



# Construcció i explotació d'un magatzem de dades del sector turístic a Catalunya

**Enric Colomé Campos**  
Grau d'Enginyeria Informàtica

**Consultor: Carles Llorach Rius**

16 de juny de 2015



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>	<i>Construcció i explotació d'un magatzem de dades del sector turístic a Catalunya</i>
<b>Nom de l'autor:</b>	<i>Enric Colomé Campos</i>
<b>Nom del consultor:</b>	<i>Carles Llorach Rius</i>
<b>Data de lliurament (mm/aaaa):</b>	<i>06/2015</i>
<b>Àrea del Treball Final:</b>	<i>Magatzem de dades</i>
<b>Titulació:</b>	<i>Grau d'Enginyeria Informàtica</i>

**Resum del Treball**

Aquest document és la memòria del treball de final de carrera del grau d'Enginyeria Informàtica a la UOC. L'objectiu d'aquest projecte és atendre les demandes de l'empresa contractant, GLTF (Grup Líder en Turisme Familiar). Aquesta empresa pertany al sector turístic i el seu àmbit d'actuació és centra en el territori català, on gestiona una xarxa d'establiments. L'objecte d'estudi són els moviments i estades dels viatgers, relacionat amb els equipaments, l'oferta i tipologia dels establiments, i la percepció de seguretat als municipis catalans.

El treball s'estructura, en termes generals, al voltant de les tres tasques en què han consistit les PACs (proves d'avaluació continuada): el pla de treball del projecte, el disseny i anàlisi i la implementació. Així doncs, es procedeix, en base a les necessitats del client, a fer una anàlisi preliminar de les dades i la planificació del projecte. Tot seguit, es detalla el disseny físic i lògic del magatzem de dades i es defineix com es farà el procés de transformació i càrrega de dades. Per últim, es detalla la fase d'implementació del magatzem de dades i l'elaboració dels informes previstos.

Les dades originals es troben en format *csv* i *xls*, raó per la qual el tractament de correcció i transformació de dades s'ha fet des del full de càlcul Excel aprofitant els recursos que ofereix el llenguatge *vbexcel* (versió de Visual Basic per a Excel). La càrrega de les dades i la creació dels informes s'han fet en l'entorn de la suite Pentaho i l'eina Workbench del gestor de base de dades MySQL.

## **Abstract**

This document is the final study report of the degree on Computer Engineering, taken at the UOC university. The aim of this project is to meet the demands of the contracting company, called GLTF (Group Leader in Familiar Tourism). This company belongs to the touristic sector and its scope is focused on the Catalan region, where it operates a network of resorts. The object of this study is the travellers' trips and stays, taking into account the facilities of the area, accommodation and types of resorts, and the perception of safety. All these elements are applied to Catalan municipalities.

The work is broadly structured around three tasks, which constituted the PACs (Continuous Assessment Tests): the project's work plan, analysis and design, and implementation. Once we know the customer's needs, the analysis of data and the planning of the project are carried out. Straightaway, the physical and logical design of the data warehouse is detailed, and the loading process is defined. The last step is to implement the data warehouse and to develop the expected reports.

The original data were saved in csv and xls format, and this is the reason why their correction and transformation has been done from the Excel spreadsheet, using the resources offered by the vbexcel (version of Visual Basic for Excel) language. The data loading process and the creation of the reports have been done in the environment of the Pentaho suite and the MySQL database manager.

## **Paraules clau**

Magatzem de dades, Data warehouse, ETL, fets, dimensions, Pentaho, MySql

## Índex

1. Introducció.....	1
1.1 Context i justificació del Treball .....	1
1.1.1 Justificació del projecte .....	1
1.1.2 Descripció del projecte.....	1
1.2 Objectius del Treball.....	2
1.2.1 Objectius generals .....	2
1.2.2. Objectius específics .....	2
1.3 Enfocament i mètode seguit .....	3
1.4 Planificació del Treball.....	5
1.4.1 Proposta d'activitats i cronograma .....	5
1.4.2 Fites a complir .....	7
1.4.3 Anàlisi de riscos .....	7
1.4.4 Diagrama de Gantt.....	9
1.5 Breu sumari de productes obtinguts .....	11
1.6 Breu descripció dels altres capítols de la memòria .....	11
2. Anàlisi i disseny .....	12
2.1 Requeriments funcionals / no funcionals .....	12
2.1.1 Requeriments funcionals.....	12
2.1.2 Requeriments no funcionals.....	13
2.1.3 Anàlisi de les fons de dades .....	14
2.2 Disseny conceptual .....	29
2.2.1 Taules de fets .....	29
2.2.2 Granularitat .....	31
2.2.3 Dimensions .....	31
2.2.4 Restriccions d'integritat.....	31

2.2.5 Viabilitat .....	32
2.3 Disseny de la BD .....	32
2.3.1 Diagrama E-R .....	32
2.3.2 Disseny lògic.....	33
2.3.3 Disseny físic.....	34
2.4 Procés d'ETL a alt nivell. ....	35
2.4.1 Extracció de dades .....	35
2.4.2 Transformació de dades .....	36
3. IMPLEMENTACIÓ.....	45
3.1 Arquitectura .....	45
3.2 Càrrega de fitxers .....	45
3.2.1 Fitxers de dades .....	46
3.2.2 Fitxers de dimensió i de fets.....	48
3.3 Informes .....	55
3.3.1 Disseny lògic dels informes.....	55
4. Conclusions.....	63
5. Glossari .....	64
6. Bibliografia.....	65
7. Annexos .....	66
7.1 Annex I .....	66
7.2 Annex II .....	74

## Llista de figures

Il·lustració 1 Fases del cicle de vida d'un SI.....	3
Il·lustració 2 Cicle de vida del projecte .....	4
Il·lustració 3 Enfocament de Kimball. ....	4
Il·lustració 4 Taula d'activitats i cronograma.....	7
Il·lustració 5 Diagrama de Ishikawa de riscos .....	7
Il·lustració 6 Diagrama de Gantt.....	10
Il·lustració 7 Llista d'informes prefixats.....	12
Il·lustració 8 Casos d'ús .....	13
Il·lustració 9 Taula viabilitat dimensions .....	32
Il·lustració 10 Taula viabilitat fets .....	32
Il·lustració 11 Diagrama E-R .....	33
Il·lustració 12 Esquema del disseny lògic.....	33
Il·lustració 13 Columnes de parts de categoria a equipaments .....	37
Il·lustració 14 Nivells 1,2,3 de categoria a equipaments .....	37
Il·lustració 15 Dades finals d'equipaments .....	37
Il·lustració 16 Transformació d'infraestructura.....	38
Il·lustració 17 Agrupació de dades a infraestructura .....	39
Il·lustració 18 Taula d'efectius de la policia local.....	40
Il·lustració 19 Correcció de l'estructura a pernoctacions .....	41
Il·lustració 20 Estimació de les dades que falten a pernoctacions .....	41
Il·lustració 21 Transposició de files a pernoctacions .....	42
Il·lustració 22 Annexió de dades i inserció de files a pernoctacions.....	42
Il·lustració 23 Valors bàsics de les files mensuals a pernoctacions .....	42
Il·lustració 24 Adaptació de la fórmula mensual a pernoctacions.....	43
Il·lustració 25 Arquitectura.....	45
Il·lustració 26 Captura del fitxer csv amb l'Spoon.....	46
Il·lustració 27 Creació de la taula de dades .....	47
Il·lustració 28 Execució de la càrrega de dades .....	47
Il·lustració 29 Comprovació de la creació del fitxer i la càrrega de dades .....	48
Il·lustració 30 Creació de la taula DimData .....	49
Il·lustració 31 Creació d'índex i FK de la taula FetsEstabliments .....	50
Il·lustració 32 Generació de la sentència d'inserció a DimEstabliments.....	52
Il·lustració 33 Inserció de dades a FetsPolisHabitants .....	53
Il·lustració 34 Informe 1 OLAP amb Saiku Analytics .....	56
Il·lustració 35 Informe de % d'ocupació per marca turística .....	57
Il·lustració 36 Gràfica de % d'ocupació per marca turística.....	58
Il·lustració 37 Informe 4 OLAP amb Saiku Analytics .....	59
Il·lustració 38 Informe Top Ten de municipis per pernoctacions .....	60
Il·lustració 39 Gràfic de Top Ten de municipis per pernoctacions .....	61
Il·lustració 40 Informe 1 .....	66
Il·lustració 41 Informe 2.....	66

Il·lustració 42 Informe 3.....	67
Il·lustració 43 Gràfica informe 3.....	67
Il·lustració 44 Informe 4.....	68
Il·lustració 45 Informe 5.....	68
Il·lustració 46 Informe 6.....	69
Il·lustració 47 Gràfica informe 6.....	69
Il·lustració 48 Informe 7.....	70
Il·lustració 49 Informe 8.....	71
Il·lustració 50 Informe 9.....	72
Il·lustració 51 Informe 10.....	73
Il·lustració 52 Gràfica informe 10.....	73
Il·lustració 53 Informe 1 Saiku de pernoctacions estimades per municipi .....	74
Il·lustració 54 Informe 4 Saiku de rànquing de municipis per equipaments.....	75
Il·lustració 55 Informe 9 amb Saiku de rati de policies per habitant .....	75
Il·lustració 56 Informe 10 amb OLAP Saiku.....	76
Il·lustració 57 Gràfic de l'informe 10 amb Saiku .....	76



# 1. Introducció

Aquest document s'emmarca dintre les activitats del Treball de Final de Grau (TFG). Concretament, correspon a la memòria final i resumeix els anteriors lliuraments de les diferents fases del projecte: el pla de treball, el disseny i l'anàlisi, i la implantació.

L'objecte del projecte és dissenyar i elaborar un magatzem de dades ([DW](#), de l'anglès *data warehouse*) a partir del material i dels requisits formulats pel client, l'empresa de turisme Grup Líder en Turisme Familiar (GLTM). Amb la creació d'aquest DW es pretén analitzar l'evolució del sector en el decurs dels darrers anys, mitjançant l'accés directe a les dades, amb la creació d'una sèrie d'informes predefinits i amb el disseny d'un model que doni suport a les necessitats de l'usuari.

## 1.1 Context i justificació del Treball

En el context de l'assignatura, amb el treball es pretén mostrar el grau d'aprofitament dels estudis tot integrant amb la seva confecció habilitats i coneixements adquirits al llarg de la carrera.

### 1.1.1 Justificació del projecte

La justificació del projecte obeeix a una doble vesant, la personal i la del propi projecte. Pel que fa a la vesant personal, la raó d'haver triat aquesta opció per al TFG és el meu interès sobre el tema, l'estadística, la mineria de dades i el fenomen [BigData](#).

D'altra part, ja sobre el projecte concret, l'empresa de turisme GLTM ha observat una evolució negativa del volum de pernoctacions en els darrers anys. Per aquesta raó, ha decidit fer un estudi amb l'objectiu d'identificar la causa de la davallada i els motius decisoris dels seus clients.

La creació del magatzem de dades és un pas previ necessari per dur a terme l'anàlisi dels factors que són objecte d'estudi. A més, s'inclou en el treball la creació d'uns informes predefinits que han estat sol·licitats pel client per tal de visualitzar l'estat de les variables considerades.

### 1.1.2 Descripció del projecte

En tant que TFG, el projecte es divideix en tres fases que es corresponen amb les primeres PACs del semestre:

- Pla de treball i anàlisi preliminar de requeriments.
- Anàlisi de requeriments i Disseny conceptual i tècnic.
- Implementació.

Aquest lliurament correspondrà a la quarta PAC, la memòria del projecte. A partir del conjunt de fitxers facilitats per l'empresa GLTF i dels requeriments del projecte, es realitzarà la implementació del magatzem de dades corporatiu. En la primera de les

fases, el pla de treball, es confeccionarà una planificació estimada de les diferents tasques que caldrà realitzar per dur-lo a terme. També es farà un anàlisi preliminar i una breu descripció dels elements identificats i de les dades operacionals, així com un resum descriptiu dels informes previstos.

En la fase següent, l'anàlisi de requeriments i disseny conceptual i tècnic, es confeccionarà i lliurarà un document amb l'anàlisi detallat dels elements descrits en l'apartat anterior. També es lliurarà un document del disseny utilitzat amb la descripció del model dimensional que donarà suport a les necessitats dels usuaris, així com el disseny a alt nivell dels procediments d'extracció de dades.

En la tercera fase, es durà a terme la implementació del magatzem de dades. Aquest apartat constarà de les següents tasques:

- Construcció del magatzem de dades.
- Instal·lació de l'eina d'exploració de dades.
- Construcció dels informes i anàlisi de la informació.

Per acabar, es confeccionarà una memòria final del projecte i se'n farà una presentació virtual.

## 1.2 Objectius del Treball

Els objectius del treball es presenten des de l'òptica general i des de l'específica d'aquest projecte en concret.

### 1.2.1 Objectius generals

Segons consta en el pla d'estudis, l'objectiu principal de l'assignatura és desenvolupar un projecte que permeti ampliar els coneixements adquirits en les assignatures de bases de dades. En el context d'aquest treball, l'objectiu és adquirir experiència en el disseny, construcció i explotació d'un magatzem de dades a partir de la informació disponible en bases de dades transaccionals de diferents formats.

Des de l'òptica del client, el magatzem de dades ha de servir, bàsicament, per a ajudar a la presa de decisions, segmentar el mercat en funció de les dades i els objectius del negoci i contribuir al coneixement de l'empresa i de l'entorn.

### 1.2.2. Objectius específics

Dintre del projecte es contemplen els següents objectius:

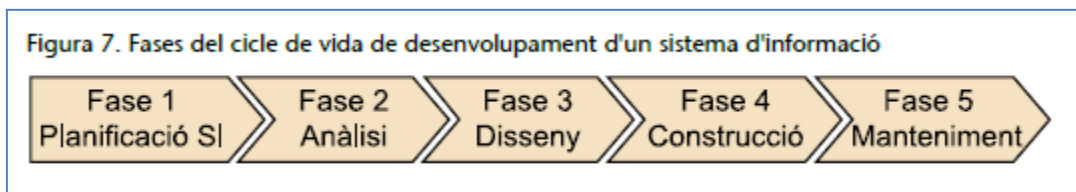
- Dissenyar correctament un magatzem de dades.
- Preparar correctament l'entorn de treball amb l'arquitectura proposada.
- Ser capaç de planificar i executar el procés de selecció, depuració i càrrega de dades de diversos orígens i amb diferents formats.
- Ser capaç de fer-ne el disseny multidimensional i utilitzar eines del tipus [OLAP](#) (*On-Line Analytical Processing*).

- Dissenyar, elaborar i presentar els informes que es demanen.
- Confeccionar la memòria final del projecte i fer-ne una presentació virtual.

### 1.3 Enfocament i mètode seguit

Atès que les dades que ens ha fet arribar el client són formats [csv](#) i [xls](#), s'ha decidit fer el tractament inicial amb el full de càlcul [Excel](#). A més, alguns dels fitxers ens han arribat amb moltes errades i en formats gens estructurats, cosa que dificulta enormement l'automatització de la transformació de les dades. Altres consideracions que hem tingut en compte per seguir aquest procediment inicial han estat que algunes fonts emprades en el treball ara són inaccessibles, que el treball és únic (no es tracta de dades generades dintre l'empresa, sinó que provenen íntegrament de fons externes), i les facilitats de treballar en l'entorn *vba excel* (Visual Basic per a Excel) i l'experiència en aquest llenguatge. Pel que fa al treball amb la base de dades i l'elaboració i publicació dels informes, hem escollit l'eina d'ETL (*Extract, Transform and Load*) Spoon i Report Designer, tots dos de l'entorn [Pentaho](#).

En termes generals, el treball ha seguit el mètode habitual de desenvolupament d'un projecte TIC. Les fases del cicle de vida del projecte es corresponen, aproximadament, amb les fases del cicle de vida (*SDLC, Systems development life cycle*) de desenvolupament d'un sistema d'informació (Ramon, 2010):

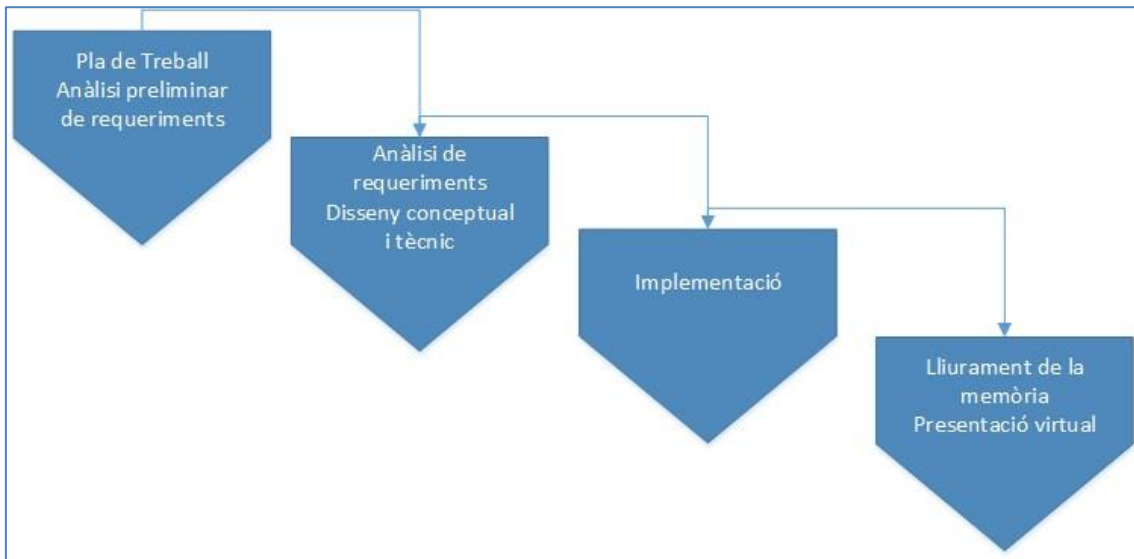


II-Il·lustració 1 Fases del cicle de vida d'un SI.  
Font: Ramon, 2012. Gestió de Projectes. UOC

En el nostre cas, les fases del projecte han vingut marcades pels lliuraments de les PACs:

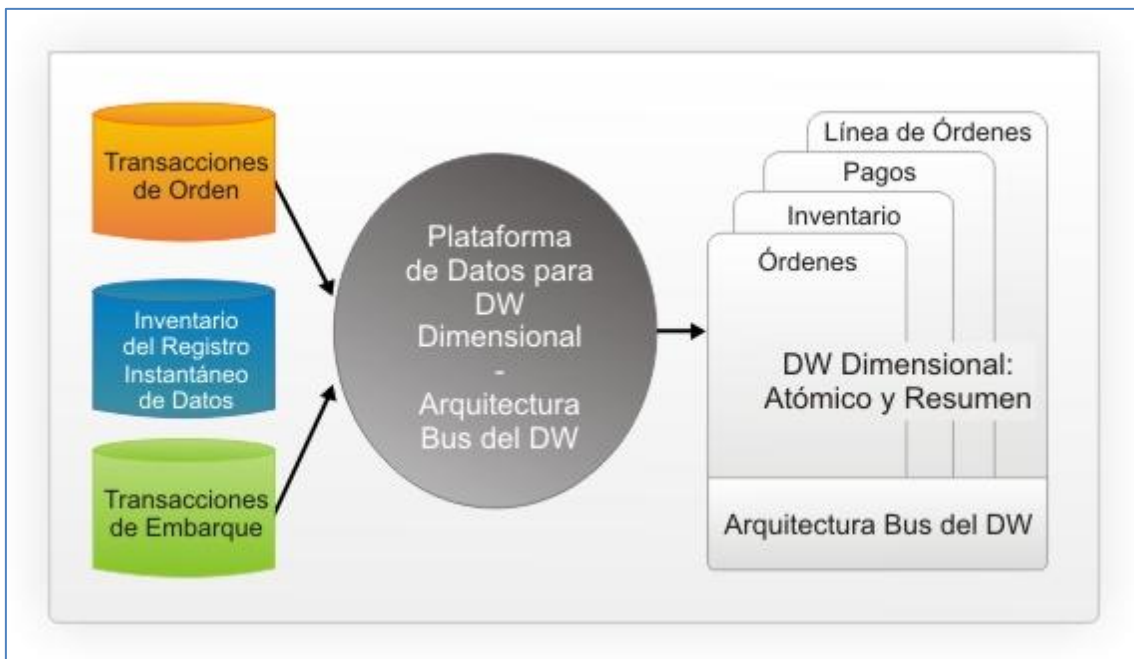
- Pla de treball i l'anàlisi preliminar de requeriments. Amb els requisits del client, es planifica el treball, es confecciona un pla de riscos i s'elabora el cronograma d'actuacions previstes.
- Anàlisi de requeriments i disseny conceptual i lògic. Tenint en compte els objectius i els requeriments del projecte, es dissenya el model de dades conceptual i lògic que determinaran el model relacional en la fase d'implementació. En aquesta fase queda instal·lat el programari (en el nostre cas, la MV Amazon amb la suite Pentaho i [MySql](#)) i s'inicia el procés d'ETL amb el tractament i transformació de les dades previ a la càrrega que es durà a terme en la fase d'implementació.
- Implementació. En aquest punt es crea el magatzem de dades: es carreguen les dades ja transformades en l'etapa anterior i s'afegeixen als fitxers segons el disseny relacional creat. Amb aquestes dades es podran crear i presentar els informes demanats.

- Lliurament de la memòria i presentació virtual. El treball es documenta convenientment per a lliurar la corresponent PAC i se'n confecciona una presentació que servirà de base per la defensa del projecte.



II-lustració 2 Cicle de vida del projecte

Per la creació del DW s'ha escollit l'enfocament de Kimball, anomenat [model dimensional](#). Aquest model inclou les dimensions de l'anàlisi i els seus atributs, els fets que es volen estudiar i les mesures associades i la relació entre les diferents taules. En aquest paradigma, el magatzem de dades és, en paraules de Kimball, "una còpia de las transaccions de dades específicament estructurada per a la consulta i l'anàlisi". Aquest és un model descendent (*Bottom-up*), en el sentit que parteix de les taules de dimensions i fets per a configurar el DW. Així doncs, el magatzem "és la unió de tots els Data Marts d'una entitat".



II-lustració 3 Enfocament de Kimball.  
Font: Espinosa, Roberto. [El rincón del BI](#)

## 1.4 Planificació del Treball

En aquest apartat es mostra el calendari confeccionat i el diagrama de Gantt associat a les activitats, amb les principals fites del projecte.

### 1.4.1 Proposta d'activitats i cronograma

En la taula següent es detalla la relació d'activitats i l'estimació de temps en dies. Tot i que la variabilitat és molt alta, es pot prendre com indicador una dedicació de 2 a 3 hores diàries.

Id	Nom de la Tasca	d	Inici	Fi	Notes
1	<b>Inici</b>	<b>2</b>	<b>25/02/15</b>	<b>26/02/15</b>	
1.1	Presentació	1	25/02/15	25/02/15	
1.2	Lectura del pla docent	1	25/02/15	25/02/15	
1.3	Descàrrega dels materials	1	26/02/15	26/02/15	Els materials de la web no són els corresponents a aquest semestre. Resta pendent.
2	<b>PAC1</b>	<b>13</b>	<b>27/02/15</b>	<b>11/03/15</b>	<b>Confecció del Pla de Treball</b>
2.1	Descàrrega de l'enunciat i els materials	1	27/02/15	27/02/15	
2.2	Descàrrega dels exemples	1	02/03/15	02/03/15	
2.3	Determinar l'estructura del treball	1	03/03/15	03/03/15	
2.4	Lectura material docent de l'assignatura	9	03/03/15	11/03/15	Introducció i mòduls 1 i 2
2.5	Lectures complementàries	8	03/03/15	11/03/15	Gestió de projectes, Competències Comunicatives
2.6	<b>Redacció del Pla de treball</b>	<b>9</b>	<b>03/03/15</b>	<b>11/03/15</b>	<b>Correspon a la PAC1</b>
2.6.1	Primera part	4	03/03/15	06/03/15	Introducció, Justificació, Objectius, Requeriments, Resultats
2.6.2	Segona part	3	07/03/15	10/03/15	Organització, Activitats i Cronograma, Conclusions, Bibliografia i Annexos
2.6.3	<i>Revisió i Lliurament</i>	<i>1</i>	<i>11/03/15</i>	<i>11/03/15</i>	<i>FITA</i>
3	<b>PAC2</b>	<b>35</b>	<b>12/03/15</b>	<b>15/04/15</b>	<b>Anàlisi i Disseny</b>
3.1	Descàrrega de l'enunciat i materials	1	12/03/15	12/03/15	
3.2	Determinar l'estructura del treball	1	13/03/15	13/03/15	
3.3	Lectura material docent de l'assignatura	34	12/03/15	14/04/15	Lectura dels mòduls 3 i 4 de l'assignatura

3.4	Lectures complementàries	34	12/03/15	14/04/15	Us de bases de dades, Disseny de BBDD, Minería de dades
3.5	<b>Elaboració de l'Anàlisi i Disseny</b>	<b>35</b>	<b>12/03/15</b>	<b>15/04/15</b>	<b>Correspon a la PAC2</b>
3.5.1	Primera part	1	13/03/15	15/03/15	Introducció i Requeriments
3.5.2	Segona part	12	16/03/15	28/03/15	Disseny BD, diagrama E-R, Model multidimensional i Informes
3.5.3	Tercera part	12	29/03/15	11/04/15	Procés ETL, automatització ETL i tractament d'errors en la càrrega
3.5.4	Quarta part	1	12/04/15	14/04/15	Revisió, conclusions i bibliografia
3.5.5	<i>Revisió i Lliurament</i>	<i>1</i>	<i>15/04/15</i>	<i>15/04/15</i>	<i>FITA</i>
4	<b>PAC3</b>	<b>43</b>	<b>16/04/15</b>	<b>28/05/15</b>	<b>Implementació</b>
4.1	Descàrrega de l'enunciat i materials	1	16/04/15	16/04/15	
4.2	Instal·lació de l'entorn virtual	1	16/04/15	16/04/15	
4.3	Lectura material docent de l'assignatura	14	17/04/15	28/05/15	Lectura del mòdul 5 de l'assignatura. Revisió dels mòduls anteriors.
4.4	Lectures complementàries	41	17/04/15	28/05/15	Revisió de lectures anteriors
4.5	<b>Implementació</b>	<b>43</b>	<b>16/04/15</b>	<b>28/05/15</b>	<b>Correspon a la PAC3</b>
4.5.1	Primera part	4	18/04/15	21/04/15	Implementació de la base de dades
4.5.2	Segona part	13	23/04/15	06/05/15	Implementació del procés ETL i automatització
4.5.3	Tercera part	13	07/05/15	20/05/15	Implementació de l'eina OLAP i els informes
4.5.4	Quarta part	5	21/05/15	26/05/15	Explicació dels passos seguits amb captures de pantalla, comentaris, suposicions fetes, justificació i MV preparada
4.5.5	<i>Revisió i Lliurament</i>	<i>1</i>	<i>27/05/15</i>	<i>27/05/15</i>	<i>FITA</i>
5	<b>PAC4</b>	<b>19</b>	<b>29/05/15</b>	<b>16/06/15</b>	<b>Memòria i Presentació</b>
5.1	Descàrrega de l'enunciat i materials	1	29/05/15	29/05/15	
5.2	Lectura del material docent de l'assignatura	18	29/05/15	16/06/15	Redacció de textos, Presentació de documents i elaboració de presentacions
5.3	Revisió i Integració dels documents anteriors	3	30/05/15	02/06/15	Documents lliurats en les PAC1, PAC2 i PAC3
5.4	<b>Redacció de la memòria final</b>	<b>3</b>	<b>03/06/15</b>	<b>07/06/15</b>	<b>Redefinir l'estructura, corregir l'estil, afegir treball futur, conclusions, bibliografia, taules i annexos. Correspon a la PAC4</b>

5.5	Elaboració de la presentació	5	08/06/15	16/06/15	Disseny de la presentació, elements i estil. Índex, descripció, plantejament del problema, desenvolupament del projecte, objectius aconseguits, demostració, millores, conclusions i bibliografia. Correspon a la PAC4
5.6	Revisió i Lliurament	1	16/06/15	16/06/15	FITA
6	Debat Virtual	5	29/06/15	03/07/15	FITA

II-lustració 4 Taula d'activitats i cronograma

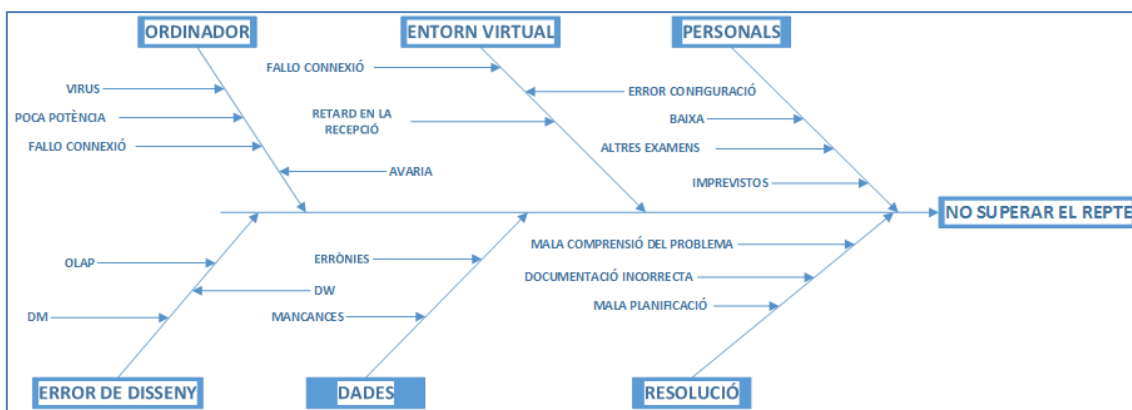
A la pràctica la programació s'ha vist alterada pel dia a dia, però en conjunt s'ha mantingut la tònica senyalada a la taula. Caldria, potser, afegir un temps suplementari de recerca en fòrums i webs sobre alguns aspectes o funcionalitats del programari.

