Font de dades		Taula de l' Staging Area	
Fitxer	Atribut	Atribut	Taula
SM_2011_2012	HOMEINSITUATION	stg_homeInstitution	stg_mobility1
	HOSTINSTITUTION	stg_hostInsittution	
	PLACEMENTENTERPRISE	stg_institutionName	
	COUNTRYOFPLACEMENT	stg_countryPlacement	
	SUBJECTAREA	stg_subjectArea	
	STUDYSTARTDATE	stg_studysStartDate	
	PLACEMENTSTARTDATE	stg_placementStartDade	
	TYPEPLACEMENTSECTOR	stg_placemenSector	
	MOBILITYTYPE	stg_mobilityType	
	AGE	stg_age	
	GENDER	stg_gender	
	NATIONALITY	stg_nationality	
	LEVELSTUDY	stg_studyLevel	
	LINGPREPARATION	stg_lingPreparation	
	STUDYGRANT	stg_studyGrant	
	PLACEMENTGRANT	stg_placementGrant	

Taula 44. SM\_2012\_2013

Font de dades		Taula de l' Staging Area	
Fitxer	Atribut	Atribut	Taula
SM_2012_2013	HOME_INSTITUTION_CDE	stg_homeInstitution	stg_mobility1
	HOST_INSTITUTION_CDE	stg_hostInsittution	
	PLACEMENT_ENTERPRISE_VALUE	stg_institutionName	
	PLACEMENT_ENTERPRISE_CTRY_CDE	stg_countryPlacement	
	STUDY_START_DATE	stg_studysStartDate	
	PLACEMENT_START_DATE	stg_placementStartDade	
	STUDENT_SUBJECT_AREA_VALUE	stg_subjectArea	
	MOBILITY_TYPE_CDE	stg_mobilityType	
	STUDENT_AGE_VALUE	stg_age	
	STUDENT_GENDER_CDE	stg_gender	
	STUDENT_NATIONALITY_CDE	stg_nationality	
	STUDENT_STUDY_LEVEL_CDE	stg_studyLevel	
	LINGUISTIC_PREPARATION_CDE	stg_lingPreparation	
	STUDY_GRANT_AMT	stg_studyGrant	
	PLACEMENT_GRANT_AMT	stg_placementGrant	

Taula 45 Taula per la càrrega del grànul

Atribut	Taula de l' Staging Area
stg_homeInstitution	stg_mobility2
stg_hostInsittution	
stg_countryPlacement	
stg_subjectArea	
stg_mobilityType	
stg_age	
stg_gender	
stg_nationality	
stg_studyLevel	
stg_lingPreparation	
stg_students	
stg_ageAvgContribution	
stg_withStudyGrant	
stg_withPlacementGrant	
stg_withTotalGrant	
Stg_sector	
stg_course	

## 7.2 Annex II: Scripts SQL DW

### 7.2.1 Annex II: Script SQL dw

```

CREATE `dw`
USE `dw`;
--
-- Table structure for table `dim_age`
DROP TABLE IF EXISTS `dim_age`;
CREATE TABLE `dim_age` (
  `dim_id` tinyint(4) unsigned NOT NULL AUTO_INCREMENT,
  `dim_age` tinyint(4) unsigned NOT NULL,

```

```

PRIMARY KEY (`dim_id`),
UNIQUE KEY `dim_agecol_UNIQUE` (`dim_age`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
--
-- Table structure for table `dim_course`
DROP TABLE IF EXISTS `dim_course`;
CREATE TABLE `dim_course` (
  `dim_id`      tinyint(4) unsigned NOT NULL AUTO_INCREMENT,
  `dim_course` varchar(9) NOT NULL,
  PRIMARY KEY (`dim_id`),
  UNIQUE KEY `dim_course_UNIQUE` (`dim_course`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
--
-- Table structure for table `dim_gender`
DROP TABLE IF EXISTS `dim_gender`;
CREATE TABLE `dim_gender` (
  `dim_id`      tinyint(4) unsigned NOT NULL AUTO_INCREMENT,
  `dim_genderShort` varchar(1) NOT NULL,
  `dim_genderLong` varchar(10) NOT NULL,
  PRIMARY KEY (`dim_id`),
  UNIQUE KEY `dim_genderShort_UNIQUE` (`dim_genderShort`),
  UNIQUE KEY `dim_genderLong_UNIQUE` (`dim_genderLong`)
) ENGINE=InnoDB AUTO_INCREMENT=3 DEFAULT CHARSET=utf8;
--
-- Table structure for table `dim_institution`
DROP TABLE IF EXISTS `dim_institution`;
CREATE TABLE `dim_institution` (
  `dim_id` int(10) unsigned NOT NULL AUTO_INCREMENT,
  `dim_countryCode` varchar(4) NOT NULL,
  `dim_countryName` varchar(50) NOT NULL,
  `dim_erasmusCode` varchar(13) DEFAULT NULL,
  `dim_institutionName` varchar(210) DEFAULT NULL,
  `dim_placementSizeShort` varchar(1) DEFAULT NULL,
  `dim_placementSizeLong` varchar(10) DEFAULT NULL,
  `dim_placementSectorShort` varchar(1) DEFAULT NULL,
  `dim_placementSectorLong` varchar(75) DEFAULT NULL,
  PRIMARY KEY (`dim_id`),
  UNIQUE KEY `dim_institutionName_UNIQUE`
(`dim_countryCode`,`dim_institutionName`,`dim_erasmusCode`,`dim_placementSectorShort`)
) ENGINE=InnoDB AUTO_INCREMENT=75066 DEFAULT CHARSET=utf8;
--
-- Table structure for table `dim_lingpreparation`
DROP TABLE IF EXISTS `dim_lingpreparation`;
CREATE TABLE `dim_lingpreparation` (
  `dim_id`      tinyint(4) unsigned NOT NULL AUTO_INCREMENT,
  `dim_lingPreparationShort` varchar(2) NOT NULL,

