

Construcció i explotació d'un magatzem de dades per a l'anàlisi d'informació sobre allotjaments turístics

M^a Carme Santacreu Abad

ETIG

Carles Llorach Rius

17 de Juny de 2013

RESUM

Aquest document correspon a la memòria del treball de final de carrera dels estudis d'enginyeria Tècnica d'informàtica de Gestió, dins l'àrea de Magatzem de Dades.

Avui en dia, poder gestionar la informació en les empreses, és una eina clau per poder subsistir en un mercat dinàmic i canviant. Saber emmagatzemar gran varietat d'informació i saber-la explotar degudament mitjançant els magatzems de dades (Data Warehouse, que a partir d'ara abreviarem com DW)) és basic actualment per al creixement i la gestió de les empreses.

La disciplina anomenada Business Intelligence (intel·ligència empresarial) ens acosta a aquest sistemes d'informació i ens ajuden a la presa de decisions.

Aquest treball, consisteix en crear i explotar un magatzem de dades des de el seu estudi inicial fins la seva implementació, passant per les diferents etapes de planificació, anàlisi, disseny i desenvolupament, i segons els requeriments proporcionats pel client final del producte.

L'objectiu és familiaritzar-se amb l'arquitectura d'un DW, que és, com dissenyar-lo partint de les necessitats del client, com detectar els indicadors de negoci, i com modelar les dades per obtenir un producte final útil i àgil dirigit a l'anàlisi de dades pròpies de les organitzacions.

Aquest objectiu es duu a terme amb un cas pràctic, que es detalla en aquesta memòria, i que fa referència a la creació i explotació d'un magatzem de dades sobre allotjaments turístics, a petició de l'Observatori Nacional d'Ocupació (ONdO), amb l'objectiu d'aprofundir en l'estudi de l'evolució d'aquests establiments a Catalunya en els últims anys.

PARAULES CLAU

Magatzem de dades, Data Warehouse, Business Intelligence, Indicadors de negoci, KPI, model multi-dimensional, Procés ETL, quadre de comandament, OLAP, SQL, Pentaho.

INDEX DE CONTINGUTS

RESUM	2
MEMÒRIA DEL PROJECTE.....	6
1.- INTRODUCCIÓ.....	6
1.1.- JUSTIFICACIÓ DEL TFC	6
1.2.- OBJECTIUS DEL TFC	7
1.3.- ENFOCAMENT I METODE SEGUIT	7
1.4.- PLANIFICACIÓ DEL PROJECTE	8
1.4.1. PLANIFICACIÓ INICIAL.....	8
1.4.2. TASQUES	9
1.4.3. DIAGRAMA DE GANNT	10
1.4.4. RISCOS	11
1.4.5. PLANIFICACIÓ FINAL.....	11
1.5.- PRODUCTES OBTINGUTS.....	12
1.6.- DESCRIPCIÓ DELS ALTRES CAPÍTOLS DE LA MEMÒRIA.....	13
2.- ANÀLISI	14
2.1.- ESTUDI DE LES DADES D'ORIGEN.....	14
2.2.- REQUERIMENTS FUNCIONALS	19
2.3.- REQUERIMENTS NO FUNCIONALS	20
2.3.1. SEURETAT.....	20
2.3.2. ESCALABILITAT	21
2.3.3. FACILITAT D'US	21
2.3.4. VIABILITAT	21
2.4.- DIAGRAMA DE CASOS D'US.....	22
2.5.- MODEL CONCEPTUAL.....	23
2.5.1. ÀNÀLISI DEL MODEL MULTI-DIMENSIONAL.....	24
2.5.2. ELS INDICADORS CLAU DE RENDIMENT (KPI).....	25
2.5.3. DIMENSIONS	26
2.5.4. FETS.....	27
2.5.5. MESURES E INDICADORS.....	27
2.5.6. GRANULITAT	28
3.- DISSENY	29
3.1.- DIAGRAMA D'ARQUITECTURA DE MAQUINARI	29
3.2.- DIAGRAMA D'ARQUITECTURA DE PROGRAMARI.....	30
3.3.- DISSENY DE LA BD I DIAGRAMA DEL MODEL FÍSIC.....	31
3.4.- CÀRREGA DE LA BD – PROCÉS ETL.....	33
3.5.- DISSENY I DESCRIPCIÓ DELS IFORMES.....	35
3.5.1. INFORME 1. Places establiments vs superfície i habitants.....	35
3.5.2. INFORME 2 . Total establiments i places ofertades.	36
3.5.3. INFORME 3 . Evolució total establiments i places ofertades.	37
3.5.4. INFORME 4 . Rànquing d'establiments per zona.	38
3.5.5. INFORME 5 . Detall evolució establiments.	39
3.5.6. INFORME 6 .Detall evolució establiments agrupat.	40
3.5.7. INFORME 7 . Detall Equipaments per província.....	41
4.- CAPTURES DE PANTALLA	42
4.1.- ACCÉS AL PROGRAMARI DE DESENVOLUPAMENT DEL PROJECTE	42

4.2.-	PROCÉS ETL.....	45
4.2.1.	JOB TRANSFORMACIONS DIMENSIONS	46
4.2.2.	JOB TRANSFORMACIONS FETS.....	48
4.2.3.	JOB CARREGA TAULES DIMENSIONS	50
4.2.4.	JOB CARREGA TAULES FETS	52
5.-	CONCLUSIONS.....	54
6.-	LINIES D'EVOLUCIÓ FUTURES	55
	GLOSSARI.....	56
	BIBLIOGRAFIA	57
6.1.-	PUBLICACIONS / LECTURES DE MATERIALS.....	57
6.2.-	REFERENCIES PROCEDENTS D'INTERNET	57
	ANNEXOS.....	58
7.-	QUALITAT DE LES DADES	58
8.-	SORTIDA INFORMES EN PDF	59
8.1.-	INFORME1	59
8.2.-	INFORME 2	60
8.3.-	INFORME 3	61
8.4.-	INFORME 4	62
8.5.-	INFORME 5	63
8.6.-	INFORME 6	64
8.7.-	INFORME 7	65
9.-	SCRIPT DE CREACIÓ DE BD	66

INDEX DE FIGURES

figura 1: Taula resum d'arxius d'origen de dades proporcionades pel client.....	14
figura 2: Taula resum d'arxius d'origen de dades d'altres fonts.....	15
figura 3: Taula de tipologies d'establiment detectades	15
figura 4: Vista d'arxiu 'Establiments. Csv'	16
figura 5: Taula d'àmbits territorials detectats	16
figura 6: Vista d'arxiu 'aec-924' de comarques	18
figura 7: Vista d'arxiu 'aec-925' de superfície de municipis.	18
figura 8: Vista d'arxiu 'població.csv'	19
figura 9: Vista d'arxiu 'PoblacióPerSexe.xls'	19
figura 10: El DW d'allotjaments turístics.....	20
figura 11: Taula resum de necessitat d'espai de BBDD	21
figura 12: Cas d'us usuari final	22
figura 13: Cas d'us key user	23
figura 14: Cas d'us usuari administrador.....	23
figura 15: Cub multidimensional.	24
figura 16: Diagrama de concepte	27
figura 17: Arquitectura de maquinari	29
Figura18: Arquitectura de programari	30
figura 22: Disseny Informe 1	35
figura 23: Disseny Informe 2	36
figura 24: Disseny Informe 3	37
figura 25: Disseny Informe 4	38
figura 26: Disseny Informe 5	39
figura 27: Disseny Informe 6	40
figura 28: Disseny Informe 7	41
figura 29: Esquema BD en MySQL workbench.....	42
figura 30: Connexió Pentaho BI user console	43
figura 31: Informes en Pentaho BI user console	43
figura 32: Connexió BD	44
figura 33: Clau per publicar en PRD	44
figura 34: Catàleg transformacions i jobs PDI.....	45
figura35 : job ETL_dimensions.....	46
figura36 : transformació TEMPS	46
figura 38: transformació Categoria Establiment.....	47
figura 39: transformació comarca.....	47
figura 41: job ETL_fets	48
figura 42: transformació equipaments	49
figura 43: transformació població	49
figura 44: transformació evolució establiments	50
figura 45: job LOAD_dimensions.....	50
figura 46: transformació LOAD TEMPS	51
figura 47: transformació LOAD categoria Equipaments.....	51
figura 48: transformació LOAD categoria Establiments	51
figura 49: transformació LOAD municipi	52
figura 50: job LOAD fets	52
figura 51: transformació LOAD població	52
figura 52: transformació LOAD equipaments	53
figura 53: transformació LOAD evolució establiments	53

MEMÒRIA DEL PROJECTE

1.- INTRODUCCIÓ

El projecte consisteix en crear i explotar un magatzem de dades sobre allotjaments turístics, a petició de l'Observatori Nacional d'Ocupació (ONdO), amb l'objectiu d'aprofundir en l'estudi de l'evolució d'aquests establiments a Catalunya en els últims anys.

1.1.- JUSTIFICACIÓ DEL TFC

L'Observatori Nacional d'Ocupació (ONdO) vol aprofundir en l'evolució del nombre d'allotjaments turístics a Catalunya en vista a la creixent demanda en els últims anys i que ofereixen unes sis-centes mil places d'ocupació a Catalunya.

En concret, el client, vol analitzar la següent informació :

- Total d'establiments
- Total de places
- % de places respecte població
- Oferta mitjana de places
- Nombre d'establiments/Nombre d'equipaments
- % de població per equipament
- Indicador d'establiments vs habitants per gènere
- Indicador de places vs persones
- Indicador d'equipaments vs població
- Quantitat de places ofertes / superfície del territori

Com a punt de partida, el client ens proporciona 3 paquets d'arxius :

- *Establiments*: 7 arxius de text (.,txt), un per cada any amb: nombre d'establiments, tipologia, places ofertes i classificació
- *Equipaments*. 1 Arxiu de text separat per coma (csv), amb la relació d'equipaments públics a data 31/12/2012.
- *Població*: 1 Arxiu de text separat per coma (csv), amb els habitants per any i la superfície de cada municipi.

Aquesta informació, està disgregada i desestructurada, pel que la seva gestió i explotació és impracticable i molt costosa.

Per aquest motiu el ONdO, ha sol·licitat una solució informàtica consistent en un Data Warehouse on emmagatzemar e integrar aquesta informació i poder-la explotar de manera àgil i útil.

1.2.- OBJECTIUS DEL TFC

L'objectiu general d'aquest treball és familiaritzar-se amb l'arquitectura d'un Data Warehouse per a donar servei a un client, des de l'estudi inicial, passant per l'anàlisi, la construcció i càrrega del magatzem de dades, i la seva explotació en forma d'informes exhaustius i útils per a la presa de decisions.

Més concretament s'estableixen com a objectius :

- ✓ Familiaritzar-se en les tecnologies de Business Intelligence i de Data Warehouse.
- ✓ Agafar experiència en l'anàlisi de la informació clau d'una organització (o part d'ella) i aprendre a detectar els KPIs (Key performance indicator) més importants i estratègic del client.
- ✓ Aprendre a crear una base de dades òptima per la seva explotació en quan a rendiment i dinamicitat, entenen com a dinamicitat la capacitat per poder-se adaptar a les futures necessitats del negoci.
- ✓ Aprendre a utilitzar les eines de tractament de dades proporcionades per la universitat.

1.3.- ENFOCAMENT I METODE SEGUIT

Aquest projecte ha estat desenvolupat amb la metodologia clàssica d'un projecte de sistemes qualsevol, és a dir, aplicant el cicle de vida tradicional o en cascada, també anomenat SDLC, on la principal característica és que cada fase és continuació de l'anterior i es basa en aquesta, però amb un enfocament clar centrat en l'emmagatzematge i accés a les dades .

Aquesta metodologia ha estat alineada amb les diferents entregues realitzades al consultor.

Les fases del cicle de vida tradicional son aquestes :

1. **Anàlisi de requeriments.** Estudi inicial de les necessitats del client i de les dades proporcionades. Planificació del projecte.
2. **Anàlisi.** Estudi exhaustiu de les dades. Model conceptual. Enfocament del model multi-dimensional.
3. **Disseny.** Model i disseny de dades. Arquitectura de maquinari i de programari.
4. **Implementació.** Creació de la base de dades, procés ETL e informes.
5. **Prova.** Càrrega de la base de dades i extracció d'informes. Revisió de la qualitat de les dades i d'accés a aquestes (optimització).

6. **Posada en producció.** Publicació del informes i entrega del projecte.

prenent principal atenció en l'emmagatzematge i accés a les dades, és a dir, en els següents 3 punts:

- Disseny i creació de la base de dades.
- Extracció i transformació de les dades proporcionades pel client per a la càrrega de la base de dades (procés ETL).
- Generació d'informes.

Cada una de les fases es detallen en els següents capítols d'aquesta memòria.

1.4.- PLANIFICACIO DEL PROJECTE

1.4.1. PLANIFICACIÓ INICIAL

La planificació, al inici del projecte, es detalla a continuació. .

Cal tenir en compte que no s'ha fet una planificació seqüencial en el temps, ja que per experiència, no es solen complir. Tasques com la lectura de materials, i els borradors de les PACs no tenen dependència directa amb la tasca anterior en el temps, si no que es van fent en paral·lel.

També destacar el punt de rectificació del document de memòria a mesura que el consultor va avaluant les entregues.

En el diagrama de Gannt no s'han inclòs la primera tasca de presentació de l'assignatura, ni la de debat virtual, ja que, en aquesta última, no es coneix la data ni el procediment amb exactitud.

En la taula s'indica, a nivell informatiu, quin és el pes de les diferents fases respecte a la totalitat del projecte.

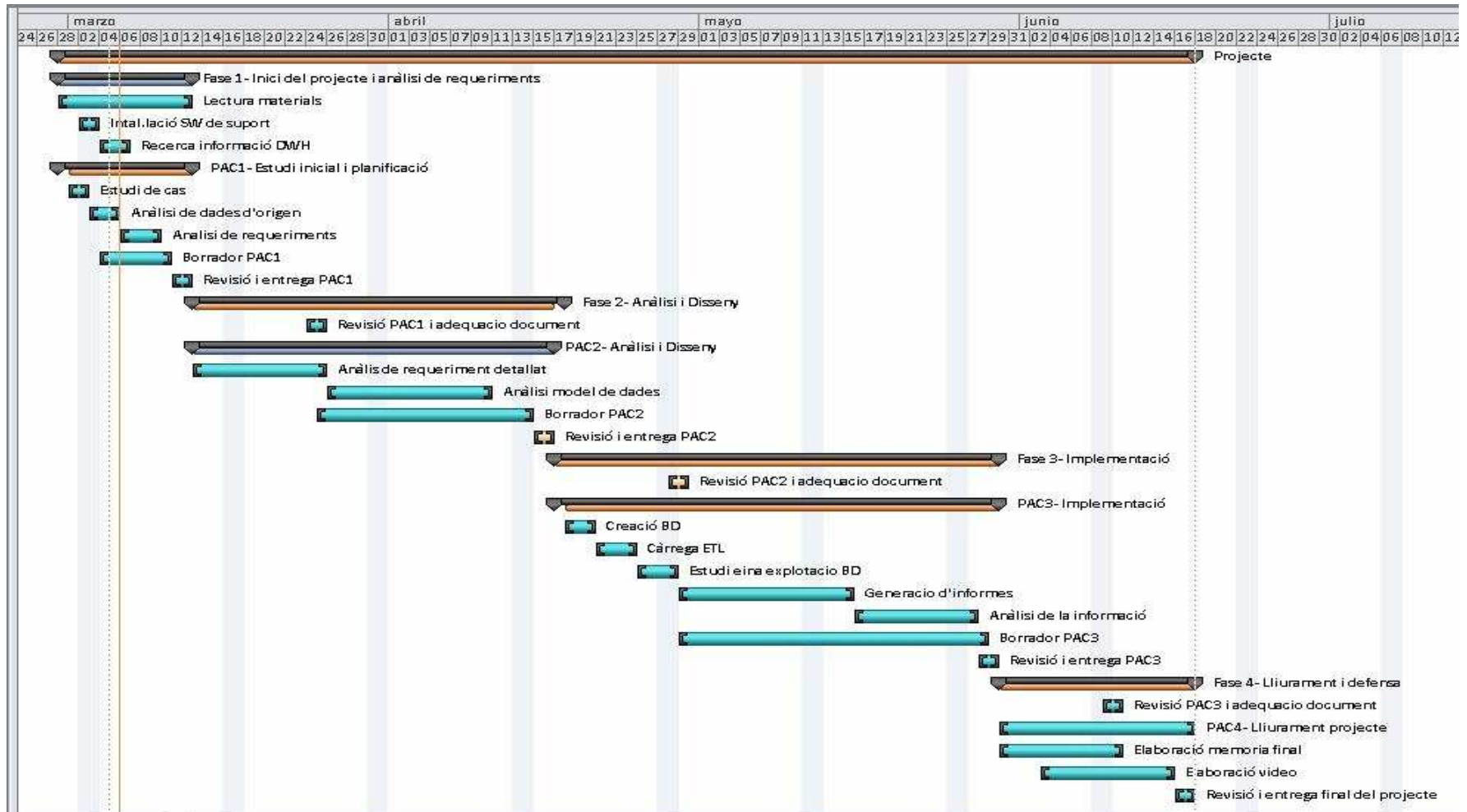
El total de dedicació estimat és de aproximadament de 2 hores diàries, 3 dies a la setmana laboral i 6 hores en cap se setmana, per tan un total de 12 hores setmanals.

Tenint en compte el període d'entregues del projecte (sense contar la defensa), del 28/2/13 fins el 17/6/13 son 15 setmanes, la dedicació total s'estima en 180 hores.

1.4.2. TASQUES

Tasca	Data inici	Data fi	Pes
Lectura de pla docent i presentació al campus	27/02/2013	27/02/2013	
PROJECTE	28/02/2013	17/06/2013	
FASE 1 del projecte	28/02/2013	12/03/2013	11,82%
Lectura materials de l'assignatura	28/02/2013	06/03/2013	
Instal·lació software de suport	02/03/2013	03/03/2013	
Recerca informació sobre DWH	02/03/2013	03/03/2013	
PAC1 - Pla de treball	28/02/2013	12/03/2013	
Estudi del cas	01/03/2013	02/03/2013	
Anàlisi de dades d'origen proporcionades	03/03/2013	05/03/2013	
Anàlisi de requeriments	06/03/2013	09/03/2013	
Borrador PAC1	04/03/2013	10/03/2013	
Revisió i entrega PAC1	11/03/2013	12/03/2013	
FASE 2 del projecte	13/03/2013	16/04/2013	31,82%
Revisió PAC1 i adequació document de projecte	24/03/2013	25/03/2013	
PAC2- Anàlisi i Disseny	13/03/2013	16/04/2013	
Anàlisi de requeriments detallat	13/03/2013	25/03/2013	
Anàlisi model de dades	26/03/2013	10/04/2013	
Borrador PAC2	25/03/2013	14/04/2013	
Revisió i entrega PAC2	15/04/2013	16/04/2013	
FASE 3 del projecte	17/04/2013	29/05/2013	39,09%
Revisió PAC2 i adequació document de projecte	28/04/2013	29/04/2013	
Instal·lació màquina virtual	17/04/2013	17/04/2013	
PAC3 - implementació	17/04/2013	29/05/2013	
Creació BD	18/04/2013	20/04/2013	
Unificació i càrrega de dades BD	21/04/2013	24/04/2013	
Estudi eina explotació BD	25/04/2013	28/04/2013	
Generació d'informes	29/04/2013	15/05/2013	
Anàlisi de la informació	16/05/2013	27/05/2013	
Borrador PAC3	29/04/2013	28/05/2013	
Revisió i entrega PAC3	28/05/2013	29/05/2013	
FASE 4 del projecte	30/05/2013	17/06/2013	17,27%
Revisió PAC3 i adequació document de projecte	09/06/2013	10/06/2013	
PAC4 - lliurament final i defensa	30/05/2013	17/06/2013	
Elaboració memòria final	30/05/2013	10/06/2013	
Elaboració vídeo presentació projecte	03/06/2013	15/06/2013	
Revisió i Entrega final projecte	16/06/2013	17/06/2013	
FASE Final del projecte	18/06/2013	09/07/2013	
Evaluació tribunal	18/06/2013	09/07/2013	
Debat virtual	18/06/2013	09/07/2013	

1.4.3. DIAGRAMA DE GANNT



1.4.4. RISCOS

Es tindran en compte les possibles incidències que puguin sorgir en el període d'elaboració del TFC, i es preveuen plans de contingència a efecte de finalitzar el projecte en les dates determinades.

