

TFC – Magatzems de Dades

Construcció i explotació d'un magatzem de dades per a l'anàlisi d'informació sobre trànsit rodat de vehicles

Memòria

Yolanda Sánchez Hernández

Enginyeria Tècnica d'Informàtica de Gestió
Universitat Oberta de Catalunya
Curs 2013-2014 / Tardor
06-01-2014

Consultor: Carles Llorach Rius

*Per a la meva família,
la meva constant.*

Agraïments

A la meva família, per la paciència i comprensió que m'han dispensat en el decurs d'aquests estudis i el seu suport incondicional en aquesta aventura en la que em vaig embarcar passats els quaranta.

A tots els consultors i companys amb els que he tingut oportunitat de treballar.

Ha sigut una llarga jornada de 5 anys, on de vegades s'ha fet molt difícil combinar família, feina i estudis sense caure en el desànim o en els dubtes sobre la capacitat per finalitzar exitosament l'enginyeria que se'm va "escapar" als divuit.

A tots, gràcies.

Resum

Aquest document recull els procediments portats a terme per cobrir totes les fases del projecte de final de carrera (TFC) de Magatzems de dades.

L'objectiu final és la construcció i explotació d'un magatzem de dades, que inclou tot un seguit de tasques que trobarem explicades en aquesta memòria orientades a:

1. Adquirir coneixements sobre els conceptes associats amb els magatzems de dades (*Data Warehouse*) i la seva utilitat.
2. Contextualitzar els magatzems de dades en la intel·ligència de negoci o *Business Intelligence*.
3. Dissenyar i implementar els processos de càrrega (ETL) necessaris per incorporar les dades proporcionades per les entitats col·laboradores.
4. Dissenyar i implementar el model multidimensional corresponent.
5. Tot i no ser un requeriment obligatori, familiaritzar-se amb el concepte "quadre de comandament".
6. Estudi i selecció de les eines adients pel projecte.

La intenció d'aquest document no és només il·lustrar el procediment seguit, sinó justificar i defensar totes les decisions preses, tot i utilitzant un llenguatge clar i concís mantenint el focus d'atenció en els processos principals.

És per aquest motiu que s'ha decidit adjuntar un conjunt de documents addicionals (veure [Annex A](#)) que detallen a fons els conceptes més complexes d'aquest document.

Índex de continguts

1. Introducció	6
1.1. Conceptes bàsics	6
1.2. Justificació del TFC	8
1.3. Descripció del projecte	9
1.4. Objectius del projecte	9
1.4.1. Generals	9
1.4.2. Específics	10
1.5. Enfocament i mètode seguit	10
1.6. Planificació del projecte	11
1.6.1. Fites a complir	11
1.6.2. Planificació inicial vs. Planificació final (detall)	11
1.6.3. Anàlisi de riscos	12
1.7. Productes obtinguts	14
1.8. Continguts de la memòria	14
2. Anàlisi	15
2.1. Requeriments de la solució	15
2.1.1. Actors	15
2.1.2. Requeriments funcionals	16
2.1.3. Requeriments no funcionals	16
2.1.4. Origen de dades	18
2.2. Casos d'ús	19
2.3. Model conceptual	25
2.3.1. Model relacional	25
2.3.2. Model multidimensional	26
3. Disseny	33
3.1. Arquitectura del projecte	33
3.1.1. Arquitectura del programari	33
3.1.2. Arquitectura del maquinari	36
3.2. Disseny de la Base de Dades relacional	37
3.2.1. Diagrama entitat-relació	37
3.2.2. Descripció detallada	38
3.3. Model multidimensional detallat	43
3.4. Disseny dels processos de càrrega (ETL)	45
3.4.1. Processos principals: Fonts de dades externes → Staging Area	48
3.4.2. Processos principals: Staging Area → ODS	51
3.4.3. Processos principals: ODS → Magatzem de dades corporatiu (DW)	53

3.5. Quadre de comandament	54
3.5.1. Dades Població	55
3.5.2. Dades Vehicles	56
3.5.3. Dades Conductors	57
3.6. Consultes i Informes	57
3.6.1. Informe dades Població (RPT_Poblacio)	57
3.6.2. Informe dades Vehicles (RPT_Vehicles)	58
3.6.3. Informe dades Conductors (RPT_Conductors)	58
3.6.4. Informe dades Radars (RPT_Radars)	58
3.6.5. Web corporativa	59
4. Conclusions	60
5. Línies d'evolució futures	60
6. Glossari	60
7. Bibliografia	62
8. Annexos	62
Annex A: Documents adjunts a la memòria	62