```

```

`dim_lingPreparationLong` varchar(10) NOT NULL,
PRIMARY KEY (`dim_id`),
UNIQUE KEY `dim_lingPreparationShort_UNIQUE` (`dim_lingPreparationShort`),
UNIQUE KEY `dim_lingPreparationLong_UNIQUE` (`dim_lingPreparationLong`)
) ENGINE=InnoDB AUTO_INCREMENT=5 DEFAULT CHARSET=utf8;
--
-- Table structure for table `dim_mobilitytype`
DROP TABLE IF EXISTS `dim_mobilitytype`;
CREATE TABLE `dim_mobilitytype` (
  `dim_id`          tinyint(4) unsigned NOT NULL AUTO_INCREMENT,
  `dim_mobilityTypeShort` varchar(1) NOT NULL,
  `dim_mobilityTypeLong` varchar(25) NOT NULL,
  PRIMARY KEY (`dim_id`),
  UNIQUE KEY `dim_mobilityTypeShort_UNIQUE` (`dim_mobilityTypeShort`),
  UNIQUE KEY `dim_mobilityTypeLong_UNIQUE` (`dim_mobilityTypeLong`)
) ENGINE=InnoDB AUTO_INCREMENT=4 DEFAULT CHARSET=utf8;
--
-- Table structure for table `dim_nationality`
DROP TABLE IF EXISTS `dim_nationality`;
CREATE TABLE `dim_nationality` (
  `dim_id`          smallint(6) unsigned NOT NULL AUTO_INCREMENT,
  `dim_countryCode` varchar(4) NOT NULL,
  `dim_countryName` varchar(50) NOT NULL,
  PRIMARY KEY (`dim_id`),
  UNIQUE KEY `dim_countryCode_UNIQUE` (`dim_countryCode`),
  UNIQUE KEY `dim_countryName_UNIQUE` (`dim_countryName`)
) ENGINE=InnoDB AUTO_INCREMENT=254 DEFAULT CHARSET=utf8;
--
-- Table structure for table `dim_studylevel`
DROP TABLE IF EXISTS `dim_studylevel`;
CREATE TABLE `dim_studylevel` (
  `dim_Id`          tinyint(4) unsigned NOT NULL AUTO_INCREMENT,
  `dim_studyLevelShort` varchar(1) NOT NULL,
  `dim_studyLevelLong` varchar(10) NOT NULL,
  PRIMARY KEY (`dim_Id`),
  UNIQUE KEY `dim_studyLevelShort_UNIQUE` (`dim_studyLevelShort`),
  UNIQUE KEY `dim_studyLevelcol_UNIQUE` (`dim_studyLevelLong`)
) ENGINE=InnoDB AUTO_INCREMENT=5 DEFAULT CHARSET=utf8;
--
-- Table structure for table `dim_subjectarea`
DROP TABLE IF EXISTS `dim_subjectarea`;
CREATE TABLE `dim_subjectarea` (
  `dim_id`          smallint(6) unsigned NOT NULL AUTO_INCREMENT,
  `dim_ISCDE97Code` varchar(4) NOT NULL,
  `dim_ISCEDescription` varchar(75) NOT NULL,
  `dim_erasmusShortCode` varchar(2) NOT NULL,