Avaria o incidència en el software instal·lat.

Tot l'entorn de treball del TFC es realitzarà en un ordinador portàtil . Si per alguna causa hagués algun incidència en el seu funcionament, s'utilitzaran altres ordinadors disponibles fins que l'incidència quedi solventada.

Puntes de feina

Pots passar que per motius laborals no sigui possible dedicar al TFC les hores previstes i per tant la planificació. En aquest cas s'intentaran recuperar les hores en els caps de setmana i festius del període.

Malaltia o altres problemes personals / familiars no previstos

S'actuarà com en el punt anterior.

Fallada en l'accés a internet en la setmana vacacional.

En principi, no hauria d'haver problema en la connexió a internet via wi-fi en el període vacacional. Tot i això, si es dones el cas, es continuaria treballant en l'anàlisi i el model de dades del projecte amb normalitat i segons planificació

Retard per aprenentatge de les eines

Al no tenir experiència amb les eines a usar en el desenvolupament, l'estudi d'aquestes poden provocar retard en les tasques planificades. En aquest cas s'intentaran recuperar les hores en els caps de setmana i festius del període, com en els casos anteriors.

1.4.5. PLANIFICACIÓ FINAL

La planificació final no ha variat respecte a les dades d'entrega de les diferents etapes del projecte, però si en el nombre de dedicació en hores, que s'ha incrementat bàsicament per l'aprenentatge de les eines de desenvolupament del projecte, tal i com es va preveure en el apartat de riscos.

1.5.- PRODUCTES OBTINGUTS

Els productes obtinguts en aquest treball son, per una banda els informes resultants de la explotació de la base de dades i entregables al client que ens ha encarregat el projecte, i per l'altra els documents entregables propis del treball de final de carrera.

Entregables del client :

Base de dades : Creada amb *MySQL workbench*. Nom del esquema : TFC








Procés ETL : Creat amb *Pentaho PDI (Kettle)*

Es compona del processos de Extracció i transformació, els Load a les taules de la base de dades, i els jobs que executen totes les transformacions.

Les transformacions i jobs estan en el catàleg propi de la eina:

Informes : Creats amb Pentaho Report Designer i publicats a la consola *Pentaho BI server user console* pel seu accés via navegador, i ubicats a la carpeta 'TFC-establiments turistics'

La relació d'informes és la següent:

-  informe 1-Places establiments vs superficie i habitants.prpt
-  informe 2-Total establiments i places ofertades.prpt
-  informe 3-Evolucio total establiments i places ofertades.prpt
-  informe 4-Ranquing establiments per zona.prpt
-  informe 5-Detall evolucio establiments.prpt
-  informe 6-Detall evolucio establiments agrupat.prpt
-  informe 7-Detall Equipaments per província.prpt

Entregables propis del TFC : entregats a l'aula del TFC, en les dates especificades en el pla docent.

PAC1 --> Estudi inicial i Pla de Treball

PAC2 --> Anàlisi i Disseny

PAC3 --> Implementació

PAC4 --> Lliurament Final, que inclou :

Memòria del projecte : msantacreua_memoria.pdf

Presentació del projecte : msantacreua_presentaciop.avi

Producte : msantacreua_producte.zip (inclou la maquina virtual amb extensió .ova)

1.6.- DESCRIPCIÓ DELS ALTRES CAPÍTOLS DE LA MEMÒRIA

A continuació es descriu breument els següents capítols de la memòria.

Anàlisi : S'analitzen les necessitats del client mostrant els diagrames d'us necessaris que detallaran les necessitats del projecte. En funció de l'analitzat s'exposarà la solució a nivell de model conceptual i es justificarà la seva elecció.

Disseny : Es mostraran els diagrames d'arquitectura de maquinari i programari, així com el disseny de la base de dades i els informes creats.

Captures de pantalla : En aquest capítol s'ha trobat oportú mostrar en detall el procés ETL i les claus d'accés al programari que s'han utilitzat.

Conclusions : Conclusions a las que s'ha arribat després de finalitzar el treball.

Línies d'evolució futura : Relació de fites pròpies del projecte no aconseguides, i com es podrien resoldre, i proposta de millor i d'evolució del magatzem de dades.

Glossari : Glosaria de termes .

Bibliografia : Bibliografia utilitzada per realitzar el treball.

Annexes : En el annexe d'aquesta memòria s'ha incorporat el control de qualitat de les dades, el script de creació de la BD, i els informes creats, en format pdf.

2.- ANALISI

2.1.- ESTUDI DE LES DADES D'ORIGEN.

Com a pas previ al anàlisi i disseny de la base de dades, hem hagut d'analitzar amb deteniment les dades d'origen proporcionades, i obtenir també, d'altres fonts de dades altra informació rellevant per al projecte com és la població de tots els municipis, els diferents àmbits territorials, etc.

El nostre client ONdO ens proporciona tota la informació rebuda en els següents fitxers:

- **Establiments:** Un arxiu per cada any amb: nombre d'establiments, tipologia, places ofertes i classificació
- **Equipaments.** Arxiu amb la relació d'equipaments públics a data 31/12/2012.
- **Població:** Un arxiu amb els habitants per any i la superfície de cada municipi.

Concretament, els arxius físics son els següents :

Arxiu de dades	Font d'Origen
poblacio.csv	IDESCAT
establiments 2012.txt	Federació Catalana d'Allotjaments Turístics
establiments 2011.txt	Federació Catalana d'Allotjaments Turístics
establiments 2010.txt	Federació Catalana d'Allotjaments Turístics
establiments 2009.txt	Federació Catalana d'Allotjaments Turístics
establiments 2008.txt	Federació Catalana d'Allotjaments Turístics
establiments 2007.txt	Federació Catalana d'Allotjaments Turístics
establiments 2006.txt	Federació Catalana d'Allotjaments Turístics
Equipaments.csv	IDESCAT

figura 1: Taula resum d'arxius d'origen de dades proporcionades pel client

Adicionalment s'extreuen d'Internet, de la pàgina oficial de IDESCAT (Institut d'Estadística de Catalunya) els següents arxius :

- **Comarques :** Relació oficial de comarques (2012) amb la província/províncies a les que pertanyen. Aquesta extracció es necessària ja que relació comarca-província no existeix en cap dels arxius subministrats pel client.
- **Municipis.** Relació oficial de municipis (2012) amb la comarca a la que pertanyen i la superfície. Aquesta extracció és necessària ja que ni l'arxiu d'equipaments subministrat ni el de població contempla la totalitat de municipis (947) , i per tant la superfície subministrada en l'arxiu de població tampoc dona la totalitat de superfície de Catalunya (aprox. 32.000 Km²).
- **Població per sexe.** Cens de població oficial per municipis (anys de 2006 a 2012), separat per sexe. Aquesta extracció és necessària ja que al arxiu de població subministrat només ve l'any 2012 separat per sexes, i a més, no estan tots els municipis. Amb la certesa que la font d'informació és fiable, prendrem aquest arxiu com el bo per carregar la base de dades en detriment del de població.csv.

Arxiu de dades	Font d'Origen	Adreça d'origen	Descripció
aec-924.xls	IDESCAT	http://www.idescat.cat/pub/aec/924	Comarques
aec-925.xls	IDESCAT	http://www.idescat.cat/pub/aec/925	Municipis
PoblacioPerSexe.xls	IDESCAT	http://www.idescat.cat/territ/BasicTerr?TC=20&V0=3&V1=0&V3=669&V4=446&P=N&PA RENT=1&CTX=B&DISTR I=TRUE&ALLINFO =TRUE&ANYS=2012&x=8&y=6	Població per sexe

figura 2: Taula resum d'arxius d'origen de dades d'altres fonts

Arxiu Establiments

1. Les dades proporcionades sobre establiments venen en varis fitxers per any.

Decisions preses → per facilitar la càrrega s'unificaran en un sol arxiu.

2. En un mateix fitxer, ens trobem varies taules d'informació correlatives d'esquerra a dreta, on la columna de la comarca està repetida en algun cas. Aquestes taules fan referència als establiments en general, amb nombre i places i el detall dels 4 tipus d'establiments.

Concretament s'extreuen 4 taules d'informació del mateix fitxer:

- 1-Establiments
- 2-Establiments-Detall establiments hotelers
- 3-Establiments-Detall càmpings
- 4-Establiments-Detall turisme Rural

Decisions preses → Com el volum de dades no és massa gran, s'extreuen manualment les categories i tipologies d'establiments en un altre arxiu, que s'utilitzaran per carregar la taula de CATEGORIA ESTABLIMENTS.

3. Relacionat amb el punt anterior, També es detecta que la tipologia dels establiments no és la mateixa per a tots els anys. La següent taula marca, per cada any, quins tipus d'establiments es detecten:

Allotjaments	2006	2007	2008	2009	2010	2011	2012
Establiments hotelers							
Estrelles d'or							
Estrelles d'argent							
Hotels							
Hostals o pensions							
Càmpings							
Luxe							
1a							
2a							
3a							
Privats							
Turisme Rural							
allotjament independent							
Casa de poble							
Casa de poble compartida							
Casa de poble independent							
Masia							
Masoveria							

figura 3: Taula de tipologies d'establiment detectades

Decisions preses → en el cas dels establiments hotelers, ja que la tipologia del últim any és 'Hotel' i Hostals o pensions', les 'Estrelles d'or' s'assumeixen com 'Hotels' i les 'Estrelles d'argent' com a 'Hostals o pensions.'

En el cas del 'Campings' la tipologia 'Privats' s'assumeix com a 'Luxe'.

En el cas de 'Turisme Rural', el 'allotjament independent' s'assumeix com a 'Casa de poble independent', i la 'Casa de poble' com a 'casa de poble compartida'.

4. Una altra de la informació que cal controlar és que es mostra informació relativa a les comarques, i seguidament dels àmbits geogràfics i províncies com a agregacions de les comarques. En la següent extracció s'emmarquen per a visualitzar millor la diferència :

Solsonès	25	5	104	134	701	2.307	660	3.668
Tarragonès	128	25	10	163	40.215	32.232	73	72.520
Terra Alta	16	3	22	41	581	561	191	1.333
Urgell	10	0	35	45	412	0	287	699
Val d'Aran	109	9	16	134	6.371	3.069	148	9.588
Vallès Occidental	49	0	3	52	5.792	0	28	5.820
Vallès Oriental	59	8	48	115	3.734	2.805	398	6.937
Catalunya	2.837	353	2.156	5.346	294.777	277.069	16.968	588.814
Metropolità	940	41	68	1.049	121.802	26.418	543	148.763
Comarques Gironines	818	139	659	1.616	79.100	133.485	5.426	218.011
Camp de Tarragona	251	50	180	481	50.639	61.416	1.471	113.526
Terres de l'Ebre	97	16	107	220	5.374	7.869	857	14.100
Ponent	91	4	157	252	4.744	1.548	1.196	7.488
Comarques Centrals	157	36	456	649	5.321	12.667	3.685	21.673
Alt Pirineu i Aran	346	56	379	781	16.460	20.676	2.679	39.815
Penedès	137	11	150	298	11.337	12.990	1.111	25.438
Barcelona	1.163	76	548	1.787	133.019	44.446	4.489	181.954
Girona	864	145	698	1.707	81.281	136.503	5.725	223.509
Lleida	423	62	607	1.092	19.958	22.461	4.305	46.724
Tarragona	387	70	303	760	60.519	73.659	2.449	136.627

figura 4: Vista d'arxiu 'Establiments. Csv'

Decisions preses → Com s'han definit com a dimensió les comarques i la seva relació amb el àmbit territorial i la província, eliminarem les files referents a aquestes agregacions perquè es facin directament als informes. D'aquesta manera es resol també directament, la diferència entre l'àmbit territorial 'Penedès' aparegut a partir del arxius dels 2011.

Àmbit territorial	2006	2007	2008	2009	2010	2011	2012
Metropolità							
Comarques Gironines							
Camp de Tarragona							
Terres de l'Ebre							
Ponent							
Comarques Centrals							
Alt Pirineu i Aran							
Penedès							

figura 5: Taula d'àmbits territorials detectats

5. També es detecta en alguns arxius hi ha taules repetides.

Decisions preses → Es necessari actuació manual per eliminar la informació duplicada que pot implicar registres duplicats en la base de dades. Aprofitarem per formatejar les dades treien tot el que no és necessari.

Actuarem:

- eliminant informació duplicada.
- eliminant columnes innecessàries (en blanc o amb capçaleres)
- eliminant les files referents a àmbits territorials i províncies.
- eliminant les columnes de totals (que es calcularan en els informes).

Arxiu Equipaments

6. El fitxer d'equipaments ve en format text separat per comes. Aquest format és fàcil d'integrar però s'ha de tenir en compte, que si en algun dels atributs, com per exemple el nom, ve alguna coma, el formateig pot no ser correcte.

Decisions preses → Revisar l'arxiu abans de carregar-lo, obrint-lo, per exemple, en Excel i visualitzant les dades dels filtres de les capçaleres. En cas de trobar alguna incoherència rectificar-la manualment.

7. Es detecta que els accents i caràcters especials no s'obren correctament.

Decisions preses → Trobar algun traductor de caràcters i/o verificar que en procés ETL s'integra correctament.

8. També es detecta que la categoria del equipament ve amb valors separats per barres.

Decisions preses → Mitjançant un extracció manual de la columna (taula dinàmica, per exemple, s'extreuen manualment les categories d'equipaments, en un altre arxiu, que s'utilitzarà per carregar la taula de CATEGORIA EQUIPAMENTS.

9. La dada de comarca no és necessari importar-la

Decisions preses → La comarca s'obtindrà de la taula MUNICIPIS en el moment d'elaborar els informes, per tant es pot eliminar. .

Arxiu Població

10. En aquest arxiu ve la superfície del municipi, però no estan tots els municipis de Catalunya, a més la separació per sexe només està pel 2012.

Decisions preses → La informació sobre població i superfície s'obtindrà d'altres arxius obtinguts directament del IDESCAT (veure Arxiu Municipis i Arxiu Població per sexe)

Arxiu Comarques (aec-924)

11. De tota la informació d'aquest arxiu només es necessita el nom de la comarca i la província per poder carregar la taula COMARCA

Decisions preses → Eliminar manualment les files de capçalera i les no necessàries i deixar només les columnes de comarca i província.

Comarca	Capital de comarca	Províncies a les quals pertanyen els municipis	Nombre de municipis	Població
Alt Camp	Valls	Tarragona	23	45.299
Alt Empordà	Figueres	Girona	68	141.517
Alt Penedès	Vilafranca del Penedès	Barcelona	27	106.252
Alt Urgell	Seu d'Urgell, la	Lleida	19	21.386
Alta Ribagorça	Pont de Suert, el	Lleida	3	4.235

figura 6: Vista d'arxiu 'aec-924' de comarques

Arxiu Municipis (aec-925)

12. De tota la informació d'aquest arxiu només es necessita el nom del municipi, la comarca a la que pertany i la superfície.

Decisions preses → Eliminar manualment les files de capçalera i les no necessàries i deixar només les columnes de municipi, comarca i superfície.

La superfície territorial que s'extraurà en els informes s'obtindrà d'aquesta taula i no pas de la de població, que no conté tots els municipis.

Municipi	Comarca	Codi	Altitud (m)	Superfície (km ²)	Població
Abella de la Conca	Pallars Jussà	250019	956	78,27	174
Abrera	Baix Llobregat	080018	105	19,94	11.870
Àger	Noguera	250024	642	160,57	593
Agramunt	Urgell	250030	337	79,65	5.633
Aguilar de Segarra	Bages	080023	480	43,32	250
Agullana	Alt Empordà	170010	166	27,73	858
Aiguafreda	Vallès Oriental	080142	404	7,9	2.478
Aiguamúrcia	Alt Camp	430017	314	72,95	906
Aiguaviva	Gironès	170025	169	13,92	774
Àitona	Segrià	250387	110	66,9	2.419
Alamús, els	Segrià	250045	212	20,53	751
Alàs i Cerc	Alt Urgell	250058	768	57,67	376

figura 7: Vista d'arxiu 'aec-925' de superfície de municipis.

Arxiu Població per sexe

13. Després d'analitzar les dades de població d'aquest arxiu es detecta que la població per municipi no es correspon amb la existent en l'arxiu població.

Exemple sobre el municipi d'Abrera:

En arxiu població

	A	J	K	L
1	Municipi	Població 2012 homes	Població 2012 dones	Extensió (km2)
2	Abrera	5.926	5.685	20

figura 8: Vista d'arxiu 'població.csv'

En arxiu extret de IDESCAT

	A	B	C	D
1	Padró municipal d'habitants per sexe. Xifres oficials			
2	Catalunya. Distribució per municipis. Any 2012.			
3				
4	Sexe			
5				
6		Homes	Dones	Total
7				
8	Abella de la Conca	98	76	174
9	Abrera	6072	5798	11870

figura 9: Vista d'arxiu 'PoblacióPerSexe.xls'

Decisions preses → Tenint en compte que la extracció és d'una font fiable, prendrem les dades de població extretes en l'arxiu PoblacioPerSexe.xls, que està separada per sexe en tots els anys i conté tots els municipis de Catalunya, enlloc de la subministrada en Poblacio.csv

2.2.- REQUERIMENTS FUNCIONALS

Els requeriments funcionals d'un sistema descriuen el que el sistema ha de fer.

En el nostre cas la solució ha de contemplar :

- ⇒ La creació d'una base de dades que emmagatzemi, a trets generals , la informació referent a Població, Equipaments i Allotjaments, en taules normalitzades .
- ⇒ Un procés que unifiqui i carregui en la base de dades, la informació rebuda del IDESCAT y de FCAT i d'altres fonts , en format .csv i .txt respectivament. (Procés ETL).
- ⇒ Eina per la explotació i anàlisi OLAP de les dades de la base de dades.

⇒ Informes varis predefinit, que mitjançant algun petició d'ordre, demani a l'usuari diferents filtres que com a mínim seran d'any, comarca/província, tipus d'establiment i categoria.

Els requeriments funcionals es poden resumir en el següent gràfic:

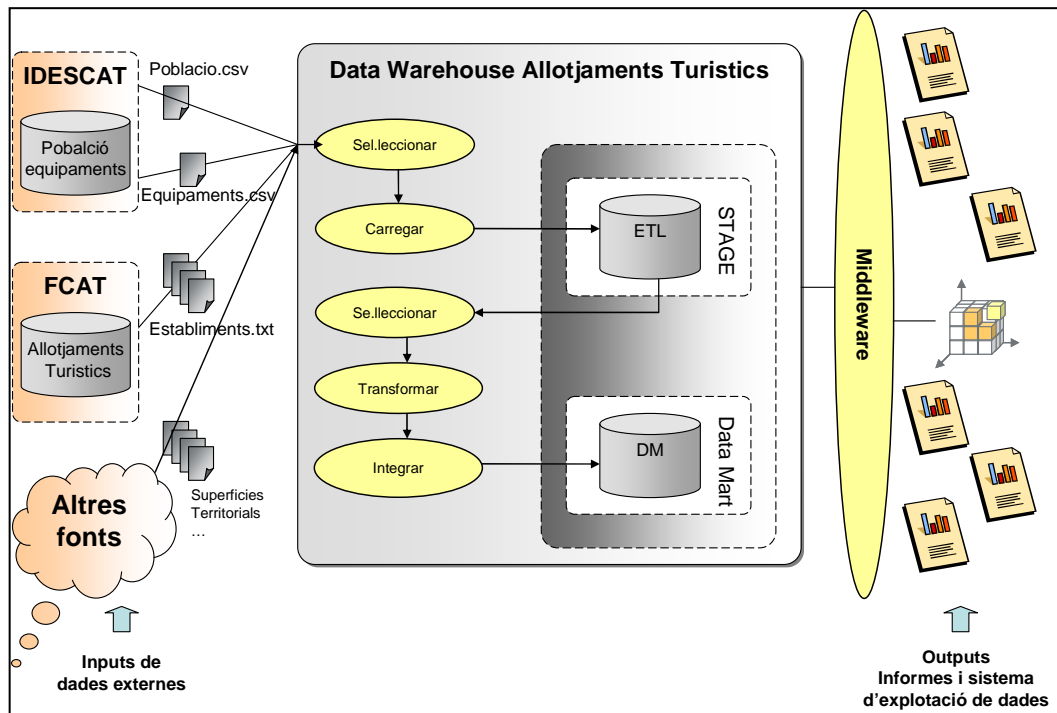


figura 10: El DW d'allotjaments turístics

2.3.- REQUERIMENTS NO FUNCIONALS

Els requeriments no funcionals, son aquells que no es refereixen directament a les funcions específiques que proporciona el sistema, sinó a les propietats d'aquest, com la fiabilitat, el temps de resposta, la capacitat d'emmagatzemar, etc.