### 1.4.2 Fites a complir

- 11/03/2015 Lliurament de la PAC1: Pla de Treball
- 15/04/2015 Lliurament de la PAC2: Anàlisi i disseny
- 28/05/2015 Lliurament de la PAC3: Implementació
- 16/06/2015 Lliurament de la PAC4: Memòria i Presentació
- 03/07/2015 Defensa del projecte

### 1.4.3 Anàlisi de riscos

S'han exclòs els riscos de la part del client. Aquest podrien ser derivats d'una infraestructura poc adequada, per falta de competències per a utilitzar la solució, per una definició ambigua o incompleta del problema que es vol solucionar, etc. S'ha fet la suposició que aquests riscos estan controlats i que no depenen de nosaltres.



II-lustració 5 Diagrama de Ishikawa de riscos

Risc 1	ORDINADOR
Descripció	Problemes amb l'entorn de l'ordinador amb què es farà el Treball. Pot ser per algun atac, per manca de potència per a la tasca plantejada, per avaria d'algun component o per fallades en la connexió a Internet

Impacte	Termini d'acabament del projecte
Probabilitat	Baixa
Acció de mitigació	Revisió o reparació. Si no és possible, substitució. Si la fallada és de connexió, treballar en un altre punt amb connexió mentre es soluciona el problema. Portar al dia les còpies de seguretat.

Risc 2.1	ENTORN VIRTUAL
Descripció	Fallades contínues en la connexió amb el servidor.
Impacte	Lliurament del projecte
Probabilitat	Mitja
Acció de mitigació	Intentar instal·lar les eines en entorn local i demanar una solució al consultor de l'assignatura.

Risc 2.2	ENTORN VIRTUAL
Descripció	El material per a preparar l'entorn no arriba a temps.
Impacte	Termini d'acabament del projecte
Probabilitat	Molt baixa
Acció de mitigació	Cercar el material per altres vies i demanar una solució al consultor de l'assignatura.

Risc 2.3	ENTORN VIRTUAL
Descripció	Configuració incompatible amb l'entorn local.
Impacte	Lliurament del projecte
Probabilitat	Molt baixa
Acció de mitigació	Intentar instal·lar les eines en entorn local i demanar una solució al consultor de l'assignatura.

Risc 3.1	PERSONALS
Descripció	Situacions personals que dificultin o, fins i tot, impedeixin seguir el ritme de treball satisfactòriament. Pot ser per qüestions de salut (malaltia o accident) o d'altres tipus de problemes que requereixin atenció exclusiva.
Impacte	Termini d'acabament del projecte
Probabilitat	Baixa
Acció de mitigació	Solucionar el problema i buscar la manera de recuperar el temps perdut, ja sigui invertint-hi més temps, ja sigui deixant alguna altra tasca menys prioritària.



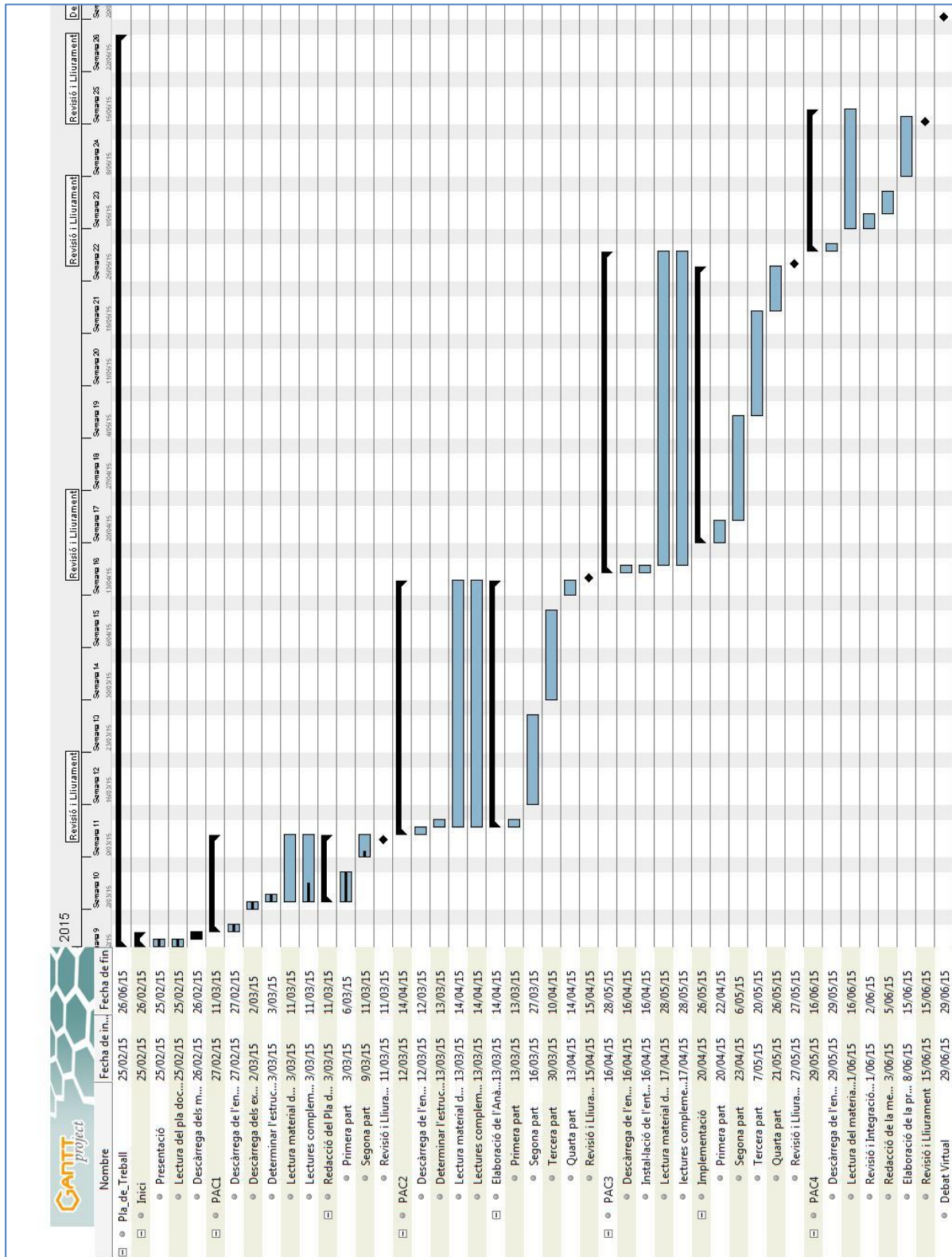
Risc 3.2	PERSONALS
Descripció	Hi ha d'altres assignatures –o tasques- que requereixen una atenció per damunt del previst, tal com un examen o el lliurament d'una pràctica final.
Impacte	Termini d'acabament del projecte
Probabilitat	Mitja
Acció de mitigació	Redistribuir el temps d'estudi per perdre el mínim temps possible i recuperar quan més aviat millor el temps perdut.

Risc 4	ERRORS DE DISSENY
Descripció	Dissenyar incorrectament el magatzem de dades, el cub OLAP o el resum DM.
Impacte	Implementació
Probabilitat	Baixa
Acció de mitigació	Consultar freqüentment els dubtes que puguin aparèixer i contrastar cada etapa del disseny amb el consultor de l'assignatura.

Risc 5	DADES
Descripció	Dades irrecuperables (o inexistents) des de la font o fitxers amb un format incomprensible.
Impacte	Disminució de la quantitat de dades disponibles per a l'anàlisi. L'impacte variarà en funció de la quantitat i importància de les dades irrecuperable.
Probabilitat	Molt baixa
Acció de mitigació	Intentar recuperar les dades des d'una altra font. Demanar una solució al consultor.

Risc 6.1	RESOLUCIÓ
Descripció	Problemes propis en la resolució del projecte, tals com mala comprensió del problema o mala planificació del temps disponible.
Impacte	Lliurament del projecte
Probabilitat	Baixa
Acció de mitigació	Portar un seguiment diari del progrés del treball i consultar sense demora els dubtes amb el consultor.

#### 1.4.4 Diagrama de Gantt



II-lustració 6 Diagrama de Gantt

## 1.5 Breu sumari de productes obtinguts

Al final del treball s'haurà generat un model de dades que permetrà obtenir informació sobre les variables d'estudi. Alhora, es deixaran predefinits una sèrie d'informes amb les consultes plantejades pel client.

En cada una de les fites temporals, el projecte ha generat els lliurables que han conformat les PACs de l'assignatura. Així doncs, al final de la primera fase s'haurà generat el pla de treball, document que conté la previsió de totes les tasques que s'hauran de fer, el pla de riscos i el material que s'haurà de produir.

En la segona PAC s'haurà lliurat un document amb l'anàlisi de requeriments funcionals i no funcionals i les fonts de dades que aporta l'empresa per a dur a terme el treball; també s'ha dut a terme en aquesta PAC l'anàlisi del model de dades -conceptual i físic- i els processos de transformació de dades previs a la càrrega al magatzem.

Del tercer lliurament s'ha obtingut el producte final: s'ha construït el magatzem de dades i s'han elaborat i mostrat els informes sol·licitats.

En el quart lliurament, el document obtingut és la present memòria, on es resumeix tot el procés del treball realitzat.

Per últim, es prepara una presentació virtual per mostrar les parts més significatives d'aquest projecte i l'informe de competències transversals.

## 1.6 Breu descripció dels altres capítols de la memòria

Els altres capítols de la memòria són:

- Anàlisi i disseny. En aquest capítol es completa l'anàlisi preliminar de requeriments, tant funcionals com no funcionals, es mostren els casos d'ús, s'analitza l'estructura i contingut dels fitxers font, es fa el disseny conceptual (determinació dels fets i les dimensions) i el disseny de la base de dades (diagrama E-R, i esquema lògic i físic), i s'explica el procés de transformació de dades.
- Implementació. En aquest capítol es mostra de l'arquitectura del sistema, en completa el procés d'ETL amb la càrrega de les dades a la base de dades i es creen i mostren, amb captures de pantalla, els informes predefinits.

## 2. Anàlisi i disseny

Aquest apartat es divideix en dos parts, l'anàlisi de requisits i el disseny del model de dades.

### 2.1 Requeriments funcionals / no funcionals

En aquest apartat s'analitzaran els dos components que fonamenten el treball: les dades originals i els lliurables. Les dades són les proporcionades per l'empresa GLTF i les que, al nostre criteri, creuem oportunes per millorar l'obtenció de dades. Val a dir que aquestes dades provenen exclusivament de bases de dades externes, és a dir, no generades dintre l'empresa. Els fitxers amb les dades presenten diferents formats, estructures i problemàtiques que es detallaran en un apartat d'anàlisi de les dades.

Els requeriments funcionals es refereixen a com interactua el sistema (què fa), mentre que els requisits no funcionals imposen restriccions a la manera de funcionar (com ho fa). Així, aquest capítol el dividirem en quatre parts: els requeriments funcionals (requisits funcionals: requisits de funcionalitat), casos d'ús, els requeriments no funcionals (requisits no funcionals) i l'anàlisi de dades (requisits funcionals: requisits de dades).

.

#### 2.1.1 Requeriments funcionals

L'usuari ha de poder obtenir, interactuant amb el sistema, tot un seguit d'informes i ha de tenir, també, accés a les dades per al seu procés (seria el cas de l'administrador del sistema) i per a la seva consulta, ja sigui directa o a través d'alguna eina que ofereixi el model dimensional que es dissenyarà en el corresponent apartat. En l'encàrrec de l'empresa es detalla que, com a mínim, del magatzem de dades s'ha de poder obtenir aquesta informació:

- Total de pernoctacions estimades per municipi
- Promig de viatgers per tipus d'establiment i comarca
- % d'ocupació per marca turística
- Rànquing de municipis per categoria d'equipaments
- Màxim i mínim d'efectius policial per tipologia d'establiment i municipi
- "Top ten" de municipis per franja de pernoctacions
- % de pernoctacions d'un municipi sobre el total
- Categorització de municipis A/B/C
- Rati de policies locals per habitant
- Distribució mensual i estacionalitat de les pernoctacions

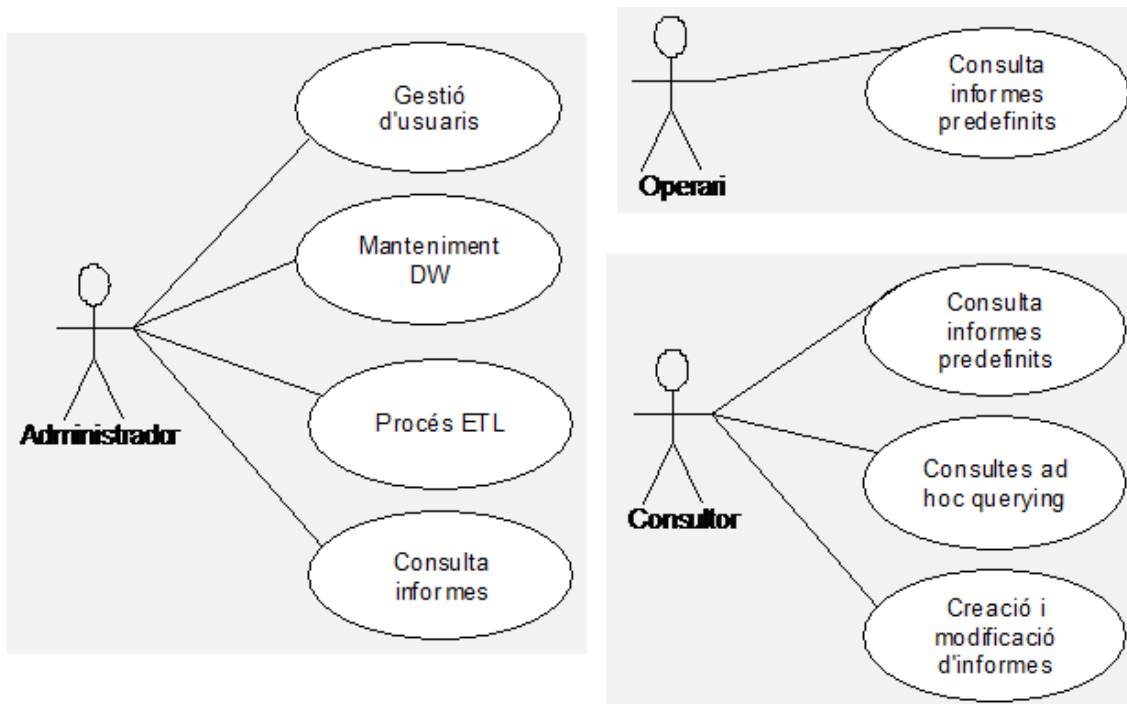
#### II-l·lustració 7 Llista d'informes prefixats

Tota aquesta informació es proporcionarà, sempre que sigui possible, dintre d'una temporalitat a nivell de mes i any. Es podrà consultar de forma agregada, per demarcació territorial i tipus d'establiment. Així, el conjunt d'informes prefixat que s'ha de proporcionar serà, en principi, el corresponent per a mostrar la informació sol·licitada. En el decurs de l'anàlisi es pot incorporar algun altre que pugui ser útil per a l'empresa.

### 2.1.1.1 Casos d'ús

Malgrat que en l'enunciat no es comenta, és convenient definir nivells d'accés pel tipus d'usuari. Es plantegen tres tipus d'usuari:

- Operari. Consulta els informes predefinitos.
- Consultor. A més de consultar els informes predefinitos, pot fer consultes *ad hoc querying* directament a les dades. Pot crear informes nous i modificar els existents.
- Administrador. A més de fer consultes i crear o modificar informes, és el responsable del manteniment del magatzem de dades, de la gestió dels perfils d'usuaris i del procés d'ETL.



Il·lustració 8 Casos d'ús

### 2.1.2 Requeriments no funcionals

Tot i que de l'enunciat no es desprenen restriccions, pot ser convenient considerar alguns requisits de qualitat del sistema, com ara:

<b>Requisit</b>	El sistema ha de ser segur per als usuaris
<b>Descripció</b>	El sistema ha de ser segur per als usuaris, amb els nivells d'accés ben definits i protegits
<b>Tipus</b>	Seguretat
<b> criteris d'acceptació</b>	Un usuari no ha de poder accedir a nivells no autoritzats

<b>Requisit</b>	El temps de resposta ha de ser proporcionat
<b>Descripció</b>	El temps de resposta ha de ser proporcional a la mida de la consulta
<b>Tipus</b>	Compliment - Rendiment

<b>Criteris d'acceptació</b>	En cap cas ha de ser superior a 1 segon per a consultes senzilles
------------------------------	---

<b>Requisit</b>	El sistema ha d'estar sempre disponible
<b>Descripció</b>	El sistema ha d'estar sempre disponible mentre hi hagi connexió
<b>Tipus</b>	Compliment - Disponibilitat
<b>Criteris d'acceptació</b>	Mentre hi hagi connexió, l'usuari ha de poder accedir a les dades sense més restriccions que les de seguretat al menys el 99% de les vegades

<b>Requisit</b>	La interfície ha de ser fàcil d'entendre i d'utilitzar
<b>Descripció</b>	La interfície ha de ser fàcil d'entendre i d'utilitzar per a un usuari sense experiència
<b>Tipus</b>	Usabilitat
<b>Criteris d'acceptació</b>	Es faran proves per a comprovar que a nivell d'usuari no ha de requerir gaire -o gens- aprenentatge previ, o en tot cas no superior a 2 sessions

<b>Requisit</b>	El sistema ha de permetre usuaris concurrents
<b>Descripció</b>	El sistema ha de permetre l'ús concurrent de la base de dades a un número suficient d'usuaris
<b>Tipus</b>	Compliment - Concurrencia
<b>Criteris d'acceptació</b>	El nombre d'accessos concurrents que ha de suportar el sistema no ha de ser inferior a 50

<b>Requisit</b>	El sistema ha de respondre als errors
<b>Descripció</b>	El sistema no ha de patir caigudes com a resultat d'errors de dades o accions dels usuaris
<b>Tipus</b>	Compliment - Robustesa
<b>Criteris d'acceptació</b>	Es faran proves d'accions i càrregues de dades incorrectes i les caigudes no han de ser superior al 1%

### 2.1.3 Anàlisi de les fons de dades

Els fitxers de dades que aporta l'empresa són:

- Infraestructura turística.
- Equipaments.
- Pernotacions.
- Polícies locals.

A part, s'han importat els fitxers:

- Població per municipi per any.
- Marques turístiques – comarques.
- Comarques i municipis.

També s'han generat el fitxer:

- Data, amb els camps d'any, mes o estacionalitat.

### 2.1.3.1 Anàlisi dels fitxers (descripció general)

En la següent taula es mostra la informació disponible sobre els fitxers origen de les dades.

equipaments.csv

<b>fitxer</b>	equipaments.csv
<b>format</b>	text pla, extensió csv amb els valors separats per comes; els registres es separen amb salt de línia LF; el camp d'adreça va entre cometes dobles quan hi ha coma en el text
<b>descripció fitxer</b>	agrupament de dades provinents de diferents taules relatives als municipis de Catalunya i els seus equipaments
<b>comentaris</b>	no s'ha pogut generar de vell nou el fitxer; la primera importació s'ha fet a un full de càlcul d'Excel, amb la intenció de visualitzar l'estructura i localitzar errades en les dades
<b>generat</b>	probablement a primers d'any
<b>origen</b>	IDESCAT - Institut d'Estadística de Catalunya
<b>estructura</b>	estructura tabular; conté 31772 files i 10 columnes
<b>mida</b>	8765kb
<b>descripció contingut</b>	conté la relació dels equipaments dels municipis de Catalunya, amb dades de contacte i localització i categoria d'equipament, aquesta última agrupada en un sol camp
<b>any</b>	probablement amb les dades de 2013
<b>resum de camps</b>	<i>nom, adreca, municipi, comarca</i> : text
	<i>codi postal, telèfon, longitud, latitud</i> : text amb format; dígit numèrics
	<i>location</i> : text amb format xml
	<i>categories</i> : text amb format; separador vertical
<b>notes</b>	Al camp de municipi es separa amb un espai el nom de l'article; hi ha errors en la latitud i longitud, en el nom, en les adreces, en el codi postal, en el camp <i>location</i> i valors nuls en el telèfon; els valors de comarca no porten article.

infraestructura turistica.xls

<b>fitxer</b>	infraestructura turistica.xls
<b>format</b>	Fitxer de full de càlcul Excel, d'extensió xls, compatible amb les darreres versions d'Office i OpenOffice, amb un full d'índex i 18 fulls de dades.
<b>descripció fitxer</b>	Conté un inventari d'establiments turístics i el nombre de places disponibles dels municipis de Catalunya, classificats per tipus d'establiment i categoria, dels anys 2011 a 2013.
<b>comentaris</b>	Per no carregar massa el document, presentem l'enllaç d'un sol any, el 2013
<b>generat</b>	23/02/2015
<b>origen</b>	IDESCAT, a partir de les dades del Departament d'Empresa i Ocupació.
<b>estructura</b>	Inclou 18 fulls de dades i un d'índex; cada 6 fulls corresponen a un tipus d'establiment; cada tres fulls corresponen als anys 2013, 2012 i 2011: el primer grup de 3 amb el nombre d'establiments per tipus i categoria i el segon grup de 3 amb el nombre de places per a aquest tipus i categoria d'establiment; cada grup de 3 fulls presenta una estructura comuna i coherent amb la resta de grups.
<b>mida</b>	1457kb
<b>notes</b>	El full índex conté errors; al camp de municipi es separen amb una coma i un espai

	el nom de l'article.
	<b>Fulls 1-3 (cont.)</b>
<b>descripció contingut</b>	establiments hotelers per categoria a cada municipi, fulls 1-3
<b>link</b>	<a href="http://www.idescat.cat/territ/BasicTerr?TC=20&amp;V0=1&amp;V1=25001&amp;V3=6033&amp;V4=6030&amp;ALLINFO=TRUE&amp;PARENT=100&amp;DISTRI=TRUE&amp;CTX=B">http://www.idescat.cat/territ/BasicTerr?TC=20&amp;V0=1&amp;V1=25001&amp;V3=6033&amp;V4=6030&amp;ALLINFO=TRUE&amp;PARENT=100&amp;DISTRI=TRUE&amp;CTX=B</a>
<b>any</b>	2011-2013
<b>resum de camps</b>	<i>ajuntament</i> : text
	<i>categories hotel (5 entrades)</i> : numèric natural
	<i>total hotels</i> : numèric natural
	<i>hostals i pensions</i> : numèric natural
	<i>total establiments</i> : numèric natural
<b>fila de totals</b>	<i>categories hotel (5 entrades) Catalunya</i> : numèric natural
	<i>total hotels Catalunya</i> : numèric natural
	<i>hostals i pensions Catalunya</i> : numèric natural
	<i>total establiments Catalunya</i> : numèric natural
<b>columnes de totals</b>	<i>total d'hotels per municipi</i> : numèric natural
	<i>total amb hostals i pensions per municipi</i> : numèric natural
<b>estructura</b>	Els tres fulls presenten la mateixa estructura tabular amb 9 columnes, 5 files de descripció, una d'encapçalament, una de separació, 947 files de dades, una en blanc de separació, una de total, i tres amb referència textual a les fonts.
<b>notes</b>	Hi ha files descriptives i de separació que no contenen dades.
	<b>Fulls 4-6 (cont.)</b>
<b>descripció contingut</b>	Places hoteleres per categoria a cada municipi, fulls 4-6
<b>link</b>	<a href="http://www.idescat.cat/territ/BasicTerr?TC=20&amp;V0=1&amp;V1=25001&amp;V3=6033&amp;V4=6031&amp;ALLINFO=TRUE&amp;PARENT=100&amp;DISTRI=TRUE&amp;CTX=B">http://www.idescat.cat/territ/BasicTerr?TC=20&amp;V0=1&amp;V1=25001&amp;V3=6033&amp;V4=6031&amp;ALLINFO=TRUE&amp;PARENT=100&amp;DISTRI=TRUE&amp;CTX=B</a>
<b>any</b>	2011-2013
<b>resum de camps</b>	<i>ajuntament</i> : text
	<i>places categories hotel (5 entrades)</i> : numèric natural
	<i>total places hotels</i> : numèric natural
	<i>places hostals i pensions</i> : numèric natural
	<i>total places</i> : numèric natural
<b>fila de totals</b>	<i>places categories hotel (5 entrades) Catalunya</i> : numèric natural
	<i>places total hotels Catalunya</i> : numèric natural
	<i>places hostals i pensions Catalunya</i> : numèric natural
	<i>places total Catalunya</i> : numèric natural
<b>columnes de totals</b>	<i>total de places a hotels per municipi</i> : numèric natural
	<i>total de places amb hostals i pensions per municipi</i> : numèric natural
<b>estructura</b>	Els tres fulls presenten la mateixa estructura tabular amb 9 columnes, 5 files de descripció, una d'encapçalament, una de separació, 947 files de dades, una en blanc de separació, una de total, i tres amb referència textual a les fonts.
<b>notes</b>	Hi ha files descriptives i de separació que no contenen dades.