```



```

PRIMARY KEY (`dim_id`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
--
-- Table structure for table `fac_mobility`
DROP TABLE IF EXISTS `fac_mobility`;
CREATE TABLE `fac_mobility` (
  `fac_courseId` tinyint(4) unsigned NOT NULL,
  `fac_ageId` tinyint(4) unsigned NOT NULL DEFAULT '0',
  `fac_genderId` tinyint(4) unsigned NOT NULL DEFAULT '0',
  `fac_studyLevelId` tinyint(4) unsigned NOT NULL DEFAULT '0',
  `fac_lingPreparationId` tinyint(4) unsigned NOT NULL,
  `fac_mobilityTypeId` tinyint(4) unsigned NOT NULL,
  `fac_nationalityId` smallint(6) unsigned NOT NULL,
  `fac_subjectAreaId` smallint(6) unsigned NOT NULL DEFAULT '0',
  `fac_homeInstitutionId` int(10) unsigned NOT NULL DEFAULT '0',
  `fac_hostInstitutionId` int(10) unsigned NOT NULL DEFAULT '0',
  `fac_students` smallint(6) DEFAULT NULL,
  `fac_ageAvgContribution` decimal(9,2) DEFAULT NULL,
  `fac_withStudyGrant` int(11) DEFAULT NULL,
  `fac_withPlacementGrant` int(11) DEFAULT NULL,
  `fac_withTotalGrant` int(11) DEFAULT NULL,
  UNIQUE KEY `pk_index`
(`fac_courseId`,`fac_ageId`,`fac_genderId`,`fac_studyLevelId`,`fac_lingPreparationId`,`fac_nationalityId`,`fac_mobilityTypeId`,`fac_subjectAreaId`,`fac_homeInstitutionId`,`fac_hostInstitutionId`),
  KEY `course_idx` (`fac_ageId`),
  KEY `gender_idx` (`fac_genderId`),
  KEY `lingPreparation_idx` (`fac_lingPreparationId`),
  KEY `nationality_idx` (`fac_nationalityId`),
  KEY `studyLevel_idx` (`fac_studyLevelId`),
  KEY `mobilityType_idx` (`fac_mobilityTypeId`),
  KEY `curs_fk_idx` (`fac_courseId`),
  KEY `hostInstitution_idx` (`fac_homeInstitutionId`),
  KEY `fk_fac_mobility_8_idx` (`fac_subjectAreaId`),
  KEY `fk_fac_mobility_10_idx` (`fac_hostInstitutionId`),
  CONSTRAINT `fk_fac_mobility_1` FOREIGN KEY (`fac_ageId`) REFERENCES `dim_age` (`dim_id`)
ON DELETE NO ACTION ON UPDATE NO ACTION,
  CONSTRAINT `fk_fac_mobility_10` FOREIGN KEY (`fac_hostInstitutionId`) REFERENCES `dim_institution` (`dim_id`) ON DELETE NO ACTION ON UPDATE NO ACTION,
  CONSTRAINT `fk_fac_mobility_2` FOREIGN KEY (`fac_courseId`) REFERENCES `dim_course` (`dim_id`) ON DELETE NO ACTION ON UPDATE NO ACTION,
  CONSTRAINT `fk_fac_mobility_3` FOREIGN KEY (`fac_genderId`) REFERENCES `dim_gender` (`dim_id`) ON DELETE NO ACTION ON UPDATE NO ACTION,
  CONSTRAINT `fk_fac_mobility_4` FOREIGN KEY (`fac_studyLevelId`) REFERENCES `dim_studylevel` (`dim_id`) ON DELETE NO ACTION ON UPDATE NO ACTION,
  CONSTRAINT `fk_fac_mobility_5` FOREIGN KEY (`fac_lingPreparationId`) REFERENCES `dim_lingpreparation` (`dim_id`) ON DELETE NO ACTION ON UPDATE NO ACTION,
  CONSTRAINT `fk_fac_mobility_6` FOREIGN KEY (`fac_mobilityTypeId`) REFERENCES `dim_mobilitytype` (`dim_id`) ON DELETE NO ACTION ON UPDATE NO ACTION,

```

```

CONSTRAINT `fk_fac_mobility_7` FOREIGN KEY (`fac_nationalityId`) REFERENCES `dim_nationality`
(`dim_id`) ON DELETE NO ACTION ON UPDATE NO ACTION,
CONSTRAINT `fk_fac_mobility_8` FOREIGN KEY (`fac_subjectareald`) REFERENCES
`dim_subjectarea` (`dim_id`) ON DELETE NO ACTION ON UPDATE NO ACTION,
CONSTRAINT `fk_fac_mobility_9` FOREIGN KEY (`fac_homeInstitutionId`) REFERENCES
`dim_institution` (`dim_id`) ON DELETE NO ACTION ON UPDATE NO ACTION
) ENGINE=InnoDB DEFAULT CHARSET=utf8;

```

## 7.2.2 Annex II: Script SQL SA

```

CREATE DATABASE `stg`
--
-- Table structure for table `stg_Sector`
DROP TABLE IF EXISTS `stg_Sector`;
CREATE TABLE `stg_Sector` (
  `stg_Id`          int(10) unsigned NOT NULL AUTO_INCREMENT,
  `stg_sectorShort` varchar(1) NOT NULL,
  `stg_sectorLong`  varchar(75) NOT NULL,
  PRIMARY KEY (`stg_Id`),
  UNIQUE KEY `stg_sectorLong_UNIQUE` (`stg_sectorLong`),
  UNIQUE KEY `stg_sectorShort_UNIQUE` (`stg_sectorShort`)
) ENGINE=InnoDB AUTO_INCREMENT=23 DEFAULT CHARSET=utf8;
--
-- Table structure for table `stg_Size`
DROP TABLE IF EXISTS `stg_Size`;
CREATE TABLE `stg_Size` (
  `stg_Id`          tinyint(3) unsigned NOT NULL AUTO_INCREMENT,
  `stg_sizeShort`   varchar(1) NOT NULL,
  `stg_sizeLong`    varchar(15) NOT NULL,
  PRIMARY KEY (`stg_Id`)
) ENGINE=InnoDB AUTO_INCREMENT=5 DEFAULT CHARSET=utf8;
--
-- Table structure for table `stg_age`
DROP TABLE IF EXISTS `stg_age`;
CREATE TABLE `stg_age` (
  `stg_id`          tinyint(3) unsigned NOT NULL AUTO_INCREMENT,
  `stg_age`         tinyint(4) NOT NULL,
  PRIMARY KEY (`stg_id`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
--
-- Table structure for table `stg_country`
DROP TABLE IF EXISTS `stg_country`;
CREATE TABLE `stg_country` (
  `stg_Id`          int(10) unsigned NOT NULL AUTO_INCREMENT,
  `stg_countryCode` varchar(4) NOT NULL,
  `stg_countryName` varchar(50) NOT NULL,
  PRIMARY KEY (`stg_Id`),