Destacarem les següents propietats del sistema :

- ✓ Seguretat
- ✓ Escalabilitat
- ✓ Facilitat d'us.
- ✓ Viabilitat

A més per a que el nostre sistema funcionari definirem una arquitectura de maquinari i de programari.

2.3.1. SEGURETAT

El sistema estarà restringit per usuari i clau d'accés .

Es preveuen tres tipus d'usuari amb tres tipus de perfil diferent per l'accés al sistema. En el següent apartat es descriuen els diferents perfils i els rols d'us del sistema per cadascun d'ells.

2.3.2. ESCALABILITAT

El sistema ha d'estar capacitat per permetre en el futur el desenvolupament de noves funcionalitats, modificacions o eliminació de funcionalitat que hagin quedat obsoletes.

En el nostre cas, dissenyarem la base de dades de manera que sigui fàcil i ràpid afegir noves dades com tipus d'establiments o categories, i/o l'àmbit d'estudi com és el cas d'afegir com atribut la comunitat autònoma per poder ampliar a Espanya l'amplitud de dades a analitzar

2.3.3. FACILITAT D'US

El nostre sistema ha de ser de fàcil us i entrenament .

A més es preveuran missatges d'error per a les operacions no permeses i filtres auto-explicatius per a la extracció dels informes.

2.3.4. VIABILITAT

La viabilitat es refereix a valorar si el model en estrella dissenyat és viable o no, calculant l'espai necessari en el sistema pel seu emmagatzemament. /veure el disseny de la base de dades previst en l'apartat corresponent).

Volum de dades

Temps → del 2006 al 2012. 7 registres.

Municipi → 947 registres

Població → 7 registres d'any x 947 registres de municipi = 6.629 registres

Categoria equipaments → 1.101 registres

Equipaments → 15800 registres dels primers 4 nivells.

Categories establiments → 10 registres.

Establiments → 7 registres d'any x 41 registres de comarca x 10 registres de categoria d'establiment : 2.870 registres.

A continuació s'avalua l'espai segons la longitud màxima de registre previst (només amb els atributs necessaris):

	Registres	Longitud en bytes	Total espai en bytes	en KB	en MB
Temps	7	4	28		
Municipi	947	148	140.156		
Categoria equipaments	220	324	71.280		
Categories establiments	10	64	640		
Establiments	2.870	50	143.500		
Població	6.629	16	106.064		
Equipaments	15.800	54	853.200		
Totals	4.054	660	1.314.868	1.284	1

figura 11: Taula resum de necessitat d'espai de BBDD

Rendiment

Segons els nostres càlculs, es necessari un espai aproximat de 1 MB per emmagatzemar les dades en la nostra base de dades. A aquest espai, s'ha d'afegir tot aquell espai que el gestor de base de dades requereix per la creació de claus primàries, claus foranes, etc.

Aquest espai necessari no representa cap problema de rendiment al sistema si tenim en compte les capacitats i rapidesa dels processadors actuals.

2.4.- DIAGRAMA DE CASOS D'US

El diagrama de casos d'us representa la forma en com el client/s (actor/s) opera amb el sistema que s'està desenvolupant.

En el nostre Data Warehouse hem definits els següents actors :

- un usuari Final que només podrà visualitzar els informes. Serà l'usuari que analitzi els informes, tant a nivell gerencial com a nivell operatiu.
- un usuari 'Key user' que administra els perfils d'usuaris, i te accés a les eines d'explotació i als informes, però no pot manipular les dades. Aquest usuari és el que dona accés a la resta d'usuaris operatius i te nivell i coneixement suficient per crear els informes que li demanin els usuaris finals.
- un usuari administrador que te accés a tot el programari, al manteniment de la base de dades i processos ETL, a les eines d'explotació i al informes. És l'usuari 'master' i és també el que fa les còpies de seguretat del sistema.

Casos d'us usuari Final.

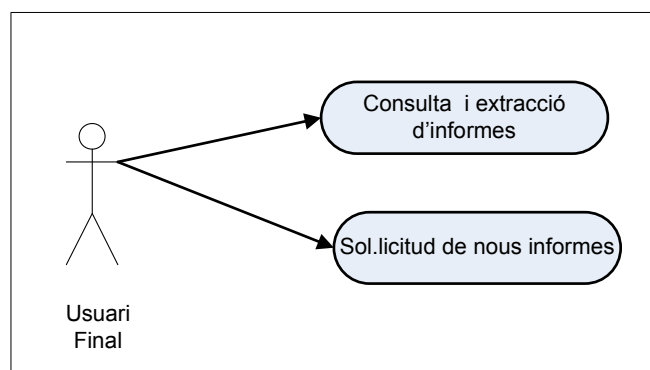


figura 12: Cas d'us usuari final

Casos d'us usuari Key User

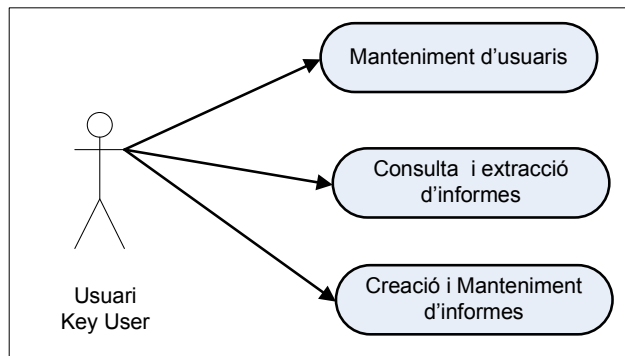


figura 13: Cas d'us key user

Casos d'us usuari Administrador

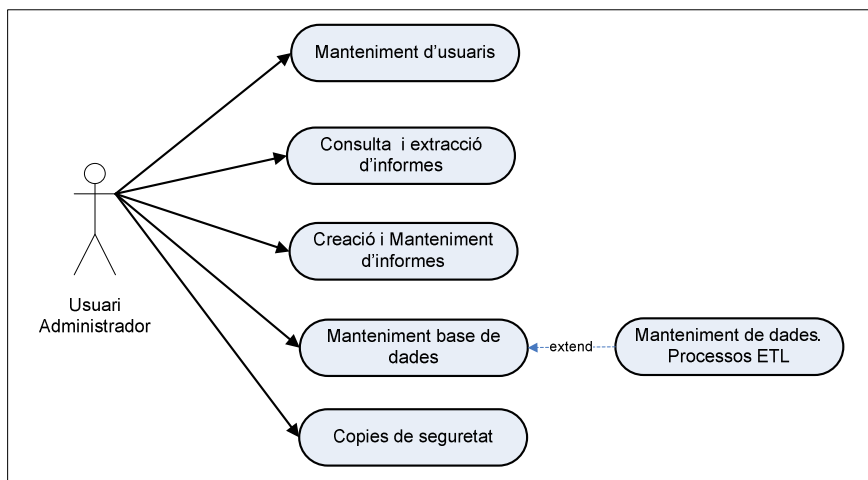


figura 14: Cas d'us usuari administrador

2.5.- MODEL CONCEPTUAL

En aquesta fase de l'anàlisi, obtindrem una estructura de la informació de la BD independent de la tecnologia que cal emprar.

El model conceptual no es més que un esquema de les dades i de les associacions o interrelacions que s'han obtingut a partir del coneixement dels requeriments i les especificacions del que es vol obtenir finalment.

Partirem de les dades d'origen que ens proporciona el client, i que s'han detallat en el estudi inicial :

- ⇒ Establiments XXXX: Un arxiu per cada any amb: nombre d'establiments, tipologia, places ofertes i classificació
- ⇒ Equipaments. Arxiu amb la relació d'equipaments públics a data 31/12/2012.
- ⇒ Població: Un arxiu amb els habitants per any i la superfície de cada municipi

A més obtindrem d'Internet la relació entre municipi, comarca, àmbit territorial i província.

2.5.1. ÀNALISI DEL MODEL MULTI-DIMENSIONAL

Analitzant les dades d'origen que ens proporciona el client, i coneixent els objectius d'anàlisi de les dades que ens planteja, el més apropiat és utilitzar una model de dades OLAP (de l'anglès, *On-Line Analytical Processing*), que és una solució utilitzada en el camp de la Intel·ligència empresarial (*Business Intelligence*) que té l'objectiu d'agilitzar la consulta de grans quantitats de dades.

Tota eina OLAP ha de ser multidimensional. La multidimensionalitat consisteix en concebre les dades a analitzar en termes de **fets** (el que volem analitzar) i **dimensions** (que utilitzarem per analitzar-los), de manera que les puguem situar en un espai n-dimensional (també anomenat cub multidimensional o hipercub).

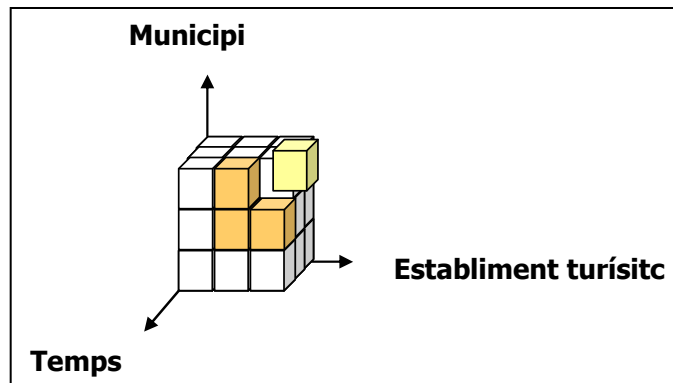


figura 15: Cub multidimensional.

Aquesta visualització en 'cub' permet realitzar combinacions de taules i camps que tinguin relacions lògiques.

- ✓ Un **fet** és un objecte d'anàlisi, és a dir '**el que volem medir o analitzar**'.
- ✓ Una **dimensió** representa un punt de vista que utilitzarem en l'anàlisi de dades, és a dir '**com ho volem medir**', i s'usen per seleccionar i agregar dades a un cert nivell de detall.
- ✓ Dins d'una dimensió podem distingir grups d'instàncies segons el seu tamany (granularitat).

Per la elaboració de les taules de dimensions i fets s'utilitzen varis mètodes. El que utilitzarem nosaltres serà el model en 'estrella'.

El model en 'estrella' és desnormalitzat (pot contenir informació repetida), amb el que s'aconsegueix minimitzar el nombre d'unions entre les taules, i en conseqüència, s'incrementa el rendiment de la consulta.

Les taules de fets són les que generalment es localitzen al centre, i que es relacionen amb les dimensions situades al voltant.

La definició de les dimensions i fets correctes és molt important ja que permetran al l'usuari construir consultes e informes amb les característiques que es requereixin i la combinació de camps de les diferents dimensions.

2.5.2. ELS INDICADORS CLAU DE RENDIMENT (KPI)

Un KPI és simplement un indicador que està vinculat a un objectiu.

En la majoria dels cassos, un KPI és l'estat d'un indicador (o càlcul associat a un cub) , és a dir , si està per sobre o per sota d'una meta pre-determinada.

Els KPI's es dissenyen per permetre que un usuari de negoci pugui saber instantàniament si està dins o fora del seu pla, sense haver de buscar informació addicional.

- ⇒ En aquesta línia es dissenyen els informes de control (dashboard), que son la síntesis de la informació que agrupa tots els indicadors necessaris per a la presa de decisions.

El nostre client , l'Observatori Nacional d'Ocupació (ONdO), ens diu que vol aprofundir en l'evolució dels establiments turístics de Catalunya, que ofereixen gairebé sis-centes mil places a Catalunya, i analitzar les possibles correlacions entre allotjaments, equipaments públics i població.

A falta de poder consultar al propi client, definirem com KPIs els següents ítems:

1-Evolució places establiments turístics vs població

Descripció: Medeix la relació entre la evolució anual de places d'establiments turístics i la població, dividida per gèneres, del mateix període.

Objectiu Estratègic: Es tracta d'avaluar l'equilibri entre el nombre de persones possibles a ocupar les places d'establiments turístics ofertades, i l'índex de població, per gènere de la zona.

Mètrica: Rati entre places d'establiments i persones per gènere.

Meta: No fixada.

2-Evolució places restabliments turístics vs equipaments públics

Descripció: Medeix la relació entre la evolució anual de places d'establiments turístics i la evolució, per al mateix període, d'equipaments públics.

Objectiu Estratègic: Es tracta d'avaluar si és necessari l'ampliació d'alguns dels equipaments públics de en base a l'increment de visitants segons la oferta de places dels establiments turístics, per zones.

Mètrica: Percentatge entre nombre de places (quantitat) d'establiments entre nombre d'equipaments.

Meta: No fixada.

3- Nombre d'establiments turístics per zona geogràfica.

Descripció: Medeix la quantitat d'establiments turístics, dividits per tipologia d'establiment, per comarca, àmbit territorial o província.

Objectiu Estratègic: Localitzar les àrees geogràfiques que concentren la major part d'establiments turístics.

Mètrica: Quantitat d'establiments per zona.

Meta: No fixada.

2.5.3. DIMENSIONS

Les dimensions, tal i com s'ha dit anteriorment, son les perspectives d'anàlisi de les variables. Son els catàlegs d'informació complementaria necessària per la presentació de les dades als usuaris, com poden ser per exemple les descripcions, noms, zones, rang de temps, etc. És doncs, la informació general complementaria a cadascun dels registres de les taules e fets, però que no formen part de l'anàlisi:

En el projecte s'han detectat les següents :

Context Temps :

Dimensió TEMPS amb els anys a processar

Context Territori :

Dimensió MUNICIPI amb tots els municipis de Catalunya extrets mitjançant un arxiu csv de la pàgina de IDESCAT a Internet amb la comarca a la que pertany, àmbit territorial i comunitat autònoma.

Context Equipaments :

Dimensió CATEGORIA EQUIPAMENTS amb la relació de les diferents categories existent en l'arxiu Establiments.

Context Establiments :

Dimensió CATEGORIA ESTABLIMENTS amb les categories d'establiments turístics existents.

2.5.4. FETS

Els fets contenen les dades pròpies per a l'anàlisi .Com s'ha dit anteriorment es volen analitzar :

El nombre i places d'establiments turístics : EVOLUCIO ESTABLIMENTS

La seva comparativa amb la població : POBLACIO

La seva comparativa amb els establiments públics existents : EQUIPAMENTS

En el següent diagrama es mostra el fets (en rosa) i la relació amb les diferents dimensions (en groc) definides.

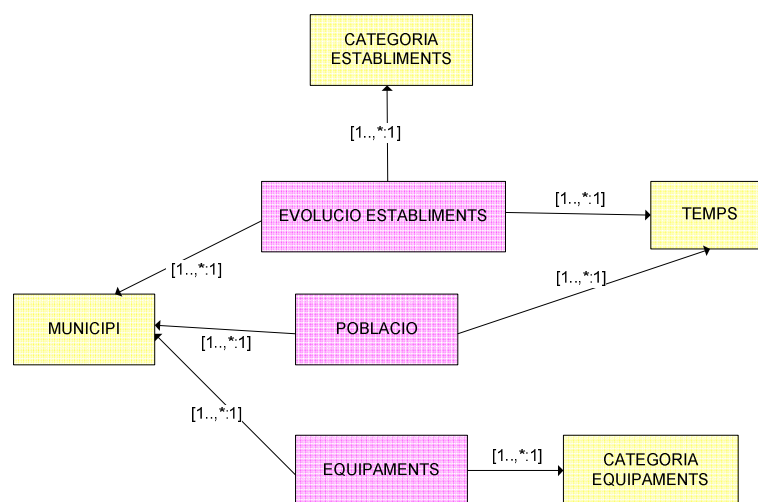


figura 16: Diagrama de concepte

2.5.5. MESURES E INDICADORS

Les mesures son les dades numèriques associades als esdeveniments. De fet son els conceptes o mides de negoci, que s'han detectat per assolir els objectius del client, i que queden reflexades en el fet/s definits.

En el nostre cas, les mesures seran les següents:

- ✓ Nombre d'establiments turístics.
- ✓ Nombre de places dels establiments.
- ✓ % de places respecte població
- ✓ % de població per equipament
- ✓ Oferta mitjana de places
- ✓ Nombre d'establiments/Nombre d'equipaments
- ✓ Indicador d'establiments vs habitants per gènere
- ✓ Indicador de places vs persones (rati)
- ✓ Indicador d'equipaments vs població
- ✓ Quantitat de places ofertes / superfície del territori (rati)

2.5.6. GRANULITAT

La granulitat és la grandària d'un objecte respecte a un altre. La granularitat defineix el nivell de detall que existeix en les dimensions i fets i determina un disseny òptim d'accés a les dades.

La granulitat també determina el volum de dades de la base de dades ja que a major granularitat (més detall) major serà el número de registres final de la taula de fet.

En aquest projecte s'ha detectat que alguns dels objectes d'anàlisi tenen granulitat (o jerarquies diferents), com pot ser el de nombre i places d'establiments turístics que va a nivell de comarca, i el d'establiments o població que ho fan a nivell de municipi.

Aquest conflicte s'ha resol creant per separat les dimensions de comarca i municipi de manera que qualsevol consulta de dades a aquest dos nivell puguin ser independents, deixant el de màxim nivell (comarca) en la taula de fet.

3.- DISSENY

3.1.- DIAGRAMA D'ARQUITECTURA DE MAQUINARI

No disposen la informació referent a la infraestructura actual del nostre client, per tant proposem com a mínim els següents elements resumits en la següent figura:

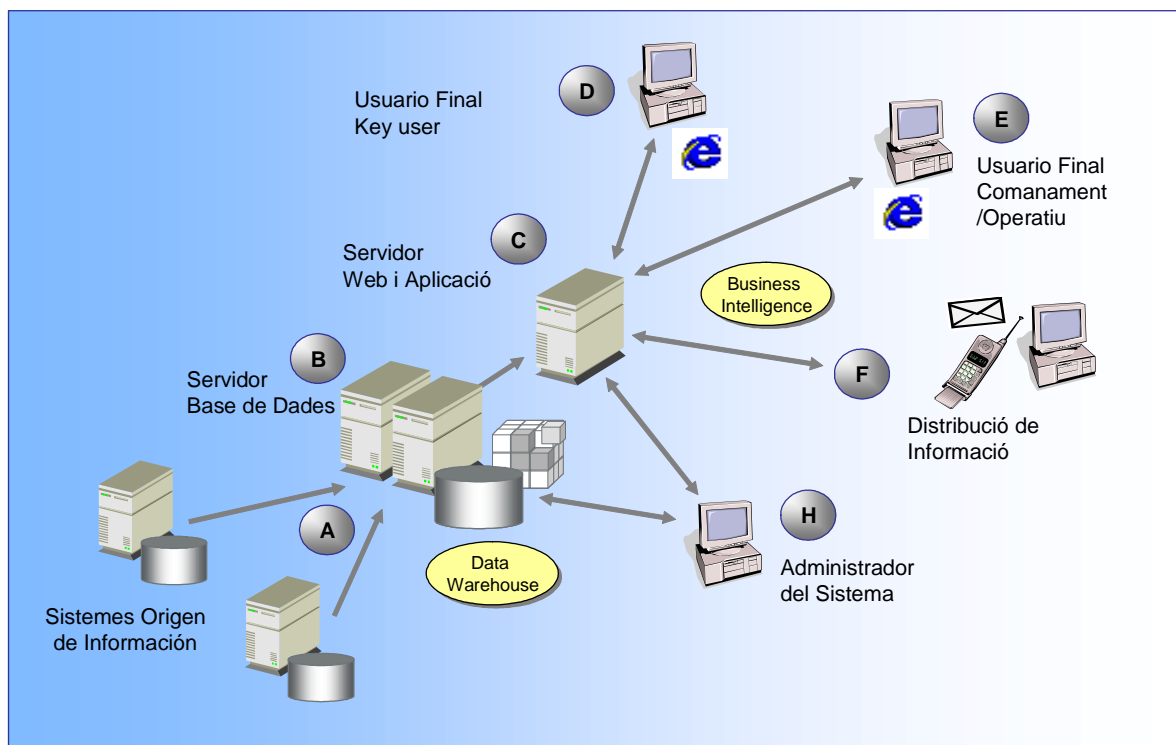


figura 17: Arquitectura de maquinari

A: les màquines pròpies de les dades d'origen.