	<b>Fulls 7-9 (cont.)</b>
<b>descripció contingut</b>	Càmpings per categoria a cada municipi, fulls 7-9
<b>link</b>	<a href="http://www.idescat.cat/territ/BasicTerr?TC=20&amp;V0=1&amp;V1=25001&amp;V3=6034&amp;V4=6035&amp;ALLINFO=TRUE&amp;PARENT=100&amp;DISTRI=TRUE&amp;CTX=B">http://www.idescat.cat/territ/BasicTerr?TC=20&amp;V0=1&amp;V1=25001&amp;V3=6034&amp;V4=6035&amp;ALLINFO=TRUE&amp;PARENT=100&amp;DISTRI=TRUE&amp;CTX=B</a>
<b>any</b>	2011-2013
<b>resum de camps</b>	<i>ajuntament</i> : text
	<i>categories càmping (4 entrades)</i> : numèric natural
	<i>total càmpings</i> : numèric natural
<b>fila de totals</b>	<i>categories càmping (4 entrades) Catalunya</i> : numèric natural
<b>columnes de totals</b>	<i>total càmpings Catalunya</i> : numèric natural
	<i>total càmpings per municipi</i> : numèric natural
<b>estructura</b>	els tres fulls presenten la mateixa estructura tabular amb 6 columnes, 3 files de descripció, una de separació, una d'encapçalament, 947 files de dades, una en blanc de separació, una de totals, i tres amb referència textual a les fonts
<b>notes</b>	hi ha files descriptives i de separació que no contenen dades
	<b>Fulls 10-12 (cont.)</b>
<b>descripció contingut</b>	places càmping per categoria a cada municipi, fulls 10-12
<b>link</b>	<a href="http://www.idescat.cat/territ/BasicTerr?TC=20&amp;V0=1&amp;V1=25001&amp;V3=6034&amp;V4=6036&amp;ALLINFO=TRUE&amp;PARENT=100&amp;DISTRI=TRUE&amp;CTX=B">http://www.idescat.cat/territ/BasicTerr?TC=20&amp;V0=1&amp;V1=25001&amp;V3=6034&amp;V4=6036&amp;ALLINFO=TRUE&amp;PARENT=100&amp;DISTRI=TRUE&amp;CTX=B</a>
<b>any</b>	2011-2013
<b>resum de camps</b>	<i>ajuntament</i> : text
	<i>places categories càmping (4 entrades)</i> : numèric natural
	<i>total places càmpings</i> : numèric natural
<b>fila de totals</b>	<i>total places categories càmping (4 entrades) Catalunya</i> : numèric natural
<b>columnes de totals</b>	<i>total places càmpings Catalunya</i> : numèric natural
<b>estructura</b>	<i>total places de càmping per municipi</i> : numèric natural
<b>estructura</b>	Els tres fulls presenten la mateixa estructura tabular amb 6 columnes, 3 files de descripció, una d'encapçalament, una de separació, 947 files de dades, una en blanc de separació, una de totals, i tres amb referència textual a les fonts.
<b>notes</b>	Hi ha files descriptives i de separació que no contenen dades.
	<b>Fulls 13-15 (cont.)</b>
<b>descripció contingut</b>	Turisme rural per categoria a cada municipi, fulls 13-15
<b>link</b>	<a href="http://www.idescat.cat/territ/BasicTerr?TC=20&amp;V0=1&amp;V1=25001&amp;V3=6037&amp;V4=6038&amp;ALLINFO=TRUE&amp;PARENT=100&amp;DISTRI=TRUE&amp;CTX=B&amp;ANYS=2013">http://www.idescat.cat/territ/BasicTerr?TC=20&amp;V0=1&amp;V1=25001&amp;V3=6037&amp;V4=6038&amp;ALLINFO=TRUE&amp;PARENT=100&amp;DISTRI=TRUE&amp;CTX=B&amp;ANYS=2013</a>
<b>any</b>	2011-2013
<b>resum de camps</b>	<i>ajuntament</i> : text
	<i>categories rural (4 entrades)</i> : numèric natural
	<i>total turisme rural</i> : numèric natural
<b>fila de totals</b>	<i>categories turisme rural (4 entrades) Catalunya</i> : numèric natural
<b>fila de totals</b>	<i>total turisme rural Catalunya</i> : numèric natural

<b>columnes de totals</b>	<i>total establiments de turisme rural per municipi</i> : numèric natural
<b>estructura</b>	els tres fulls presenten la mateixa estructura tabular amb 6 columnes, 4 files de descripció, una d'encapçalament, una de separació, 947 files de dades, una en blanc de separació, una de totals, i tres amb referència textual a les fonts
<b>notes</b>	hi ha files descriptives i de separació que no contenen dades
	<b>Fulls 16-18 (cont.)</b>
<b>descripció contingut</b>	places turisme rural per categoria a cada municipi, fulls 16-18
<b>link</b>	<a href="http://www.idescat.cat/territ/BasicTerr?TC=20&amp;V0=1&amp;V1=25001&amp;V3=6037&amp;V4=6039&amp;ALLINFO=TRUE&amp;PARENT=100&amp;DISTRI=TRUE&amp;CTX=B&amp;ANYS=2013">http://www.idescat.cat/territ/BasicTerr?TC=20&amp;V0=1&amp;V1=25001&amp;V3=6037&amp;V4=6039&amp;ALLINFO=TRUE&amp;PARENT=100&amp;DISTRI=TRUE&amp;CTX=B&amp;ANYS=2013</a>
<b>any</b>	2011-2013
<b>resum de camps</b>	<i>ajuntament</i> : text
	<i>places categories rural (4 entrades)</i> : numèric natural
	<i>total places turisme rural</i> : numèric natural
<b>fila de totals</b>	<i>total places categories càmping (4 entrades) Catalunya</i> : numèric natural
	<i>total places càmpings Catalunya</i> : numèric natural
<b>columnes de totals</b>	<i>total places de turisme rural per municipi</i> : numèric natural
<b>estructura</b>	els tres fulls presenten la mateixa estructura tabular amb 6 columnes, 4 files de descripció, una d'encapçalament, una de separació, 947 files de dades, una en blanc de separació, una de totals, i tres amb referència textual a les fonts
<b>notes</b>	hi ha files descriptives i de separació que no contenen dades
	<b>Full 19 (cont.)</b>
<b>descripció contingut</b>	full d'índex, full 19
<b>notes</b>	taula d'índex del contingut del fitxer; no conté dades a tractar

polícies locals.xls

<b>fitxer</b>	polícies locals.xls
<b>format</b>	fitxer de full de càlcul Excel, d'extensió xls, compatible amb les darreres versions d'Office i OpenOffice, amb un full d'índex i 3 fulls de dades
<b>descripció fitxer</b>	conté el nombre d'efectius de la policia local per municipi classificats per graduació, dels anys 2011 a 2013
<b>comentaris</b>	per no fer massa espès el document, presentem el link d'un sol any, el 2013
<b>generat</b>	25/02/2015
<b>origen</b>	IDESCAT – Institut d'Estadística de Catalunya
<b>link</b>	<a href="http://www.idescat.cat/territ/BasicTerr?TC=20&amp;V0=3&amp;V1=3&amp;V3=1031&amp;V4=272&amp;P=N&amp;PARENT=25&amp;CTX=B&amp;DISTRI=TRUE&amp;ALLINFO=TRUE&amp;ANYS=2013&amp;x=9&amp;y=9">http://www.idescat.cat/territ/BasicTerr?TC=20&amp;V0=3&amp;V1=3&amp;V3=1031&amp;V4=272&amp;P=N&amp;PARENT=25&amp;CTX=B&amp;DISTRI=TRUE&amp;ALLINFO=TRUE&amp;ANYS=2013&amp;x=9&amp;y=9</a>
<b>estructura</b>	inclou tres fulls de dades i un full d'índex; els fulls corresponen als anys 2011,2012 i 2013; els tres fulls presenten la mateixa estructura tabular, amb 11 columnes i 957 files, de les quals n'hi ha 4 a l'inici que són descriptives, una conté els encapçalaments, una separadora, 947 contenen dades, una separadora al final, una de totals per columna, una separadora i una que fa referència a les fonts

<b>mida</b>	347kb
<b>descripció contingut</b>	nombre d'efectius de policies locals per graduació i municipi, amb una columna de total per municipi i una fila resum de total per columna
<b>any</b>	2011-2013
<b>resum de camps</b>	<i>ajuntament</i> : text
	<i>graduacions (8 entrades)</i> : numèric
	<i>total (per municipi)</i> : numèric natural
	<i>per mil (ràtio per municipi)</i> : numèric real
<b>fila de resums</b>	<i>total per graduacions (Catalunya)</i> : numèric natural
	<i>total d'efectius (Catalunya)</i> : numèric natural
	<i>per mil (ràtio d'efectius per habitant a Catalunya)</i> : numèric real
<b>columnes calculades</b>	<i>total d'efectius per graduació</i> : numèric natural
	<i>per mil (ràtio d'efectius per habitants al municipi)</i> : numèric real
<b>notes</b>	al camp de municipi es separa amb una coma i un espai el nom de l'article; hi ha files descriptives i de separació que no contenen dades

pernoctaciones.csv

<b>fitxer</b>	pernoctaciones.csv
<b>format</b>	text pla, extensió csv amb els valors separats per tabulador i el final de registre marcat amb salt de línia LF; pel tipus d'exportació en què esta fet el fitxer, els tabuladors consecutius desapareixen, cosa que altera la estructura de les dades
<b>descripció fitxer</b>	dades agregades referents al moviment turístic a Catalunya
<b>comentaris</b>	aquest fitxer no presenta cap estructura coherent; a l'anar a l'origen es veu que és un agregat de 4 taules (zonas turísticas, viajeros y pernoctaciones, residencia y periodos) i això permet deduir-ne l'estructura lògica, de manera que s'ha fet una importació a un full de càlcul i s'ha manipulat manualment per donar-li aquesta estructura i visualitzar les dades
<b>origen</b>	INE - Instituto Nacional de Estadística
<b>link</b>	<a href="http://www.ine.es/jaxiT3/Tabla.htm?t=2039">http://www.ine.es/jaxiT3/Tabla.htm?t=2039</a>
<b>estructura</b>	aquesta taula presenta un primer nivell d'agregació amb viatgers i pernoctacions; per a cada categoria, hi ha un segon nivell amb residents a Espanya i residents fora d'Espanya; per últim, per a cada categoria presenta les dades per mes dels anys 2012 a 2014 per a cada zona turística de Catalunya; hi ha 145 columnes i 19 files, de les que les 3 primeres són descriptives, les 3 següents d'encapçalament i les 13 restants de dades
<b>mida</b>	10kb
<b>descripció contingut</b>	les dades corresponen al nombre de viatgers i pernoctacions per procedència, presentades per mes i any entre 2012 i 2014 per a cada marca turística
<b>resum de camps</b>	<i>camps agregats, nivell 1: viatgers, pernoctacions.</i>
	<i>camps agregats nivell 2: residents a Espanya, residents fora d'Espanya</i>
	<i>mes i any (36x4 entrades)</i> : numèric natural
	<i>marca turística</i> : text
<b>notes</b>	les marques de Costa del Maresme i Costa del Garraf estan unificades sota el nom de Costa de Barcelona; amb tot, apareix un Costa de Barcelona 2015 sense

	dades; la marca Pirineo Catalán es substitueix per Pirineus; amb tot, apareix Pirineo Catalán sense dades; de les zones turístiques Terres de l'Ebre, Vall d'Aran, Terres de Lleida i Catalunya Central manquen les dades del període 2012
--	--

#### mturistiques.csv

<b>fitxer</b>	mturistiques.csv
<b>format</b>	text pla, extensió csv amb els valors separats per punt i coma; els registres es separen amb retorn de carro CR LF
<b>descripció fitxer</b>	Comarques de les marques turístiques de Catalunya
<b>generat</b>	09/04/2015
<b>origen</b>	IDESCAT - Institut d'Estadística de Catalunya
<b>link</b>	<a href="http://www.idescat.cat/codis/?id=50&amp;n=37">http://www.idescat.cat/codis/?id=50&amp;n=37</a>
<b>estructura</b>	estructura tabular; conté 54 files i 3 columnes
<b>mida</b>	2kb
<b>descripció contingut</b>	conté la relació de les marques turístiques de Catalunya i les comarques que les integren, amb els seus respectius codis
<b>any</b>	2014
<b>resum de camps</b>	<i>nivell, codi, nom: text</i>
<b>notes</b>	venen junts en la mateixa columna els nivells de marca i comarca; els valors de comarca no porten article

#### comarques.csv

<b>fitxer</b>	comarques.csv
<b>format</b>	text pla, extensió csv amb els valors separats per punt i coma; els registres es separen amb salt de línia LF
<b>descripció fitxer</b>	Comarques de Catalunya
<b>generat</b>	09/04/2015
<b>origen</b>	IDESCAT - Institut d'Estadística de Catalunya
<b>link</b>	<a href="http://www.idescat.cat/codis/?id=50&amp;n=10&amp;lang=es">http://www.idescat.cat/codis/?id=50&amp;n=10&amp;lang=es</a>
<b>estructura</b>	estructura tabular; conté 992 files i 3 columnes
<b>mida</b>	30kb
<b>descripció contingut</b>	conté la relació de les comarques de Catalunya i els municipis que les integren, amb els seus respectius codis
<b>any</b>	2014
<b>resum de camps</b>	<i>nivell, codi, nom: text</i>

<b>notes</b>	venen junts en la mateixa columna els nivells de municipi i comarca; els valors de comarca no porten article; els valors de municipi porten l'article (quan hi és) separat per coma i espai
--------------	---

#### habitantspermunicipis.xls

<b>fitxer</b>	habitantspermunicipis.xls
<b>format</b>	fitxer de full de càlcul Excel, d'extensió xls, compatible amb les darreres versions d'Office i OpenOffice, amb un full d'índex i 4 fulls de dades
<b>descripció fitxer</b>	conté el nombre d'habitants de Catalunya per municipi, dels anys 2011 a 2014
<b>comentaris</b>	per no fer massa espès el document, presentem el link d'un sol any, el 2014
<b>generat</b>	10/04/2015
<b>origen</b>	IDESCAT – Institut d'Estadística de Catalunya
<b>link</b>	<a href="http://www.idescat.cat/territ/BasicTerr?TC=20&amp;V0=3&amp;V1=0&amp;V3=669&amp;V4=446&amp;P=N&amp;PARENT=1&amp;CTX=B&amp;DISTRI=TRUE&amp;ALLINFO=TRUE&amp;ANYS=2014&amp;x=13&amp;y=11">http://www.idescat.cat/territ/BasicTerr?TC=20&amp;V0=3&amp;V1=0&amp;V3=669&amp;V4=446&amp;P=N&amp;PARENT=1&amp;CTX=B&amp;DISTRI=TRUE&amp;ALLINFO=TRUE&amp;ANYS=2014&amp;x=13&amp;y=11</a>
<b>estructura</b>	inclou 4 fulls de dades i un full d'índex; els fulls corresponen als anys 2011,2012,2013 i 2014; els tres fulls presenten la mateixa estructura tabular, amb 4 columnes i 957 files, de les quals n'hi ha 3 a l'inici que són descriptives, una conté els encapçalaments, una separadora, 947 contenen dades, una separadora al final, una de totals per columna, una separadora i una que fa referència a les fonts
<b>mida</b>	497kb
<b>descripció contingut</b>	nombre d'habitants per sexe i total dels municipis de Catalunya dels anys 2011 a 2014
<b>any</b>	2011-2014
<b>resum de camps</b>	<i>municipi</i> : text
	<i>homes</i> : numèric natural
	<i>dones</i> : numèric natural
	<i>total</i> : numèric natural
<b>columnes calculades</b>	<i>total</i> : numèric natural
<b>notes</b>	al camp de municipi es separa amb una coma i un espai el nom de l'article; hi ha files descriptives i de separació que no contenen dades

#### 2.1.3.2 Anàlisi de les dades (característiques i incidències)

En aquest apartat es mostren els camps en detall, així com les errades, incidències i possibles problemes identificats en els fitxer d'origen de les dades.

#### equipaments

<b>fitxer</b>	equipaments.csv		
	S'han detectat 707 registres duplicats		
<b>detall de camps</b>	<b>camp</b>	<b>tipus/mida</b>	<b>principals errors detectats</b>
	<i>nom</i>	text: [2:145]	parts repetides dintre el text

	<i>adreca</i>	text: [1:112]	adreces aparentment incompletes; signes incorrectes;
	<i>municipi</i>	text: [3:43]	
	<i>comarca</i>	text: [5:17]	
	<i>cp</i>	text: [1-2,4-5]	codis postals aparentment incorrectes; quan falta pot estar entrat com "0" (zero) o com "-" (guió)
	<i>telefon</i>	text: [0:12]	registres buits; dades incorrectes
	<i>longitud</i>	text: [5:11]	formats irregulars; dades incoherents
	<i>latitud</i>	text: [5:17]	formats irregulars; dades incoherents
	<i>location</i>	text: [61:83]	formats irregulars; dades incoherents
	<i>categories</i>	text: [30:382]	

#### infraestructura turística

<b>fitxer</b>	infraestructura turística		
<b>fulls 1-3</b>	establiments hotelers per categoria a cada municipi		
<b>detall de camps</b>	<b>camp</b>	<b>tipus/mida/valor</b>	<b>errors detectats</b>
	<i>municipi</i>	text: [3:43] []	
	<i>1estrella</i>	num enter: [] [0:35]	
	<i>2estrella</i>	num enter: [] [0:39]	
	<i>3estrella</i>	num enter: [] [0:116]	
	<i>4estrella</i>	num enter: [] [0:149]	
	<i>5estrella</i>	num enter: [] [0:36]	
	<i>totalhotel</i>	num enter: [] [0:365]	
	<i>hostal</i>	num enter: [] [0:238]	
<i>total</i>	num enter: [] [0:603]		

<b>fitxer</b>	infraestructura turística		
<b>fulls 4-6</b>	places en establiments hotelers per categoria a cada municipi		
<b>detall de camps</b>	<b>camp</b>	<b>tipus/mida/valor</b>	<b>errors detectats</b>
	<i>municipi</i>	text: [3:43] []	
	<i>1estrella</i>	num enter: [] [0:2278]	
	<i>2estrella</i>	num enter: [] [0:3529]	
	<i>3estrella</i>	num enter: [] [0:14967]	
	<i>4estrella</i>	num enter: [] [0:33695]	
	<i>5estrella</i>	num enter: [] [0:9725]	
	<i>totalhotel</i>	num enter: [] [0:63445]	
	<i>hostal</i>	num enter: [] [0:5683]	
<i>total</i>	num enter: [] [0:69128]		

<b>fitxer</b>	infraestructura turística		
<b>fulls 7-9</b>	establiments de càmping per categoria a cada municipi		
<b>detall de camps</b>	<b>camp</b>	<b>tipus/mida/valor</b>	<b>errors detectats</b>
	<i>municipi</i>	text: [3:43] []	
	<i>luxe</i>	num enter: [] [0:1]	

	<i>primera</i>	num enter: [] [0:5]	
	<i>segona</i>	num enter: [] [0:12]	
	<i>tercera</i>	num enter: [] [0:4]	
	<i>total</i>	num enter: [] [0:12]	

<b>fitxer</b>	infraestructura turística		
<b>fulls 10-12</b>	places en establiments de càmping per categoria a cada municipi		
<b>detall de camps</b>	<b>camp</b>	<b>tipus/mida/valor</b>	<b>errors detectats</b>
	<i>municipi</i>	text: [3:43] []	
	<i>luxe</i>	num enter: [] [0:3069]	
	<i>primera</i>	num enter: [] [0:11199]	
	<i>segona</i>	num enter: [] [0:11169]	
	<i>tercera</i>	num enter: [] [0:4299]	
	<i>total</i>	num enter: [] [0:15276]	

<b>fitxer</b>	infraestructura turística		
<b>fulls 13-15</b>	establiments de turisme rural per categoria a cada municipi		
<b>detall de camps</b>	<b>camp</b>	<b>tipus/mida/valor</b>	<b>errors detectats</b>
	<i>municipi</i>	text: [3:43] []	
	<i>compartida</i>	num enter: [] [0:12]	
	<i>independent</i>	num enter: [] [0:39]	
	<i>masia</i>	num enter: [] [0:6]	
	<i>masoveria</i>	num enter: [] [0:25]	
	<i>total</i>	num enter: [] [0:53]	

<b>fitxer</b>	infraestructura turística		
<b>fulls 16-18</b>	places en establiments de turisme rural per categoria a cada municipi		
<b>detall de camps</b>	<b>camp</b>	<b>tipus/mida/valor</b>	<b>errors detectats</b>
	<i>municipi</i>	text: [3:43] []	
	<i>compartida</i>	num enter: [] [0:118]	
	<i>independent</i>	num enter: [] [0:180]	
	<i>masia</i>	num enter: [] [0:78]	
	<i>masoveria</i>	num enter: [] [0:173]	
	<i>total</i>	num enter: [] [0:316]	

polícia local.xls

<b>fitxer</b>	polícia local.xls		
<b>fulls 1-3</b>	efectius de la policia local per graduació a cada municipi		
<b>detall de camps</b>	<b>camp</b>	<b>tipus/mida/valor</b>	<b>errors detectats</b>
	<i>municipi</i>	text: [3:43] []	

<i>superintendent</i>	num enter: [] [0:1]	
<i>intendent major</i>	num enter: [] [0:7]	
<i>intendent</i>	num enter: [] [0:20]	
<i>inspector</i>	num enter: [] [0:20]	
<i>sotsinspector</i>	num enter: [] [0:24]	
<i>sargent</i>	num enter: [] [0:69]	
<i>caporal</i>	num enter: [] [0:268]	
<i>agent</i>	num enter: [] [0:2507]	
<i>total</i>	num enter: [] [0:2857]	
<i>per mil</i>	num real, 2 decimals: [] [0:5.02]	

#### pernoctacions

<b>fitxer</b>	pernoctacions		
	<b>NOTA:</b> les entrades per any/mes i categoria es representaran amb un sol camp per any per a cada agregació (un total de 12 camps, 2 nivell1 x2nivell2 x3anys) per no fer un document que, amb 146 entrades, seria molt extens i no aportaria més informació rellevant que la que es presenta amb aquest resum		
<b>detall de camps</b>	<b>camp</b>	<b>tipus/mida/valor</b>	<b>errors detectats</b>
	<i>viatgers</i>	agregat: nivell1	
	<i>pernoctacions</i>	agregat: nivell1	
	<i>espanyols</i>	agregat: nivell2	
	<i>estrangers</i>	agregat: nivell2	
	<i>zona turística</i>	text: [8:33] []	zones sense dades
	2014M12-2014M1 (1.1)	num enter: [] [4006:191770]	
	2013M12-2013M1 (1.1)	num enter: [] [4504:196048]	
	2012M12-2012M1 (1.1)	num enter: [] [25448:182582]	no hi ha dades a les zones TT de l'Ebre, TT de Lleida, Catalunya Central i Vall d'Aran
	2014M12-2014M1 (1.2)	num enter: [] [248:596346]	
	2013M12-2013M1 (1.2)	num enter: [] [562:602601]	
	2012M12-2012M1 (1.2)	num enter: [] [1916:567520]	no hi ha dades a les zones TT de l'Ebre, TT de Lleida, Catalunya Central i Vall d'Aran
	2014M12-2014M1 (2.1)	num enter: [] [7408:687027]	
	2013M12-2013M1 (2.1)	num enter: [] [8546:599123]	
	2012M12-2012M1 (2.1)	num enter: [] [38415:645373]	no hi ha dades a les zones TT de l'Ebre, TT de Lleida, Catalunya Central i Vall d'Aran
	2014M12-2014M1 (2.2)	num enter: [] [691:1837696]	
2013M12-2013M1 (2.2)	num enter: [] [741:1788755]		



	2012M12- 2012M1 (2.2)	num enter: [] [3874:1639175]	no hi ha dades a les zones TT de l'Ebre, TT de Lleida, Catalunya Central i Vall d'Aran
--	--------------------------	------------------------------	--

mturistica.csv

<b>fitxer</b>	mturistica.csv		
	marca i comarca estan en el mateix camp; els codis respectius, també		
<b>detall de camps</b>	<b>camp</b>	<b>tipus/mida</b>	<b>principals errors detectats</b>
	<i>nivell</i>	text: [7:15]	
	<i>Codi</i>	text: [1:4]	
	<i>nom</i>	text: [5:17]	

comarques.csv

<b>fitxer</b>	comarques.csv		
	municipi i comarca estan en el mateix camp; els codis respectius, també		
<b>detall de camps</b>	<b>camp</b>	<b>tipus/mida</b>	<b>principals errors detectats</b>
	<i>nivell</i>	text: [7:8]	
	<i>Codi</i>	text: [2:6]	
	<i>nom</i>	text: [3:43]	

habitants\_per\_municipis.csv

<b>fitxer</b>	habitants_per_municipis.csv		
<b>fulls 1-3</b>	població a cada municipi		
<b>detall de camps</b>	<b>camp</b>	<b>tipus/mida/valor</b>	<b>errors detectats</b>
	<i>municipi</i>	text: [3:43] []	
	<i>homes</i>	num enter: [] [8:769819]	
	<i>dones</i>	num enter: [] [9:851124]	
	<i>total</i>	num enter: [] [24:1620943]	

### 2.1.3.3 Anàlisi dels continguts (possibles errors i proposta de solució)

Els errors poden ser:

- errors de taula, com ara format incorrecte o incoherent, corrupció de dades...
- errors de camp o registre, què poden ser:
  - errors físics, com ara repetits, camps o registres buits, signes incorrectes...
  - errors lògics, com ara correspondències trencades, mides incorrectes, camps mal calculats...

Per localitzar els errors més comuns, s'ha fet un examen visual de les dades amb el suport d'algunes funcions de full de càlcul, consultant màxims, mínims o longituds, per exemple. Amb això s'han pogut detectar estructures mal definides, detectar camps i registres sense dades, duplicitats, dades incoherents, formats o signes incorrectes. També s'ha pogut observar la correctesa dels camps calculats.

## Equipaments:

S'han detectat 707 registres duplicats. Es proposa eliminar els duplicats.

- **NOM.**

- Es pot veure que hi alguns valors de molta longitud. Mirant el contingut, s'observen repeticions de part del text en el contingut d'alguns registres. En una primera aproximació podem veure que hi ha errors d'aquest tipus al menys a partir d'una longitud de 90 caràcters, tot i que pot ser que n'hi hagi en longituds inferiors. Aprofitant les funcions de l'Excel, es pot intentar localitzar aquestos registres i mirar de corregir l'error, ja sigui amb l'ús de funcions, ja sigui manualment.
- Es pot veure un ús inconsistent de majúscules i minúscules i d'abreviatures. Pel que fa a les majúscules es pot optar per passar-ho tot a majúscules. Pel que fa a les abreviatures, s'haurien de identificar i substituir per la paraula o abreviatura correcta, sempre que això sigui viable.
- Hi ha algun valor està tallat o incomplet. Es proposa, si es pot, completar-lo manualment.

- **ADRECA.**

- Hi ha dades aparentment incompletes, amb un sol número o una sola lletra. Aquestes dades no es poden verificar sense adreçar-se al cadastre del mateix municipi, per tant, aquestes dades no es verificaran perquè ultrapassen l'àmbit del treball, no aporten informació significativa per a l'estudi i no generen errors.
- En canvi, hi ha un seguit de registres que presenten un error en l'Excel per la inclusió de un signe reservat. Aquest signe, el signe "=" (igual) seguit d'un guió "-", no forma part de la nomenclatura habitual de les adreces a Catalunya i és fàcil d'eliminar amb les funcions del mateix full de càlcul.
- Hi ha un ús inconsistent de majúscules. Es proposa passar al format d'inicial amb majúscula.
- Hi ha petites errades de signes separadors, falta de coma entre carrer i número o carrers sense número. Atès que són errades de difícil correcció automàtica i que no és informació requerida per a l'estudi, es proposa no actuar.