```

```

    UNIQUE KEY `stg_countryName_UNIQUE` (`stg_countryName`),
    UNIQUE KEY `stg_countryCode_UNIQUE` (`stg_countryCode`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `stg_course`
DROP TABLE IF EXISTS `stg_course`;
CREATE TABLE `stg_course` (
  `stg_id`      tinyint(4) unsigned NOT NULL AUTO_INCREMENT,
  `stg_course` varchar(9) NOT NULL,
  PRIMARY KEY (`stg_id`),
  UNIQUE KEY `stg_course_UNIQUE` (`stg_course`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
--
-- Table structure for table `stg_gender`
DROP TABLE IF EXISTS `stg_gender`;
CREATE TABLE `stg_gender` (
  `stg_id`      tinyint(3) unsigned NOT NULL AUTO_INCREMENT,
  `stg_genderShort` varchar(1) NOT NULL,
  `stg_genderLong` varchar(10) NOT NULL,
  PRIMARY KEY (`stg_id`),
  UNIQUE KEY `stg_genderLong_UNIQUE` (`stg_genderLong`),
  UNIQUE KEY `stg_genderShort_UNIQUE` (`stg_genderShort`)
) ENGINE=InnoDB AUTO_INCREMENT=3 DEFAULT CHARSET=utf8;
--
-- Table structure for table `stg_institution`
DROP TABLE IF EXISTS `stg_institution`;
CREATE TABLE `stg_institution` (
  `stg_id`      int(11) unsigned NOT NULL AUTO_INCREMENT,
  `stg_institutionName` varchar(210) NOT NULL,
  `stg_countryCode` varchar(4) NOT NULL,
  `stg_countryName` varchar(50) DEFAULT NULL,
  `stg_erasmusCode` varchar(13) DEFAULT NULL,
  `stg_placementSizeShort` varchar(1) DEFAULT NULL,
  `stg_placementSizeLong` varchar(10) DEFAULT NULL,
  `stg_placementSectorShort` varchar(1) DEFAULT NULL,
  `stg_placementSectorLong` varchar(75) DEFAULT NULL,
  PRIMARY KEY (`stg_id`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
--
-- Table structure for table `stg_lingpreparation`
DROP TABLE IF EXISTS `stg_lingpreparation`;
CREATE TABLE `stg_lingpreparation` (
  `stg_id`      tinyint(3) unsigned NOT NULL AUTO_INCREMENT,
  `stg_lingPreparationShort` varchar(2) NOT NULL,
  `stg_lingPreparationLong` varchar(10) NOT NULL,

```

```

PRIMARY KEY (`stg_id`,`stg_lingPreparationLong`),
UNIQUE KEY `stg_lingPreparationLong_UNIQUE` (`stg_lingPreparationLong`)
) ENGINE=InnoDB AUTO_INCREMENT=5 DEFAULT CHARSET=utf8;
--
-- Table structure for table `stg_mobilitytype`
DROP TABLE IF EXISTS `stg_mobilitytype`;
CREATE TABLE `stg_mobilitytype` (
  `stg_id`          tinyint(3) unsigned NOT NULL AUTO_INCREMENT,
  `stg_mobilityTypeCode`  varchar(1) NOT NULL,
  `stg_mobilityTypeDescription`  varchar(25) NOT NULL,
  PRIMARY KEY (`stg_id`,`stg_mobilityTypeCode`),
  UNIQUE KEY `stg_mobilityTypeDescription_UNIQUE` (`stg_mobilityTypeDescription`)
) ENGINE=InnoDB AUTO_INCREMENT=4 DEFAULT CHARSET=utf8;
--
-- Table structure for table `stg_studylevel`
DROP TABLE IF EXISTS `stg_studylevel`;
CREATE TABLE `stg_studylevel` (
  `stg_id`          tinyint(3) unsigned NOT NULL AUTO_INCREMENT,
  `stg_studyLevel`  varchar(1) NOT NULL,
  `stg_studyLevelDescription`  varchar(15) NOT NULL,
  PRIMARY KEY (`stg_id`),
  UNIQUE KEY `stg_studyLevelDescription_UNIQUE` (`stg_studyLevelDescription`),
  UNIQUE KEY `stg_studyLevel_UNIQUE` (`stg_studyLevel`)
) ENGINE=InnoDB AUTO_INCREMENT=5 DEFAULT CHARSET=utf8;
--
-- Table structure for table `stg_subjectArea`
DROP TABLE IF EXISTS `stg_subjectArea`;
CREATE TABLE `stg_subjectArea` (
  `stg_id`          smallint(6) unsigned NOT NULL AUTO_INCREMENT,
  `stg_ISCED97Code`  varchar(4) NOT NULL,
  `stg_ISCED97Description`  varchar(75) NOT NULL,
  `stg_erasmusShortCode`  varchar(2) NOT NULL,
  PRIMARY KEY (`stg_id`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
--
-- Table structure for table `stg_mobility1`
DROP TABLE IF EXISTS `stg_mobility1`;
CREATE TABLE `stg_mobility1` (
  `stg_id` int(11) NOT NULL AUTO_INCREMENT,
  `stg_homeInstitution`  varchar(13) NOT NULL,
  `stg_hostInstitution`  varchar(13) DEFAULT NULL,
  `stg_placement`  varchar(210) DEFAULT NULL,
  `stg_placementCountry`  varchar(4) DEFAULT NULL,
  `stg_subjectArea`  varchar(4) DEFAULT NULL,
  `stg_mobilityType`  varchar(1) DEFAULT NULL,
  `stg_age`  tinyint(4) DEFAULT NULL,