B: un servidor per a la base de dades

C: un servidor on instal·lar el programari propi del Business Intelligence

D : una màquina/es per al usuari final que farà de key user

E : una màquina /es per al usuari final de comandament u operatiu.

F : una màquina o perifèric que s'utilitzi per a la distribució dels informes

H : una màquina per l'administrador del sistema.

3.2.- DIAGRAMA D'ARQUITECTURA DE PROGRAMARI

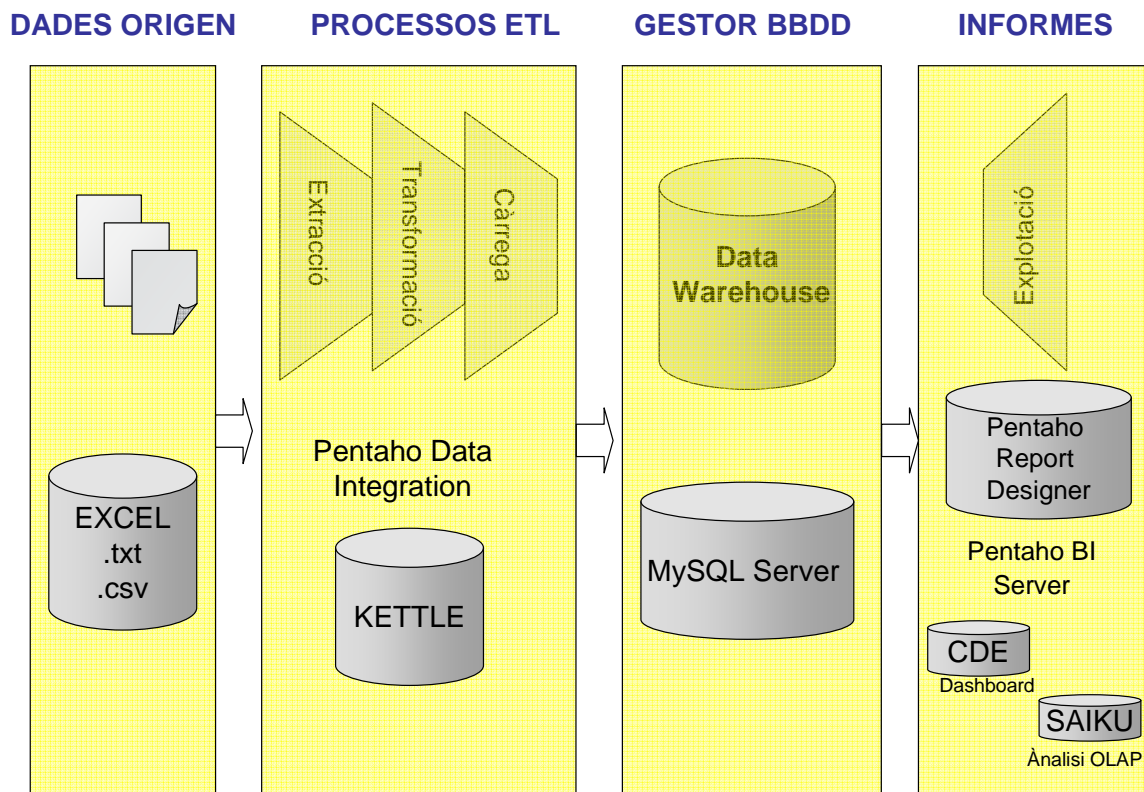


Figura18: Arquitectura de programari

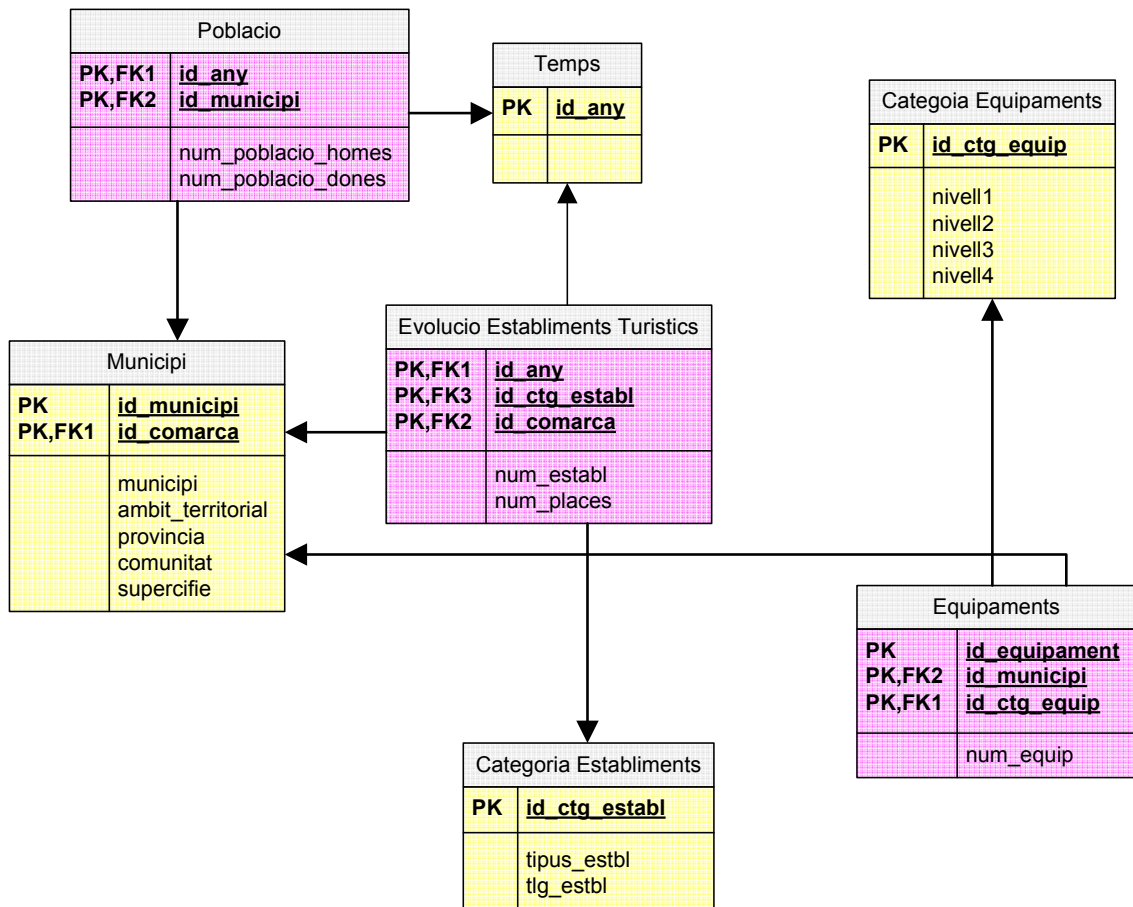
A continuació es mostren els components de programari que formen part de l'arquitectura del projecte:

- ORACLE VM VIRTUALBOX : Maquina virtual
- Pentaho Data Integration v.4.4.0 - KETTLE (Keettle Extraction TRansportation Transformation Loading Environment) : Eina per a processos ETL
- MySQL Server 5.5 : gestor de base de dades.
- Pentaho Report Designer : Editor d'informes.
- Pentaho BI Server CE 3.:plataforma de BI (Business Intelligence) amb la infraestructura necessària per crear solucions d'intel·ligència empresarial. Inclou una consola per poder accedir als informes via navegador.
- Pentaho Community Dashboard Editor (CDE) : Editor de informes de quadres de comandament.
- Pentaho SAIKU : Eina per l'anàlisi OLAP

3.3.- DISSENY DE LA BD I DIAGRAMA DEL MODEL FÍSIC

Tal i com s'ha definit anteriorment, en l'anàlisi de la base de dades del nou DW, l'esquema en estrella es el més adequat, amb un model de dades que te les taules de fets (el que volem medir o analitzar), rodejades de les taules de dimensions (com ho volem medir).

Primer mostrarem el diagrama entitar-relació i a continuació el detall de les taules definides



A més , les taula de fets estaran relacionades per claus foranes a les taules de Dimensions.

A continuació es detallen les taules amb els seus atributs, primer les Dimensions i per finalitzar les taules de Fets

Les DIMENSIONS

TEMPS DIMENSIO

Atribut	Tipus	Descripció	Possibles valors	Origen	Long
id_any	NUM	identificador d'any	2013, 2012, 2010...	entrada manual	4

MUNICIPI DIMENSIO

Atribut	Tipus	Descripció	Possibles valors	Origen	Long
id_municipi	NUM	identificador de municipi			4
Municipi	STRING	Non Municipi	Barcelona, Hospitalet de Llobregat...	fitxer csv IDESCAT extern	50
Comarca	STRING	Nom Comarca	Baix Llobregat, Berguedà...	fitxer csv IDESCAT extern	30
Ambit_territorial	STRING	Àmbit territorial	Àre metroolitana, Penedès	fitxer csv IDESCAT extern	20
Provincia	STRING	Provincia	Barcelona, Tarragona...	fitxer csv IDESCAT extern	20
Comunitat	STRING	Comunitat autonoma	Catalunya	fitxer csv IDESCAT extern	20
Superficie	NUM	Superficie municipi en KM 2			4
					148

CATEGORIA EQUIPAMENTS DIMENSIO

Atribut	Tipus	Descripció	Possibles valors	Origen	Long
id_ctg equip	NUM	identificador de categoria d'equipament			4
nivell_1	STRING	Categoria de nivell 1	Equipaments	equipaments.csv	80
nivell_2	STRING	Categoria de nivell 2	Cultura	equipaments.csv	80
nivell_3	STRING	Categoria de nivell 3	Museus	equipaments.csv	80
nivell_4	STRING	Categoria de nivell 4	Arqueologia	equipaments.csv	80
					324

CATEGORIA ESTABLIMENTS DIMENSIO

Atribut	Tipus	Descripció	Possibles valors	Origen	Long
id_ctg_establ	NUM	identificador de categoria d'establiment			4
tipus_establ	STRING	Tipus establiment	Hotel, Camping, Turisme rural...	establiments XXXX.txt	30
tig_establ	STRING	Tipologia establiment	luxe, hostel, masia independent...		30
					64

i els FETS

EVOLUCIO ESTABLIMENTS FET

Atribut	Tipus	Descripció	Possibles valors	Origen	Long
id_any	NUM	identificador d'any		FK amb TEMPS	4
Comarca	STRING	Comarca		FK amb MUNICIPI	30
id_ctg_establ	NUM	Categoria establiments		FK amb CATEGORIA ESTABLIMENTS	4
num_establ	NUM	Nombre d'establiments		establiments XXXX.txt	6
num_places	NUM	Nombre de places		establiments XXXX.txt	6
					50

POBLACIO FET

Atribut	Tipus	Descripció	Possibles valors	Origen	Long
id_any	NUM	identificador d'any	2013, 2012, 2010...	FK amb TEMPS	4
id_municipi	NUM	identificador de municipi		FK amb MUNICIPI	4
num_poblacio_dones	NUM	Nombre d'habitants homes del municipi		PoblacioPerSexe.xls	4
num_poblacio_homes	NUM	Nombre d'habitants dones del municipi		PoblacioPerSexe.xls	4
					16

EQUIPAMENTS FET

Atribut	Tipus	Descripció	Possibles valors	Origen	Long
id_municipi	NUM	identificador municipi		FK amb MUNICIPI	4
id_ctg equip	NUM	identificador categoria equipament		FK amb CATEGORIA EQUIPAMENTS	4
num equipaments	STRING	Sumatori del nombre d'equipaments de la ...		equipaments.csv	50
					54

Cal destacar que les claus primàries de les taules es codifiquen com a numeradors seqüencials en format integer. D'aquesta manera reduïm espai en la base de dades e independitzem l'origen de les dades de l'atribut identificador. Només hi ha una excepció d'aquest cas i és l'atribut comarca de la taula Evolució establiments que es deixa amb el valor original a fi d'evitar els joins entre taules.

En el disseny de les taules s'han eliminat els atributs original que no s'usaran pròpiament, com son les dades dels equipaments, deixant només un numerador que contabilitza en nombre en el procés de transformació.

De la mateixa manera, s'han reduït el nombre de nivells de les categories dels equipaments ja no es necessitava tant nivell de detall.

3.4.- CÀRREGA DE LA BD – PROCÉS ETL

Tal i com ja s'ha comentat anteriorment, la base de dades es carregarà en el procés ETL, que extraurà / seleccionarà del dades d'origen, les transformarà i les carregarà (Load) a les taules definides.

De com es transformaran les dades s'ha parlat en el capítol 2.1 (Anàlisi- Estudi dades d'origen), i en el capítol (4.2-Captures de pantalla-procés ETL) es detalla el procés ETL seguit.

Finalment, l'esquema de base de dades creat a MySQL workbench ha estat el següent

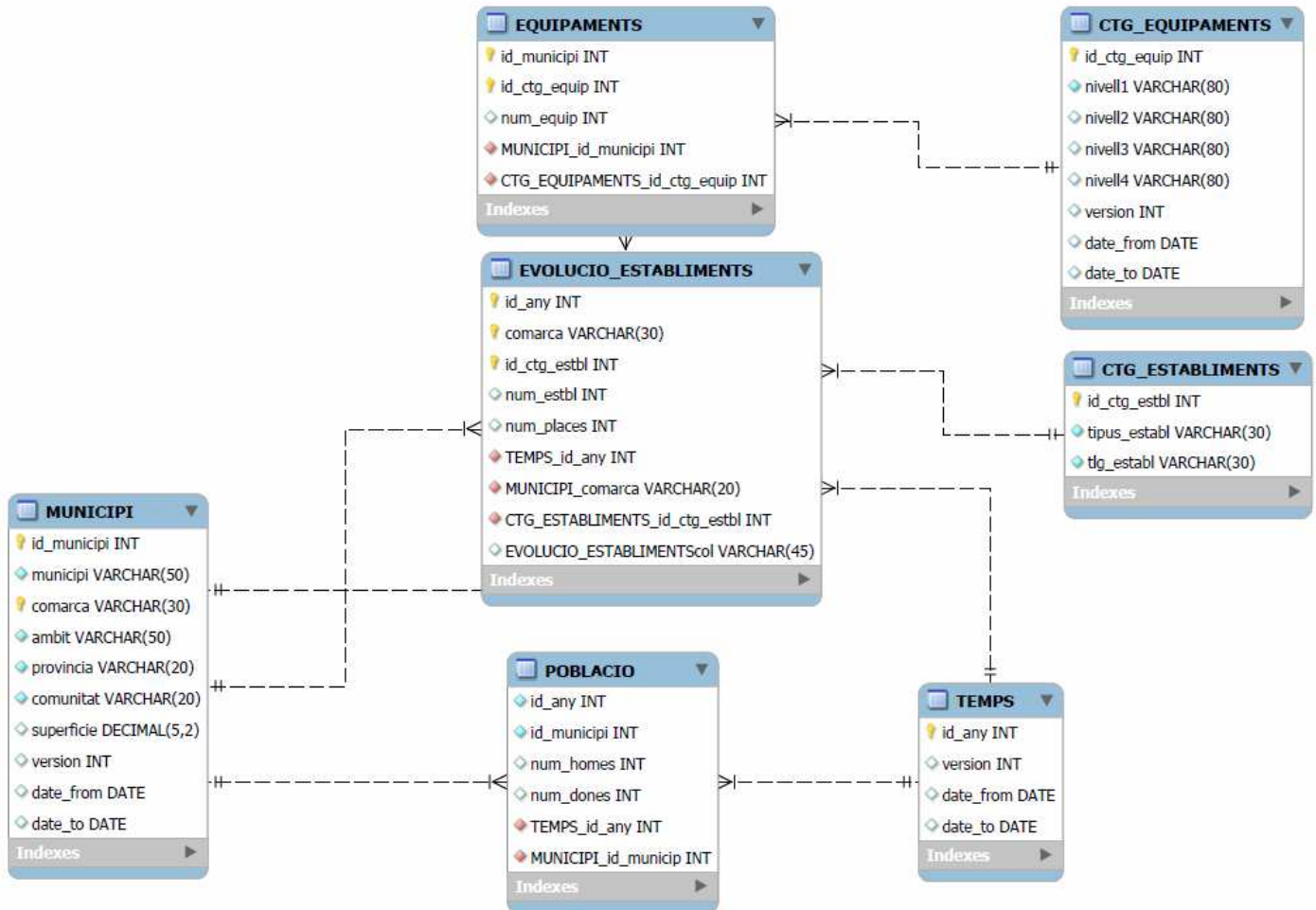


figura 21: Esquema de base de dades

Les claus primàries i les forànies entre les taules son les definies i descrites en e model de dades.

3.5.- DISSENY I DESCRIPCIÓ DELS IFORMES

A continuació es detallen els informes dissenyats, amb la seva descripció, els elements que els componen i quin paràmetre són necessaris. En el capítol de Annexes d'aquest document s'incorporen els informes en pdf.

3.5.1. INFORME 1. Places establiments vs superfície i habitants.

Informe executiu, d'una sola plana, que inclou gràfics de varis tipus i detall de dades, de totals de places d'establiments ofertades a Catalunya, població, equipaments i superfície territorial.

Paràmetres : cap

Elements :

- Gràfic de línies amb l'evolució del nombre de places ofertades en cada província, per any
- Detall de la oferta mitjana de places per província de tots els anys amb el total de Catalunya.
- Gràfic de línies amb l'evolució de la població de cada província, per any.
- Detall de la població mitjana per província i sexe de tots els anys amb el total de Catalunya.
- Gràfic de barres amb el total d'equipaments per província
- Gràfic circular de superfície per província amb total de superfície de Catalunya

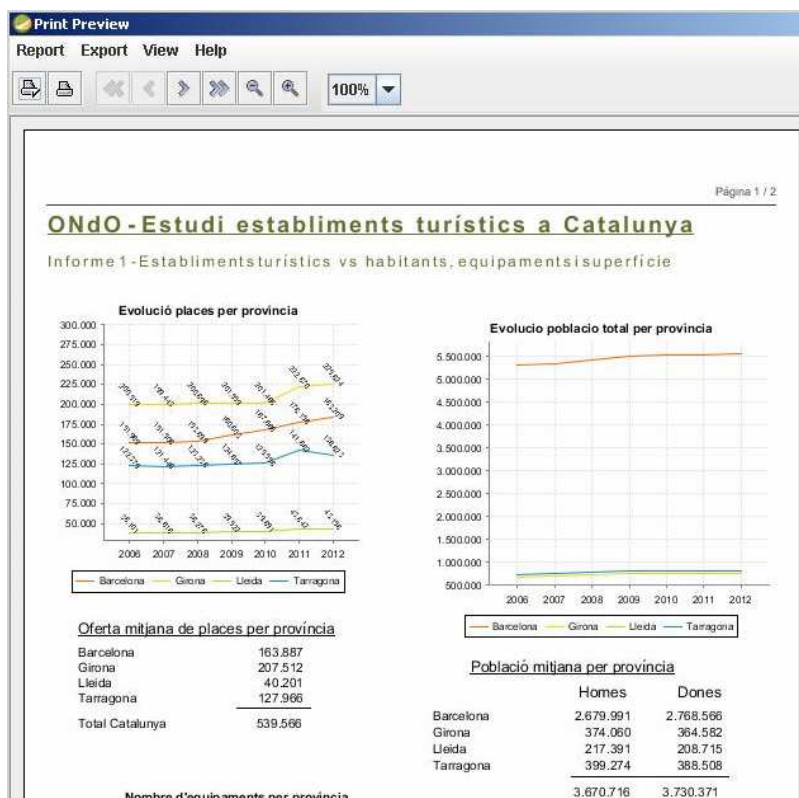


figura 22: Disseny Informe 1

3.5.2. INFORME 2 . Total establiments i places ofertades.

Informe executiu, d'una sola plana, que inclou gràfics circulars de totals d'establiments i places ofertades a Catalunya en el any especificat.