- **MUNICIPI.** Tot i no haver-hi apreciat errades, es vol fer notar que en aquest camp l'article del poble, quan hi és, està en minúscules i separat per dos espais en blanc. Cal anar en conte perquè és possible que en algun altre fitxer aquest camp tingui un separador diferent (p.e. en el fitxer d'infraestructura turística). Es proposa igualar el format i substituir el primer espai separador per una coma.

- **COMARCA.** No s'aprecien errors en aquest camp.

- **CP.** El codi postal no podrà substituir al nom de la població, atès que en poblacions grans n'hi ha més d'un. Per aquest motiu no és una dada necessària per a l'estudi. De totes maneres, es proposa unificar el criteri dels codis no entrats i canviar el guió per un zero i intentar donar valor amb d'altres registres de la mateixa població amb valor.

- **TELEFON.**

- Aquest camp presenta moltes errades de dades incompletes que difícilment es poden corregir. D'altra banda, tampoc és una dada rellevant en l'estudi encomanat.

- Al marge d'aquests, però, n'hi ha força registres sense valor, cosa que pot provocar problemes en el tractament de les dades. Per solucionar aquest possible problema, es proposa donar valor als registres que no en tenen tal com ho hem fet amb els codis postals, afegint-hi un zero.
- El format és inconsistent, amb espais entre números o sense. Es proposa unificar el format eliminant els espais en blanc.
- **LONGITUD.** Catalunya està situada entre 0 i 4 graus de longitud, per tant, tots els valors que estiguin fora d'aquest rang seran incorrectes. D'altra banda, hi ha valors negatius, valors en format numèric i valors en format de text. Es proposa unificar el format dels valors que ho permeten, en especial els valors negatius i els numèrics - tot i que això no garanteix que el valor resultant sigui correcte- per tal d'evitar possibles errades en el tractament de les dades. Els valors incorrectes es poden negligir doncs no afecten al resultat de l'estudi. Això no obstant, si els canvis fets a location ho permeten, és possible recuperar les dades d'aquest camp. Això no garanteix la certesa de totes les dades, però en permetrà un tractament uniforme
- **LATITUD.** Catalunya està situada entre 40 i 43 graus de latitud nord, per tant, tots els valors que estiguin fora d'aquest rang seran incorrectes. Hi ha molt pocs valors amb format numèric que es poden resoldre manualment, sense que això suposi la correcció de la dada. Els valors incorrectes es poden negligir doncs no afecten al resultat de l'estudi. Això no obstant, si els canvis fets a location ho permeten, és possible recuperar les dades d'aquest camp. Això no garanteix la certesa de totes les dades, però en permetrà un tractament uniforme.
- **LOCATION.** Aquest valor està lligat al la longitud i la latitud. Es presenten els dos valors i el valor 0.0 separats per una coma. Tots tres estan encerclats pel que semblen etiquetes XML. Es podria buscar alguna relació entre longitud i latitud i location, però sembla que molts dels errors que estan en algun dels dos primers camps, també hi són en aquest. Per tant, aquest camp no ens serveix completament per corregir els anteriors ni els anteriors ens serveixen per corregir aquest. No sembla que hi hagi dades que puguin provocar errors en el tractament de les dades, de manera que els valors incorrectes es poden negligir doncs no afecten al resultat de l'estudi. També es pot plantejar l'opció d'eliminar el camp, doncs és redundant. De totes maneres, de cara al futur, s'ha fet una recuperació basada en els camps longitud i latitud, els valors i la longitud. Això no garanteix la certesa de totes les dades, però en permetrà un tractament uniforme. Alhora, aquest camp corregit ens permetrà recuperar valors de longitud i latitud.
- **CATEGORIES.** Tot i que en aquest camp no s'hi aprecien errades, s'ha de fer notar que la separació no és estrictament jeràrquica, de manera que un equipament pertany a totes les categories que mostra el camp. Caldrà tenir en compte aquest extrem quan s'empri aquesta dada.

#### Infraestructura turística (Fulls 1-3, 4-6, 7-9, 10-12, 13-15, 16-18)

- **MUNICIPI.** Tot i que no s'aprecien errors en aquest camp, caldrà fer notar que els articles de les poblacions, quan hi són, van separats per una coma i un espai. Cal anar en conte perquè és possible que en algun altre fitxer aquest camp tingui un separador diferent (p.e. en el fitxer d'equipaments).
- No s'aprecien errors en les dades numèriques. Tot i això, cal tenir en compte que en el full hi ha files que no contenen dades i no són l'encapçalament (files

descriptives o separadores). A més, existeix una fila final calculada amb el total de cada columna que, en principi, no ha de ser carregada.

#### Policia local

##### Fulls 1-3

- **MUNICIPI.** Tot i que no s'aprecien errors en aquest camp, caldrà fer notar que els articles de les poblacions, quan hi són, van separats per una coma i un espai. Cal anar en conte perquè és possible que en algun altre fitxer aquest camp tingui un separador diferent (p.e. en el fitxer d'equipaments).
- No s'aprecien errors en les dades numèriques. Tot i això, cal tenir en compte que en el full hi ha files que no contenen dades i no són l'encapçalament (files descriptives o separadores). A més, existeix una fila final calculada amb el total de cada columna que, en principi, no ha de ser carregada.

#### Pernoctacions

En els comentaris de la descripció del fitxer s'ha explicat que sense la visualització de l'origen de les dades a la web de l'INE no s'hagués pogut reconstruir.

- **ZONA TURÍSTICA.** Hi ha 4 zones turístiques sense dades: Costa del Garraf, Costa Barcelona 2015, Costa Barcelona-Maresme, Pirineo Catalán. Aquestes files es poden suprimir, doncs en el cas de C. del Garraf i C. Barcelona-Maresme s'ha fet una reagrupació en Costa Barcelona; Costa Barcelona 2015 no entra en el període d'estudi i Pirineo Catalán està com Pirineus. Es dona el format en català, corregint un doble espai a la marca Catalunya Central (Catalunya per Catalunya, Val d'Aran per Vall d'Aran).
- Els camps corresponents a les dades del mesos de l'any 2012 (2012M12-2012M1) de totes les categories (viatgers - espanyols, viatgers - estrangers, pernoctacions - espanyols i pernoctacions - estrangers) de les zones Terres de l'Ebre, Terres de Lleida, Catalunya Central i Vall d'Aran no contenen valor. Atès que en origen aquestes dades no hi són i per tal d'evitar errades en el tractament de les dades, es proposa omplir aquestes faltes amb la taxa de variació respecte a l'any anterior que presenten el conjunt de valors de les altres zones.
- A part del senyalat, no s'aprecien errors en les dades numèriques. Tot i això, cal tenir en compte que en el full hi ha dades agregades, hi ha files que no contenen dades i no són l'encapçalament (files descriptives ).

#### Marques\_turístiques

- Els camps nivell, Codi i nom inclouen tant marques com comarques. Es proposa desfer el fitxer en tres taules, una amb codi i descripció de marca turística, una amb codi i descripció de comarca i una tercera relacionant per codi la marca amb les respectives comarques.

#### Comarques\_municipis

- Els camps nivell, Codi i nom inclouen tant municipis com comarques. Es proposa desfer el fitxer en tres taules, una amb codi i descripció de municipi, una amb codi i descripció de comarca i una tercera relacionant per codi els municipis amb les respectives comarques

habitants\_per\_municipi

- **MUNICIPI.** Tot i que no s'aprecien errors en aquest camp, caldrà fer notar que els articles de les poblacions, quan hi són, van separats per una coma i un espai. Cal anar en conte perquè és possible que en algun altre fitxer aquest camp tingui un separador diferent (p.e. en el fitxer d'equipaments).
- No s'aprecien errors en les dades numèriques. Tot i això, cal tenir en compte que en el full hi ha files que no contenen dades i no són l'encapçalament (files descriptives o separadores). A més, existeix una fila final calculada amb el total de cada columna que, en principi, no ha de ser carregada.

## 2.2 Disseny conceptual

Si bé inicialment es va considerar fer servir una taula desnormalitzada única, l'anàlisi més profund aconsella una òptica diferent, fent servir taules de fets i dimensions amb un disseny multidimensional. Pel tipus i estructura de les taules, el model serà d'estrella; a part, aquest model pot aportar més avantatges que un model de floquet de neu per la seva senzillesa.

D'altra banda, hem de senyalar que les claus primàries de les taules seran claus subrogades (*RowID*), és a dir, claus numèriques i seqüencials que s'assignaran de manera automàtica quan es generen nous registres. Aquest tipus de claus permet separar la gestió dels índexs del tipus de dades de la taula i és més eficient.

### 2.2.1 Taules de fets

El fets que s'han identificat per a donar resposta a les peticions de l'empresa seran:

- establiments
- places
- pernoctacions
- viatgers
- efectius
- habitants
- equipaments

Es plantegen quatre taules de fets:

1. Establiments, amb nombre d'establiments i places

any	mesTemporada	marca	comarca	municipi	tipus	categoria	establiments	places
idData		idZona			idEstabliment		establiments	places

2. Pernoctacions, amb nombre de pernoctacions i de viatgers

any	mesTemporada	marca	comarca	municipi	viatgers	pernoctacions
idData		idZona			viatgers	pernoctacions

3. Equipaments

marca	comarca	municipi	nivell1	nivell2	nivell3	*
-------	---------	----------	---------	---------	---------	---

idZona	idEquipament	*
--------	--------------	---

(\*) Aquest és un camp calculat amb el recompte d'equipaments per zona i nivell

#### 4. Polícies, amb el nombre de polícies i d'habitants

marca	comarca	municipi	efectius	habitants
idZona			efectius	habitants

### 2.2.1.1 Mesures

Les mesures que es corresponen amb les taules de fets:

- establiments , numèric, nombre d'establiments
- places , numèric, nombre de places
- pernoctacions , numèric, nombre de pernoctacions
- viatgers , numèric, nombre de viatgers
- efectius , numèric, nombre de polícies locals
- equipaments , numèric, recompte d'equipaments
- habitants, numèric, nombre d'habitants

Aquestes mesures s'ajusten al que es demana als informes:

- (1) (places del municipi \* total de pernoctacions / places totals )
- (2) (nombre d'establiments per tipus de la comarca \* total viatgers / total d'establiments)
- (3) (places de la marca \* total de pernoctacions / places totals(1))
- (4) recompte i ordenació per total de categoria per municipi
- (5) màxim – mínim (polícies per municipi / tipus d'establiment per municipi)
- (6) los 10 primers municipis (places del municipi \* total de pernoctacions / places totals(1)) categoritzats per franges
- (7) ((places del municipi \* total de pernoctacions / places totals(1))/places totals)x100
- (8) (categorització per número d'habitants 80/15/5)
- (9) (rati per mil (promig de tots els ratis per mil de comarca o marca))
- (10) (distribució de les pernoctacions per any i zona entre els mesos i les temporades)

### 2.2.1.2 Atributs de les taules de fets

S'identifiquen els següents atributs de la taula de les taules de fets:

- Establiments
  - establiments
  - places
- Pernoctacions
  - pernoctacions
  - viatgers
- Polícies
  - efectius
  - habitants

- Equipaments

### 2.2.1.3 Definició de les cel·les

Les cel·les definides seran les que corresponen als quatre fets identificats:

- Establiments
  - establiments
  - places
- Pernoctacions
  - pernoctacions
  - viatgers
- Polícies
  - efectius
  - habitants
- Equipaments

### 2.2.2 Granularitat

La granularitat, pel que fa al temps, es demana, quan sigui possible, per mes i any. Així, per als establiments, places, pernoctacions i viatgers, serà per any i mes o temporada. Pel que fa a les pernoctacions, viatgers, efectius policials, equipaments, establiments i oferta de places, la granularitat espacial serà a nivell de municipi.

### 2.2.3 Dimensions

S'identifiquen les següents dimensions:

- Data, referència temporal.
- Zona, referència geogràfica.
- Equipament, referent als equipaments.
- Establiment, referent a l'oferta turística.

#### 2.2.3.1 Atributs de cada dimensió

Per a cada dimensió, s'identifiquen els següents atributs:

- Data: id, Any, mes1, mes2.
- Zona: id, Marca turística, comarca, municipi.
- Equipament: id, Nivell1, nivell2, nivell3.
- Establiment: id, Tipus, categoria.

#### 2.2.3.2 Jerarquia d'agregació d'atributs de dimensió

Per a cada dimensió es mostra la jerarquia d'agregació d'atributs:

- Data: Any, mes1+mes2.
- Zona: Marca turística, comarca, municipi.
- Equipament: Nivell1, nivell2, nivell3.
- Establiment: Tipus, categoria.

### 2.2.4 Restriccions d'integritat

Restriccions sobre les dimensions:

- Zona. El codi 'marca+comarca+municipi' és únic i no conté buits.
- Establiments. El codi 'tipus+categoria' és únic i no conté buits.

- Equipaments. Tal com s'han dissenyat els nivells, el codi 'nivell1+nivell2+nivell3' és únic i no conté buits.
- Data. Caldrà verificar la disponibilitat de les dates, pel que fa als anys i la granularitat de mesTemporada.

## 2.2.5 Viabilitat

Dimensions			
Taules	mida dels camps	numero de registres	Total
<b>Data</b>	3x4=12	50	600
<b>Estacionalitat</b>	3x4=12	20	240
<b>Zona</b>	3x50 + 4 = 154	1000	154000
<b>Establiments</b>	2x25 + 4 = 54	14	756
<b>Equipaments</b>	4+50+100+100=254	300	76200
		càlcul aprox. de nivells diferents	231796

II-lustració 9 Taula viabilitat dimensions

Fets			
Taules	mida dels camps	numero de registres	Total
<b>FetEstabliment</b>	5x4=20	14x1000x50=700000	14000000
<b>FetEquipament</b>	3x4=12	1000x35=35000	420000
		s'ha calculat el promig d'equipaments per municipi	
<b>FetPernoctacions</b>	4x4=16	50x1000=50000	800000
<b>FetPoliciesHabitants</b>	3x4=12	1000	12000
			15232000

II-lustració 10 Taula viabilitat fets

En total, s'estima un volum de 15463796 bytes, clarament per baix d'1Gb, per tant és viable.

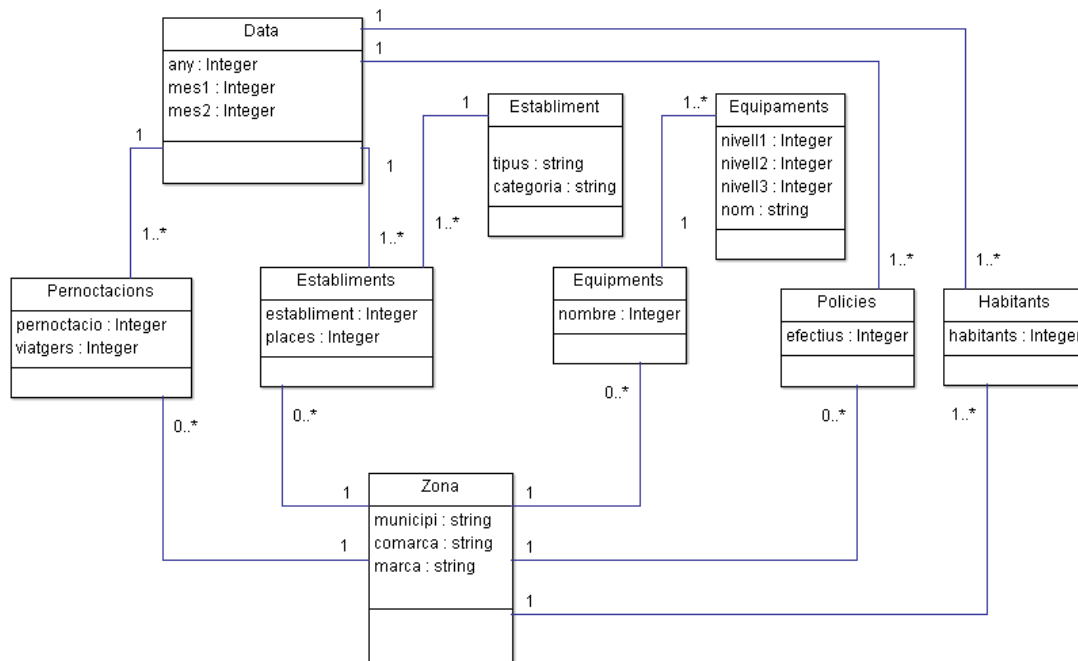
## 2.3 Disseny de la BD

Dintre d'aquest apartat es mostren el diagrama d'entitat-relació (E-R), l'esquema del disseny físic i el del nivell lògic.

### 2.3.1 Diagrama E-R

Es mostra el diagrama d'entitat-relació ([ER](#), *Entity Relationship*) per a representar les entitats del model i les seves relacions i propietats.

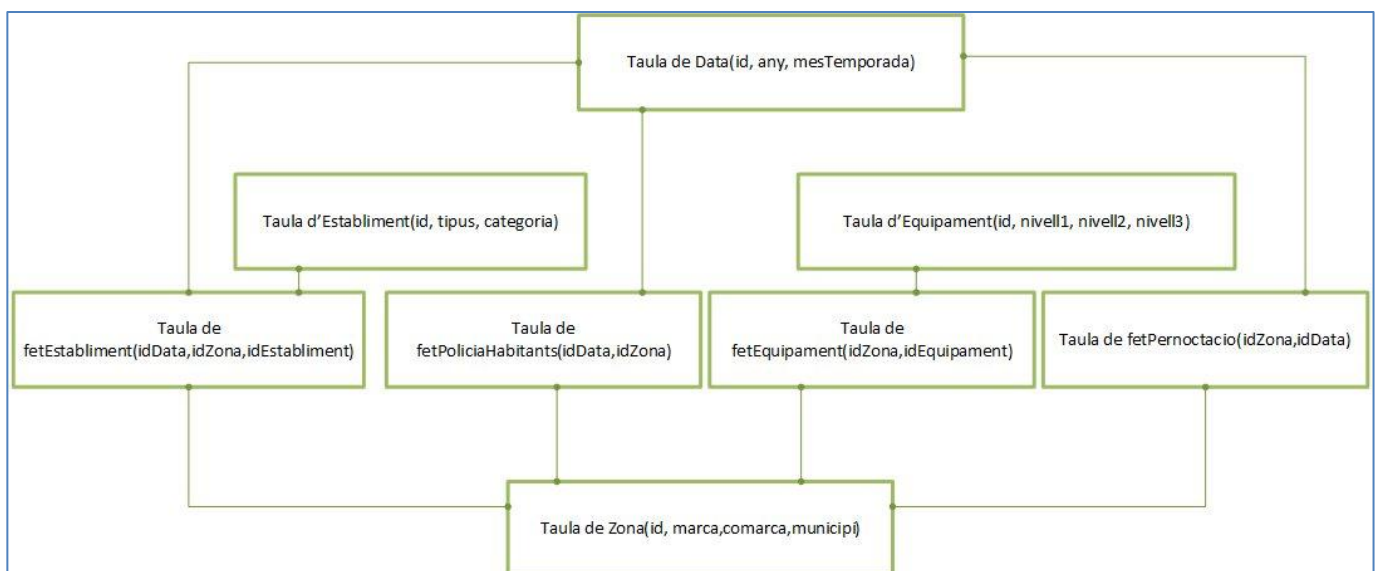




II-Iustració 11 Diagrama E-R

### 2.3.2 Disseny lògic

El disseny escollit és en estrella. El nom d'*estrella* ve donat per la forma de l'esquema de com es relacionen les taules entre elles, per contraposició al de *floquet de neu*, que té una forma fractal.



II-Iustració 12 Esquema del disseny lògic

### 2.3.3 Disseny físic

Es mostra el disseny físic de la base de dades, amb la descripció de les taules, els camps (tipus, mida, claus, restriccions), els índexs i les restriccions de taula.

Taula		Data			
Camps	Tipus	Mida	PK	FK	Not Null
id	int	3	Sí	No	Sí
any	int	4	No	No	Sí
mes1	int	2	No	No	No
mes2	int	2	No	No	No

Taula		Zona			
Camps	Tipus	Mida	PK	FK	Not Null
id	int	5	Sí	No	Sí
marca	varchar	50	No	No	Sí
comarca	varchar	50	No	No	Sí
municipi	varchar	50	No	No	Sí

Taula		Establiments			
Camps	Tipus	Mida	PK	FK	Not Null
id	int	5	Sí	No	Sí
tipus	varchar	25	No	No	Sí
categoria	varchar	25	No	No	Sí

Taula		Equipaments			
Camps	Tipus	Mida	PK	FK	Not Null
id	int	5	Sí	No	Sí
nivell1	varchar	50	No	No	Sí
nivell2	varchar	100	No	No	Sí
nivell3	varchar	100	No	No	No

Taula		FetsEstabliments			
Camps	Tipus	Mida	PK	FK	Not Null
iddata	int	3	Sí	No	Sí
idZona	int	5	Sí	Sí	Sí
idEstabliments	int	5	Sí	Sí	Sí
establiments	int	5	No	No	No
places	int	7	No	No	No

Taula		FetsEquipaments			
Camps	Tipus	Mida	PK	FK	Not Null
idZona	int	5	Sí	Sí	Sí
idEquipaments	int	5	Sí	Sí	Sí
equipaments	int	5	No	No	No

Taula	FetsPernoctacions				
Camps	Tipus	Mida	PK	FK	Not Null
idData	int	3	Sí	Sí	Sí
idZona	int	5	Sí	Sí	Sí
pernoctacions	int	7	No	No	No
viatgers	int	7	No	No	No

Taula	FetsPoliciesHabitants				
Camps	Tipus	Mida	PK	FK	Not Null
idData	int	5	Sí	Sí	Sí
idZona	int	5	Sí	Sí	Sí
efectiu	int	5	No	No	No
habitants	int	7	No	No	No

## 2.4 Procés d'ETL a alt nivell.

Per a fer el procés de transformació es disposava de diferents opcions: utilitzar *scripts* de SQL, l'automatització amb l'eina Spoon, desenvolupar una eina pròpia amb algun llenguatge de programació adient, o fer servir les facilitats dels fulls de càlcul. S'han valorat les següents premisses:

- El procés és únic, no es repetirà amb els mateixos paràmetres per un altre estudi.
- Totes les dades són externes, és a dir, generades fora de l'empresa. Alguns dels fitxers originals no s'han pogut tornar a generar o recuperar des del lloc original.
- Les dades es donen en format csv o xls, formats aptes per al treball amb un full de càlcul.
- Es té un bon domini de les funcionalitats del full de càlcul Excel.

Amb aquest arguments, s'ha descartat l'opció de generar una eina pròpia per l'alt cost i la nul·la reutilització. L'ús d'*scripts* de SQL no soluciona el principal problema, que és disposar d'una estructura apta per a taules relacionals. L'opció de fer servir Spoon s'ha valorat en deteniment; s'ha rebutjat per què té un alt cost en temps respecte a la nul·la reutilització i el fet que s'hagi de combinar amb processos manuals previs per preparar una estructura que pugui ser manipulada. Per últim, l'opció escollida ha estat l'ús de les funcionalitats del mateix full de càlcul i el llenguatge de programació incorporat *vbexcel* (Visual Basic Excel) per la proximitat amb els formats i les possibilitats que hi ofereix en aquest cas concret. Així doncs, en aquest apartat es duu a terme el procés de transformació i es deixa el camí aplanat per a fer la càrrega de dades en la següent fase del projecte.

### 2.4.1 Extracció de dades

Una part del procés d'extracció de dades arriba ja fet per l'empresa mentre que una part, les dades que s'han d'aconseguir per a completar la informació, s'han de cercar i extraure de vell nou. Les taules que s'han afegit són:

- Habitants per municipi.
- Comarques per marca turística.
- Municipis per comarques.

## 2.4.2 Transformació de dades

La transformació de les dades pot incloure els processos de neteja, correcció, compleció, estandardització i codificació, entre d'altres.

A partir de l'anàlisi de les dades fet en el capítol corresponent, apliquem les correccions prèvies a la transformació de dades necessàries per a la càrrega. Aquest procés seria el tractament d'errors detectats segons les propostes plantejades. Posteriorment es procedirà a generar un fitxer temporal amb l'estructura i les dades necessàries per a la següent fase.

### 2.4.2.1 Equipaments

1. Es procedeix a eliminar els 707 registres duplicats sobre el fitxer original.
2. NOM: S'han canviat a majúscules. S'han eliminat les parts repetides que presentaven alguns registres.
3. CP: Cavi de “-“ per “0”. Canvi manual de codis postals amb dades que no corresponen a les províncies catalanes (cerca a la web [www.correos.es](http://www.correos.es)). Substitució del guió pel cp que mostra la mateixa població en un altre registre. Això deixa 7 registres que s'han buscat i canviat manualment.
4. MUNICIPI: S'han substituït els dos espais que separaven el nom del municipi del seu article (quan hi era) per una coma i un espai.
5. ADRECA: S'han eliminat els signes =- que donaven error.
6. TELEFON: S'han omplert els valors buits amb un zero. S'ha unificat el format eliminant espais entre números.
7. LOCATION: S'han fet canvis tenint en compte el format, els valors i la longitud dels camps latitud i longitud, posant zero davant el punt i igualant la precisió a 9 dígits decimals retallant o reomplint amb zeros respectivament.
8. LOGITUD: S'han substituït les dades a partir dels canvis fets en el camp location, separant els valors corresponents ja formatats.
9. LATITUD: S'han substituït les dades a partir dels canvis fets en el camp location, separant els valors corresponents ja formatats.
10. EQUIPAMENTS: S'han extret tres nivells de la cadena d'equipaments: el primer nivell correspon a la segona entrada, doncs la primera és “Equipament” per a tots els registres i per tant no aporta res de nou; el tercer nivell correspon a la última entrada amb valor, què serà com a mínim la tercera entrada atès que a partir d'aquesta ja n'hi ha sense valor; la segona entrada serà tot el comprés entre la primera i la tercera, incloent-hi aquesta quan és la darrera amb valor.

Tot aquest procés s'ha fet utilitzant funcions del'Excel o creant-hi macros (petits programes amb una funció molt concreta) amb el *vbexcel*. De tot aquest procés es mostra tot seguit la part més significativa: la incorporació de la marca turística (que prové del fitxer de marca turística i comarca) i el tractament de les categories d'equipament. Amb les funcions de l'Excel s'ha segmentat la cadena d'equipaments pel separador “|” i s'ha assignat cada part a una columna.