```

```

`stg_gender` varchar(1) DEFAULT NULL,
`stg_nationality` varchar(4) DEFAULT NULL,
`stg_studyLevel` varchar(1) DEFAULT NULL,
`stg_lingPreparation` varchar(2) DEFAULT NULL,
`stg_studyGrant` decimal(10,2) DEFAULT NULL,
`stg_placementGrant` decimal(10,2) DEFAULT NULL,
`stg_studyStartDate` varchar(7) DEFAULT NULL,
`stg_placementStartDate` varchar(7) DEFAULT NULL,
`stg_course` varchar(9) DEFAULT 'null',
`stg_placementSector` varchar(1) DEFAULT NULL,
PRIMARY KEY (`stg_id`)
) ENGINE=InnoDB AUTO_INCREMENT=267183 DEFAULT CHARSET=utf8;
-
-- Table structure for table `stg_mobility2`
DROP TABLE IF EXISTS `stg_mobility2`;
CREATE TABLE `stg_mobility2` (
  `stg_id` int(11) NOT NULL AUTO_INCREMENT,
  `stg_homeInstitution` varchar(13) DEFAULT NULL,
  `stg_hostInstitution` varchar(13) DEFAULT NULL,
  `stg_placement` varchar(210) DEFAULT NULL,
  `stg_placementCountry` varchar(4) DEFAULT NULL,
  `stg_subjectArea` varchar(4) DEFAULT NULL,
  `stg_mobilityType` varchar(1) DEFAULT NULL,
  `stg_age` tinyint(4) DEFAULT NULL,
  `stg_gender` varchar(1) DEFAULT NULL,
  `stg_nationality` varchar(4) DEFAULT NULL,
  `stg_studyLevel` varchar(1) DEFAULT NULL,
  `stg_lingPreparation` varchar(2) DEFAULT NULL,
  `stg_students` smallint(6) DEFAULT NULL,
  `stg_ageAvgContribution` decimal(9,2) DEFAULT NULL,
  `stg_withStudyGrant` tinyint(1) unsigned DEFAULT '0',
  `stg_withPlacementGrant` tinyint(1) unsigned DEFAULT '0',
  `stg_course` varchar(9) DEFAULT NULL,
  `stg_withTotalGrant` tinyint(1) unsigned DEFAULT '0',
  `stg_placementSector` varchar(1) DEFAULT NULL,
  PRIMARY KEY (`stg_id`)
) ENGINE=InnoDB AUTO_INCREMENT=226318 DEFAULT CHARSET=utf8;

```

### 7.3 Annex III: Cub OLAP TFG\_ErasmusMobility

```

=<Schema name="TFG_ErasmusMobility">
=<Dimension visible="true" highCardinality="false" name="Age">
=<Hierarchy visible="true" hasAll="true" primaryKey="dim_id">
<Table name="dim_age" />

```

```

    <Level name="Age" visible="true" column="dim_age" type="Numeric"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never" />
  </Hierarchy>
</Dimension>
_ <Dimension visible="true" highCardinality="false" name="Course">
_ <Hierarchy visible="true" hasAll="true" primaryKey="dim_id">
  <Table name="dim_course" />
    <Level name="Course" visible="true" column="dim_course" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never" />
  </Hierarchy>
</Dimension>
_ <Dimension visible="true" highCardinality="false" name="Gender">
_ <Hierarchy name="Gender" visible="true" hasAll="true" primaryKey="dim_id">
  <Table name="dim_gender" />
    <Level name="Gender" visible="true" column="dim_genderLong" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never" />
  </Hierarchy>
</Dimension>
_ <Dimension visible="true" highCardinality="false" name="LingualPreparation">
_ <Hierarchy name="LingualPreparation" visible="true" hasAll="true"
primaryKey="dim_id">
  <Table name="dim_lingpreparation" />
    <Level name="LingualPreparation" visible="true"
column="dim_lingPreparationLong" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never" />
  </Hierarchy>
</Dimension>
_ <Dimension visible="true" highCardinality="false" name="MobilityType">
_ <Hierarchy name="MobilityType" visible="true" hasAll="true" primaryKey="dim_id">
  <Table name="dim_mobilitytype" />
    <Level name="MobilityType" visible="true" column="dim_mobilityTypeLong"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never" />
  </Hierarchy>
</Dimension>
_ <Dimension visible="true" highCardinality="false" name="Nationality">
_ <Hierarchy name="Code" visible="true" hasAll="true" primaryKey="dim_id">
  <Table name="dim_nationality" />
    <Level name="Code" visible="true" column="dim_countryCode" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never" />
  </Hierarchy>
_ <Hierarchy name="Name" visible="true" hasAll="true" primaryKey="dim_id">
  <Table name="dim_nationality" />
    <Level name="Name" visible="true" column="dim_countryName" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never" />

```

```

</Hierarchy>
</Dimension>
= <Dimension visible="true" highCardinality="false" name="StudyLevel">
= <Hierarchy name="StudyLevel" visible="true" hasAll="true" primaryKey="dim_Id">
  <Table name="dim_studylevel" />
    <Level name="StudyLevel" visible="true" column="dim_studyLevelLong"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never" />
  </Hierarchy>
</Dimension>
= <Dimension visible="true" highCardinality="false" name="SubjectArea"
caption="Subject Area">
= <Hierarchy name="ByISCDE97Code" visible="true" hasAll="true"
primaryKey="dim_id">
  <Table name="dim_subjectarea" />
    <Level name="SubjectAreaCode" visible="true" column="dim_ISCDE97Code"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Subject Area ISCDE97 Code" />
  </Hierarchy>
= <Hierarchy name="ByISCDE97Description" visible="true" hasAll="true"
primaryKey="dim_id">
  <Table name="dim_subjectarea" />
    <Level name="SubjectAreaDescription" visible="true"
column="dim_ISCEDescription" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never" caption="Subject Area Description" />
  </Hierarchy>
</Dimension>
= <Dimension visible="true" highCardinality="false" name="HomeInstitution"
caption="Home Institution">
= <Hierarchy name="ByHomeCountryCode" visible="true" hasAll="true"
primaryKey="dim_id">
  <Table name="dim_institution" />
    <Level name="HomeCountryCode" visible="true" column="dim_countryCode"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Home Country Code" />
  </Hierarchy>
= <Hierarchy name="ByHomeCountryName" visible="true" hasAll="true"
primaryKey="dim_id">
  <Table name="dim_institution" />
    <Level name="HomeCountryName" visible="true" column="dim_countryName"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Home Country Name" />
    <Level name="HomeInstitutionName" visible="true"
column="dim_institutionName" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never" caption="Home Institution Name" />
  </Hierarchy>
</Dimension>