Paràmetres : any

Elements :

- Gràfic circular de total de nombre d'establiments per tipus d'establiment.
- Gràfic circular de total de places d'establiments ofertades, per tipus d'establiment.
- Grup de 3 gràfics circulars de totals d'establiments, un per cada tipus d'establiments, i amb la tipologia d'aquests.
- Grup de 3 gràfics circulars de totals de places d'establiments, un per cada tipus d'establiments, i amb la tipologia d'aquests.

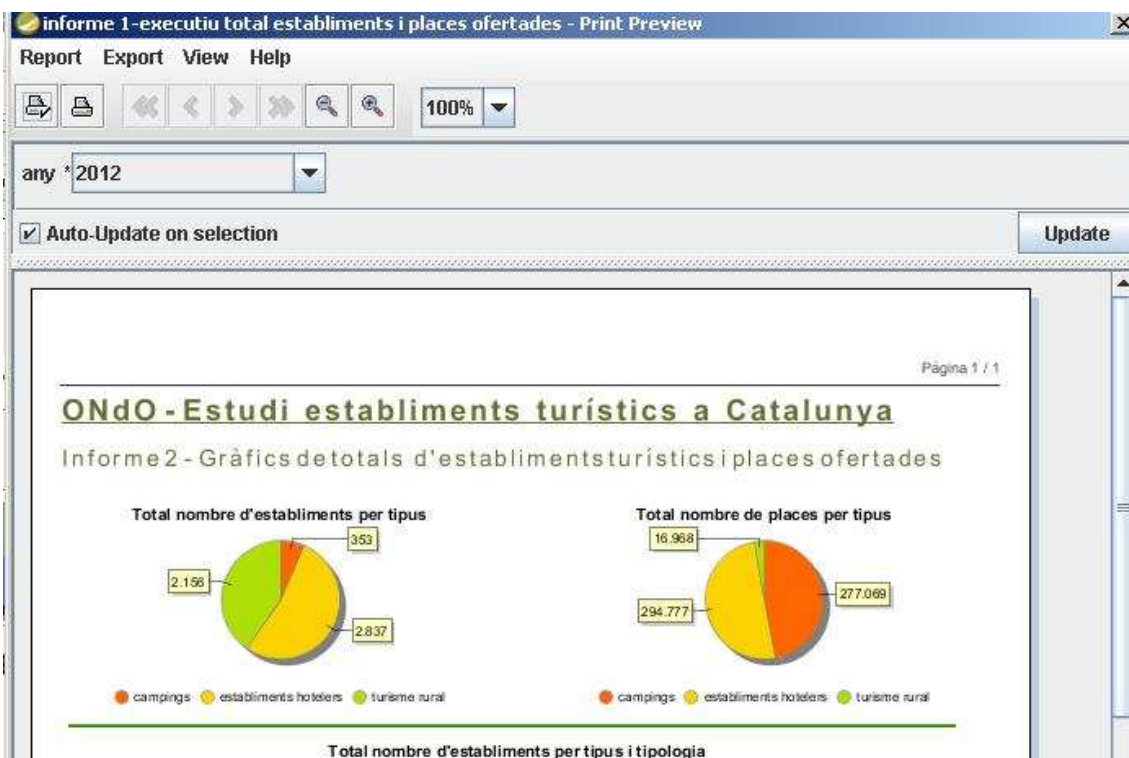


figura 23: Disseny Informe 2

3.5.3. INFORME 3 . Evolució total establiments i places ofertades.

Informe executiu, d'una sola plana, que inclou gràfics de barrera de la evolució d'establiments i places ofertades a Catalunya.

Paràmetres : cap

Elements :

- Gràfic de barres de total de nombre d'establiments per any.
- Gràfic de barres circular de total de places d'establiments ofertades per any
- Gràfic de barres de nombre d'establiments i tipus d'establiments per any
- Gràfic de barres de places d'establiments i tipus d'establiments per any.
- Gràfic de barres de nombre d'establiments per província i any.
- Gràfic de barres de places d'establiments per província i any.



figura 24: Disseny Informe 3

3.5.4. INFORME 4 . Rànquing d'establiments per zona.

Informe operatiu amb gràfics i detall del rànquing de places d'establiments per comarca, àmbit territorial i província.

Paràmetres : any

Elements :

- Gràfic de barres ordenades per nombre de places de major a menor, per comarca de l'any seleccionat.
- Gràfic de barres de nombre de places per àmbit territorial de l'any seleccionat. Cal destacar que en aquest gràfic les barres no estan ordenades pel valor absolut de tot el àmbit territorial, si no que es situa en posició relativa al rànquing per comarques.
- Gràfic de barres de nombre de places per província de l'any seleccionat. Cal destacar que en aquest gràfic, al igual que l'anterior, les barres no estan ordenades pel valor absolut de tota la província , si no que es situa en posició relativa al rànquing per comarques.
- Detall de nombre d'establiments i places ofertades ordenades pel nombre de places.

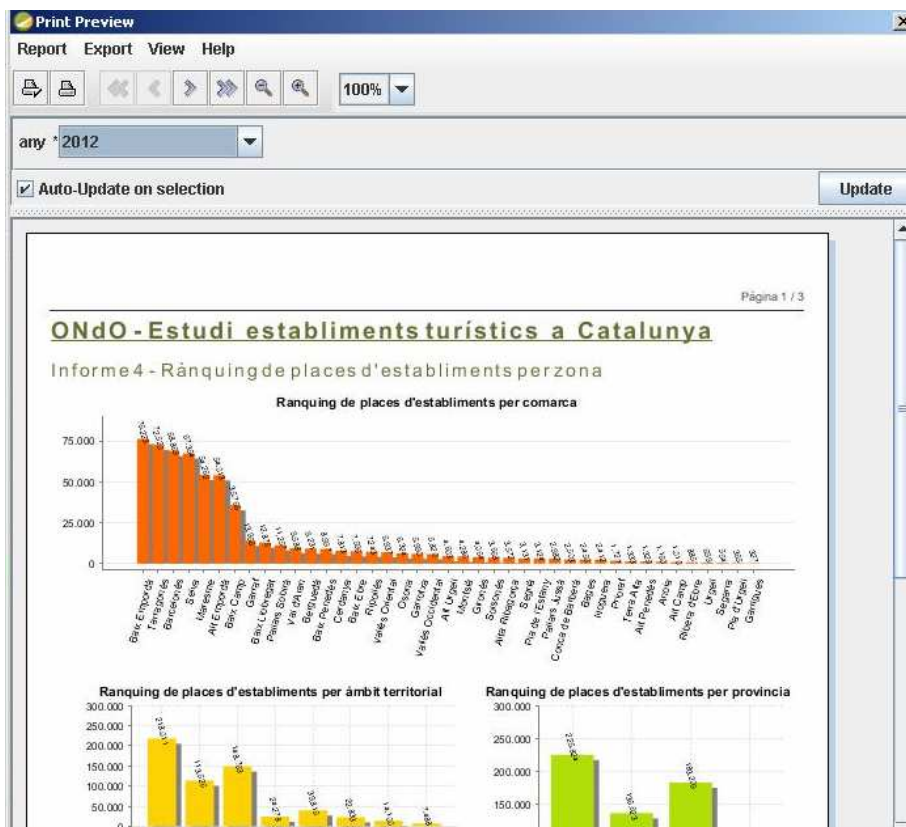


figura 25: Disseny Informe 4

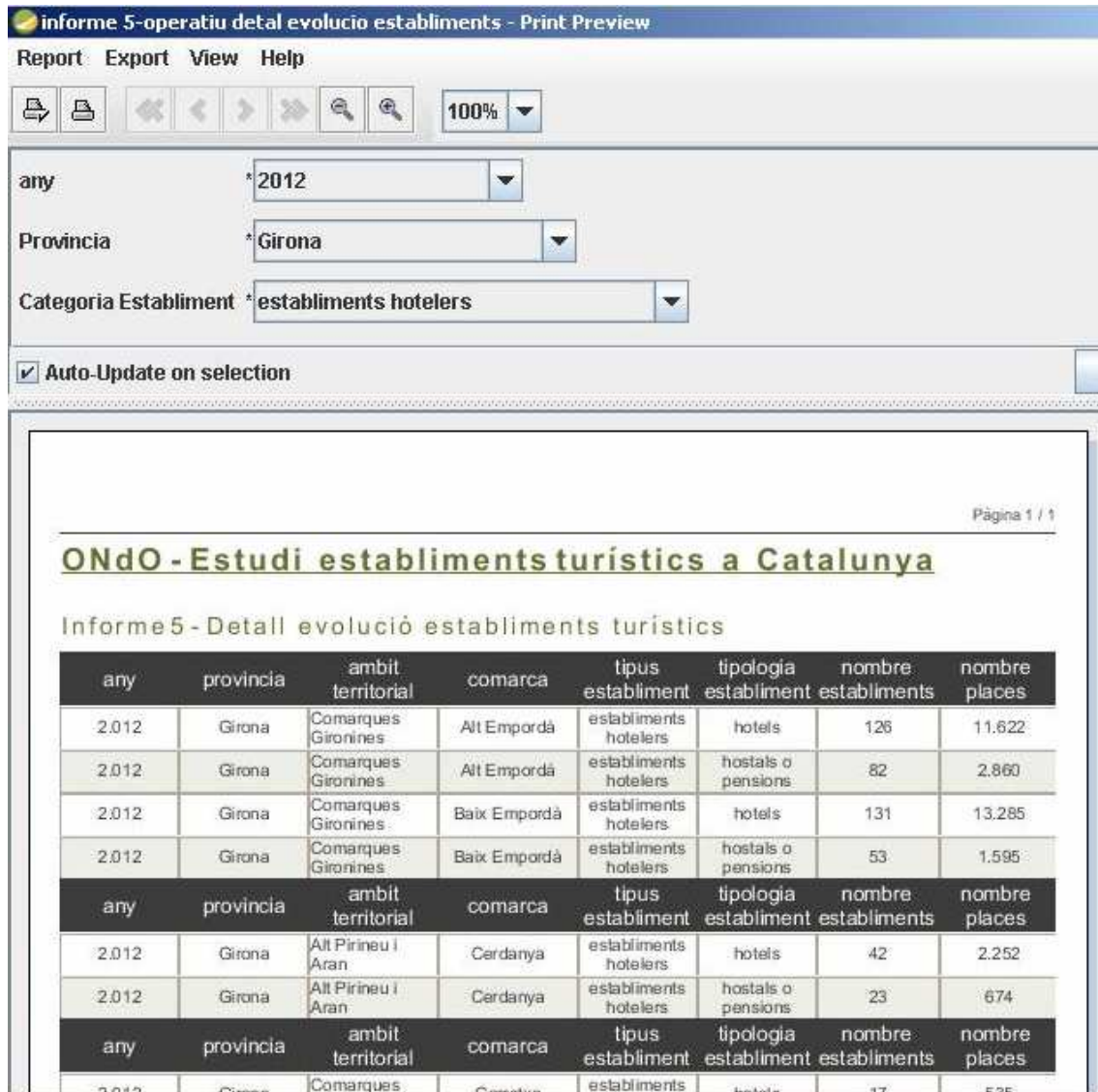
3.5.5. INFORME 5 . Detall evolució establiments.

Informe operatiu amb el detall dels establiments i places ofertades a Catalunya.

Paràmetres : any, província i categoria d'establiment

Elements :

- Detall del nombre d'establiments i places ofertades per als paràmetres introduïts, agrupats per àmbit territorial .



Informe 5-operatiu detal evolucio establiments - Print Preview

Report: Export View Help

any: 2012
 Província: Girona
 Categoria Establiment: establiments hotelers

Auto-Update on selection

Pàgina 1 / 1

ONdO - Estudi establiments turístics a Catalunya

Informe5 - Detall evolució establiments turístics

any	província	àmbit territorial	comarca	tipus establiment	tipologia establiment	nombre establiments	nombre places
2012	Girona	Comarques Gironines	Alt Empordà	establiments hotelers	hotels	126	11.622
2012	Girona	Comarques Gironines	Alt Empordà	establiments hotelers	hostals o pensions	82	2.860
2012	Girona	Comarques Gironines	Baix Empordà	establiments hotelers	hotels	131	13.285
2012	Girona	Comarques Gironines	Baix Empordà	establiments hotelers	hostals o pensions	53	1.595
any	província	àmbit territorial	comarca	tipus establiment	tipologia establiment	nombre establiments	nombre places
2012	Girona	Alt Pirineu i Aran	Cerdanya	establiments hotelers	hotels	42	2.252
2012	Girona	Alt Pirineu i Aran	Cerdanya	establiments hotelers	hostals o pensions	23	674
any	província	àmbit territorial	comarca	tipus establiment	tipologia establiment	nombre establiments	nombre places
2012	Girona	Comarques	Garrotxa	establiments	hotels	17	5.35

figura 26: Disseny Informe 5

3.5.6. INFORME 6 .Detall evolució establiments agrupat.

Informe operatiu amb el detall del nombre d'establiments i places ofertades a Catalunya agrupades per Província, Àmbit territorial i Comarca.

Paràmetres : any.

Elements :

- Detall del nombre d'establiments i places ofertades per al paràmetre introduïts, agrupats per província, àmbit territorial i comarca .

Pàgina 1 / 7

ONdO - Estudi establiments turístics a Catalunya

Informe 6 - Detall evolució establiments turístics, agrupat

Barcelona

Comarques Centrals

Anoia

Nombre establiments	Nombre de places ofertades	Tipus establiment
1	90	campings
14	704	establiments hotelers
61	387	turisme rural

Bages

Nombre establiments	Nombre de places ofertades	Tipus establiment
3	504	campings
35	1.369	establiments hotelers
84	614	turisme rural

Berguedà

Nombre establiments	Nombre de places ofertades	Tipus establiment
19	7.491	campings
53	1.353	establiments hotelers
130	1.175	turisme rural

Osona

figura 27: Disseny Informe 6

3.5.7. INFORME 7 . Detall Equipaments per província.

Informe operatiu amb el detall del tipus d'equipaments per província a Catalunya.

Paràmetres : cap

Elements :

- Gràfic d'àrees amb el volum d'equipaments per província segons el nivell.
- Detall del nombre d'equipaments de segon nivell per província

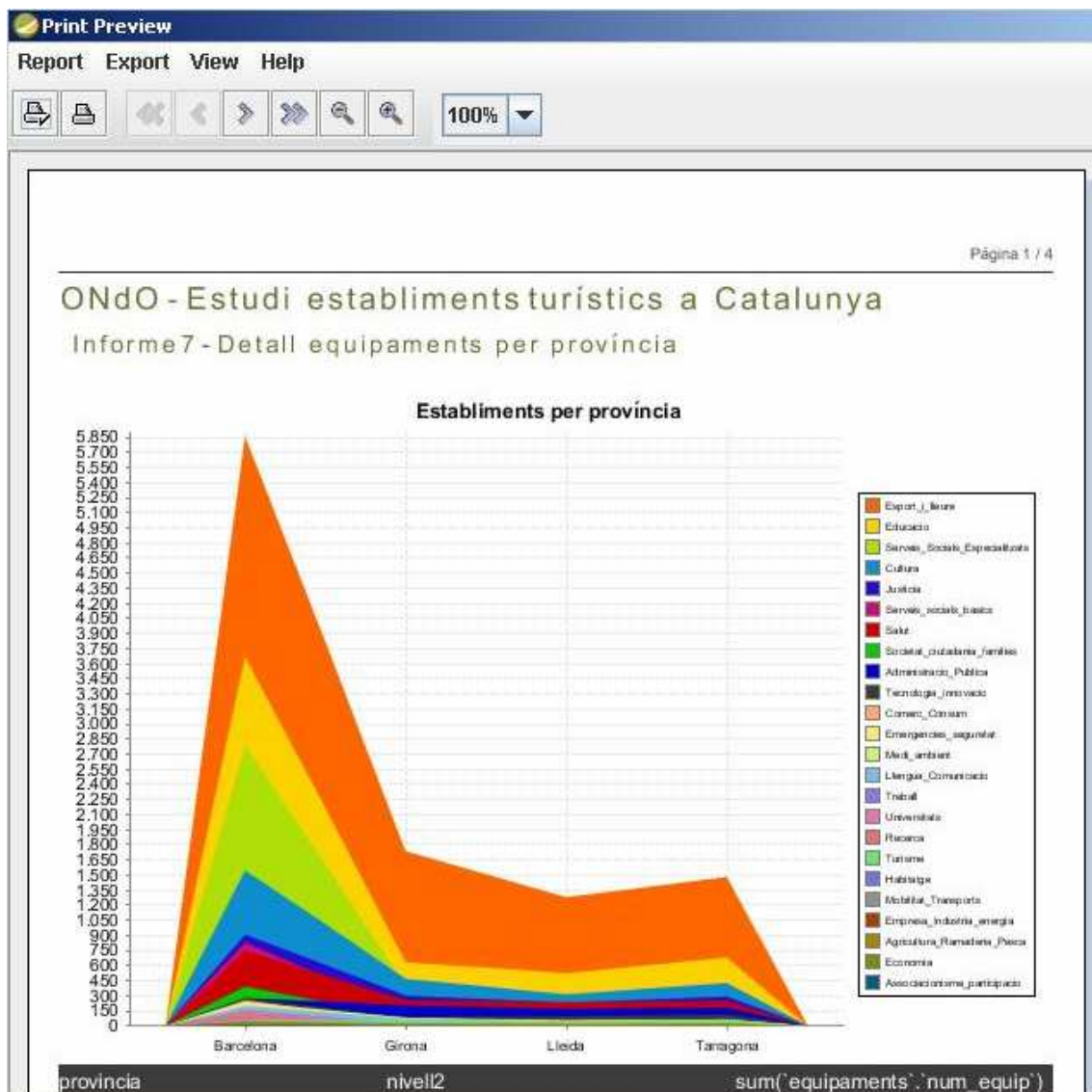


figura 28: Disseny Informe 7

4.- CAPTURES DE PANTALLA

En aquest capítol s'intenta mostrar per una banda, com s'han accedit al programari amb el que s'ha desenvolupat el projecte, amb els usuaris i claus d'accés, i per una altra com s'ha dut a terme el procés ETL.