U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
1	categories											
2	Equipaments	Educacio	Ensenyament:Cicles_forr	Ensenyaments_d	Tecnic_esportiu_Cicles_formatius	Educacio_segund	1r_cicle_educaci	Batxillerat		2n_cicle_educaci	Tecnic_esportiu_Educacio_primaria	
3	Equipaments	Educacio	Ensenyament:Cicles_forr	Cicles_formatius	Educacio_segund	1r_cicle_educaci	Batxillerat		Programes_de_q	2n_cicle_educaci	Educacio_primaria	
4	Equipaments	Educacio	Ensenyament:Batxillerat	2n_cicle_educaci	Ensenyaments_d	Tecnic_esportiu_Educacio_primari	Cicles_formatius	Tecnic_esportiu_Cicles_formatius	Educacio_secundaria_obligatoria			
5	Equipaments	Educacio	Ensenyament:Cicles_forr	Ensenyaments_d	Tecnic_esportiu_Cicles_formatius	Educacio_segund	Batxillerat		2n_cicle_educaci	Tecnic_esportiu_Educacio_primaria		
6	Equipaments	Educacio	Ensenyament:Cicles_forr	Ensenyaments_d	Tecnic_esportiu_Cicles_formatius	Educacio_segund	Batxillerat		2n_cicle_educaci	Tecnic_esportiu_Educacio_primaria		
7	Equipaments	Educacio	Ensenyament:Cicles_forr	Cicles_formatius	Educacio_segund	1r_cicle_educaci	Formacio_person	Batxillerat		2n_cicle_educaci	Educacio_primaria	
8	Equipaments	Educacio	Ensenyament:Cicles_forr	Ensenyaments_d	Tecnic_esportiu_Cicles_formatius	Educacio_segund	Batxillerat		Programes_de_qualificacio_professional_inicial_POPI			
9	Equipaments	Educacio	Ensenyament:Cicles_forr	Ensenyaments_d	Tecnic_esportiu_Cicles_formatius	Educacio_segund	Batxillerat		Programes_de_qualificacio_professional_inicial_POPI			
10	Equipaments	Educacio	Ensenyament:Cicles_forr	Educacio_segund	1r_cicle_educaci	Batxillerat		Programes_de_q	2n_cicle_educaci	Educacio_primaria		
11	Equipaments	Educacio	Ensenyament:Cicles_forr	Cicles_formatius	Educacio_segund	1r_cicle_educaci	Batxillerat		2n_cicle_educaci	Educacio_primaria		
12	Equipaments	Educacio	Ensenyament:Cicles_forr	Cicles_formatius	Educacio_segund	1r_cicle_educaci	Batxillerat		2n_cicle_educaci	Educacio_primaria		
13	Equipaments	Educacio	Ensenyament:Cicles_forr	Cicles_formatius	Educacio_segund	1r_cicle_educaci	Batxillerat		2n_cicle_educaci	Educacio_primaria		
14	Equipaments	Educacio	Ensenyament:Cicles_forr	Cicles_formatius	Educacio_segund	1r_cicle_educaci	Batxillerat		2n_cicle_educaci	Educacio_primaria		
15	Equipaments	Educacio	Ensenyament:Cicles_forr	Cicles_formatius	Educacio_segund	1r_cicle_educaci	Batxillerat		2n_cicle_educaci	Educacio_primaria		
16	Equipaments	Educacio	Ensenyament:Cicles_forr	Cicles_formatius	Educacio_segund	1r_cicle_educaci	Batxillerat		2n_cicle_educaci	Educacio_primaria		
17	Equipaments	Educacio	Ensenyament:Cicles_forr	Cicles_formatius	Educacio_segund	1r_cicle_educaci	Batxillerat		2n_cicle_educaci	Educacio_primaria		
18	Equipaments	Educacio	Ensenyament:Cicles_forr	Ensenyaments_d	Tecnic_esportiu_Cicles_formatius	Educacio_segund	Batxillerat		Tecnic_esportiu_grau_superior			
19	Equipaments	Educacio	Ensenyament:Cicles_forr	Ensenyaments_d	Tecnic_esportiu_Cicles_formatius	Educacio_segund	Batxillerat		Tecnic_esportiu_grau_superior			

### II-lustració 13 Columnes de parts de categoria a equipaments

Per fer els nivells, s'ha considerat no tenir en compte la primera part, atès que és la mateixa per a tots els registres i no aporta informació. Tal com es va comentar en el lliurament anterior, s'ha decidit crear el nivell 1 a partir de la segona columna, el nivell 2 amb les dades de la tercera columna i el nivell 3, de l'última. Si tan sols hi ha dos columnes, els nivells 2 i 3 coincideixen. Prèviament, s'ha comprovat que les columnes que correspondran als nivells 1 i 2 no presenten cel·les buides.

Código		Controles													XML	Modificar								
A02		=>S!(A12>"";A2;S!(AK2>"";AK2;S!(AJ2>"";AJ2;S!(AI2>"";AI2;S!(AH2>"";AH2;S!(AG2>"";AG2;S!(AF2>"";AF2;S!(AE2>"";AE2;S!(AD2>"";AD2;S!(AC2>"";AC2;S!(AB2>"";AB2;S!(AA2>"";AA2;S!(Z2>"";Z2;S!(Y2>"";Y2;S!(X2>"";X2;W2)))))))))																						
E	F	G	H	I	J	U	V	W	X	Y	Z	A	A	A	A	A	A	A	A	A	AM	AN	AO	AP
1	comarca	marca_turística	telefon	longitud	latitud	Locati	category	categories													nivell1	nivell2	nivell3	
2	Pallars Jussà	Pirineus	I252378	1.162.94	42.171	<Poin	Equipan	Equipan	Esport_i	Installac	Altres_espais_petits										Esport_i	lleure	Instal·lacions_esportives	Altres_espais_petits
3	Pallars Jussà	Pirineus		0	1.089.2	42.166	<Poin	Equipan	Equipan	Esport_i	Installac	Area_d_activitat_esportiva									Esport_i	lleure	Instal·lacions_esportives	Area_d_activitat_esportiva
4	Baix Llobregat	Costa Barcelona	7700325	1.901.3	41.516	<Poin	Equipan	Equipan	Adminis	Ens_Loci	Ajuntaments										Administracio	Ens_Locals	Ajuntaments	
5	Baix Llobregat	Costa Barcelona		0	1.899.6	41.517	<Poin	Equipan	Equipan	Comerc	Mercats_municipals										Comerc_Consum	Mercats_municipals	Mercats_municipals	
6	Baix Llobregat	Costa Barcelona		0	1.900.8	41.515	<Poin	Equipan	Equipan	Cultura	AteneusCentresCivicsCasesCultura										Cultura	AteneusCentresCivicsCases	AteneusCentresCivicsCasesC	
7	Baix Llobregat	Costa Barcelona	7700881	1.901.3	41.521	<Poin	Equipan	Equipan	Cultura	Biblioteques	publiques										Cultura	Biblioteques	publiques	
8	Baix Llobregat	Costa Barcelona	7702702	1.896.3	41.521	<Poin	Equipan	Equipan	Educaci	Ensenya	Musica										Educacio	Ensenyaments_de_regim_e	Musica	
9	Baix Llobregat	Costa Barcelona	7703458	1.900.5	41.521	<Poin	Equipan	Equipan	Educaci	Ensenya	1r_cicle_educacio_infantil_llar_infa	Educacio									Educacio	Ensenyaments_de_regim_gi	1r_cicle_educacio_infantil_ll	
10	Baix Llobregat	Costa Barcelona		0	1.895.2	41.522	<Poin	Equipan	Equipan	Educaci	Ensenya	1r_cicle_educacio_infantil_llar_infa	Educacio								Ensenyaments_de_regim_gi	1r_cicle_educacio_infantil_ll		
11	Baix Llobregat	Costa Barcelona	7703211	1.902.5	41.526	<Poin	Equipan	Equipan	Educaci	Ensenya	Educaci	Batxillerat									Educacio	Ensenyaments_de_regim_g	Batxillerat	
12	Baix Llobregat	Costa Barcelona	7701645	1.900.5	41.521	<Poin	Equipan	Equipan	Educaci	Ensenya	2n_cicle_educacio_primaria										Educacio	Ensenyaments_de_regim_g	Educacio_primaria	

### II-lustració 14 Nivells 1,2,3 de categoria a equipaments

Finalment es transporten les dades a un nou full i es passen a format csv. Així es deixen preparades per al procés de càrrega en la fase següent.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	municipi	comarca	marca_turística	nivell1	nivell2	nivell3	nom	adreca	cp	telefon	longitud	latitud	Location			
2	Abella de la	Pallars Jussà	Pirineus	Esport_i	lleu	Instal·lacions	Altres_espai	ANTIC CEIP	CAfores	s/n	25651	973252378	1.162.906.43	42.171.550.5	<Point><coordinates>1.162906430,42.171550579,0.0</cc	
3	Abella de la	Pallars Jussà	Pirineus	Esport_i	lleu	Instal·lacions	Area_d_activ	AREA ESCAL	Abella de la		25651		0	1.089.229.62	42.160.953.5	<Point><coordinates>1.089229625,42.160953574,0.0</cc
4	Abrera	Baix Llobregat	Costa Barcel	Administrac	Ens_Locals	Ajuntaments	AJUNTAMEN	de la Constit			8630	937700325	1.901.306.68	41.516.031.9	<Point><coordinates>1.901306682,41.516031966,0.0</cc	
5	Abrera	Baix Llobregat	Costa Barcel	Comerc	Con	Mercats_mu	Mercats_mu	MERCAT MUI	del Torren		8630		0	1.899.664.57	41.517.179.3	<Point><coordinates>1.899664577,41.517179361,0.0</cc
6	Abrera	Baix Llobregat	Costa Barcel	Cultura	AteneusCen	AteneusCen	ASSOCIACIÓ	Lluís Compai			8630		1	1.900.829.50	41.515.920.1	<Point><coordinates>1.900829508,41.515920195,0.0</cc
7	Abrera	Baix Llobregat	Costa Barcel	Cultura	Biblioteques	publiques	BIBLIOTECA	Federico Gar			8630	937700881	1.901.321.94	41.521.527.4	<Point><coordinates>1.901321940,41.521527401,0.0</cc	
8	Abrera	Baix Llobregat	Costa Barcel	Educacio	Ensenyamen	Musica	ESCOLA DE	Mp, del Rebat			8630	937702702	1.896.352.67	41.521.041.5	<Point><coordinates>1.896352671,41.521041540,0.0</cc	
9	Abrera	Baix Llobregat	Costa Barcel	Educacio	Ensenyamen	1r_cicle_educ	LLAR D'INFA	Federico Gar			8630	937703458	1.900.534.70	41.521.403.6	<Point><coordinates>1.900534700,41.521403686,0.0</cc	
10	Abrera	Baix Llobregat	Costa Barcel	Educacio	Ensenyamen	1r_cicle_educ	LLI PRIVADA	de la Purlo			8630		0	1.895.269.03	41.522.367.0	<Point><coordinates>1.895269031,41.522367070,0.0</cc

### II-lustració 15 Dades finals d'equipaments

El fitxer resultant queda així:

equipaments.csv

fitxer	equipaments.csv		
	taula de 31064 files i 13 columnes		
detall de camps	<b>camp</b>	<b>tipus/mida</b>	
	<i>municipi</i>	text:	[3:43]
	<i>comarca</i>	text:	[5:17]
	<i>marca</i>	text:	[8:33]
	<i>nivell1</i>	text:	[5:30]
	<i>nivell2</i>	text:	[6:74]

	<i>nivell3</i>	text: [3:75]	
	<i>nom</i>	text: [2:130]	
	<i>adreca</i>	text: [1:112]	
	<i>cp</i>	text: [4:5]	
	<i>telefon</i>	text: [1:10]	
	<i>longitud</i>	text: [13]	
	<i>latitud</i>	text: [14]	
	<i>location</i>	text: [70]	

### 2.4.2.2 Infraestructura turística

S'eliminarà la taula d'índex i es desdoblen les taules en dos parts, una per als establiments i una altra per a les places. Tot i que aquesta segregació no era necessària, es va valorar treballar amb grups més petits de dades per tal de detectar possibles errades. S'aplanen cadascuna de les dos parts agrupant els camps en una sola taula, eliminant les columnes de totals i afegint una columna al final de cada sèrie amb l'any corresponent. S'eliminen les capçaleres innecessàries, files de separació i les files de totals. Com sigui que el número de camps és diferent per a cada tipus d'establiment, tant la taula de places com la d'establiments es desdoblaran en tres noves taules temporals, una per a cada tipus, que ha de servir per a agrupar les dades en dos noves taules amb l'estructura:

municipi	tipus	categoria	any	places
----------	-------	-----------	-----	--------

municipi	tipus	categoria	any	establiments
----------	-------	-----------	-----	--------------

Aquest pas és va fer amb una macro senzilla.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
municipi	d'1 estr	de 2 estr	de 3 estr	de 4 estr	de 5 estr	pensions	any	municipi	Luxe	Primera	Segona	Tercera	any	
Abella de la C	0	0	0	0	0	0	2013	Abella de la C	0	0	0	0	2	
Abdera	39	0	0	0	0	0	2013	Abdera	0	0	0	0	2	
Ager	27	30	0	0	0	0	14	2013	Ager	0	0	582	0	2
Agramunt	18	36	0	0	0	0	0	2013	Agramunt	0	0	0	0	2
Aguilar de Se	0	0	0	0	0	0	0	2013	Aguilar de Se	0	0	0	0	2
Agullana	0	0	0	0	0	0	0	2013	Agullana	0	0	0	0	2
Aiguafreda	0	0	0	0	0	0	0	2013	Aiguafreda	0	0	0	0	2
Aiguamúrcia	8	26	0	30	0	0	0	2013	Aiguamúrcia	0	0	0	0	2
Aiguaviva	0	0	0	0	0	0	0	2013	Aiguaviva	0	0	0	0	2
Aitona	0	0	0	0	0	0	0	2013	Aitona	0	0	0	0	2
Altamús, els	0	0	0	0	0	0	0	2013	Altamús, els	0	0	0	0	2
Alàs i Cerc	0	0	23	0	0	0	0	2013	Alàs i Cerc	0	0	0	0	2
Albagés, l'	0	0	0	0	0	0	0	2013	Albagés, l'	0	0	0	0	2
Albanyà	0	0	0	0	0	0	12	2013	Albanyà	0	318	0	0	2
Albatàrec	0	0	0	0	0	0	0	2013	Albatàrec	0	0	0	0	2
Albesa	0	0	0	0	0	0	0	2013	Albesa	0	0	0	0	2
Albi, l'	0	0	0	0	0	0	0	2013	Albi, l'	0	0	0	0	2
Albinyana	0	0	20	0	0	0	0	2013	Albinyana	0	0	0	0	2
Albiol, l'	0	0	0	0	0	0	0	2013	Albiol, l'	0	0	0	0	2
Albons	0	0	0	64	0	0	0	2013	Albons	0	0	0	0	2
Alcanar	0	0	100	0	0	0	62	2013	Alcanar	0	0	1572	0	2
Alcanó	0	0	0	0	0	0	0	2013	Alcanó	0	0	0	0	2
Alcarràs	150	40	0	0	0	0	0	2013	Alcarràs	0	0	0	0	2
Alcoletge	0	0	0	0	0	0	0	2013	Alcoletge	0	0	0	0	2
Alcover	86	0	0	0	0	0	5	2013	Alcover	0	0	0	0	2
Aldea, l'	0	52	0	0	0	0	18	2013	Aldea, l'	0	0	0	0	2

```

p(7) = 6
z(7) = 7
a(7) = 2013
f(8) = "Full 17"
p(8) = 6
z(8) = 7
a(8) = 2012
f(9) = "Full 18"
p(9) = 6
z(9) = 7
a(9) = 2011
s = 0
For j = 1 To 9
  If j = 1 Then g = 1 Else g = 2
  For c = g To p(j) + 1
    s = s + 1
    For k = 1 To 949
      If c = (p(j) + 1) Then
        n = a(j)
      Else
        n = Sheets(f(j)).Cells(k + z(j) - 3, c).Value
      End If
      Sheets("resum").Cells(k, s).Value = n
    Next k
  Next c
Next j
End Sub

```

II-Il·lustració 16 Transformació d'infraestructura

En un segon pas, també amb una macro molt simple, es va generar la taula ordenada amb les dades desitjades.

	A	B	C	D	E	F	G	H
1	marca	comarca	municipi	tipus	categoria	any	places	establiments
2	Barcelona	Barcelonès	Badalona	camping	Luxe	2011	0	0
3	Barcelona	Barcelonès	Badalona	camping	Primera	2011	0	0
4	Barcelona	Barcelonès	Badalona	camping	Segona	2011	0	0
5	Barcelona	Barcelonès	Badalona	camping	Tercera	2011	0	0
6	Barcelona	Barcelonès	Badalona	hotel	d'1 estrella	2011	0	0
7	Barcelona	Barcelonès	Badalona	hotel	de 2 estrelles	2011	0	0
8	Barcelona	Barcelonès	Badalona	hotel	de 3 estrelles	2011	84	1
9	Barcelona	Barcelonès	Badalona	hotel	de 4 estrelles	2011	0	0
10	Barcelona	Barcelonès	Badalona	hotel	de 5 estrelles	2011	0	0
11	Barcelona	Barcelonès	Badalona	hotel	pensions	2011	216	7
12	Barcelona	Barcelonès	Badalona	rural	compartida	2011	0	0
13	Barcelona	Barcelonès	Badalona	rural	independent	2011	0	0
14	Barcelona	Barcelonès	Badalona	rural	Masia	2011	0	0
15	Barcelona	Barcelonès	Badalona	rural	Masoveria	2011	0	0
16	Barcelona	Barcelonès	Barcelona	camping	Luxe	2011	0	0
17	Barcelona	Barcelonès	Barcelona	camping	Primera	2011	0	0
18	Barcelona	Barcelonès	Barcelona	camping	Segona	2011	0	0
19	Barcelona	Barcelonès	Barcelona	camping	Tercera	2011	0	0
20	Barcelona	Barcelonès	Barcelona	hotel	d'1 estrella	2011	2003	33
21	Barcelona	Barcelonès	Barcelona	hotel	de 2 estrelles	2011	3260	36
22	Barcelona	Barcelonès	Barcelona	hotel	de 3 estrelles	2011	13387	109
23	Barcelona	Barcelonès	Barcelona	hotel	de 4 estrelles	2011	29526	125
24	Barcelona	Barcelonès	Barcelona	hotel	de 5 estrelles	2011	9725	24
25	Barcelona	Barcelonès	Barcelona	hotel	pensions	2011	5401	213
26	Barcelona	Barcelonès	Barcelona	rural	compartida	2011	0	0

```

colany(2) = 6
colany(3) = 6
pricol(1) = 1
pricol(2) = 10
pricol(3) = 17
j = 1
'per als 3 tipus
For a = 1 To 3
  tipo = tipus(a)
  'per a cada categoria
  For b = 1 To numcol(a)
    cate = Sheets(10).Cells(1, pricol(a) + b)
    'per a cada registre
    For c = 2 To 2842
      muni = Sheets(10).Cells(c, pricol(a))
      anyo = Sheets(10).Cells(c, pricol(a) + colany(a) - 1)
      places = Sheets(10).Cells(c, pricol(a) + b)
      j = j + 1
      Sheets(11).Cells(j, 1) = muni
      Sheets(11).Cells(j, 2) = tipo
      Sheets(11).Cells(j, 3) = cate
      Sheets(11).Cells(j, 4) = anyo
      Sheets(11).Cells(j, 5) = places
    Next c
  Next b
Next a
End Sub

```

II-il·lustració 17 Agrupació de dades a infraestructura

Per últim, es consolida tot amb una nova taula i, a partir de les taules de comarques i de marques turístiques, s'afegeix les columnes de comarca i marca. Amb això es té la següent estructura preparada per al procés de càrrega de dades. En el moment de la càrrega, per tenir una dimensió temporal comuna amb la taula d'equipaments, s'afegeix la columna de mes/Temporada. És un camp fictici que s'ompli amb el promig mensual de cada registre.

marca	comarca	municipi	tipus	categoria	any	places	establiments
-------	---------	----------	-------	-----------	-----	--------	--------------

El fitxer quedaria així:

infraestructura turística

<b>fitxer</b>	infraestructura turística		
	taula de 39775 files i 8 columnes		
<b>detall de camps</b>	<b>camp</b>	<b>tipus/mida/valor</b>	<b>comentaris</b>
	<i>municipi</i>	text: [3:43]	
	<i>comarca</i>	text: [5:17]	
	<i>marca</i>	text: [8:33]	
	<i>tipus</i>	text:[5:7]	
	<i>categoria</i>	text:[4:14]	
	<i>any</i>	enter:[2011:2013][4:4]	
	<i>places</i>	enter:[1:5][0:33695]	
<i>establiments</i>	enter: [1:3][0:238]		

Aquest fitxer es passa a format csv i es deixa enllestit per al procés de càrrega de dades de la fase posterior.

2.4.2.3 Policia local

S'eliminarà la taula d'índex. Aplanarem el fitxer agrupant els camps en una sola taula, eliminant les columnes de graduació i la columna de ràtio i afegint una columna al final de cada registre amb l'any corresponent. S'eliminen les capçaleres innecessàries, files de separació i les files de totals. Tot i que les graduacions no són necessàries, també es podria plantejar deixar-les per a un futur. S'afegeixen els camps de marca i comarca, provinents dels camps de marca i comarca provinents del fitxer de zones a partir dels fitxers de marques turístiques i comarques i del fitxer comarques i municipis.

Código	Controles	XML	Modificar										
B6													
<pre> =BUSCARV(BUSCARV(BUSCARV(C6;C:\Users\Usuari\Documents\uoc\datawarehouse\pac2\comarques-municipis.xlsx)[comarques!\$B\$2:\$C\$948;2;FALSO]);C:\Users\Usuari\Documents\uoc\datawarehouse\pac2\comarques-municipis.xlsx)[comarques!\$K\$2:\$L\$948;2;FALSO];C:\Users\Usuari\Documents\uoc\datawarehouse\pac2\comarques-municipis.xlsx)[comarques!\$F\$2:\$G\$942;2;FALSO] </pre>													
A	B	C	L	N	O	P	Q	R	S	T	U	V	W
1	marca	comarca	municipi	total	any								
2	Pinneus	Pallars Jussà	Abella de la Conca	0	2013								
3	Costa Barcelona	Baix Llobregat	Abrebra	21	2013								
4	Terres de Lleida	Noguera	Ager	0	2013								
5	Terres de Lleida	Urgell	Agramunt	7	2013								
6	Catalunya Central	Bages	Agullar de Segarra	0	2013								
7	Costa Brava	Alt Empordà	Agullana	0	2013								
8	Catalunya Central	Valles Oriental	Aiguafreda	0	2013								
9	Costa Daurada	Alt Camp	Aiguamúrcia	0	2013								
10	Costa Brava	Gironès	Aiguaviva	0	2013								
11	Terres de Lleida	Segrià	Aitona	0	2013								
12	Terres de Lleida	Segrià	Alamús, els	0	2013								
13	Pinneus	Alt Urnell	Alàs i Cerc	0	2013								

II-Il·lustració 18 Taula d'efectius de la policia local

L'estructura serà:

marca	comarca	municipi	efectius	any
-------	---------	----------	----------	-----

La taula quedaria així preparada per al següent procés de càrrega de dades:

policia local.xls

fitxer	policia local.xls		
	taula de 2842 files i 5 columnes		
detall de camps	camp	tipus/mida/valor	comentaris
	<i>municipi</i>	text: [3:43]	
	<i>comarca</i>	text: [5:17]	
	<i>marca</i>	text: [8:33]	
	<i>any</i>	enter: [4:4] [2011:2013]	l'any 2011 no està present a totes les taules
	<i>efectius</i>	enter: [1:4] [0:2857]	

#### 2.4.2.4 Pernoctacions

S'eliminen les files descriptives i de separació. Es substitueixen els valors nuls de l'any 2012 per el valor de l'anterior més/menys la variació entre els dos anys al conjunt de les altres zones. Es transposa el contingut, es crea un camp d'any i un camp de mes i per a obtenir aquesta estructura:

any	mes	marca	pernoctacions	procedència	actes
-----	-----	-------	---------------	-------------	-------

on marca són les marques turístiques, pernoctacions pot ser viatgers o pernoctacions, residència pot ser espanyol o estranger, actes són el número de pernoctacions o viatgers per període mensual, any i mes són els del període estudiat. Per normalitzar les dades de pernoctacions respecte les zones, es farà una distribució a partir de les dades de places disponibles per municipi: per a cada municipi es calcula "total de les places del municipi \* (total de pernoctacions del període a totes les marques per mes / total de places disponibles entre tots els municipis aquest any)". El mateix per als viatgers (tot i que aquest també es podria fer per establiments): per a cada municipi es calcula "total de les places del municipi \* (total de viatgers del període a totes les marques per mes / total de places disponibles entre tots municipis aquest any)". No s'han considerat les dades de pernoctacions del 2014 perquè no tenim dades d'establiments d'aquest any. En canvi, les dades de l'any 2011 de pernoctacions i





dimensions, compensaria el crear alguna mena de programa per tal d'automatitzar el procés.

80	Barcelona	viajeros	estangers	2012M08	548.123	2012	08													
81	Barcelona	viajeros	estangers	2012M07	567.520	2012	07													
82	Barcelona	viajeros	estangers	2012M06	492.513	2012	06													
83	Barcelona	viajeros	estangers	2012M05	493.299	2012	05													
84	Barcelona	viajeros	estangers	2012M04	503.437	2012	04													
85	Barcelona	viajeros	estangers	2012M03	415.489	2012	03													
86	Barcelona	viajeros	estangers	2012M02	325.605	2012	02													
87	Barcelona	viajeros	estangers	2012M01	261.718	2012	01													
88	Barcelona	pernoctacions	espanyols	2014M12	260.595	2014	12													
89	Barcelona	pernoctacions	espanyols	2014M11	246.432	2014	11													
90	Barcelona	pernoctacions	espanyols	2014M10	240.116	2014	10													
91	Barcelona	pernoctacions	espanyols	2014M09	246.484	2014	09													
92	Barcelona	pernoctacions	espanyols	2014M08	227.196	2014	08													
93	Barcelona	pernoctacions	espanyols	2014M07	266.764	2014	07													
94	Barcelona	pernoctacions	espanyols	2014M06	237.667	2014	06													
95	Barcelona	pernoctacions	espanyols	2014M05	234.132	2014	05													
96	Barcelona	pernoctacions	espanyols	2014M04	243.827	2014	04													
97	Barcelona	pernoctacions	espanyols	2014M03	266.519	2014	03													
98	Barcelona	pernoctacions	espanyols	2014M02	200.063	2014	02													

II-Il·lustració 21 Transposició de files a pernoctacions

Després, es va procedir a fer una fusió de les dades que interessaven amb el fitxer d'establiments, concretament els camps d'any, marca, comarca, municipi i places. En aquesta primera aproximació les dades encara són anuals. Com que caldrà encabir els mesos, es va fer servir una macro senzilla per tal de generar l'espai i omplir les dades.

II-Il·lustració 22 Annexió de dades i inserció de files a pernoctacions

Fet el lloc, es passa a omplir les dades també amb una macro molt simple.

II-Il·lustració 23 Valors bàsics de les files mensuals a pernoctacions

Per últim, s'incorporen els mesos i es modifica la fórmula per tal que el càlcul de les pernoctacions i viatgers de cada municipi respecte al nombre de places sigui mensual. La fórmula relaciona el total de pernoctacions del període de totes les marques entre el

total de places disponibles de totes les marques i pel total de les places del municipi. Cal senyalar que les places són les mateixes tot l'any (no hi ha dades mensuals).

Código		Controles			XML		Modificar							
G5		=ENTERO((SUMAR.SI.CONJUNTO(Hoja1!\$E\$16:\$E\$1311;Hoja1!\$B\$16:\$B\$1311;Hoja2!\$G\$1;Hoja1!\$F\$16:\$F\$1311;Hoja2!\$A\$5;Hoja1!\$G\$16:\$G\$1311;Hoja2!\$B\$5)/SUMAR.SI.CONJUNTO(\$F\$2:\$F\$34093;\$A\$2:\$A\$34093;\$A\$5;\$B\$2:\$B\$34093;\$B\$5))*\$F\$5)												
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	any	mes	marca	comarca	municipi	places	pernoctacions	viajeros						
2	2014		12 Barcelona	Barcelonès	Badalona		528	1770	833					
3	2014		11 Barcelona	Barcelonès	Badalona		528	1807	791					
4	2014		10 Barcelona	Barcelonès	Badalona		528	3697	1330					
5	2014		9 Barcelona	Barcelonès	Badalona		528	5395	1544					
6	2014		8 Barcelona	Barcelonès	Badalona		528	7527	2036					
7	2014		7 Barcelona	Barcelonès	Badalona		528	6424	1801					
8	2014		6 Barcelona	Barcelonès	Badalona		528	5177	1638					
9	2014		5 Barcelona	Barcelonès	Badalona		528	4148	1469					

II-Il·lustració 24 Adaptació de la fórmula mensual a pernoctacions

El resultat d'aquest full es passa a format csv i es deixa preparat per procedir a la càrrega de dades en la següent fase.