```

```

= <Dimension visible="true" highCardinality="false" name="HostInstitution"
caption="Host Institution">
= <Hierarchy name="ByHostCountryCode" visible="true" hasAll="true"
primaryKey="dim_id">
  <Table name="dim_institution" />
  <Level name="HostCountryCode" visible="true" column="dim_countryCode"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Host Country Code" />
</Hierarchy>
= <Hierarchy name="ByHostCountryName" visible="true" hasAll="true"
primaryKey="dim_id">
  <Table name="dim_institution" />
  <Level name="HostCountryName" visible="true" column="dim_countryName"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Host Country Name" />
  <Level name="HostInstitutionName" visible="true" column="dim_institutionName"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Host Institution Name" />
</Hierarchy>
= <Hierarchy name="ByPlacementSectorLong" visible="true" hasAll="true"
primaryKey="dim_id">
  <Table name="dim_institution" />
  <Level name="PlacementSectorLong" visible="true"
column="dim_placementSectorLong" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never" caption="Placement Sector" />
</Hierarchy>
= <Hierarchy name="ByPlacementSizeLong" visible="true" hasAll="true"
primaryKey="dim_id">
  <Table name="dim_institution" />
  <Level name="PlacementSizeLong" visible="true"
column="dim_placementSizeLong" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never" caption="Placement Size" />
</Hierarchy>
</Dimension>
= <Cube name="TFG_ErasmusMobility" visible="true" cache="true" enabled="true">
  <Table name="fac_mobility" />
  <DimensionUsage source="Age" name="Age" visible="true" foreignKey="fac_ageId"
highCardinality="false" />
  <DimensionUsage source="Course" name="Course" visible="true"
foreignKey="fac_courseId" highCardinality="false" />
  <DimensionUsage source="Gender" name="Gender" visible="true"
foreignKey="fac_genderId" highCardinality="false" />
  <DimensionUsage source="LingualPreparation" name="LingPreparation"
visible="true" foreignKey="fac_lingPreparationId" highCardinality="false" />
  <DimensionUsage source="MobilityType" name="MobilityType" visible="true"
foreignKey="fac_mobilityTypeId" highCardinality="false" />
  <DimensionUsage source="Nationality" name="Nationality" visible="true"
foreignKey="fac_nationalityId" highCardinality="false" />

```



```

    <DimensionUsage source="StudyLevel" name="StudyLevel" visible="true"
foreignKey="fac_studyLevelId" highCardinality="false" />
    <DimensionUsage source="SubjectArea" name="SubjectArea" visible="true"
foreignKey="fac_subjectareaId" highCardinality="false" />
    <DimensionUsage source="HomeInstitution" name="HomeInstitution" visible="true"
foreignKey="fac_homeInstitutionId" highCardinality="false" />
    <DimensionUsage source="HostInstitution" name="HostInstitution" visible="true"
foreignKey="fac_hostInstitutionId" highCardinality="false" />
    <Measure name="Students" column="fac_students" formatString="#,###"
aggregator="sum" caption="Students No." />
    <Measure name="AgeAvgContribution" column="fac_ageAvgContribution"
formatString="#,00" aggregator="sum" visible="false" />
    <Measure name="withPlacementGrant" column="fac_withPlacementGrant"
formatString="#,###" aggregator="sum" caption="Placement Grants" />
    <Measure name="withStudyGrant" column="fac_withStudyGrant"
formatString="#,###" aggregator="sum" caption="Study Grants" />
    <Measure name="withTotalGrant" column="fac_withTotalGrant" formatString="#"
aggregator="sum" caption="Total Grants" />
    <CalculatedMember name="AverageAge" formatString="#,00" caption="Average Age"
formula="[Measures].[AgeAvgContribution]/[Measures].[Students]"
dimension="Measures" visible="true" />
</Cube>
</Schema>

```

## 7.4 Annex IV: Reports Saiku

### Top 10 instituciones emissoras

```

WITH
SET [~COLUMNS] AS
    {[Course].Members}
SET [~ROWS_HomeInstitution_HomeInstitution.ByHomeCountryName] AS
    {[HomeInstitution.ByHomeCountryName].[HomeInstitutionName].Members}
SET [~ROWS_HomeInstitution_HomeInstitution.ByHomeCountryCode] AS
    {[HomeInstitution.ByHomeCountryCode].[HomeCountryCode].Members}
SELECT
NON EMPTY CrossJoin([~COLUMNS], {[Measures].[Students]}) ON COLUMNS,
NON EMPTY
TopCount(NonEmptyCrossJoin([~ROWS_HomeInstitution_HomeInstitution.ByHomeCountryName],
[~ROWS_HomeInstitution_HomeInstitution.ByHomeCountryCode]), 10, [Measures].[Students]) ON
ROWS
FROM [TFG_ErasmusMobility]