4.1.- ACCÉS AL PROGRAMARI DE DESENVOLUPAMENT DEL PROJECTE

A continuació es detallen els usuaris i claus d'accés utilitzats en la implementació del sistema :

MySQL workbench

Usuari : root

Clau : tfcmmd

Esquema del projecte : TFC

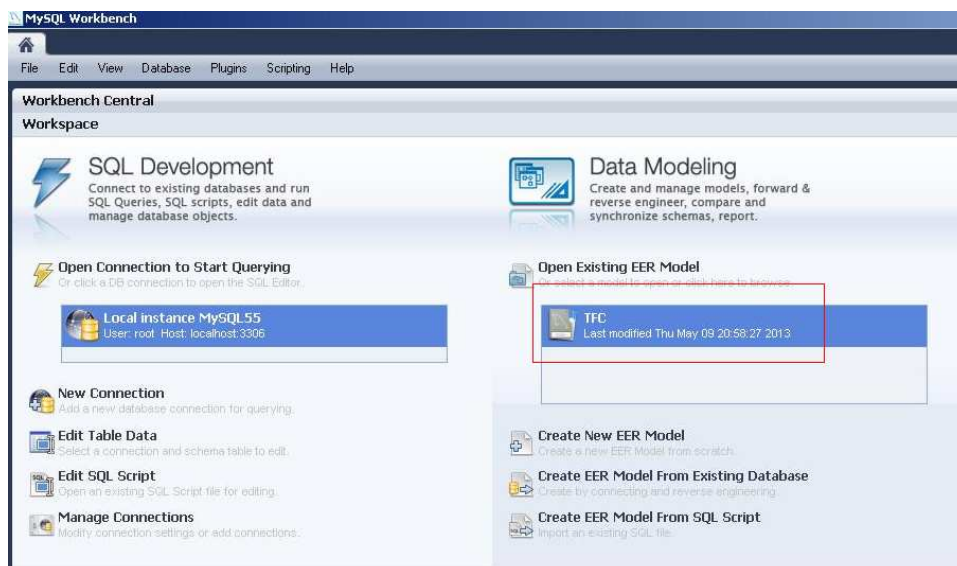


figura 29: Esquema BD en MySQL workbench

Pentaho Data Integration/PDI/Spoon

Repositori : ETL

Usuari : admin

Clau : admin

Les transformacions i jobs estan en el catàleg propi de la eina que es detallen en el proper capítol.

Pentaho BI server user console



figura 30: Connexió Pentaho BI user console

Consola localhost:8099

Usuari : admin
Clau : password

Portal localhost:8080

Usuari :tfcmmmd
Clau :tfcmmmd

Accesos als informes via navegador en la carpeta TFC-establiments turistics

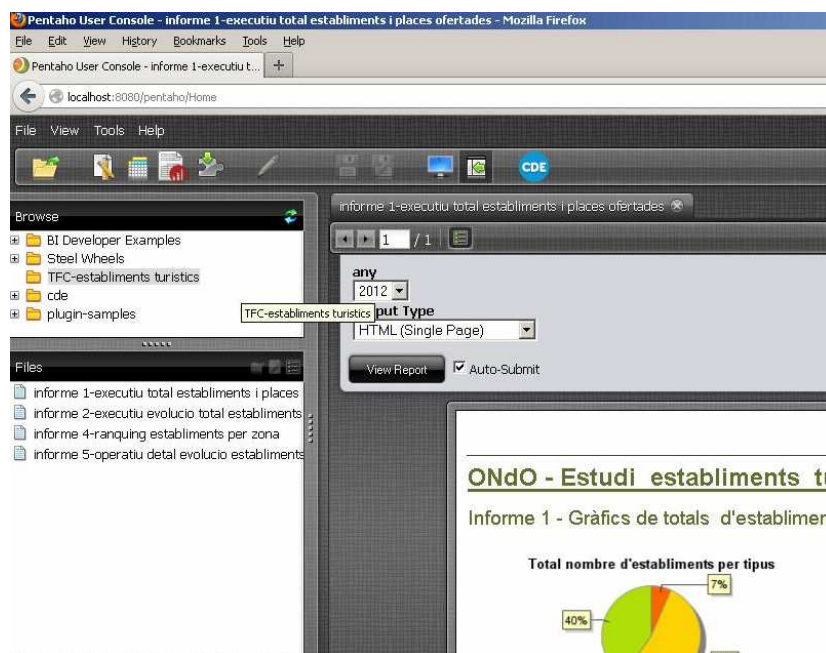


figura 31: Informes en Pentaho BI user console

Connexió a base de dades

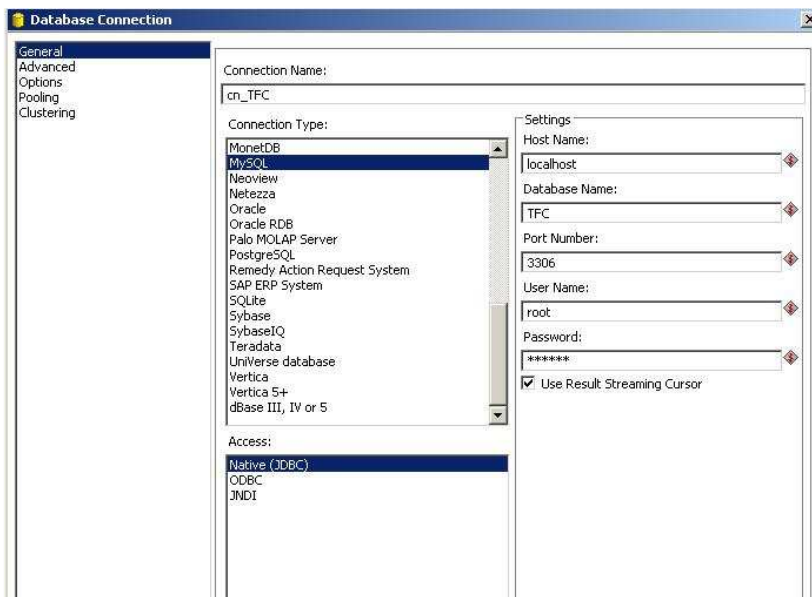


figura 32: Connexió BD

Port : 3306
Usuari : root
Clau : tfcmmd

Pentaho Report designer

Clau per publicar els informes : tfccsa

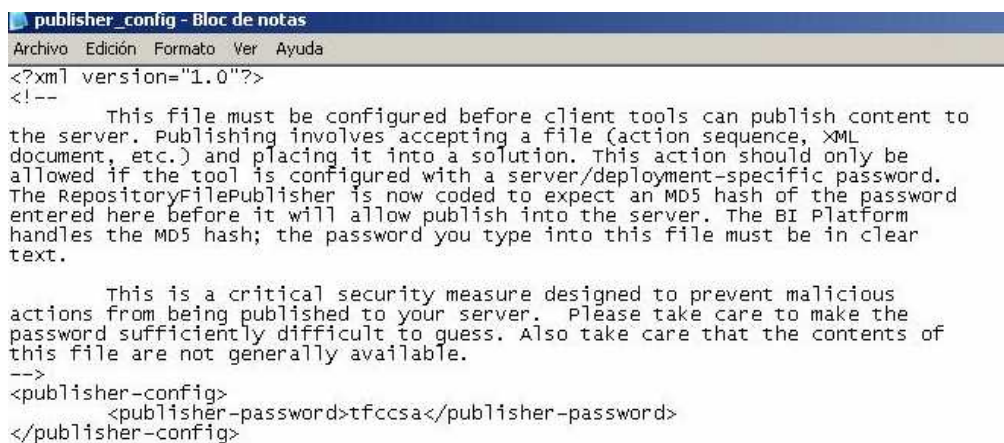


figura 33: Clau per publicar en PRD

4.2.- PROCÉS ETL

El procés ETL és l'encarregat de Extreure i Transformar les dades d'entrada al sistema, y carregar-les (Load) a la base de dades.

El procés s'efectua en 4 passes :

1- En la fase d'anàlisi alguns dels arxius d'entrada s'han importat a excel i/o han sofert alguns canvis com eliminació de files innecessàries. S'aprofiten doncs aquests arxius com a entrada al procés ETL.

2- Els arxius de entrada a transformar en troben a la carpeta C:\ETL\INPUTS. A partir d'aquests arxius, es creen els processos de transformació i s'extreuen els arxius de sortida, que es guarden en la carpeta C:\ETL\OUTPUTS en format txt separats per punt i coma , i que ja estan formatejats per poder-los carregar directament a la base de dades.

3-A partir dels arxius transformats es creen i s'executen els passos corresponents de LOAD a les taules.

4-Finalment es creen els jobs que executen les transformacions.

- El procés ETL d'aquest projecte està dissenyat per fer una càrrega inicial de les dades (tal i com es va dir en el document d'anàlisi) i no contempla processos incrementals de dades, pel que el jobs no es programen amb tal efecte.

Els processos i jobs es guarden en el catàleg de la eina PDI :

Selecció del catàleg

Selecció del objecte

Nombre	Tipo	Usuario	Fecha cambio
job_load_taulas_fets	job	admin	2013/05/09 19:33:14
Job_ETL_dimenions	job	admin	2013/05/09 19:18:51
job_ETL_fets	job	admin	2013/05/09 19:20:58
job_load_taulas	job	admin	2013/05/09 19:29:43
job_load_taulas_dimensions	job	admin	2013/05/09 19:53:38
load_CTG_EQUIPAMENTS	transformation	admin	2013/05/08 20:09:29
load_CTG_ESTABLIMENTS	transformation	admin	2013/05/05 19:07:54
load_EQUIPAMENTS	transformation	admin	2013/05/08 19:19:24
load_EVOL_ESTABL	transformation	admin	2013/05/09 18:59:24
load_MUNICIPI	transformation	admin	2013/05/08 20:09:30
load_POBLACIO	transformation	admin	2013/05/08 20:09:31
load_TEMPS	transformation	admin	2013/05/08 20:09:25
trf_categoriaEquipaments	transformation	admin	2013/04/30 18:52:08
trf_categoriaEstabliments	transformation	admin	2013/05/09 19:11:41
trf_comarca	transformation	admin	2013/04/29 21:14:13
trf_equipaments	transformation	admin	2013/05/08 19:25:09
trf_evolutioEstabliments	transformation	admin	2013/05/09 18:57:16
trf_municipis	transformation	admin	2013/05/08 19:02:59
trf_poblacio	transformation	admin	2013/05/09 19:53:15
trf_temps	transformation	admin	2013/04/29 21:16:28

figura 34: Catàleg transformacions i jobs PDI

els jobs_XXX son els procediments que executen el processos ETL, un darrera l'altre.
els load_XXX son els processos de LOAD a les taules
els trf_XXX son els processos d'extracció i transformació de les dades.

A continuació es detalla per ordre els procés seguit :

4.2.1. JOB TRANSFORMACIONS DIMENSIONS

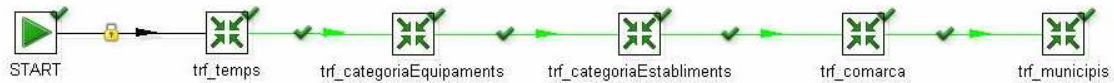


figura35 : job ETL_dimensions

Aquest job executa els processos de transformació associats a les taules de dimensions.

Els passos estan enllaçats de manera que si un falla, els següents continuen.

Els arxius de sortida es guarden la taula C:\TFC\OUTPUTS

A continuació es detalla el procés de cada transformació.

Transformació 1 : trf_temps

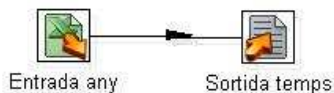


figura36 : transformació TEMPS

Extracció : Arxiu Z-temps.txt crear manualment amb els anys que van del 2006 al 2012

Transformació :

1. sortida en format txt separat per punt i coma

Transformació 2 : trf_ctg equipaments

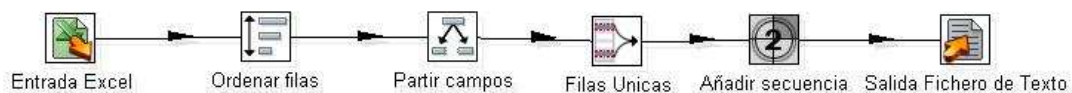


figura 37: transformació Categoría Equipaments

Extracció : Arxiu Equipaments_unicode.xls importat el arxiu 'Equipaments.csv' proporcional pel client.

Transformació :

2. Ordenar files per columna 'categories'
3. Partir la columna 'categories' pel caràcter '|'

4. Deixar només 4 nivells de categories i resumir files amb les mateixes categories deixant-ne una.
5. Afegir clau de seqüència
6. sortida en format txt separat per punt i coma

Transformació 3 : trf ctg establiments



figura 38: transformació Categoria Establiment

Extracció : Arxiu Z-CategoriesEstabliments.xls creat manualment, ja que segons l'indicat en el apartat de 'decisiones preses' de la PAC2, s'unifiquen les categories que varien segons l'any.

Transformació :

7. Afegir clau de seqüència
8. Sortida en format txt separat per punt i coma

Transformació 4 : trf comarca



figura 39: transformació comarca

Extracció : Arxiu aec-924.xls extret del la web d'IDESCAT

Transformació :

9. Afegir clau de seqüència
10. Afegir comunitat autònoma com a constant
11. Sortida en format txt separat per punt i coma (arxiu temporal)

Transformació 5 : trf_municipi



figura 40: transformació municipi

Extracció : Arxiu aec-925.xls extret del la web d'IDESCAT

Transformació :

12. Afegir clau de seqüència
13. Afegir comarca, àmbit territorial i comunicat amb la informació del arxiu temporal 'comarca'.
14. Substituir comes per blanc en el nom del municipi
15. Sortida en format txt separat per punt i coma

4.2.2. JOB TRANSFORMACIONS FETS

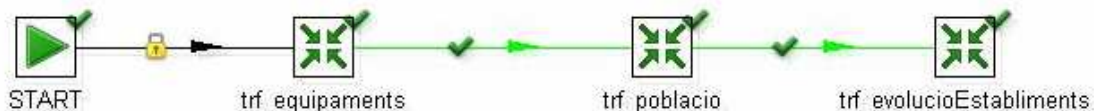


figura 41: job ETL_fets

Aquest job executa els processos de transformació associats a les taules de fets.

Els passos estan enllaçats de manera que si un falla, els següents continuen.

Els arxius de sortida es guarden la taula C:\TFC\OUTPUTS

A continuació es detalla el procés de cada transformació.

Transformació 6 : trf equipaments



figura 42: transformació equipaments

Extracció : Arxiu Equipaments.csv subministrat pel client

Transformació :

16. Ordenar per municipi i categoria d'equipaments.
17. Partir la columna 'categories' pel caràcter '|'
18. Deixar només 4 nivells de categories i resumir files amb les mateixes categories deixant-ne una.
19. Buscar el id del municipi en l'arxiu de municipis creat anteriorment.
20. Buscar el id de categoria d'equipament en l'arxiu ctg_equipaments creat anteriorment.
21. Sortida en format txt separat per punt i coma

Transformació 7 : trf poblacio

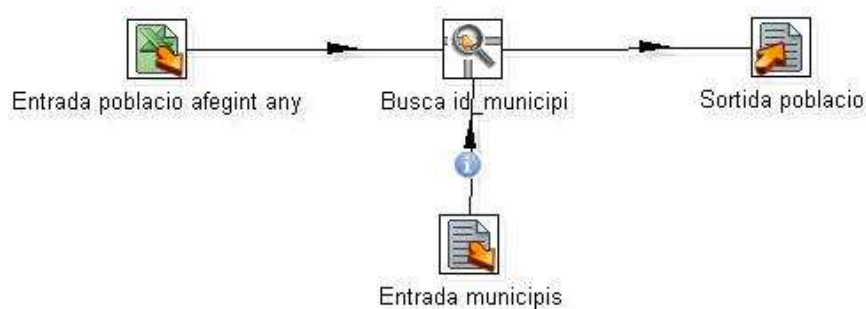


figura 43: transformació població

Extracció : Arxiu PoblacioPerSexe.xls extret de IDESCAT

Transformació :

22. Afegir nom de pestanya identificador d'any, com a columna
23. Buscar el id del municipi en l'arxiu de municipis creat anteriorment.
24. Sortida en format txt separat per punt i coma

Transformació 8 : trf evolucioEstabliments

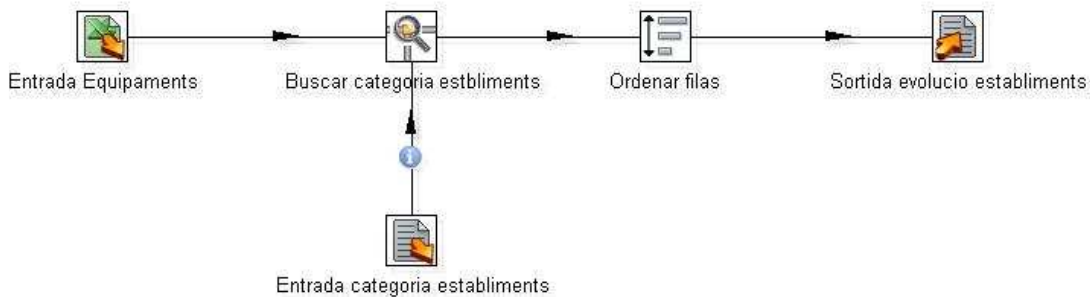


figura 44: transformació evolució establiments

Extracció : Arxiu unificat establiments AAAA proporcionat pel client.

Transformació :

25. Afegir nom de pestanya identificador d'any, com a columna
26. Buscar el id de la categoria d'establiment en l'arxiu de ctg_establiments creat anteriorment.
27. ordenar files per any, categoria i comarca
28. Sortida en format txt separat per punt i coma

4.2.3. JOB CARREGA TAULES DIMENSIONS

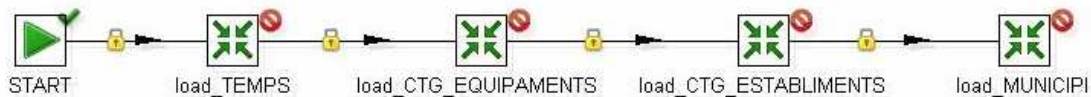


figura 45: job LOAD_dimensions

Aquest job executa els processos de càrrega de les taules de dimensions a partir dels arxius txt transformats en els processos anteriors. Tal i com indica la imatge, i degut a que és un procés de càrrega inicial (només es pot executar un cop), el DROP de les taules de dimensions i posterior recàrrega no es poden fer si no abans de buidar les corresponents taules de fets, amb la que hi ha integritat referencial.

Els passos estan enllaçats de manera que si un falla, els següents continuen.

A continuació es detalla el procés de cada transformació.

Transformació 9 : load temps



figura 46: transformació LOAD TEMPS

Extracció : Arxiu temps.txt de C:\TFC\OUTPUTS

Transformació :

29. Inserció en la taula TEMPS

Transformació 10 : load ctg equipaments

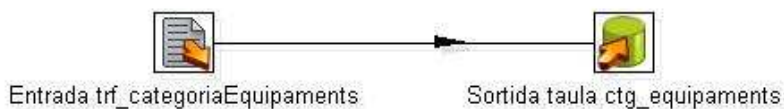


figura 47: transformació LOAD categoria Equipaments

Extracció : Arxiu categoriesEquipaments.txt de C:\TFC\OUTPUTS

Transformació :

30. Inserció en la taula CTG_EQUIPAMENTS

Transformació 11 : load ctg establiments

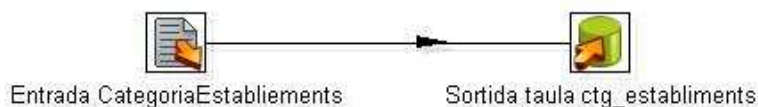


figura 48: transformació LOAD categoria Establiments

Extracció : Arxiu categoriaEstabliments.txt de C:\TFC\OUTPUTS

Transformació :

31. Inserció en la taula CTG_ESTABLIMENTS

Transformació 12 : load municipi



figura 49: transformació LOAD municipi

Extracció : Arxiu municipi.txt de C:\TFC\OUTPUTS

Transformació :

- 32. Inserció en la taula MUNICIPI

4.2.4. JOB CARREGA TAULES FETS



figura 50: job LOAD fets

Aquest job executa els processos de càrrega de les taules de fets a partir dels arxius txt transformats en els processos anteriors. A diferència de la càrrega de les taules de dimensions, aquestes es poden executar tantes vegades com es vulgui.

Els passos estan enllaçats de manera que si un falla, els següents continuen.

A continuació es detalla el procés de cada transformació.