La taula quedaria així:

pernoctacions.csv

fitxer	pernoctacions.csv		
	taula de 34093 files i 8 columnes		
detall de camps	camp	tipus/mida/valor	comentaris
	any	enter:[4:4][2012:2014]	l'any 2014 no està present a totes les taules
	mes	enter:[1:2][1:12]	
	municipi	text: [3:43]	
	comarca	text: [5:17]	
	marca	text: [8:33]	
	places	enter:[1:5][0:69128]	
	pernoctacions	enter:[1:5][0:80802]	
viatgers	enter:[1:5][0:21300]		

#### 2.4.2.5 Habitants per municipi

S'eliminarà la taula d'índex. Aplanarem el fitxer agrupant els camps en una sola taula, mantenint la columna de total, eliminant les columnes de sexe. S'eliminen les capçaleres innecessàries, files de separació i les files de totals. S'afegeixen els camps de marca i comarca. L'estructura serà:

any	marca	comarca	municipi	habitants
-----	-------	---------	----------	-----------

La taula quedaria així:

habitants\_per\_municipis.csv

fitxer	habitants_per_municipis.csv		
	taula de 3789 files i 5 columnes		
detall de camps	camp	tipus/mida/valor	comentaris
	any	enter:[4:4][2011:2014]	el rang d'anys és superior al de la resta de taules

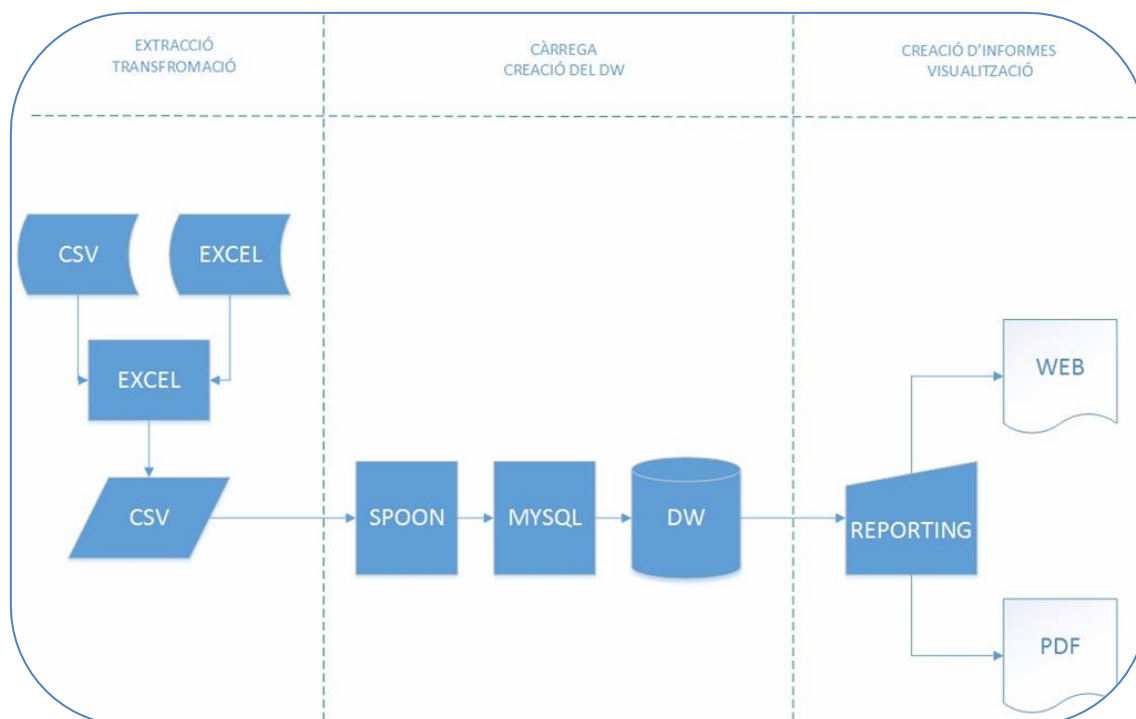
	<i>municipi</i>	text: [3:43]	
	<i>comarca</i>	text: [5:17]	
	<i>marca</i>	text: [8:33]	
	<i>habitants</i>	enter:[2:7][19:1620943]	

### 3. IMPLEMENTACIÓ

En aquesta fase es procedeix a la creació i càrrega de dades de les taules de la base de dades des dels fitxers de dades temporals preparats en l'etapa anterior. Fet el DW, es preparen i mostren els informes predefinits segons els requeriments del projecte.

#### 3.1 Arquitectura

En la següent il·lustració es mostra l'arquitectura de procés del projecte. Es parteix de fonts csv (format de text estructurat) i xls (full de càlcul Excel de Microsoft Office) per efectuar les transformacions amb el full de càlcul Excel. Per portabilitat es torna a passar al format csv (amb l'Spoon no caldria, però s'ha fet una petita abstracció i s'ha considerat que el format csv és més compatible) i es porta a l'entorn de [Pentaho](#). Amb l'eina [Kettle - Spoon de Pentaho](#) es creen i carreguen els csv a fitxers de [MySQL](#). A partir d'aquests fitxers mestres, des de [MySQL WorkBench](#) es crea i s'omple de dades l'estrella del magatzem de dades. Amb les dades disponibles, ja es pot procedir a crear els informes amb l'eina [Pentaho Report Designer](#) i publicar-los.



Il·lustració 25 Arquitectura

#### 3.2 Càrrega de fitxers

La càrrega dels fitxers al DW es farà amb l'eina Spoon. Amb aquesta eina també es podria fer el procés de transformació dels fitxers per preparar-los per a la càrrega. Com ja s'ha explicat prèviament, per fer els processos de transformació s'han utilitzat les funcions de l'Excel i s'han creat noves macros amb *vbexcel*. D'aquesta manera, els

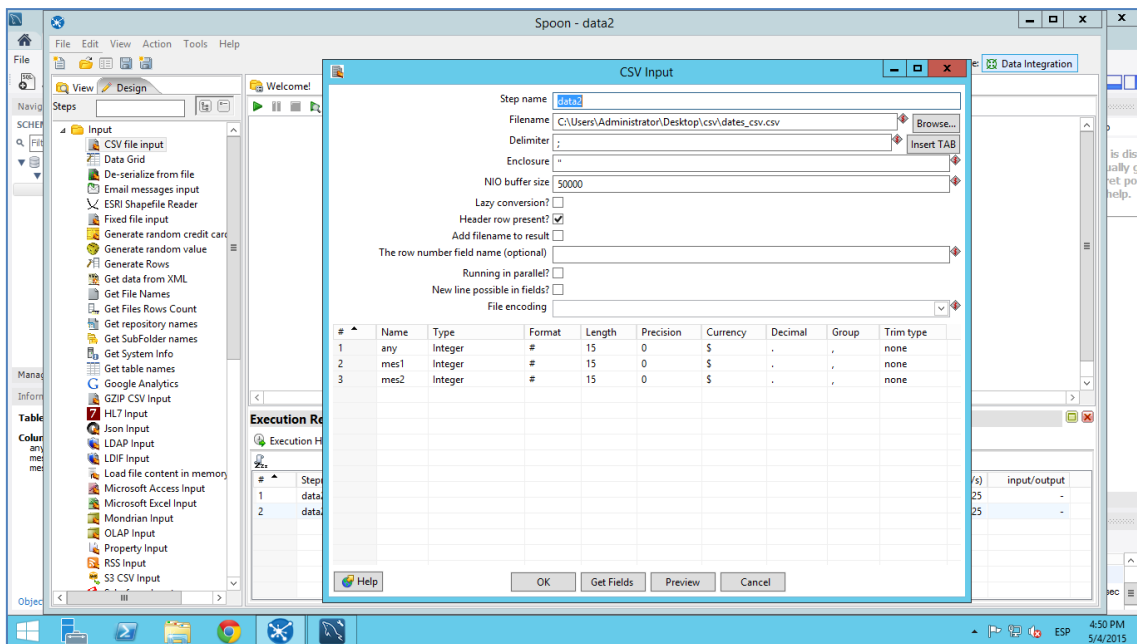
fitxers que ara es carreguen serviran per a donar les dades a la base de dades; després, es generaran els fitxers de dimensions i de fets.

### 3.2.1 Fitxers de dades

Tots els fitxers s'han convertit a format csv i s'han pujat a una carpeta anomenada "csv" creada a tal efecte a l'escriptori de l'entorn de treball. D'aquesta manera, el format és uniforme i es simplifica el procés de càrrega.

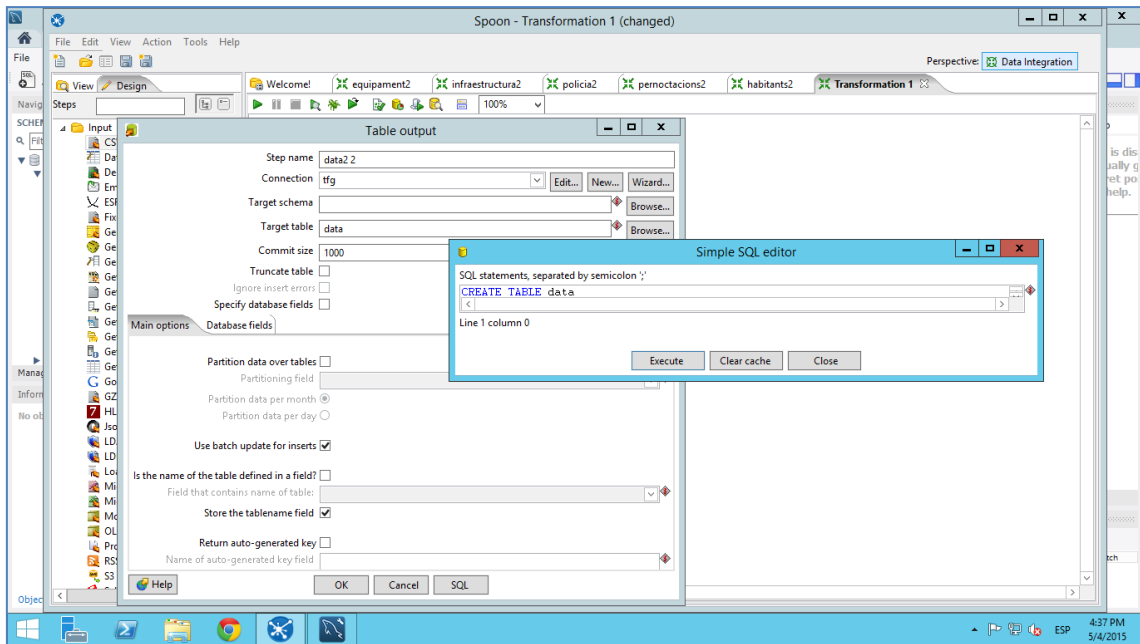
El procés de càrrega de dades és exactament igual per a tots els fitxers. Aquí s'il·lustra amb el fitxer de dates "dates.csv". L'eina emprada és Spoon, de la suite Pentaho. Els fitxers creats seran del tipus Sql, concretament MySql.

En primer lloc, des de l'Spoon es prepara el fitxer csv a importar, especificant la localització i el tipus de dades tal com s'ha detallat en la descripció de l'estructura física dels fitxers. Aquest pas, comú a la resta de fitxers a carregar, genera un fitxer MySql des d'on es procedirà a crear els fitxers de l'estrella del magatzem de dades. Així, doncs, és un pas intermedi previ a la creació de les taules de treball definitives.



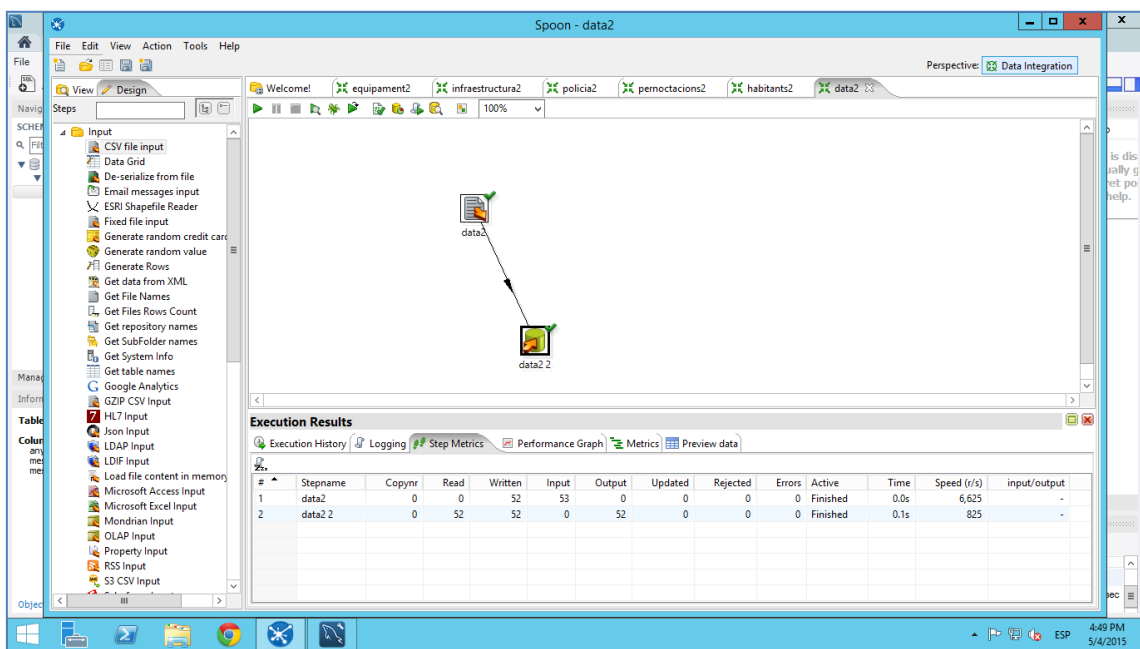
II-l·lustració 26 Captura del fitxer csv amb l'Spoon

Es prepara la connexió i es crea la taula dates. Aquest procés es fa automàticament des de l'Spoon i genera el fitxer sql que rebrà les dades de dates.csv



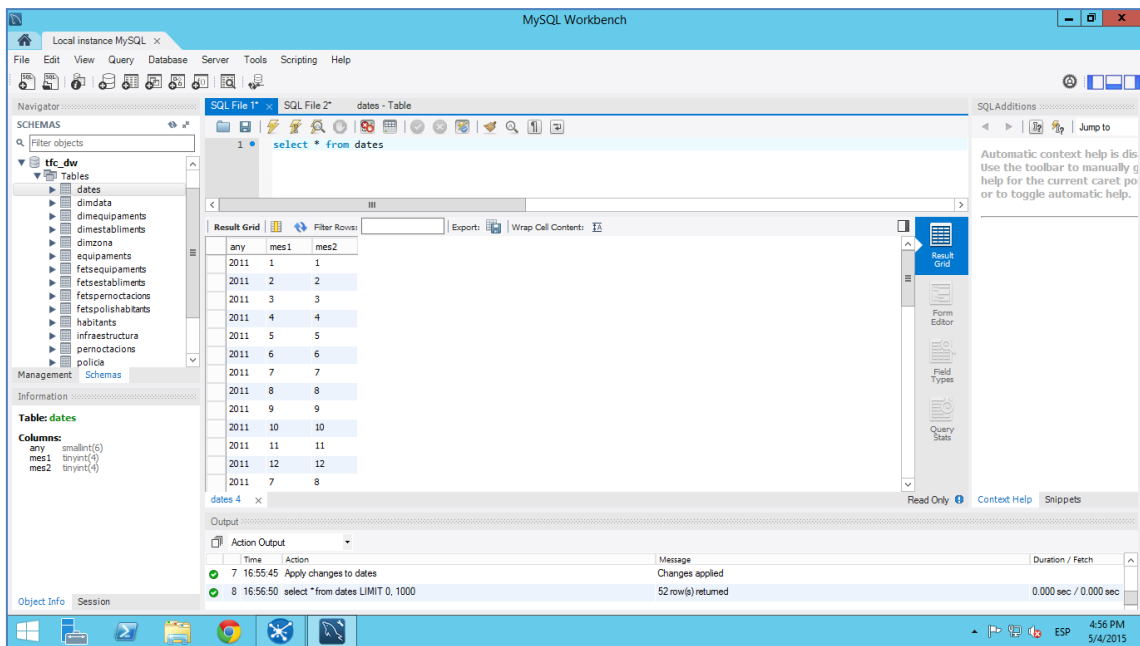
II-Il·lustració 27 Creació de la taula de dades

Finalment, executem la càrrega. El mateix Spoon ens indica el nombre de registres creats i, si és el cas, els errors ocorreguts en el procés.



II-Il·lustració 28 Execució de la càrrega de dades

Després de cada acció, s'obri des del Workbench de MySql el fitxer creat i se'n comprova el resultat.



II-Il·lustració 29 Comprovació de la creació del fitxer i la càrrega de dades

Aquests passos es repeteixen, a més del fitxer de dates exposat, per als fitxers d'infraestructura turística, equipaments, policia local, habitants per municipi i pernoctacions.

### 3.2.2 Fitxers de dimensió i de fets

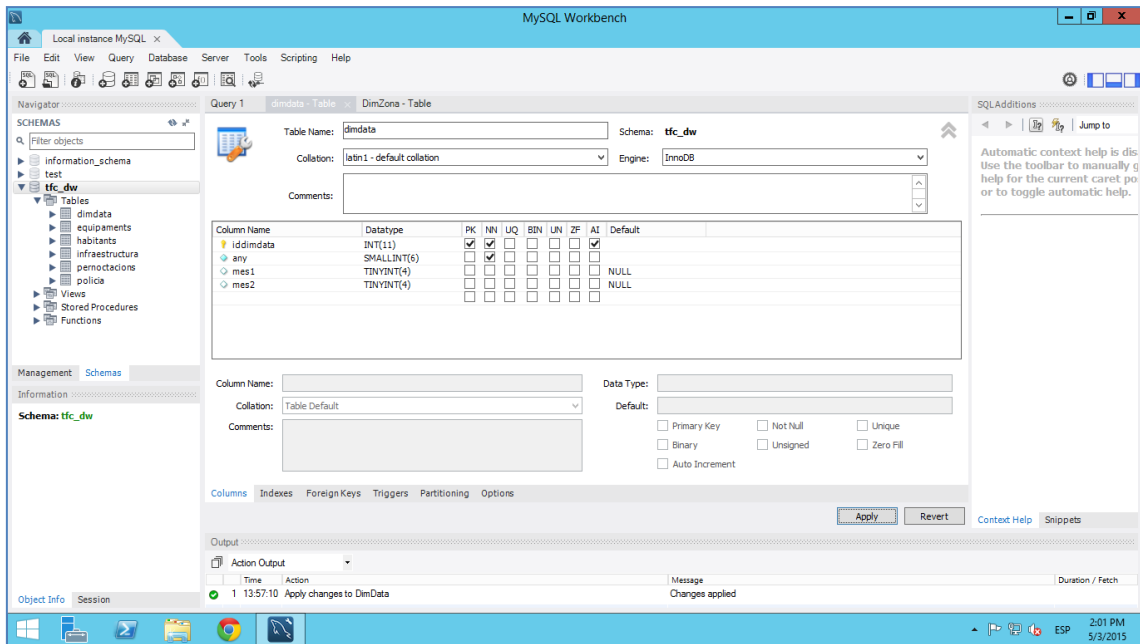
Finalitzada aquesta càrrega des dels fitxers origen ja depurats d'errors, es procedeix a generar les taules segons el disseny físic plantejat.

En primer lloc crearem les taules de dimensió i de fets des del MySql Workbench.

#### 3.2.2.1 Fitxers de dimensió

En aquest apartat es mostren les sentències *sql* emprades per a la creació de les diferents taules de dimensions. Com exemple es mostra la imatge del Workbench corresponent a la taula de dates. Per a la resta de fitxers, tan sols es mostra la instrucció corresponent.





II-Il·lustració 30 Creació de la taula DimData

Sentència de creació de la taula DimData:

```
CREATE TABLE `tfc_dw`.`dimdata` (
  `iddimdata` INT NOT NULL AUTO_INCREMENT,
  `any` SMALLINT NOT NULL,
  `mes1` TINYINT NULL,
  `mes2` TINYINT NULL,
  PRIMARY KEY (`id`));
```

Sentència de creació de la taula DimZona:

```
CREATE TABLE `tfc_dw`.`dimzona` (
  `idDimZona` INT NOT NULL AUTO_INCREMENT,
  `marca` VARCHAR(50) NOT NULL,
  `comarca` VARCHAR(50) NOT NULL,
  `municipi` VARCHAR(50) NOT NULL,
  PRIMARY KEY (`idDimZona`));
```

Sentència de creació de la taula DimEstabliments:

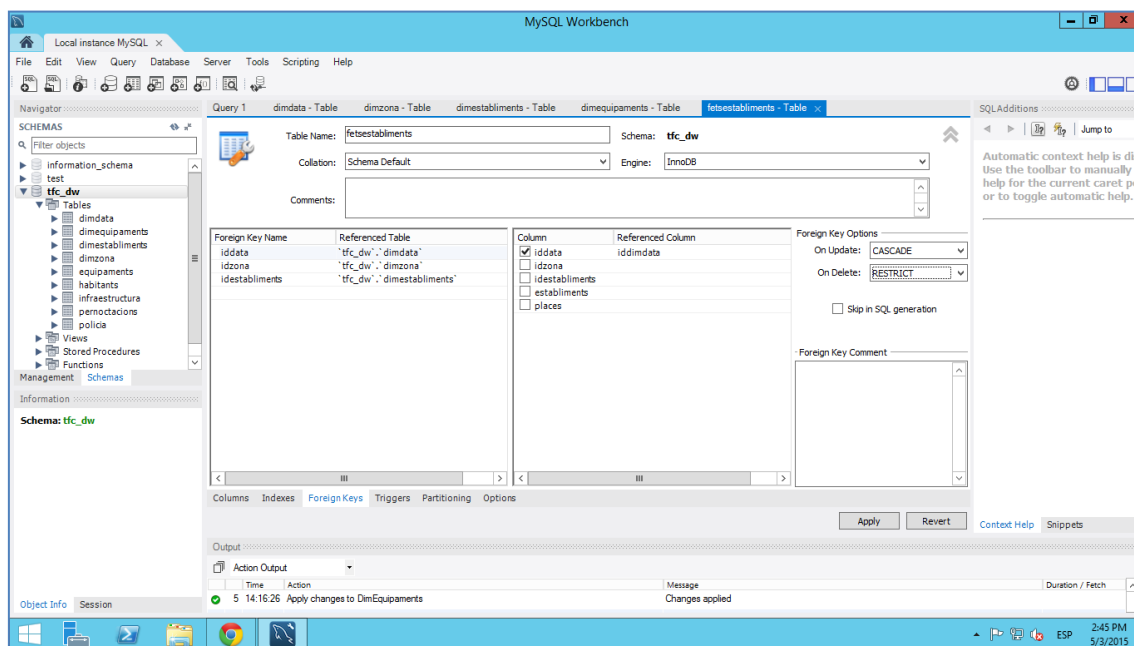
```
CREATE TABLE `tfc_dw`.`dimestabliments` (
  `idDimEstabliments` INT NOT NULL AUTO_INCREMENT,
  `tipus` VARCHAR(25) NOT NULL,
  `categoria` VARCHAR(25) NOT NULL,
  PRIMARY KEY (`idDimEstabliments`));
```

Sentència de creació de la taula DimEquipaments:

```
CREATE TABLE `tfc_dw`.`dimequipaments` (
  `iddimEquipaments` INT NOT NULL AUTO_INCREMENT,
  `nivell1` VARCHAR(50) NOT NULL,
  `nivell2` VARCHAR(100) NOT NULL,
  `nivell3` VARCHAR(100) NULL,
  PRIMARY KEY (`iddimEquipaments`));
```

### 3.2.2.2 Fitxers de fets

De manera similar, es procedeix amb les taules de fets. En aquest cas, però, també es creen relacions entre les taules, per tant cal definir, a més, aquestes relacions i les claus foranies. Com exemple, es mostra la pantalla de creació de les claus foranes i índexs de la taula FetsEstabliments. Segons el model presentat, aquesta taula es relaciona amb les dimensions de data, zona i establiments.