```

### Top 10 instituciones receptoras

```

WITH
SET [~COLUMNS] AS
    {[Course].Members}
SET [~ROWS_HomeInstitution_HomeInstitution.ByHomeCountryName] AS

```

```

    {[HomeInstitution.ByHomeCountryName].[HomeInstitutionName].Members}
SET [~ROWS_HomeInstitution_HomeInstitution.ByHomeCountryCode] AS
    {[HomeInstitution.ByHomeCountryCode].[HomeCountryCode].Members}
SELECT
NON EMPTY CrossJoin([~COLUMNS], {[Measures].[Students]}) ON COLUMNS,
NON EMPTY
TopCount(NonEmptyCrossJoin([~ROWS_HomeInstitution_HomeInstitution.ByHomeCountryName],
[~ROWS_HomeInstitution_HomeInstitution.ByHomeCountryCode]), 10, [Measures].[Students]) ON
ROWS
FROM [TFG_ErasmusMobility]

```

#### **Distribució en % per nacionalitat de l'estudiant**

```

WITH member [Measures].[Students %] AS
'[Measures].[Students]/([Measures].[Students],
[Nationality.Name].Parent)', format_string='0.00%'
SELECT
NON EMPTY NonEmptyCrossJoin({[Measures].[Students %]}, [Course].Members) ON COLUMNS,
NON EMPTY Order({[Nationality.Name].[Name].Members},[Measures].[Students %], BDESC) ON ROWS
FROM [TFG_ErasmusMobility]

```

#### **Distribució en % per àrea de coneixement**

```

WITH member [Measures].[Students %] AS
'[Measures].[Students]/([Measures].[Students],
[SubjectArea.ByISCDE97Description].Parent)', format_string='0.00%'
SELECT
NON EMPTY NonEmptyCrossJoin({[Measures].[Students %]}, [Course].Members) ON COLUMNS,
NON EMPTY
Order([SubjectArea.ByISCDE97Description].[SubjectAreaDescription].Members,[Measures].[Students
%],DESC) ON ROWS
FROM [TFG_ErasmusMobility]

```

#### **Evolució comparativa per curs**

```

WITH
SET [~ROWS] AS
    {[Gender].Members}
SET [~COLUMNS] AS
    Hierarchize({[Course].Members})
SELECT
NON EMPTY CrossJoin([~COLUMNS], {[Measures].[Students]}) ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [TFG_ErasmusMobility]

```

#### **Edat mitjana per nacionalitat receptora**

```

WITH
member [Measures].[Average Age] AS '[Measures].[AverageAge]', format_string="0.00"
SET [~ROWS] AS

```

```

    Order({[HostInstitution.ByHostCountryName].[HostCountryName].Members}, [Measures].[Average
Age], DESC)
SELECT
NonEmptyCrossJoin({[Measures].[Average Age]}, [Course].Members) ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [TFG_ErasmusMobility]

```

#### **Edat mitjana per nacionalitat emissora.**

```

WITH
member [Measures].[Average Age] AS '[Measures].[AverageAge]', format_string="0.00"
SET [~ROWS] AS
    Order({[HomeInstitution.ByHomeCountryName].[HomeCountryName].Members},
[Measures].[AverageAge], DESC)
SELECT
NonEmptyCrossJoin({[Measures].[Average Age]}, [Course].Members) ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [TFG_ErasmusMobility]

```

#### **Percentatge de beques per nacionalitat emissora**

```

WITH member [Measures].[Students %] AS
'[Measures].[withTotalGrant]/([Measures].[withTotalGrant],
[HomeInstitution.ByHomeCountryName].Parent)', format_string='0.00%'

SELECT
NON EMPTY CrossJoin({[Measures].[Students %] }, [Course].Members) ON COLUMNS,
NON EMPTY
Order({[HomeInstitution.ByHomeCountryName].[HomeCountryName].Members},[Measures].[Students
%],DESC) ON ROWS
FROM [TFG_ErasmusMobility]

```

#### **Percentatge amb beca per nacionalitat receptora**

```

WITH member [Measures].[Students %] AS
'[Measures].[withTotalGrant]/([Measures].[withTotalGrant],
[HostInstitution.ByHostCountryName].Parent)', format_string='0.00%'

SELECT
NON EMPTY CrossJoin({[Measures].[Students %] }, [Course].Members) ON COLUMNS,
NON EMPTY
Order({[HostInstitution.ByHostCountryName].[HostCountryName].Members},[Measures].[Students
%],DESC) ON ROWS
FROM [TFG_ErasmusMobility]

```

#### **Percentatge en pràctiques per àrea de coneixement**

```

WITH
member [Measures].[%Students] AS
'[Measures].[Students]/([Measures].[Students],
[SubjectArea.ByISCDE97Description].Parent)', format_string='0.00%'

```

```

SET [~ROWS] AS
    {[SubjectArea.ByISCDE97Description].[SubjectAreaDescription].Members}
SELECT
NON EMPTY CrossJoin({[Measures].[Students], [Measures].[%Students]}, [Course].[Course].Members)
ON COLUMNS,
NON EMPTY Order([~ROWS], [Measures].[%Students], DESC) ON ROWS
FROM [TFG_ErasmusMobility]
WHERE CrossJoin({[MobilityType].[Placement mobility]},
Except({[HostInstitution.ByPlacementSectorLong].[PlacementSectorLong].Members},
{[HostInstitution.ByPlacementSectorLong].[No effects]}))