Transformació 13 : load poblacio



figura 51: transformació LOAD població

Extracció : Arxiu poblacio.txt de C:\TFC\OUTPUTS

Transformació :

- 33. Busca identificador de relació de id_any en la taula TEMPS.
- 34. Busca identificador de relació de id_municipi en la taula MUNICIPI
- 35. Inserció en la taula POBLACIO

Transformació 14 : load equipaments



figura 52: transformació LOAD equipaments

Extracció : Arxiu equipaments.txt de C:\TFC\OUTPUTS

Transformació :

36. Busca identificador de relació de id_municipi en la taula MUNICIPI.
37. Busca identificador de relació de id_ctg equip en la taula ID_CTG_EQUIPAMENTS
38. Inserció en la taula EQUIPAMENTS

Transformació 15 : load evolució establiments



figura 53: transformació LOAD evolució establiments

Extracció :

Arxiu evolucionEstabliments.txt de C:\TFC\OUTPUTS

Transformació :

39. Busca identificador de relació de id_any en la taula TEMPS.
40. Busca identificador de relació de comarca en la taula MUNICIPI.
41. Busca identificador de relació de id_ctg_establ en la taula ID_CTG_ESTABLIMENTS
42. Inserció en la taula EVOLUCIO_ESTABLIMENTS

5.- CONCLUSIONS

Al inici del projecte partia del coneixement de la importància que te avui en dia el control de la informació per les empreses i que cada cop més aquestes opten per organitzar-se i explotar-se les dades mitjançant Data Warehouses, amb el que tenia principal interès en saber com era l'arquitectura d'aquest tipus de sistemes, i com s'implementen.

Amb el cas pràctic treballat he fet palès que per a que les dades es converteixen en informació útil per a la presa de decisions, han de convertir-se en informació, i per això és necessari aplicar processos de neteja i homogeneïtzació i organitzar-los de manera que s'optimitzi el procés d'anàlisi i sigui útil per a la presa de decisions.

En la etapa d'anàlisi i disseny he pogut comprovar la importància d'un bon anàlisi i model de dades i la conseqüència que te aquesta en el posterior desenvolupament. Un mal disseny pot provocar un desenvolupament complex i difícil de mantenir.

Un punt important en aquest fase, dins dels meus objectius, ha estat saber que eren els índex de rendiment (KPI), terme que he sentit anomenar amb força freqüència, i ser capaç de trobar varis indicadors de negoci adequats a la problemàtica d'aquest treball.

L'enfocament dels KPI han estat, en el meu cas, la base dels informes de control executiu, encarat a un perfil d'usuari gerencial que només vol tenir amb d'un cop de vista, com van evolucionant els indicadors, i poder decidir d'actuar com cregui necessari.

El estudi de les eines pentaho ha estat un punt feble que ha provocat no tenir el temps suficient per crear els informes amb el detall que es requeria. La creació del procés ETL ha estat costós, i el de creació del informes més lent del que es preveia.

Finalment, i degut als poc dots de comunicació que tinc (reconeguda i contrastada en varis cursos que he fet), la presentació virtual ha estat també un esforç considerable dins del projecte. Ha estat una bona experiència el fet de tenir que gravar el vídeo fent una síntesi de tot el treball realitzat i poder-lo reproduir per fer-hi una auto-crítica que valoro molt positivament. Espero doncs que sigui entenable i amena.

Per concluir, analitzant els resultats obtinguts durant les diferents fases del projecte i comparant-los amb els objectius propis marcats a l'inici, puc dir que s'han assolit en tot el seu conjunt i que la experiència em serà molt vàlida per portar en pràctica en la vida laboral.

6.- LINIES D'EVOLUCIÓ FUTURES

Arribat al final del projecte, el client hauria de donar la seva acceptació a fi de poder donar el projecte per tancat e iniciar la fase de manteniment i garantia, però a falta de no tenir aquesta possibilitat, i analitzant el estat actual amb el que es deixa, es considera que en el futur s'hauria d'actuar en els següents punts :

- De cara a la fase de manteniment del sistema s'hauria de crear documentació d'usuari i de manteniment a fi de que futurs equips poguessin prendre el relleu d'administració i gestió del DW.
- El procés ETL de càrrega de la base de dades està enfocada a una càrrega inicial de les dades. S'hauria d'adaptar per poder permetre el manteniment d'aquesta segons es vagin subministrant informació d'anys futurs.
- Referent també al procés ETL caldria revisar el procés d'actualització de les taules dimensions a partir de la de fets., i també la creació d'arxius .bat que permetin una millor automatització del projecte.
- Resoldre els problemes tècnics que han provocat que algunes joins entre taules hagin estat massa complexes, i permetre mostrar més indicadors en format rati.
- Permetre i formar al usuari final crear-se els seus propis informes o taules de control mitjançant les eines Ad-Hoc, SAIKU i CDE subministrades en la consola de Pentaho BI, estructurant les dades del DW en un cub OLAP.
- Crear tots els rols i perfils d'usuari definits, a fi de controlar l'accés al sistema, tant a nivell d'administració com a nivell d'usuari final.

GLOSSARI

Business Intelligence (o intel·ligència empresarial) : fa referència a l'ús de dades d'una empresa per a facilitar la presa de decisions

Data Warehouse (o magatzem de dades) : és una base de dades amb la informació històrica d'una organització dissenyada i estructurada per a realitzar-hi consultes eficientment.,

DSS (Decision support system): Un DSS és un sistema informàtic utilitzat per servir de recolzament al procés de presa de decisions.

EIS (Executive information system) : Un Sistema de Informació Executiva és una eina de Intel·ligència empresarial (Business Intelligence, BI), orientada a usuaris de nivell gerencial, que permet monitoritzar l'estat de las variables d'una àrea o unitat de la empresa a partir de informació interna i externa.

ETL (Extract, Transform and Load) : Procés que permet a les organitzacions a moure dades desde múltiples fonts, formatejar-los ,netejar-los i carregar-los en una base de dades, datamart o data warehouse, per analitzar en un procés de negoci.

KPIs (Key performance indicator) Indicador vinculat a un objectiu.

Legacy : Un sistema heretat (o sistema legacy) és un sistema informàtic (que ha quedat antiquat però continua sent utilitzat per l'usuari i que no es vol o no es pot reemplaçar o actualitzar de forma senzilla.

SDLC (systems development life cycle): Metodologia de desenvolupament de projectes informàtics caracteritzats en que cada fase és continuació de l'anterior i es basa en aquesta.

OLAP (On-Line Analytical Processing). : OLAP es el acrònim en anglès de processament analític. És una solució utilitzada en el camp de la anomenada Intel·ligència empresarial (o Business Intelligence) el objectiu del qual és agilitzar la consulta de grans quantitats de dades.

OLTP (OnLine Transaction Processing): OLTP es la sigla en anglès de Processament de Transaccions en Línea. És un tipus de processament que facilita i administra aplicacions transaccionals, usualment per entrada de dades i recuperació i processament de transaccions (gestor transaccional).

ROLAP (Relational Online Analytical Processing): ROLAP significa Processament Analític OnLine Relacional, es a dir, es tracta de sistemes i eines OLAP construïts sobre una base de dades relacional.

BIBLIOGRAFIA

6.1.- PUBLICACIONS / LECTURES DE MATERIALS

Alfons Bataller Díaz. "Gestió i desenvolupament de projectes. Conceptes i suggeriments." . P08/19018/00444. Universitat Oberta de Catalunya.

Nita Sáenz Higuera. Rut Vidal Oltra. "Redacció de textos científicotècnics". P08/19018/00445. Universitat Oberta de Catalunya.

Roser Beneito Montagut. "Presentació i Elaboració de documents". P08/19018/00446 .Universitat Oberta de Catalunya.

Alberto Abelló Gamazo. "Disseny multidimensional". P02/11048/00011. Universitat Oberta de Catalunya.

Josep Lluís Cano . "**Business Intelligence : Competir con Información**". [en línia]. http://itemsweb.esade.edu/biblioteca/archivo/Business_Intelligence_competir_con_info_rmacion.pdf

6.2.- REFERENCIES PROCEDENTS D'INTERNET

Martínez Orol, Alfredo (03/2007). **OLAP y el diseño de cubos**. [en línia]. <http://www.gestiopolis.com/canales8/ger/olap-online-analytic-processing.htm>

Pau Urquizu. **Business intelligence fácil**. [en línia]. <http://www.businessintelligence.info/>

BI-Latino. [en línia]. <http://www.bi-spain.com/>

Fernando Berzal. **El modelo multidimensional Data Warehousing**. [en línia]. <http://elvex.ugr.es/idbis/db/docs/intro/F%20Modelo%20multidimensional.pdf>

Oracle **The Multidimensional Data Model**. [en línia]. http://docs.oracle.com/cd/B12037_01/olap.101/b10333/multimodel.htm

Msdn.Microsoft. **Bases de datos de modelos multidimensionales**. [en línia]. <http://msdn.microsoft.com/es-es/library/ms174953.aspx>

Roberto Espinosa. **El rincón del BI** . [en línia]. <http://churriwifi.wordpress.com>

Pentaho. **Pentaho community**. [en línia]. <http://community.pentaho.com>

ANNEXOS

7.- QUALITAT DE LES DADES

A mida que s'anava desenvolupant el procés ETL es van anar detectant i retocant manualment la denominació d'algun municipi per motius de grafia (comes, apòstrofes) i en algun cas el propi nom (per exemple, en l'arxiu d'equipaments existia 'Cabrera d'Igualada' i al de municipis 'Cabrera d'anoia') que es van anar resolent.

Finalment s'ha comprovat la qualitat de les dades carregades a la base de dades, i posteriorment extretes en els informes, auditant la informació dels arxius transformats amb els arxius d'origen subministrats, per una banda, i informació rellevant obtinguda per Internet. Això és:

s'ha revisat mitjançant la informació recuperada d'Internet que :

- hi han 497 municipis en la taula MUNICIPI
- hi han 4 províncies 41 comarques i 8 àmbits territorials en la taula MUNICIPI
- que el total de superfície (sumatori dels parcials per municipi de tot Catalunya) es de aprox. 32.114 km² en la taula MUNICIPI
- que la població total (sumatori dels parcials per municipi) d'un any concret(2012, per exemple) és de 7.570.908 habitants, aprox. en la taula POBLACIO.

S'ha revisat mitjançant els arxius proporcionats pel client :

S'han fet consultes simples a la base de dades per comparar els totals obtinguts amb els totals dels arxius proporcionals, per exemple : dels arxiu establiments : tenim els totals per tipus d'establiment, àmbit geogràfic, província, i Catalunya, i que hem eliminat en el procés de transformació.

Solsonès	25	5	104	134	701	2.307	660	3.668
Tarragonès	128	25	10	163	40.215	32.232	73	72.520
Terra Alta	16	3	22	41	581	561	191	1.333
Urgell	10	0	35	45	412	0	287	699
Val d'Aran	109	9	16	134	6.371	3.069	148	9.588
Vallès Occidental	49	0	3	52	5.792	0	28	5.820
Vallès Oriental	59	8	48	115	3.734	2.805	398	6.937
Catalunya	2.837	353	2.156	5.346	294.777	277.069	16.968	588.814
Metropolità	940	41	68	1.049	121.802	26.418	543	148.763
Comarques Gironines	818	139	659	1.616	79.100	133.485	5.426	218.011
Camp de Tarragona	251	50	180	481	50.639	61.416	1.471	113.526
Terres de l'Ebre	97	16	107	220	5.374	7.869	857	14.100
Ponent	91	4	157	252	4.744	1.548	1.196	7.488
Comarques Centrals	157	36	456	649	5.321	12.667	3.685	21.673
Alt Pirineu i Aran	346	56	379	781	16.460	20.676	2.679	39.815
Penedès	137	11	150	298	11.337	12.990	1.111	25.438
Barcelona	1.163	76	548	1.787	133.019	44.446	4.489	181.954
Girona	864	145	698	1.707	81.281	136.503	5.725	223.509
Lleida	423	62	607	1.092	19.958	22.461	4.305	46.724
Tarragona	387	70	303	760	60.519	73.659	2.449	136.627

Els totals de les consultes agregades a la base de dades han coincidit.

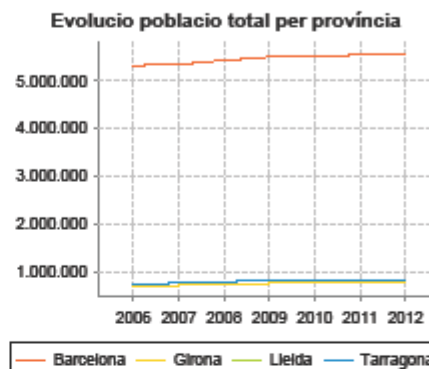
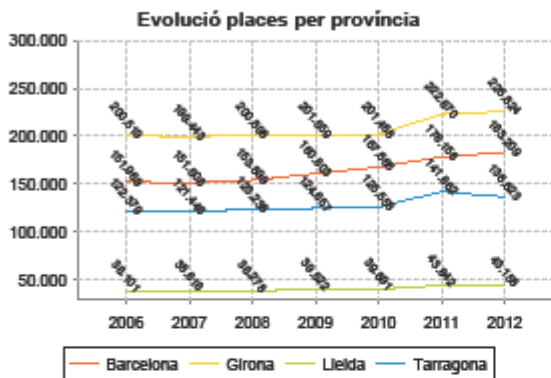
S'ha procedit de la mateixa manera per verificar les dades de població i equipaments, contabilitzant els totals prèvia extracció en Excel.

8.- SORTIDA INFORMES EN PDF

8.1.- INFORME1

ONdO - Estudi establiments turístics a Catalunya

Informe 1 - Establiments turístics vs habitants, equipaments i superfície

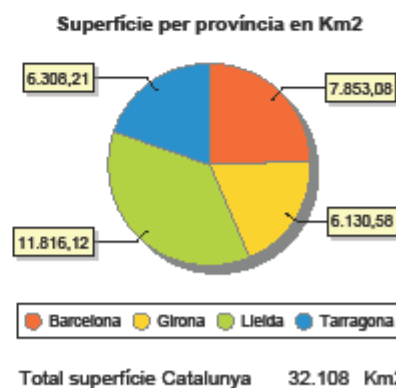
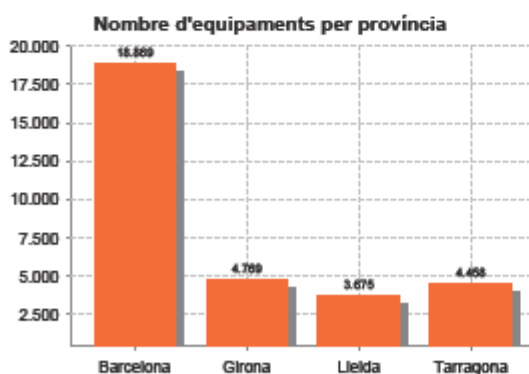


Oferta mitjana de places per província

Barcelona	163.887
Girona	207.512
Lleida	40.201
Tarragona	127.966
Total Catalunya	539.566

Població mitjana per província

	Homes	Dones
Barcelona	2.679.991	2.768.582
Girona	374.060	384.582
Lleida	217.391	208.715
Tarragona	399.274	388.508
Total Catalunya	3.670.716	3.730.371

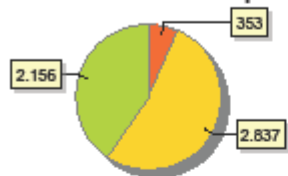


8.2.- INFORME 2

ONdO - Estudi establiments turístics a Catalunya

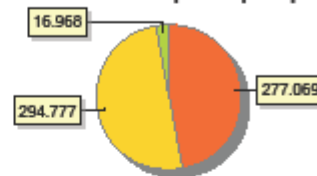
Informe 2 - Gràfics de totals d'establiments turístics i places ofertades

Total nombre d'establiments per tipus



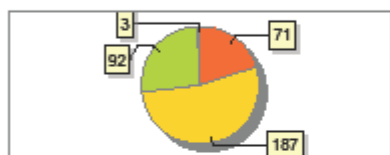
campings establiment hotelers turisme rural

Total nombre de places per tipus

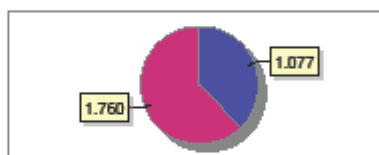


campings establiment hotelers turisme rural

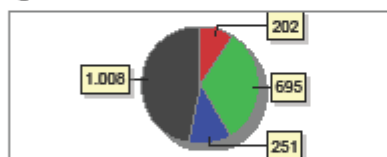
Total nombre d'establiments per tipus i tipologia



campings



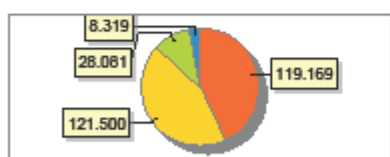
establiment hotelers



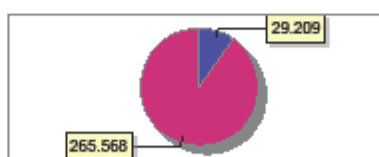
turisme rural

- 1a
- 2a
- 3a
- luxe
- hostals o pensions
- hotels
- casa de poble compartida
- casa de poble independent
- masia
- masoveria

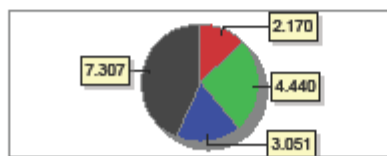
Total nombre d'establiments per tipus i tipologia



campings



establiment hotelers



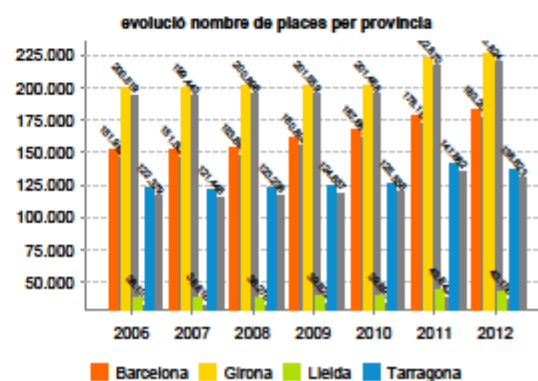
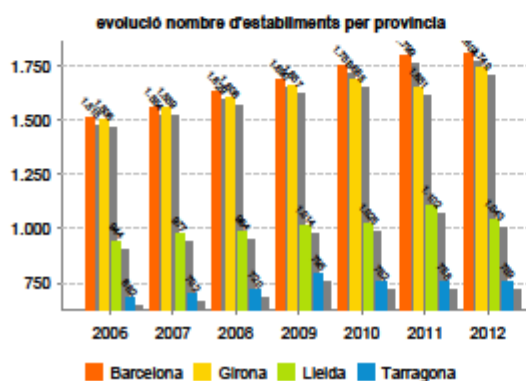
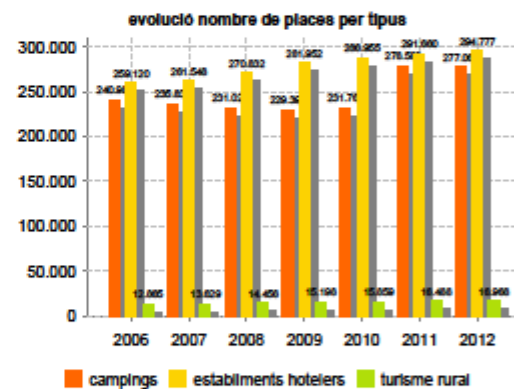
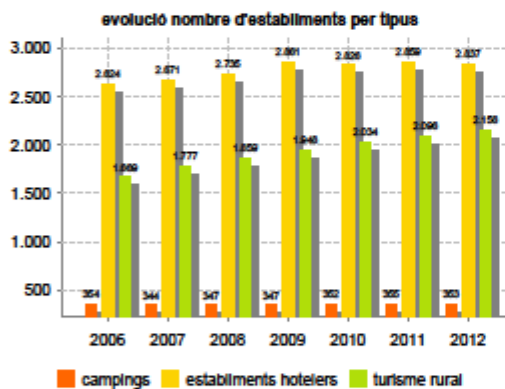
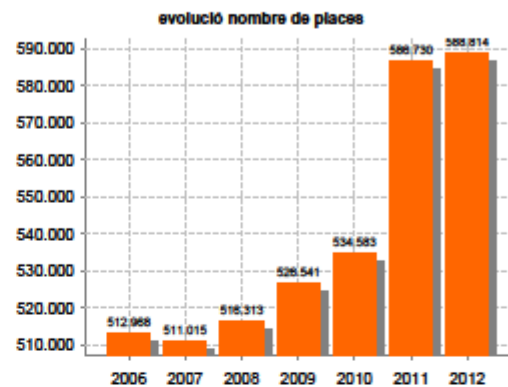
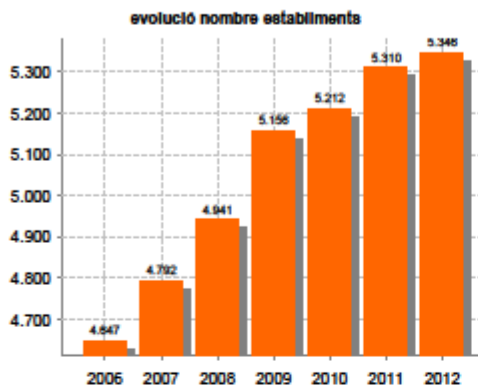
turisme rural