Il·lustració 31 Creació d'índex i FK de la taula FetsEstabliments

La sentència *sql* de creació de la taula i relacions:

```
CREATE TABLE `tfc_dw`.`fetsestabliments`(  
  `iddata` INT NOT NULL,  
  `idzona` INT NOT NULL,  
  `idestabliments` INT NOT NULL,  
  `establiments` INT NULL DEFAULT 0,  
  `places` INT NULL DEFAULT 0,  
  PRIMARY KEY (`iddata`, `idzona`, `idestabliments`),  
  INDEX `idzona_idx` (`idzona` ASC),  
  INDEX `idestabliments_idx` (`idestabliments` ASC),  
  INDEX `iddata_idx` (`iddata` ASC),  
  CONSTRAINT `iddata`  
    FOREIGN KEY (`iddata`)  
    REFERENCES `tfc_dw`.`dimdata` (`iddimdata`)  
    ON DELETE RESTRICT  
    ON UPDATE CASCADE,  
  CONSTRAINT `idzona`  
    FOREIGN KEY (`idzona`)  
    REFERENCES `tfc_dw`.`dimzona` (`iddimzona`)  
    ON DELETE RESTRICT  
    ON UPDATE CASCADE,  
  CONSTRAINT `idestabliments`  
    FOREIGN KEY (`idestabliments`)
```

```
REFERENCES `tfc_dw`.`dimestabliments` (`idDimEstabliments`)
ON DELETE RESTRICT
ON UPDATE CASCADE);
```

Sentència *sql* de creació i relacions de la taula FetsEquipaments (segons el model presentat, aquesta taula es relaciona amb les dimensions de zona i equipaments):

```
CREATE TABLE `tfc_dw`.`fetsequipaments` (
  `idzona` INT NOT NULL,
  `idequipaments` INT NOT NULL,
  `equipaments` SMALLINT NULL DEFAULT 0,
  PRIMARY KEY (`idzona`, `idequipaments`),
  INDEX `idequipaments_idx` (`idequipaments` ASC),
  INDEX `idzona_idx` (`idzona` ASC),
  CONSTRAINT `idzona_fk`
    FOREIGN KEY (`idzona`)
    REFERENCES `tfc_dw`.`dimzona` (`idDimZona`)
    ON DELETE RESTRICT
    ON UPDATE CASCADE,
  CONSTRAINT `idequipaments_fk`
    FOREIGN KEY (`idequipaments`)
    REFERENCES `tfc_dw`.`dimequipaments` (`iddimEquipaments`)
    ON DELETE RESTRICT
    ON UPDATE CASCADE);
```

Sentència *sql* de creació i relacions de la taula FetsPernoctacions (segons el model presentat, aquesta taula es relaciona amb les dimensions de zona i dates):

```
CREATE TABLE `tfc_dw`.`fetspernoctacions` (
  `iddata` INT NOT NULL,
  `idzona` INT NOT NULL,
  `pernoctacions` INT NULL DEFAULT 0,
  `viatgers` INT NULL DEFAULT 0,
  PRIMARY KEY (`iddata`, `idzona`),
  INDEX `idzona_fk_per_idx` (`idzona` ASC),
  INDEX `iddata_fk_per_idx` (`iddata` ASC),
  CONSTRAINT `iddata_fk_per`
    FOREIGN KEY (`iddata`)
    REFERENCES `tfc_dw`.`dimdata` (`iddimdata`)
    ON DELETE RESTRICT
    ON UPDATE CASCADE,
  CONSTRAINT `idzona_fk_per`
    FOREIGN KEY (`idzona`)
    REFERENCES `tfc_dw`.`dimzona` (`idDimZona`)
    ON DELETE RESTRICT
    ON UPDATE CASCADE);
```

Sentència *sql* de creació i relacions de la taula FetsPolisHabitants (segons el model presentat, aquesta taula es relaciona amb les dimensions de zona i dates):

```
CREATE TABLE `tfc_dw`.`fetspolishabitants` (
  `iddata` INT NOT NULL,
  `idzona` INT NOT NULL,
  `efectius` SMALLINT NULL DEFAULT 0,
  `habitants` INT NULL DEFAULT 0,
```

```

PRIMARY KEY (`iddata`, `idzona`),
INDEX `idzona_fk_poli_idx` (`idzona` ASC),
INDEX `iddata_fk_poli_idx` (`iddata` ASC),
CONSTRAINT `iddata_fk_poli`
  FOREIGN KEY (`iddata`)
  REFERENCES `tfc_dw`.`dimdata` (`iddimdata`)
  ON DELETE RESTRICT
  ON UPDATE CASCADE,
CONSTRAINT `idzona_fk_poli`
  FOREIGN KEY (`idzona`)
  REFERENCES `tfc_dw`.`dimzona` (`idDimZona`)
  ON DELETE RESTRICT
  ON UPDATE CASCADE);

```

### 3.2.2.3 Càrrega de dades de fitxers de dimensions

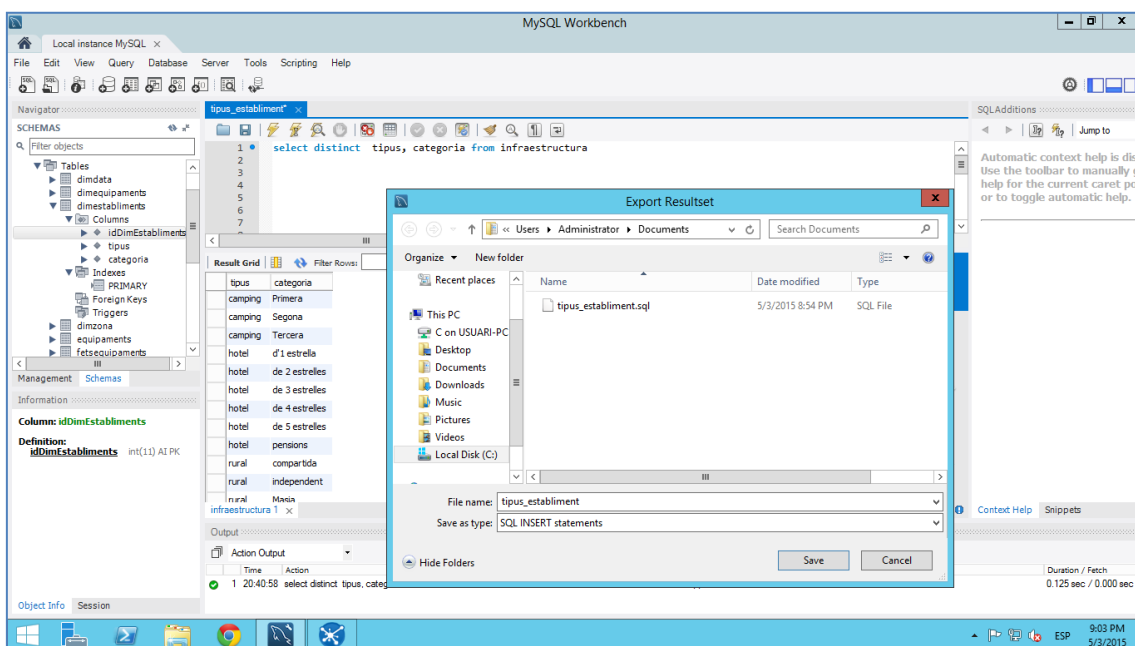
Es procedeix a omplir els fitxers amb les dades de les taules prèviament creades a partir de la importació dels fitxers origen. Això es farà de dos maneres diferents: en la primera, es farà servir la funcionalitat del Workbench per generar una instrucció d'inserció i es mostrarà el procediment amb l'exemple de la taula DimEstabliments; en la segona, es farà a través de la corresponent instrucció *sql*.

#### Taula DimEstabliments

Es fa una consulta sobre la taula infraestructura per extraure i comprovar els valors dels diferents tipus d'establiment i categoria:

```
SELECT DISTINCT tipus, categoria FROM infraestructura
```

Es genera un *SQL INSERT* des del mateix Workbench per omplir la taula de DimEstabliments. Tot i que no és l'única manera de fer-ho, aquest procediment simplifica l'escritura de la sentència *sql*.



II-Il·lustració 32 Generació de la sentència d'inserció a DimEstabliments

```
INSERT INTO `dimestabliments` (`tipus`, `categoria`) VALUES ('camping', 'Luxe');
```

```

INSERT INTO `dimestabliments` (`tipus`,`categoria`) VALUES ('camping','Primera');
INSERT INTO `dimestabliments` (`tipus`,`categoria`) VALUES ('camping','Segona');
INSERT INTO `dimestabliments` (`tipus`,`categoria`) VALUES ('camping','Tercera');
INSERT INTO `dimestabliments` (`tipus`,`categoria`) VALUES ('hotel','1 estrella');
INSERT INTO `dimestabliments` (`tipus`,`categoria`) VALUES ('hotel','2 estrelles');
INSERT INTO `dimestabliments` (`tipus`,`categoria`) VALUES ('hotel','3 estrelles');
INSERT INTO `dimestabliments` (`tipus`,`categoria`) VALUES ('hotel','4 estrelles');
INSERT INTO `dimestabliments` (`tipus`,`categoria`) VALUES ('hotel','5 estrelles');
INSERT INTO `dimestabliments` (`tipus`,`categoria`) VALUES ('hotel','pensions');
INSERT INTO `dimestabliments` (`tipus`,`categoria`) VALUES ('rural','compartida');
INSERT INTO `dimestabliments` (`tipus`,`categoria`) VALUES ('rural','independent');
INSERT INTO `dimestabliments` (`tipus`,`categoria`) VALUES ('rural','Masia');
INSERT INTO `dimestabliments` (`tipus`,`categoria`) VALUES ('rural','Masoveria');

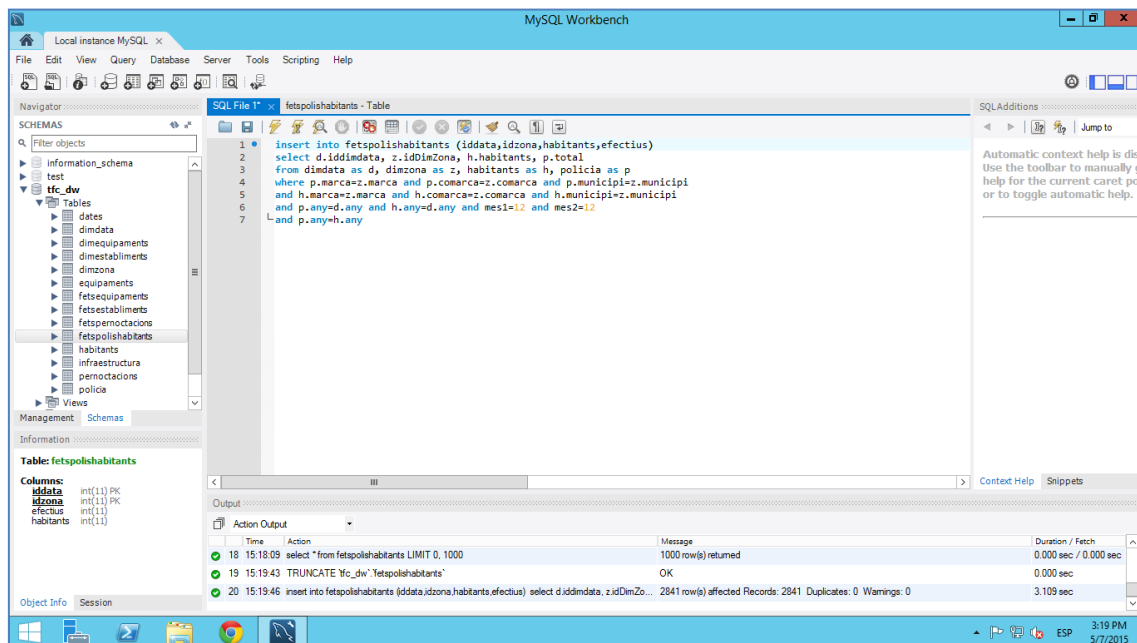
```

De manera similar es procedeix amb les taules de dimensió DimEquipaments i DimZona. En el cas de dimEquipaments, amb la consulta “SELECT DISTINCT nivell1,nivell2,nivell3 FROM equipaments;” per extraure els diferents tipus d’establiments; en el cas de DimZona, amb la consulta “SELECT DISTINCT marca,comarca,municipi FROM habitants;” per extraure els tres nivells de zona.

En canvi, per a la taula DimData es procedeix a fer la inserció directa des de la taula de dates: “INSERT INTO dimdata (any,mes1,mes2) SELECT \* FROM dates;”

### 3.2.2.4 Càrrega de dades de fitxers de fets

Per fer la càrrega de les dades dels fitxers de fets s’han utilitzat sentències *sql* escrites en el Workbench. Es mostra la pantalla d’inserció de dades a la taula FetsPolisHabitants.



II-Il·lustració 33 Inserció de dades a FetsPolisHabitants

La sentència *sql* d’inserció de dades per a FetsPolisHabitants:

```

INSERT INTO fetspolishabitants (iddata,idzona,habitants, efectius)
SELECT d.iddimdata, z.idDimZona, h.habitants, p.total
FROM dimdata AS d, dimzona AS z, habitants AS h, policia AS p
WHERE p.marca=z.marca AND p.comarca=z.comarca AND p.municipi=z.municipi
AND h.marca=z.marca AND h.comarca=z.comarca AND h.municipi=z.municipi
AND p.any=d.any AND h.any=d.any AND d.mes1=12 AND d.mes2=12
AND p.any=h.any

```

Per a la resta de taules de fets, el procediment és similar. Tot seguit es detallen les sentències.

Per a la taula de FetsEstabliments:

```

INSERT INTO fetsestabliments (iddata,idzona,idestabliments,establiments,places)
SELECT d.iddimdata, z.iddimzona, t.iddimestabliments,e.establiments, e.places
FROM dimdata d, dimzona z, dimestabliments t, infraestructura e
WHERE z.marca=e.marca AND z.comarca=e.comarca AND z.municipi=e.municipi
AND e.tipus=t.tipus AND e.categoria=t.categoria
AND e.any=d.any AND d.mes1=12 AND d.mes2=12

```

Per a la taula de FetsPernoctacions:

```

INSERT INTO fetspernoctacions (iddata, idzona,pernoctacions,viatgers)
SELECT d.iddimdata, z.iddimzona, p.pernoctacions, p.viajeros
FROM dimdata d, dimzona z, pernoctacions p
WHERE p.marca=z.marca and p.comarca=z.comarca and p.municipi=z.municipi
AND p.any=d.any AND p.mes=d.mes1 AND p.mes=d.mes2

```

Per a la taula de FetsEquipaments:

```

INSERT INTO FetsEquipaments (idzona,idequipaments,equipaments)
SELECT z.iddimzona, q.iddimequipaments, count(*)
FROM dimequipaments q,dimzona z, equipaments e
WHERE e.municipi=z.municipi AND e.comarca=z.comarca AND e.marca=z.marca
AND e.nivell1=q.nivell1 AND e.nivell2=q.nivell2 AND e.nivell3=q.nivell3
GROUP BY q.iddimEquipaments,z.iddimzona

```

## 3.3 Informes

Principalment, els informes es generen i es publiquen des de l'eina Design Report de la *suite* Pentaho. En aquest apartat es comenta el disseny lògic i físic dels informes, es mostra com són les sentències *sql* que els generen i una captura de pantalla com exemple. La resta de les captures de pantalla dels diferents informes es mostren en l'annex.

A banda, s'ha fet una anàlisi [OLAP](#) (*Online analytical processing*) amb el motor [ROLAP](#) (*Relational Online analytical processing*) Mondrian i el visor Saiku, tots dos de la BI Pentaho, del qual es presenten algunes de les consultes i informes generats. Aquestes consultes es formulen amb el llenguatge de consultes [MDX](#) (*multi-dimensional expressions*) a través del constructor [Schema Workbench](#).

### 3.3.1 Disseny lògic dels informes

Es descriuen en text els diferents informes predefinits que caldrà implementar.

#### Total de pernoctacions estimades per municipi

Es farà el càlcul repartint per places totals; es disposa de la relació de marques amb comarques, aquestes amb municipis i pernoctacions per marques, i tot per anys. Es presenta agrupat i ordenat per municipi per als tres anys amb dades 2012,2013 i 2014.

La consulta és:

```
SELECT z.municipi, d.any, sum(fp.pernoctacions) pernoctacions
FROM fetspernoctacions as fp, dimzona as z, dimdata as d
WHERE fp.iddata=d.iddimdata and fp.idzona=z.idDimZona
GROUP BY d.any, z.municipi
ORDER BY z.municipi, d.any
```

Paral·lelament, s'han implementat alguns informes en un OLAP de Pentaho amb les eines Schema Workbench i Saiku. Es mostraran en l'annex II. Aquest en pot ser una mostra.

Consulta Saiku:

```
SELECT
NON EMPTY {Hierarchize({[Measures].[pernoctes]})} ON COLUMNS,
NON EMPTY CrossJoin([data].[any].Members, [New Dimension 0.New
Hierarchy 0].[New Level 0].Members) ON ROWS
FROM [pernocta1]
```

any	municipi	pernoctes
2012	Abella de la Conca	420
	Abdera	4.584
	Àger	50.734
	Agramunt	6.989
	Aguilar de Segarra	755
	Agullana	1.091
	Aiguafreda	0
	Aiguamúrcia	21.078
	Aiguaviva	0
	Aitona	0
	Alamús, els	0
	Alàs i Cerc	2.777
	Albagés, l'	755

II-lustració 34 Informe 1 OLAP amb Saiku Analytics

#### Promig de viatgers per tipus d'establiment i comarca

Les comarques es relacionen amb municipis i municipis amb tipus d'establiments, viatgers amb zones turístiques i zones turístiques amb comarques. El promig és de viatgers per nombre d'establiments. En la consulta es comprova la divisió per zero per evitar sortides NULL. Es mostren també el resultats iguals a zero per donar tota la informació. Es reparteix el total de viatgers a cada comarca i any pel nombre d'establiments de cada tipus i categoria a cada comarca i any. Les dates disponibles de nombre de viatgers són 2012 a 2014 mentre que d'establiments són de 2011 a 2013, per tant es mostraran els anys 2012 i 2013.

La consulta és:

```
SELECT z.comarca as comarca,
CONCAT(e.tipus," ",e.categoria) as establiment, d.any,
round(if(sum(fe.establiments)>0,sum(fp.viatgers)/sum(fe.establiments),0),2) as
promig
FROM fetsestabliments fe
JOIN fetspernoctacions fp on fp.idzona=fe.idzona
JOIN dimzona z on z.idDimZona=fe.idzona
JOIN dimestabliments e on e.idDimEstabliments=fe.idestabliments
JOIN dimdata d on d.idDimdata=fe.iddata
WHERE fp.iddata=fe.iddata
GROUP BY z.comarca,fe.idEstabliments,d.any
```



### % d'ocupació per marca turística

Es té la relació de marques amb comarques, aquestes amb municipis i municipis amb establiments, viatgers amb marques, i tot per anys . El promig és de pernoctacions per places. Surt de fer pernoctacions/places i any dividit per 365 dies, atès que les pernoctacions són una dada diària. Es presenta en % per marca i any. Disposem de dades 2012 i 2013 atès que no hi ha dades de places per al 2014 ni de pernoctacions per al 2011. D'aquest informe s'ha considerat poder complementar-lo amb una gràfica de línies. S'ha donat en format horitzontal per tal de veure el nom de les marques complet.

La consulta serà:

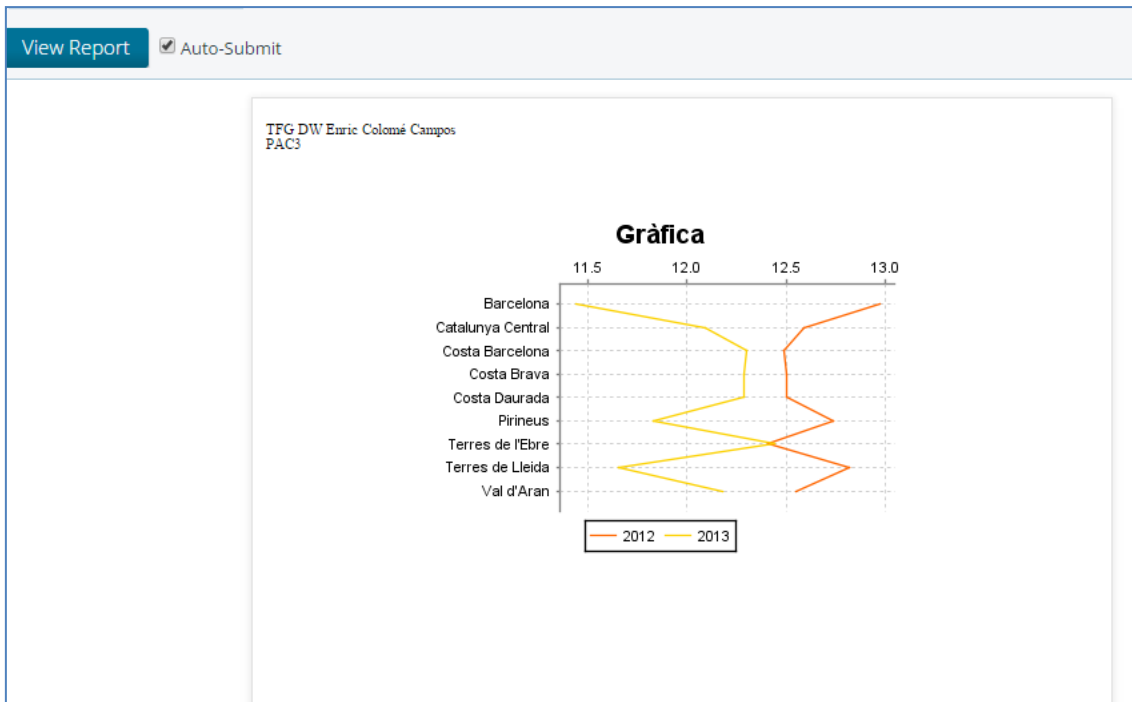
```
SELECT z.marca, d.any,
round((((sum(fp.pernoctacions)/sum(fe.places))/365)*100,2) as percentatge
FROM dimzona z, fetspernoctacions fp, fetsestabliments fe, dimdata d
WHERE z.idDimZona=fp.idzona and z.idDimZona=fe.idzona and
fe.idzona=fp.idzona
and d.iddimdata=fe.iddata and d.iddimdata=fp.iddata
GROUP BY z.marca, d.any
```

TFG-DW Enric Colomé  
Campos  
PACS

Informe de % d'ocupació per Marca turística

Marca	any	percentatge
Barcelona	2,012	12,97
	2,013	11,44
Catalunya Central	2,012	12,59
	2,013	12,09
Costa Barcelona	2,012	12,49
	2,013	12,30
Costa Brava	2,012	12,50
	2,013	12,29
Costa Daurada	2,012	12,50
	2,013	12,50

II-lustració 35 Informe de % d'ocupació per marca turística



II-Il·lustració 36 Gràfica de % d'ocupació per marca turística

#### Rànquing de municipis per categoria d'equipaments

El rànquing s'establirà per número d'equipaments per nivell de categoria. En aquest informe no es considera la data, atès que no la tenim en el fitxer d'equipaments.

El rànquing s'estableix pel nombre d'equipaments de cada categoria a cada municipi. S'utilitzarà la categoria del nivell 1+2+3 que defineix el tipus d'equipament.

A la versió OLAP/Saiku d'aquest informe es pot veure amb claredat alguna de les avantatges dels models OLAP, enfocats a la presentació de les dades: és pot afegir qualsevol nivell d'equipaments tan sols arrossegant el camp a l'apartat de files i es refà la sortida amb la distribució pel nivell afegit.

La consulta és:

```
SELECT z.municipi, sum(fq.equipaments) as nombre
FROM fetsequipaments fq
JOIN dimzona z on z.idDimZona=fq.idzona
JOIN dimequipaments q on q.iddimEquipaments=fq.idequipaments
GROUP BY z.municipi
ORDER BY nombre desc
```

La consulta en OLAP/Saiku:

```
SELECT
NON EMPTY {Hierarchize({[Measures].[equipaments]})} ON COLUMNS,
NON EMPTY Order({Hierarchize({[zona].[municipis].Members})},
[Measures].[equipaments], BDESC) ON ROWS
FROM [ranquing equipaments]
```

Saiku Analytics

Cubes: ranquing equipaments

Dimensions:

- equipaments
  - (All)
  - nivell 1
  - nivell 2
  - nivell 3
- zona
  - (All)
  - municipi

Measures:

- Measures
  - equipaments

Columns: equipaments

Rows: municipi

Filter:

municipi	equipaments
Barcelona	4.613
Lleida	641
Sabadell	615
Terrassa	522
Girona	518
Hospitalet de Llobregat, l'	507
Badalona	479
Tarragona	479
Mataró	351
Reus	336
Manresa	291
Sant Boi de Llobregat	251
Santa Coloma de Gramenet	245
Sant Cugat del Vallès	242

Il·lustració 37 Informe 4 OLAP amb Saiku Analytics

Màxim i mínim d'efectius policial per tipologia d'establiment i municipi

Es disposa de les dades de policia i d'establiment per municipi. Es repartiran policies per nombre d'establiments per tipus i categoria. El màxim i mínim serà el municipi amb més / menys efectius per a cada tipus / categoria. Es fa a partir del nombre d'efectius per nombre d'establiments. S'exclouen els mínims d'efectius i establiments iguals a zero perquè no aporten informació, atès que de totes les categories hi haurà al menys un municipi que no tindrà oferta i n'hi ha pobles sense efectius.

La consulta és:

```
SELECT d.any, de.tipus, de.categoria,
round(max(fp.efectius/fe.establiments),0) as maxim,
round(min(fp.efectius/fe.establiments),0) as minim
FROM fetspolishabitants fp, fetsestabliments fe
JOIN dimestabliments de on de.idDimEstabliments=fe.idestabliments
JOIN dimzona z on z.idDimZona=fe.idzona
JOIN dimdata d on d.idDimdata=fe.iddata
WHERE fp.iddata=fe.iddata and fp.idzona=fe.idzona and fe.places>0 and
fp.efectius>0
GROUP BY fe.iddata, fe.idestabliments
```

"Top ten" de municipis per franja de pernотacions

Els municipis amb més pernотacions; es disposa de la relació de marques amb comarques, aquestes amb municipis i pernотacions per marques, i tot per anys. Es

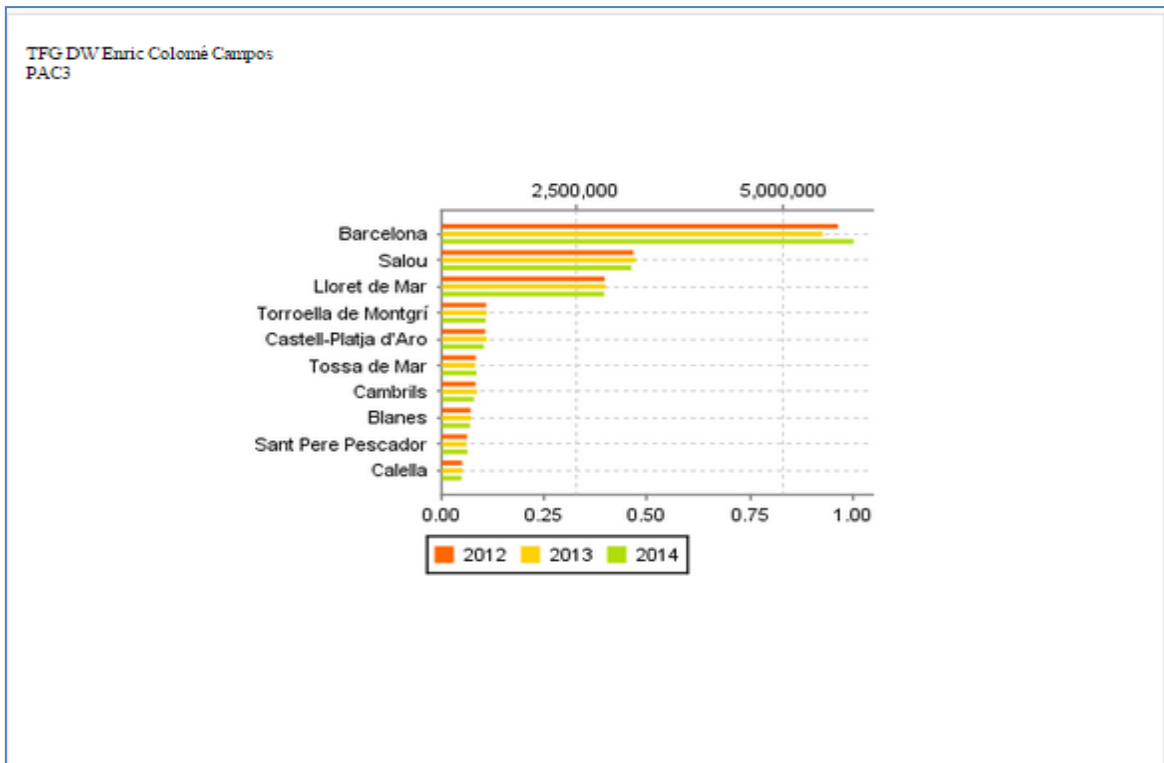
fa un informe dels 10 municipis amb més pernoctacions per a cadascun dels tres anys disponibles, de 2012 a 2014. En aquest informe s'inclou un gràfic de barres horitzontal amb els valors de cada municipi per a cada any de la sèrie.

La consulta és:

```
SELECT round(h.conta,0) ordre, h.municipi, h.any, h.pnc pernoctes
FROM (SELECT if(@d!=h.any,@c:=1,@c:=@c+1) conta,
if(@d!=h.any,@d:=h.any,@d) dia, @c,h.municipi, h.any, h.pnc
FROM (SELECT sum(pernoctacions) AS pnc, d.any, z.municipi
FROM (SELECT pernoctacions,idzona,iddata FROM fetspernoctacions force
index(iddata) ) fp
JOIN dimdata d on d.iddimdata=fp.iddata
JOIN dimzona z on z.iddimzona=fp.idzona
GROUP BY d.any,fp.idzona ORDER BY d.any,pnc DESC ) h
JOIN (SELECT @d:=0,@c:=0) k ) AS h WHERE conta<=10
```

Informe de Top ten de Municipis per franja de Pernoctacions		
<b>Any: 2,012</b>		
ordre	municipi	pernoctes
1	Barcelona	5,658,784
2	Salou	3,195,089
3	Lloret de Mar	2,848,631
4	Torroella de Montgri	1,420,957
5	Castell-Platja d'Aro	1,409,457
6	Tossa de Mar	1,296,019
7	Cambrils	1,293,226
8	Blanes	1,233,158
9	Sant Pere Pescador	1,190,314
10	Calella	1,128,481
<b>Any: 2,013</b>		
ordre	municipi	pernoctes
1	Barcelona	5,472,153
2	Salou	3,221,657
3	Lloret de Mar	2,857,082
4	Castell-Platja d'Aro	1,427,489
5	Torroella de Montgri	1,427,322
6	Cambrils	1,312,651
7	Tossa de Mar	1,288,642
8	Blanes	1,239,786
9	Sant Pere Pescador	1,185,890
10	Calella	1,134,184
<b>Any: 2,014</b>		
ordre	municipi	pernoctes
1	Barcelona	5,845,407
2	Salou	3,168,515

Il·lustració 38 Informe Top Ten de municipis per pernoctacions



II-Il·lustració 39 Gràfic de Top Ten de municipis per pernотacions

% de pernотacions d'un municipi sobre el total

Es disposa de la relació de marques amb comarques, aquestes amb municipis i municipis amb establiments, i pernотacions per marques, i tot per anys. El percentatge es farà en tant per mil atès la diferència de rang entre el total i les pernотacions a cada municipi. Els anys disponibles són de 2012 a 2014.