```

### **Percentatge en pràctiques per sector empresarial**

```

WITH
member [Measures].[Students %] AS
'[Measures].[Students]/([Measures].[Students],
[HostInstitution.ByPlacementSectorLong].Parent)', format_string='0.00%'
SET [~HostInstitution_HostInstitution.ByPlacementSizeLong_(All)] AS
    {[HostInstitution.ByPlacementSizeLong].[All HostInstitution.ByPlacementSizeLongs]}
SET [~HostInstitution_HostInstitution.ByPlacementSizeLong_PlacementSizeLong] AS
    Except({[HostInstitution.ByPlacementSizeLong].[PlacementSizeLong].Members},
{[HostInstitution.ByPlacementSizeLong].[No effects]})
SET [~ROWS] AS
    {[HostInstitution.ByPlacementSectorLong].[PlacementSectorLong].Members}
SELECT
NON EMPTY CrossJoin({ [Measures].[Students %]}, [Course].Members) ON COLUMNS,
NON EMPTY Order([~ROWS], [Measures].[Students %],DESC) ON ROWS
FROM [TFG_ErasmusMobility]
WHERE CrossJoin({[MobilityType].[Placement mobility]},
[~HostInstitution_HostInstitution.ByPlacementSizeLong_PlacementSizeLong])

```

### **Percentatge en pràctiques per mida empresa receptora**

```

WITH
member [Measures].[Students %] AS
'[Measures].[Students]/([Measures].[Students],
[HostInstitution.ByPlacementSizeLong].Parent)', format_string='0.00%'
SET [~ROWS] AS
    {[HostInstitution.ByPlacementSizeLong].[PlacementSizeLong].Members}
SELECT
NON EMPTY CrossJoin({ [Measures].[Students %]}, [Course].Members) ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [TFG_ErasmusMobility]
WHERE CrossJoin({[MobilityType].[Placement mobility]},
Except({[HostInstitution.ByPlacementSectorLong].[PlacementSectorLong].Members},
{[HostInstitution.ByPlacementSectorLong].[No effects]}))

```

### **Percentatge segons la preparació en llengües**

WITH

```

member [Measures].[%Students] AS
'[Measures].[Students]/([Measures].[Students],
[LingPreparation.LingualPreparation].Parent)', format_string='0.00%'
SET [~COLUMNS_Course_Course] AS
    {[Course].[Course].Members}
SET [~COLUMNS_Gender_Gender.Gender] AS
    {[Gender].[Gender].Members}
SET [~ROWS] AS
    {[LingPreparation.LingualPreparation].[LingualPreparation].Members}
SELECT
NON EMPTY CrossJoin(NonEmptyCrossJoin([~COLUMNS_Course_Course],
[~COLUMNS_Gender_Gender.Gender]), {[Measures].[%Students]}) ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [TFG_ErasmusMobility]

```

## 8 Bibliografia

1. Rodríguez, José Ramón, Mariné Jové, José (2014). *Planificació del projecte*, Mòduls UOC de Gestió de Projectes. Barcelona: Universitat Oberta de Catalunya.
2. Webs de la Suite Pentaho, data consulta 24/09/15,  
<http://www.pentaho.com/>  
<https://en.wikipedia.org/wiki/Pentaho>  
<http://mondrian.pentaho.com/documentation>
3. Abelló Gamazo, Alberto (2012), *Disseny multidimensional*. Mòduls UOC de Gestió de Projectes. Barcelona: Universitat Oberta de Catalunya.
4. NACE codes, Web de la UE, consultada el 20/09/15.  
[http://ec.europa.eu/environment/emas/pdf/general/nacecodes\\_en.pdf/](http://ec.europa.eu/environment/emas/pdf/general/nacecodes_en.pdf/)
5. Data Warehouse Design Considerations, web de Microsoft (consulta 21/09/15).  
<https://technet.microsoft.com/en-us/library/aa902672%28v=sql.80%29.aspx>
6. Design the Dimensional Model, web de Kimball Group (consulta 21/09/15)  
<http://www.kimballgroup.com/2001/03/design-tip-21-declaring-the-grain/>
7. Non functional requirement (consulta 10/10/2015).  
[https://en.wikipedia.org/wiki/Non-functional\\_requirement](https://en.wikipedia.org/wiki/Non-functional_requirement)
8. Eiden, Petter L: H (2005) *Quantification and Traceability of Requirements* (consulta 11/10/2015)  
<http://www.idi.ntnu.no/grupper/su/fordypningsprosjekt-2005/eide-fordyp05.pdf>
9. Dades obertes gen cat(20/10/2015)  
[http://dadesobertes.gencat.cat/ca/que\\_es\\_l'open\\_data/concepte/](http://dadesobertes.gencat.cat/ca/que_es_l'open_data/concepte/)
10. MDX Data Definition Staments (consulta 21/10/2015).  
<https://msdn.microsoft.com/en-us/library/ms144926.aspx/>
11. Facultad de Ingeniería de la Universidad de la República – Uruguay. *Tutorial MDX* (data accés 01/12/2015).  
[www.fing.edu.uy/inco/grupos/csi/esp/Cursos/cursos\\_act/2005/DAP\\_SistDW/Material/2-SDW-Laboratorio1-2005.pdf](http://www.fing.edu.uy/inco/grupos/csi/esp/Cursos/cursos_act/2005/DAP_SistDW/Material/2-SDW-Laboratorio1-2005.pdf).
12. Microsoft Systems Journal Homepage (1999). *Manipulate and Query OLAP Data Using ADOMD and Multidimensional Expressions* (data d'accés 06/12/2015).  
<http://www.microsoft.com/msj/0899/mdx/mdx.aspx>