- 1a
- 2a
- 3a
- luxe
- hostals o pensions
- hotels
- casa de poble compartida
- casa de poble independent
- masia
- masoveria

8.3.- INFORME 3

ONdO - Estudi establiments turístics a Catalunya

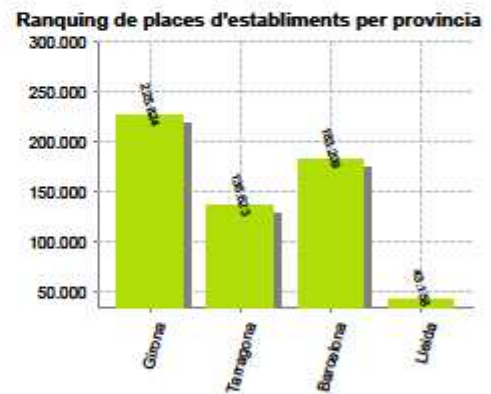
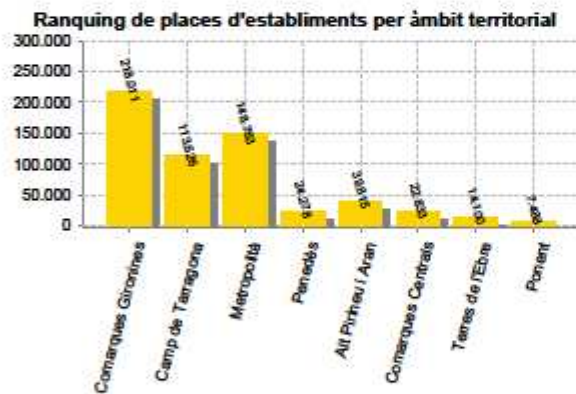
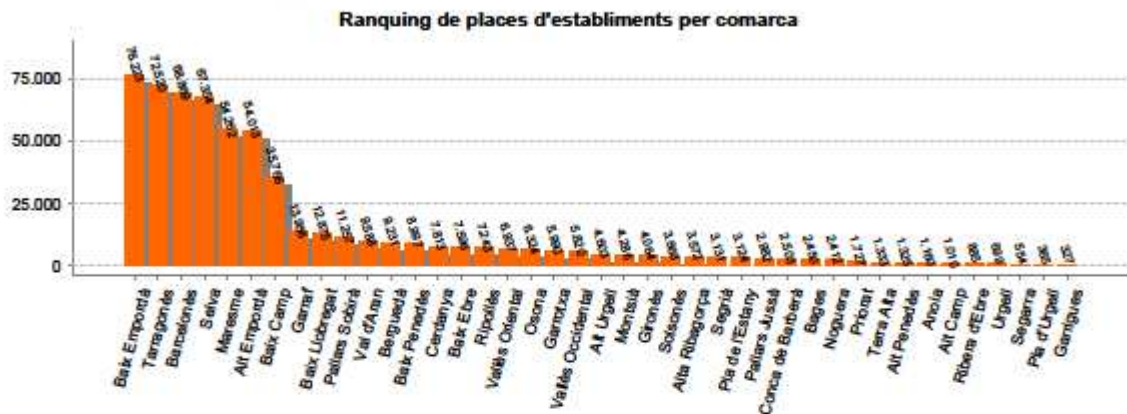
Informe 2 - Gràfics d'evolució d'establiments turístics



8.4.- INFORME 4

ONdO - Estudi establiments turístics a Catalunya

Informe 4 - Rànquing de places d'establiments per zona



any	província	àmbit territorial	comarca	nombre establiments	nombre places
2.012	Girona	Comarques Gironines	Baix Empordà	306	76.223
2.012	Tarragona	Camp de Tarragona	Tarragonès	183	72.520
2.012	Barcelona	Metropolità	Barcelonès	579	68.869
2.012	Girona	Comarques Gironines	Selva	336	67.354
2.012	Barcelona	Metropolità	Maresme	211	54.262
2.012	Girona	Comarques Gironines	Alt Empordà	391	54.013
2.012	Tarragona	Camp de Tarragona	Baix Camp	124	35.766
2.012	Barcelona	Penedès	Garraf	73	13.956

25 mayo 2013

8.5.- INFORME 5

ONdO - Estudi establiments turístics a Catalunya

Informe 5 - Detall evolució establiments turístics

any	provincia	ambit territorial	comarca	tipus establiment	tipologia establiment	nombre establiments	nombre places
2.012	Girona	Comarques Gironines	Alt Empordà	establiments hotelers	hotels	126	11.622
2.012	Girona	Comarques Gironines	Alt Empordà	establiments hotelers	hostals o pensions	82	2.860
2.012	Girona	Comarques Gironines	Baix Empordà	establiments hotelers	hotels	131	13.285
2.012	Girona	Comarques Gironines	Baix Empordà	establiments hotelers	hostals o pensions	53	1.595
any	provincia	ambit territorial	comarca	tipus establiment	tipologia establiment	nombre establiments	nombre places
2.012	Girona	Alt Pirineu i Aran	Cerdanya	establiments hotelers	hotels	42	2.252
2.012	Girona	Alt Pirineu i Aran	Cerdanya	establiments hotelers	hostals o pensions	23	874
any	provincia	ambit territorial	comarca	tipus establiment	tipologia establiment	nombre establiments	nombre places
2.012	Girona	Comarques Gironines	Garrotxa	establiments hotelers	hotels	17	535
2.012	Girona	Comarques Gironines	Garrotxa	establiments hotelers	hostals o pensions	25	497
2.012	Girona	Comarques Gironines	Gironès	establiments hotelers	hotels	28	2.435
2.012	Girona	Comarques Gironines	Gironès	establiments hotelers	hostals o pensions	22	511
2.012	Girona	Comarques Gironines	Pla de l'Estany	establiments hotelers	hotels	6	181
2.012	Girona	Comarques Gironines	Pla de l'Estany	establiments hotelers	hostals o pensions	7	165
2.012	Girona	Comarques Gironines	Ripollès	establiments hotelers	hotels	28	1.224
2.012	Girona	Comarques Gironines	Ripollès	establiments hotelers	hostals o pensions	37	884
2.012	Girona	Comarques Gironines	Selva	establiments hotelers	hotels	173	39.977
2.012	Girona	Comarques Gironines	Selva	establiments hotelers	hostals o pensions	85	3.329

Report Footer

8.6.- INFORME 6

Pàgina 1 / 7

ONdO - Estudi establiments turístics a Catalunya

Informe 6 - Detall evolució establiments turístics, agrupat

Barcelona

Comarques Centrals

Anoia

Nombre establiments	Nombre de places ofertades	Tipus establiment
1	90	campings
14	704	establiments hotelers
61	387	turisme rural

Bages

Nombre establiments	Nombre de places ofertades	Tipus establiment
3	504	campings
35	1.389	establiments hotelers
84	614	turisme rural

Berguedà

Nombre establiments	Nombre de places ofertades	Tipus establiment
19	7.491	campings
53	1.353	establiments hotelers
130	1.175	turisme rural

Osona

Nombre establiments	Nombre de places ofertades	Tipus establiment
9	3.141	campings
51	2.040	establiments hotelers
130	1.137	turisme rural

Metropolità

Baix Llobregat

Nombre establiments	Nombre de places ofertades	Tipus establiment
2	3.042	campings
88	9.613	establiments hotelers
1	4	turisme rural

Barcelonès

Nombre establiments	Nombre de places ofertades	Tipus establiment
568	66.687	establiments hotelers

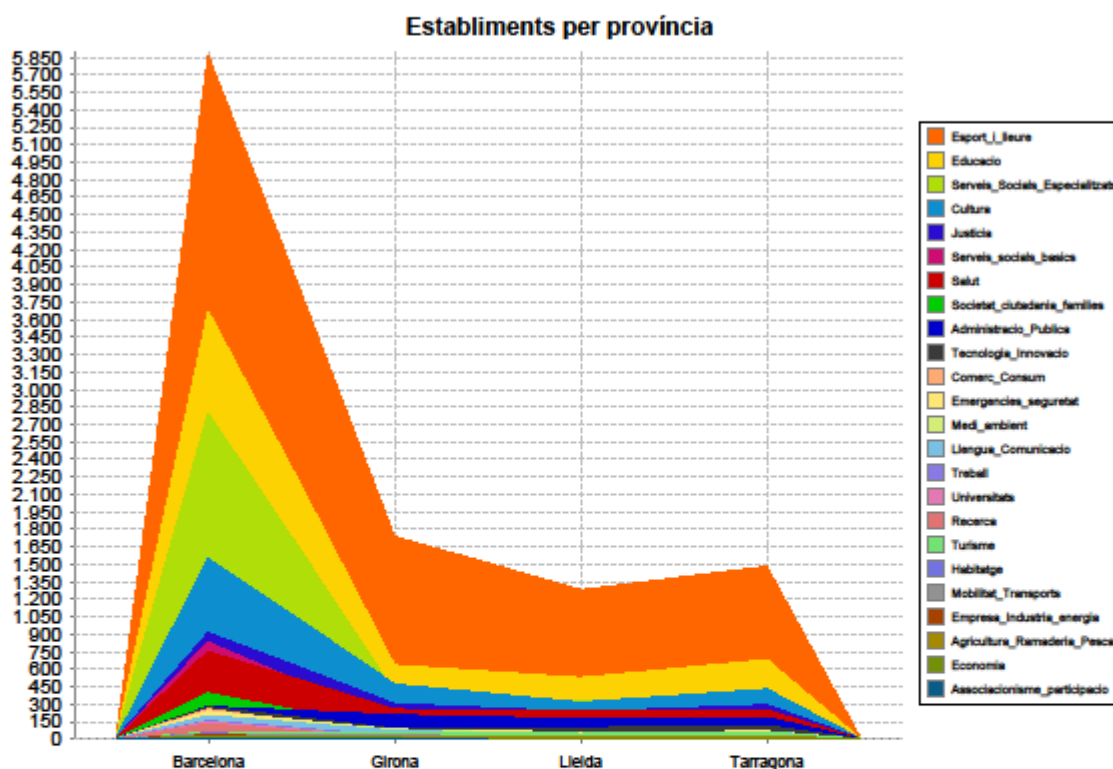
Maresme

26 mayo 2013

8.7.- INFORME 7

ONdO - Estudi establiments turístics a Catalunya

Informe 7 - Detall equipaments per província



província	nivell2	sum('equipaments'.num Equip')
Barcelona	Esport_i_lleure	5.881
Barcelona	Educacio	3.668
Barcelona	Serveis_Socials_Especialitzats	2.791
Barcelona	Cultura	1.545
Barcelona	Justicia	909
Barcelona	Serveis_socials_basics	829
Barcelona	Salut	748
Barcelona	Societat_ciudadania_families	391
Barcelona	Administracio_Publica	278
Barcelona	Tecnologia_Innovacio	275

25 mayo 2013

9.- SCRIPT DE CREACIÓ DE BD

```

SET @OLD_UNIQUE_CHECKS=@@UNIQUE_CHECKS, UNIQUE_CHECKS=0;
SET @OLD_FOREIGN_KEY_CHECKS=@@FOREIGN_KEY_CHECKS, FOREIGN_KEY_CHECKS=0;
SET @OLD_SQL_MODE=@@SQL_MODE, SQL_MODE='TRADITIONAL,ALLOW_INVALID_DATES';

CREATE SCHEMA IF NOT EXISTS `TFC` DEFAULT CHARACTER SET utf8 COLLATE utf8_spanish2_ci ;
USE `TFC` ;

-----
-- Table `TFC`.`TEMPS`
-----
DROP TABLE IF EXISTS `TFC`.`TEMPS` ;
CREATE TABLE IF NOT EXISTS `TFC`.`TEMPS` (
  `id_any` INT NOT NULL ,
  PRIMARY KEY (`id_any`))
ENGINE = InnoDB;

-----
-- Table `TFC`.`MUNICIPI`
-----
DROP TABLE IF EXISTS `TFC`.`MUNICIPI` ;
CREATE TABLE IF NOT EXISTS `TFC`.`MUNICIPI` (
  `id_municipi` INT NOT NULL AUTO_INCREMENT ,
  `municipi` VARCHAR(50) NOT NULL ,
  `comarca` VARCHAR(20) NOT NULL ,
  `ambit` VARCHAR(50) NOT NULL ,
  `provincia` VARCHAR(20) NOT NULL ,
  `comunitat` VARCHAR(20) NOT NULL ,
  `superficie` DECIMAL(5,2) NULL ,
  PRIMARY KEY (`id_municipi`, `comarca`))
ENGINE = InnoDB;

-----
-- Table `TFC`.`CTG_EQUIPAMENTS`
-----
DROP TABLE IF EXISTS `TFC`.`CTG_EQUIPAMENTS` ;
CREATE TABLE IF NOT EXISTS `TFC`.`CTG_EQUIPAMENTS` (
  `id_ctg equip` INT NOT NULL AUTO_INCREMENT ,
  `nivell1` VARCHAR(80) CHARACTER SET 'utf8' COLLATE 'utf8_spanish_ci' NOT NULL ,
  `nivell2` VARCHAR(80) NULL ,
  `nivell3` VARCHAR(80) NULL ,
  `nivell4` VARCHAR(80) NULL ,
  PRIMARY KEY (`id_ctg equip`))
ENGINE = InnoDB;

-----
-- Table `TFC`.`CTG_ESTABLIMENTS`
-----
DROP TABLE IF EXISTS `TFC`.`CTG_ESTABLIMENTS` ;
CREATE TABLE IF NOT EXISTS `TFC`.`CTG_ESTABLIMENTS` (
  `id_ctg_estbl` INT NOT NULL AUTO_INCREMENT ,
  `tipus_establ` VARCHAR(20) NOT NULL ,
  `tlg_establ` VARCHAR(20) CHARACTER SET 'utf8' COLLATE 'utf8_spanish_ci' NOT NULL ,
  PRIMARY KEY (`id_ctg_estbl`))
ENGINE = InnoDB;

-----
-- Table `TFC`.`POBLACIO`
-----
DROP TABLE IF EXISTS `TFC`.`POBLACIO` ;
CREATE TABLE IF NOT EXISTS `TFC`.`POBLACIO` (
  `id_any` INT NOT NULL ,
  `id_municipi` INT NOT NULL ,
  `num_homes` INT NULL ,
  `num_dones` INT NULL ,
  `TEMPS_id_any` INT NOT NULL ,
  `MUNICIPI_id_municip` INT NOT NULL ,
  INDEX `fk_POBLACIO_TEMPS_idx` (`TEMPS_id_any` ASC) ,
  INDEX `fk_POBLACIO_MUNICIPI1_idx` (`MUNICIPI_id_municip` ASC) ,
  CONSTRAINT `fk_POBLACIO_TEMPS`
    FOREIGN KEY (`TEMPS_id_any`)

```

```

REFERENCES `TFC`.`TEMPS` (`id_any` )
ON DELETE NO ACTION
ON UPDATE NO ACTION,
CONSTRAINT `fk_POBLACIO_MUNICIPIL`
FOREIGN KEY (`MUNICIPI_id_municipip` )
REFERENCES `TFC`.`MUNICIPI` (`id_municipi` )
ON DELETE NO ACTION
ON UPDATE NO ACTION)
ENGINE = InnoDB;

-----
-- Table `TFC`.`EQUIPAMENTS`
-----
DROP TABLE IF EXISTS `TFC`.`EQUIPAMENTS` ;
CREATE TABLE IF NOT EXISTS `TFC`.`EQUIPAMENTS` (
  `id_municipi` INT NOT NULL ,
  `id_ctg equip` INT NOT NULL ,
  `num equip` INT NULL ,
  `MUNICIPI_id_municipi` INT NOT NULL ,
  `CTG_EQUIPAMENTS_id_ctg equip` INT NOT NULL ,
  PRIMARY KEY (`id_municipi`, `id_ctg equip`),
  INDEX `fk_EQUIPAMENTS_MUNICIPIL_idx` (`MUNICIPI_id_municipi` ASC) ,
  INDEX `fk_EQUIPAMENTS_CTG_EQUIPAMENTS1_idx` (`CTG_EQUIPAMENTS_id_ctg equip` ASC) ,
  CONSTRAINT `fk_EQUIPAMENTS_MUNICIPIL`
  FOREIGN KEY (`MUNICIPI_id_municipi` )
  REFERENCES `TFC`.`MUNICIPI` (`id_municipi` )
  ON DELETE NO ACTION
  ON UPDATE NO ACTION,
  CONSTRAINT `fk_EQUIPAMENTS_CTG_EQUIPAMENTS1`
  FOREIGN KEY (`CTG_EQUIPAMENTS_id_ctg equip` )
  REFERENCES `TFC`.`CTG_EQUIPAMENTS` (`id_ctg equip` )
  ON DELETE NO ACTION
  ON UPDATE NO ACTION)
ENGINE = InnoDB;

-----
-- Table `TFC`.`EVOLUCIO_ESTABLIMENTS`
-----
DROP TABLE IF EXISTS `TFC`.`EVOLUCIO_ESTABLIMENTS` ;
CREATE TABLE IF NOT EXISTS `TFC`.`EVOLUCIO_ESTABLIMENTS` (
  `id_any` INT NOT NULL ,
  `comarca` VARCHAR(20) NOT NULL ,
  `id_ctg_estl` INT NOT NULL ,
  `num_estbl` INT NULL ,
  `num_places` INT NULL ,
  `TEMPS_id_any` INT NOT NULL ,
  `MUNICIPI_id_municipi` INT NOT NULL ,
  `MUNICIPI_comarca` VARCHAR(20) NOT NULL ,
  `CTG_ESTABLIMENTS_id_ctg_estbl` INT NOT NULL ,
  PRIMARY KEY (`id_any`, `comarca`, `id_ctg_estl`),
  INDEX `fk_EVOLUCIO_ESTABLIMENTS_TEMPS1_idx` (`TEMPS_id_any` ASC) ,
  INDEX `fk_EVOLUCIO_ESTABLIMENTS_MUNICIPIL_idx` (`MUNICIPI_id_municipi` ASC,
`MUNICIPI_comarca` ASC) ,
  INDEX `fk_EVOLUCIO_ESTABLIMENTS_CTG_ESTABLIMENTS1_idx`
(`CTG_ESTABLIMENTS_id_ctg_estbl` ASC) ,
  CONSTRAINT `fk_EVOLUCIO_ESTABLIMENTS_TEMPS1`
  FOREIGN KEY (`TEMPS_id_any` )
  REFERENCES `TFC`.`TEMPS` (`id_any` )
  ON DELETE NO ACTION
  ON UPDATE NO ACTION,
  CONSTRAINT `fk_EVOLUCIO_ESTABLIMENTS_MUNICIPIL`
  FOREIGN KEY (`MUNICIPI_id_municipi`, `MUNICIPI_comarca` )
  REFERENCES `TFC`.`MUNICIPI` (`id_municipi`, `comarca` )
  ON DELETE NO ACTION
  ON UPDATE NO ACTION,
  CONSTRAINT `fk_EVOLUCIO_ESTABLIMENTS_CTG_ESTABLIMENTS1`
  FOREIGN KEY (`CTG_ESTABLIMENTS_id_ctg_estbl` )
  REFERENCES `TFC`.`CTG_ESTABLIMENTS` (`id_ctg_estbl` )
  ON DELETE NO ACTION
  ON UPDATE NO ACTION)
ENGINE = InnoDB;

SET SQL_MODE=@OLD_SQL_MODE;
SET FOREIGN_KEY_CHECKS=@OLD_FOREIGN_KEY_CHECKS;
SET UNIQUE_CHECKS=@OLD_UNIQUE_CHECKS;

```