La consulta és:

```
SELECT z.municipi,d.any,
(sum(fp.pernотacions)*1000/
(SELECT sum(pernотacions) FROM fetspernотacions ft
JOIN dimdata t on t.iddimdata=ft.iddata WHERE t.any=d.any) ) as percentatge
FROM fetspernотacions fp JOIN dimzona z on z.idDimZona=fp.idzona
JOIN dimdata d on fp.iddata=d.iddimdata
GROUP BY z.municipi,d.any
```

Categorització de municipis A/B/C

Categorització A/B/C (principi de Pareto o regla 80/20) dels municipis per nombre d'habitants. Es fa la categorització 80/15/5 on A és els que apleguen un 80% de la població, B són els que apleguen un 15% i C, el 5% restant.

La consulta és:

```
SELECT d.any,z.municipi,fh.habitants as habitants,
habitants*100/(select sum(habitants) FROM fetspolishabitants ft
JOIN dimdata as t on t.iddimdata=ft.iddata WHERE t.any=d.any) as percent,
(SELECT if(percent>0.15,"A",if(percent>0.03,"B","C")))
FROM fetspolishabitants fh
JOIN dimzona z on fh.idzona=z.idDimZona
```

```
JOIN dimdata d on d.iddimdata=fh.iddata
ORDER BY d.any,percent desc
```

#### Rati de policies locals per habitant

Aquesta dada ens venia donada amb les dades originals. Això no obstant, es fa el càlcul de vell nou amb les dades del padró d'habitants.

La consulta és:

```
SELECT z.municipi, d.any,efectius*1000/habitants ratixmil
FROM fetspolishabitants fh
JOIN dimdata d on d.iddimdata=fh.iddata
JOIN dimzona z on z.idDimZona=fh.idzona
```

La consulta en Saiku és:

```
SELECT
NON EMPTY {Hierarchize({[Measures].[rati per mil]})} ON COLUMNS,
NON EMPTY filter(CrossJoin([data].[any].Members,
[zona].[municipi].Members),[Measures].[rati per mil]>0) ON ROWS
FROM [rati polis per habitant]
```

#### Distribució mensual i estacionalitat de les pernoctacions

El mes és directe per zona turística. Es farà en base a la distribució mensual de les pernoctacions per any. A l'informe s'ha incorporat un gràfic en format horitzontal de barres apilades.

La consulta és:

```
SELECT dimdata.any,monthname(concat(any,"-",mes1,"-","1")) as mes,
sum(pernoctacions) as dormides
FROM fetspernoctacions JOIN dimdata on iddimdata=iddata GROUP BY iddata
```

La consulta OLAP/Saiku és:

```
SELECT
NON EMPTY {Hierarchize({[Measures].[dormides]})} ON COLUMNS,
NON EMPTY {Hierarchize({[data].[any].Members}, {[data].[mes].Members})}
ON ROWS
FROM [pernocta_mes]
```

## 4. Conclusions

Aquest treball ha estat molt profitós en molts d'aspectes. D'una banda, des de l'òptica de la comunicació i la gestió de projectes, ha estat un exercici molt complet. S'han hagut de posar en pràctica molts conceptes estudiats al llarg de la carrera en diverses assignatures, com Gestió de projectes o Competències comunicatives. És un treball molt extens que cal estructurar adequadament i on és necessari observar tot un seguit de regles per obtenir un producte ben acabat.

També ha estat molt interessant poder posar en pràctica les lliçons apreses en les diferents assignatures relacionades amb la mineria de dades i bases de dades, com Minería de dades, Ús de bases dades o Disseny de bases de dades. He après el que és un model multidimensional i com plantejar el disseny de dades enfocat a la consulta enfront el disseny enfocat a les transaccions.

Aconseguir informació de cada una dels programes utilitzats i la seva operativa ha estat la part més dura del treball; malgrat tot, penso que ha valgut la pena el temps invertit, tant pel coneixement directe, com per l'adquisició d'habilitats. A l'altra banda, les investigacions i consultes sobre el model de dades han estat la part més motivant del treball. Ha estat una bona experiència del que, probablement, sigui habitual en un entorn de treball real.

Gràcies a aquesta inversió de temps en conèixer les eines, al final he agafat un cert domini –el just dintre les necessitats del projecte- que m'han animat a afegir-hi una anàlisi OLAP. De fet, ja hi estava contemplat, però el vaig haver d'arraconar per donar prioritat a punts més urgents.

Queda, per a línies futures, veure com relacionar els DW amb la mineria de dades i amb l'estadística, un camp que considero molt interessant. També pot ser molt profitós veure com es relaciona aquest model amb les noves eines que venen de la mà del *Big Data*.

## 5. Glossari

**Magatzem de dades:** El magatzem de dades és una col·lecció de dades orientades al tema, integrades, no volàtils i historiades, organitzades per a donar suport a processos d'ajuda a la decisió (Rius, 2012).

**Sistema Gestor de Bases de Dades:** Col·lecció de dades estructurades, organitzades i relacionades entre elles i el conjunt de programes que hi accedeixen i les gestionen.

**Multidimensionalitat:** La multidimensionalitat consisteix simplement a concebre les dades que volem analitzar en termes de fets i dimensions d'anàlisi, de manera que les podem situar en un espai n-dimensional (Rius, 2012).

**Pentaho BI Suite:** És un conjunt de programes lliures per generar intel·ligència empresarial (*Business Intelligence*). [Pentaho](#) es defineix como una plataforma de BI "orientada a la solució" i "centrada amb processos" que inclou tots els principals components requerits per implementar solucions basades en processos.



## 6. Bibliografia

BERNABEU, DARIO. *HEFESTO* (2009) Córdoba (Argentina): Licencia de Documentación Libre de GNU. [en línea] <https://sites.google.com/site/magm33332/hefesto>

CORONEL, ADRIANO, I ALTRES. (2012). “Gestión de riesgos”. *Gestión de proyectos* (Part I, secció 9).[en línea]. Madrid: Wiki EOI. [Consultat: 7 de març de 2015] [http://www.eoi.es/wiki/index.php/GESTI%C3%93N\\_DE\\_RIESGOS\\_en\\_Gesti%C3%B3n\\_de\\_proyectos](http://www.eoi.es/wiki/index.php/GESTI%C3%93N_DE_RIESGOS_en_Gesti%C3%B3n_de_proyectos)

CUENCA, MARIA JOSEP; NICOLUA, FRANCESCA. (2010). *Competència comunicativa per a professionals de les TIC*. Barcelona: Eureka Media

ESPINOSA, ROBERTO. (2010). “15.2.Kimball vs Inmon. Ampliación de conceptos del Modelado Dimensional.”. *El rincón del BI*. [en línea] [Consultat 11 de juny de 2015] <https://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimENSIONAL/>

PRADEL, JORDI; RAYA, JOSÉ. *Enginyeria de requisits*. (2012) Barcelona: Eureka Media

RAMÓN, JOSÉ; MARINÉ, PERE. (2010). *Gestió de projectes*. Barcelona: Eureka Media

RIUS, ÀNGELS (COORD.); SERRA, MONTSE; ABELLÓ, ALBERTO, I ALTRES. *Data warehouse*. (2012) Barcelona: Eureka Media

RIVADERA, GUSTAVO. (2010). “Cuadernos de la Facultad n. 5, 2010”. *La metodología de Kimball para el diseño de almacenes de datos (Data warehouses)*. [en línea] Salta: UCasal. [Consultat: 14 d’abril de 2015] <http://www.ucasal.edu.ar/hm/ingenieria/cuadernos/archivos/5-p56-rivadere-formateado.pdf>

### Consultes en línia:

Instituto Nacional de Estadística (INE) [<http://www.ine.es/>]

Institut d’Estadística de Catalunya (IDESCAT) [<http://www.idescat.cat/>]

### Consultes de manuals en línia:

<http://dev.mysql.com/doc/refman/5.7/en/index.html>

<http://www.mysql.com/products/workbench/>

<http://wiki.pentaho.com/display/EAI/02.+Spoon+Introduction>

<http://wiki.pentaho.com/display/EAI/MySQL>

<http://wiki.pentaho.com/display/EAI/Pentaho+Data+Integration+%28Kettle%29+Tutorial>

## 7. Annexos

### 7.1 Annex I

Captures de pantalla dels diferents informes des de Design Report

View Report  Auto-Submit

TFG DW Enric Colomé Campos  
PAC3

### Informe de Total de Pernoctacions estimades per Municipi i Any

<b>Municipi:</b> Abella de la Conca	any	pernoctacions
	2,012	420
	2,013	413
	2,014	418
<b>Municipi:</b> Abrera	any	pernoctacions
	2,012	4,584
	2,013	5,869
	2,014	3,293
<b>Municipi:</b> Àger	any	pernoctacions
	2,012	59,734
	2,013	58,924
	2,014	60,540
<b>Municipi:</b> Agramunt	any	pernoctacions
	2,012	6,989
	2,013	6,960
	2,014	7,013

II-lustració 40 Informe 1

TFG DW Enric Colomé Campos  
PAC3

### Informe de Promig de Viatgers per tipus d'Establiment i Comarca

**Comarca:** Alt Camp

<b>Establiment :</b> camping Luxe	any	promig
	2,012	0.00
	2,013	0.00
<b>Establiment :</b> camping Primera	any	promig
	2,012	0.00
	2,013	0.00
<b>Establiment :</b> camping Segona	any	promig
	2,012	0.00
	2,013	0.00
<b>Establiment :</b> camping Tercera	any	promig
	2,012	0.00
	2,013	0.00
	any	promig

II-lustració 41 Informe 2

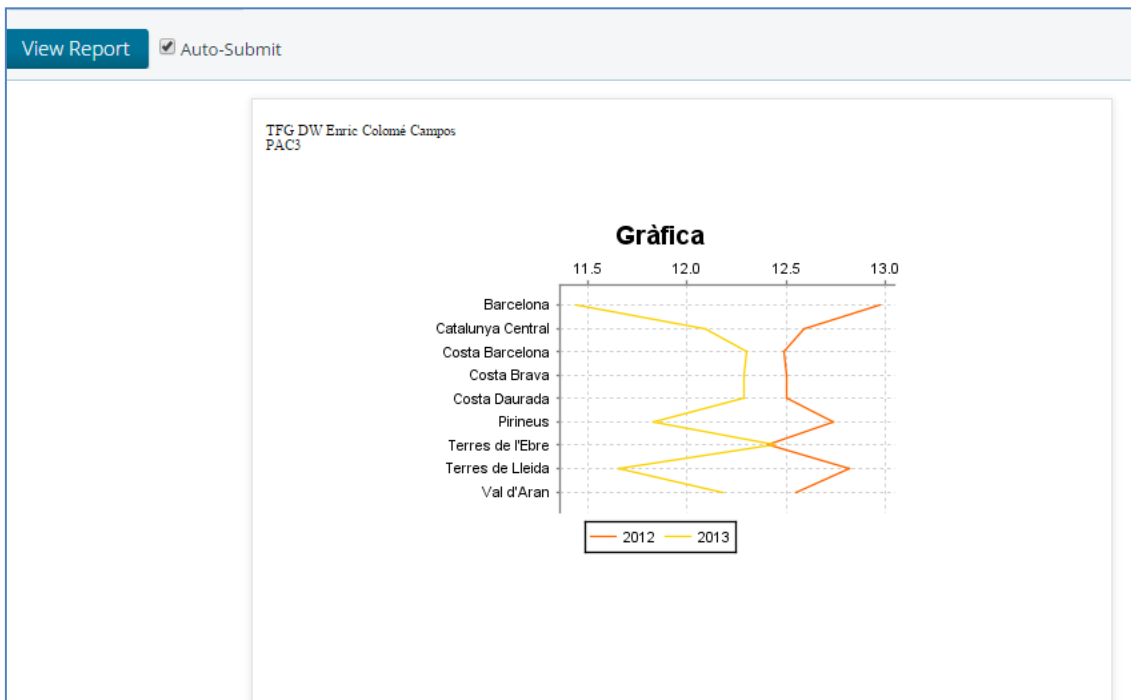
View Report  Auto-Submit

TFG DW Enric Colomé Campos  
PAC3

### Informe de % d'ocupació per Marca turística

Marca	any	percentatge
Barcelona	2,012	12.97
	2,013	11.44
Catalunya Central	2,012	12.59
	2,013	12.09
Costa Barcelona	2,012	12.49
	2,013	12.30
Costa Brava	2,012	12.50
	2,013	12.29
Costa Daurada	2,012	12.50
	2,013	12.50

II-lustració 42 Informe 3



II-lustració 43 Gràfica informe 3

## Informe de Rànquing de Municipis per Categoria d'Equipaments

municipi	nombre
Barcelona	4,613
Lleida	641
Sabadell	615
Terrassa	522
Girona	518
Hospitalet de Llobregat, l'	507
Badalona	479
Tarragona	479
Mataró	351
Reus	336
Manresa	291
Sant Boi de Llobregat	251
Santa Coloma de Gramenet	245
Sant Cugat del Vallès	242
Granollers	227
Vilanova i la Geltrú	217
Cerdanyola del Vallès	216
Vic	201
Igualada	187
Comellà de Llobregat	183

Il·lustració 44 Informe 4

## Informe de Màxim i Mínim d'Efectius policials per Tipologia d'Establiment i Municipi

Any: 2,011

Tipus: camping

Categoria: Luxe

	maxim	minim
Categoria: Primera	79	13

Categoria: Primera

	maxim	minim
Categoria: Segona	100	4

Categoria: Segona

	maxim	minim
Categoria: Tercera	238	1

Categoria: Tercera

	maxim	minim
	119	5

Il·lustració 45 Informe 5

## Informe de Top ten de Municipis per franja de Pernoctacions

Any: 2,012		
ordre	municipi	pernoctes
1	Barcelona	5,658,784
2	Salou	3,195,089
3	Lloret de Mar	2,848,631
4	Torroella de Montgrí	1,420,957
5	Castell-Platja d'Aro	1,409,457
6	Tossa de Mar	1,296,019
7	Cambrils	1,293,226
8	Blanes	1,233,158
9	Sant Pere Pescador	1,190,314
10	Calella	1,128,481

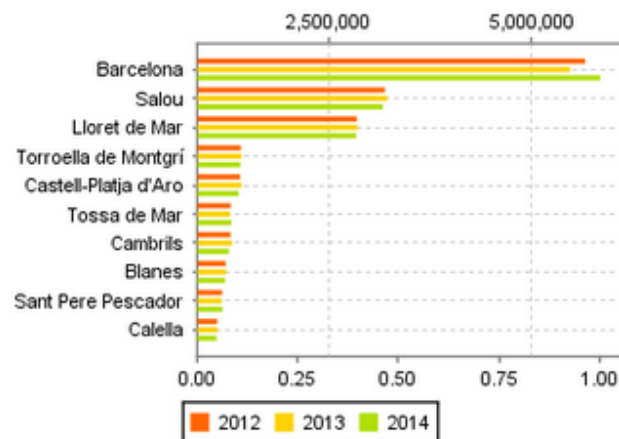
Any: 2,013		
ordre	municipi	pernoctes
1	Barcelona	5,472,153
2	Salou	3,221,657
3	Lloret de Mar	2,857,082
4	Castell-Platja d'Aro	1,427,489
5	Torroella de Montgrí	1,427,322
6	Cambrils	1,312,651
7	Tossa de Mar	1,288,642
8	Blanes	1,239,786
9	Sant Pere Pescador	1,185,890
10	Calella	1,134,184

Any: 2,014		
ordre	municipi	pernoctes
1	Barcelona	5,845,407
2	Salou	3,168,515

II-lustració 46 Informe 6

TFG-DW Enric Colomé Campos  
PAC3



II-lustració 47 Gràfica informe 6

TFG DW Enric Colomé  
Campos  
PAC3

### Informe de % (per mil) de Pernoctacions d'un Municipi sobre el Total

**Municipi:** Abella de la Conca

any	percentatge
2,012	0.0085
2,013	0.0084
2,014	0.0084

**Municipi:** Abrera

any	percentatge
2,012	0.0923
2,013	0.1191
2,014	0.0658

**Municipi:** Àger

any	percentatge
2,012	1.2030
2,013	1.1954
2,014	1.2106

**Municipi:** Agramunt

any	percentatge
-----	-------------

II-lustració 48 Informe 7

## Informe de Categorització ABC de Municipis per Habitants

Any: 2,011

categoria	habitants	municipi	percent
A	1,615,448	Barcelona	21.4261
A	256,065	Hospitalet de Llobregat, l'	3.3963
A	219,786	Badalona	2.9151
A	213,697	Terrassa	2.8343
A	207,721	Sabadell	2.7551
A	138,416	Lleida	1.8358
A	134,085	Tarragona	1.7784
A	123,868	Mataró	1.6429
A	120,824	Santa Coloma de Gramenet	1.6025
A	106,709	Reus	1.4153
A	96,722	Girona	1.2829
A	87,243	Cornellà de Llobregat	1.1571
A	83,337	Sant Cugat del Vallès	1.1053
A	82,860	Sant Boi de Llobregat	1.0990
A	76,589	Manresa	1.0158
A	73,979	Rubí	0.9812
A	66,905	Vilanova i la Geltrú	0.8874
A	64,737	Viladecans	0.8586
A	63,499	Prat de Llobregat, el	0.8422

II-lustració 49 Informe 8

## Informe de Rati de Polícies locals per Habitant

### Municipi: Abella de la Conca

any	ratixmil
2,011	0.0000
2,012	0.0000
2,013	0.0000

### Municipi: Abrera

any	ratixmil
2,011	1.7225
2,012	1.7692
2,013	1.7420

### Municipi: Àger

any	ratixmil
2,011	0.0000
2,012	0.0000
2,013	0.0000

### Municipi: Agramunt



## Informe de distribució Mensual i Estacionalitat de les Pernoctacions

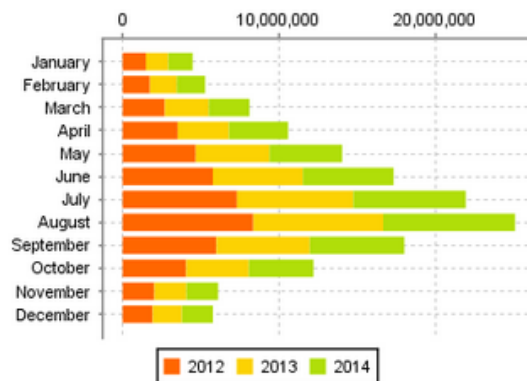
Any: 2,012

mes	pernoctes
January	1,493,525
February	1,757,839
March	2,702,565
April	3,528,056
May	4,679,489
June	5,774,915
July	7,314,964
August	8,364,288
September	6,003,658
October	4,070,202
November	2,036,559
December	1,926,402

Any: 2,013

mes	pernoctes
January	1,428,974
February	1,710,339
March	2,807,651
April	3,273,358
May	4,711,352

Il·lustració 51 Informe 10



Il·lustració 52 Gràfica informe 10

## 7.2 Annex II

Captures de pantalla d'alguns informes amb OLAP/Saiku.

The screenshot shows the Saiku Analytics interface. The main window title is 'Saiku Analytics'. The interface is divided into several sections:

- Cubes:** A dropdown menu showing 'pernoctacions per municipi'.
- Dimensions:** A tree view showing 'data' (with sub-items '(All)' and 'any') and 'zona' (with sub-items '(All)' and 'municipi').
- Measures:** A tree view showing 'Measures' (with sub-item 'pernoctes').
- Columns:** A dropdown menu showing 'pernoctes'.
- Rows:** Two dropdown menus showing 'any' and 'municipi'.
- Filter:** An empty dropdown menu.

The main content area displays a table with the following data:

any	municipi	pernoctes
2012	Abella de la Conca	420
	Abredera	4.584
	Àger	59.734
	Agramunt	6.989
	Aguilar de Segarra	755
	Agullana	1.091
	Aiguafreda	0
	Aiguamúrcia	21.078
	Aiguaviva	0
	Aitona	0
	Alamús, els	0
	Alàs i Cerc	2.777
	Albagés, l'	755

II-lustració 53 Informe 1 Saiku de pernoctacions estimades per municipi

Saiku Analytics

**Cubes**  
ranquing equipments

**Dimensions**

- equipaments
  - (All)
  - nivell 1
  - nivell 2
  - nivell 3
- zona
  - (All)
  - municipi

**Measures**

- Measures
  - equipaments

**Columns** equipments

**Rows** municipi

**Filter**

municipi	equipaments
Barcelona	4.613
Lleida	641
Sabadell	615
Terrassa	522
Girona	518
Hospitalet de Llobregat, l'	507
Badalona	479
Tarragona	479
Mataró	351
Reus	336
Manresa	291
Sant Boi de Llobregat	251
Santa Coloma de Gramenet	245
Sant Cugat del Vallès	242

II-lustració 54 Informe 4 Saiku de rànkung de municipis per equipaments

Saiku Analytics

**Cubes**  
Rati polícies habitant

**Dimensions**

- data
  - (All)
  - any
- zona
  - (All)
  - municipi

**Measures**

- Measures
  - efectius
  - habitants
  - rati per mil

```

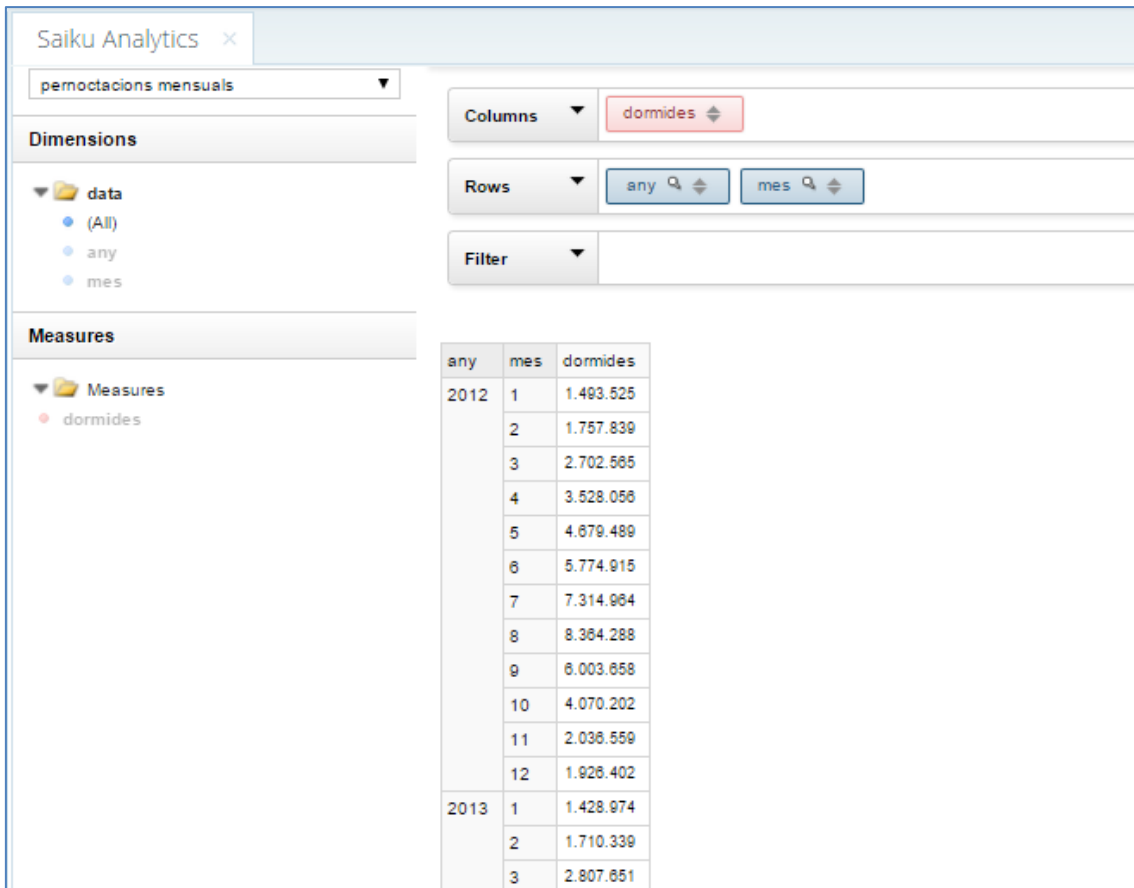
1 SELECT
2 NON EMPTY (Hierarchize({[Measures].[rati per mil]})) ON COLUMNS,
3 NON EMPTY filter(CrossJoin([data].[any].Members, [zona].[municipi].Members), [Measures].[rati per mil]>0) ON ROWS
4 FROM [rati polis per habitant]

```

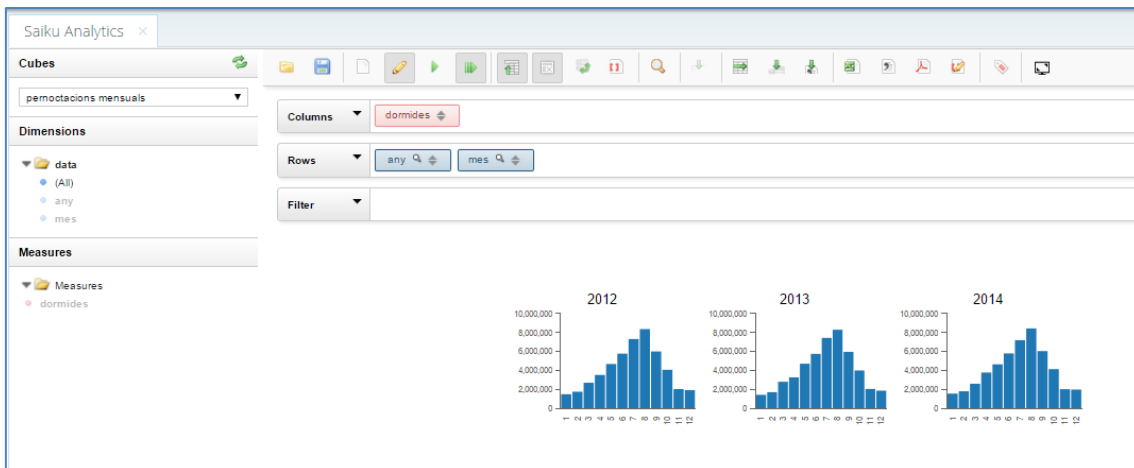
4, 30

any	municipi	rati per mil
2011	Abreña	1,723
	Agramunt	1,238
	Alcanar	1,604
	Alcarràs	1,317
	Alella	2,09
	Almacelles	0,6
	Altafulla	3,769
	Ametlla de Mar, l'	2,876
	Ametlla del Vallès, l'	2,406
	Ampostà	1,539
	Anglès	1,597
	Arboç, l'	2,343
	Arbúcies	1,341
	Arenys de Mar	2,355
	Arenys de Munt	1,765
	Argentona	1,953
	Artés	1,614

II-lustració 55 Informe 9 amb Saiku de rati de polícies per habitant



II-lustració 56 Informe 10 amb OLAP Saiku



II-lustració 57 Gràfic de l'informe 10 amb Saiku