Relació de Figures i Taules

Figura 1. Components d'un magatzem de dades	7
Taula 1. Ajustos planificació fase d'Inici	11
Taula 2. Ajustos planificació fase 1: Pla de Treball	11
Taula 3. Ajustos planificació fase 2: Anàlisi i Disseny	11
Taula 4. Ajustos planificació fase 3: Implementació	12
Taula 5. Ajustos planificació fase 4: Memòria i Presentació virtual	12
Taula 6. Riscos associats al desenvolupament del projecte	12
Taula 7. Riscos associats al funcionament del producte final	13
Taula 8. Anàlisi per acceptació/rebuig d'eines per confecció d'informes	13
Figura 2. Diagrama de casos d'ús	20
Figura 3. Model conceptual BD relacional	25
Taula 9. Comparativa solucions ROLAP/MOLAP	28
Figura 4. Esquema en estrella per dades Població (cub IndPoblacio)	28
Figura 5. Esquema en estrella per dades Conductors (cub IndConductors)	29
Figura 6. Esquema en estrella per dades Vehicles (cub IndVehicles)	30
Figura 7. Esquema en estrella per dades Radars (cub IndRadars)	31
Figura 8. Diagrama arquitectura del programari	33
Figura 9. Diagrama arquitectura del maquinari	36

Figura 10. Base de Dades Relacional	37
Figura 11. Diagrama d'entitat-relació (DER)	37
Figura 12. Objectes DW_staging	38
Figura 13. Objectes DW_ODS	40
Figura 14. Objectes DW	42
Figura 15. Estrella detallada dades de Població	43
Figura 16. Estrella detallada dades de Vehicles	43
Figura 17. Estrella detallada dades de Conductors	44
Figura 18. Estrella detallada dades de Radars	44
Figura 19. Exemple de diagrama de flux SSIS	45
Figura 20. Diagrama de processos de càrrega (paquets SSIS) de la Staging Area	46
Figura 21. Procés de càrrega SSAS	47
Figura 22. Procés de càrrega RPT	47
Taula 10. Staging tables volumètric worksheet	48
Figura 23. Procés de càrrega base per dades Ubicació (ETL_Municipis)	49
Figura 24. Procés de càrrega base per dades a analitzar (ETL_Entitats)	50
Figura 25. Procés de càrrega de l'ODS des de la Staging Area per dades base	51
Figura 26. Procés de càrrega de l'ODS des de la Staging Area per dades anàlisi	52
Figura 27. Procés de càrrega del Magatzem de Dades corporatiu des de l'ODS	53
Figura 28. Exemple de taules dinàmiques (cub <i>offline</i> IndPoblacio.cub)	54
Figura 29. Quadre de Comandament dades Població	55
Figura 30. Quadre de Comandament dades Vehicles	56
Figura 31. Quadre de Comandament dades Conductors	57
Figura 32. Informe dades Població	57
Figura 33. Informe dades Vehicles	58
Figura 34. Informe dades Conductors	58
Figura 35. Informe dades Radars	58
Figura 36. Pàgina principal web corporativa	59
Figura 37. Pàgina accés privat a web corporativa	59

1. Introducció

Aquest treball de final de carrera (en endavant TFC), proposa la construcció i explotació d'un magatzem de dades o *Data Warehouse* amb l'objectiu d'analitzar la informació relativa a l'evolució del parc de vehicles a Catalunya.

Es planteja, doncs, un cas (en endavant projecte) en el qual l'estudiant adopta el paper de consultor extern de l'empresa que encarrega el projecte (FECRES). Aquest projecte es detalla a l'enunciat del TFC .

1.1. Conceptes bàsics

El magatzem de dades o Data Warehouse (DW)

*W.H.Immon, considerat el "pare" dels magatzems de dades, ho va definir com:
"Conjunt de dades orientades per temes, integrades, que canvien en el temps i no volàtils, que tenen per objectiu donar suport a la presa de decisions"*

- Orientades per temes: es pot analitzar un determinat tema (ex: vendes).
- Integrades: el DW integra dades procedents de diverses fonts amb formats diferents, unificant-les per interpretar-les de forma única mitjançant l'ús de processos ETL.
- Canvien en el temps: es guarden dades històriques (ex: dades últims dotze mesos).
- No volàtils: les dades no canviaran, només poden ser consultades.

R.Kimball, promotor de l'enfocament dimensional i amb una visió més funcional, ho va descriure com: "Còpia de les dades transaccionals específicament estructurada per la consulta i l'anàlisi"

La intel·ligència de negoci o Business Intelligence (BI)

Howard Dresden, analista de Gartner, ho va definir com: "Conceptes i mètodes per millorar les decisions de negoci mitjançant l'ús de sistemes de suport basats en fets."

El magatzem de dades forma part de la intel·ligència de negoci o Business Intelligence, entesa com a conjunt de metodologies.

A dia d'avui, el BI ha esdevingut una eina estratègica fonamental per a la presa de decisions.

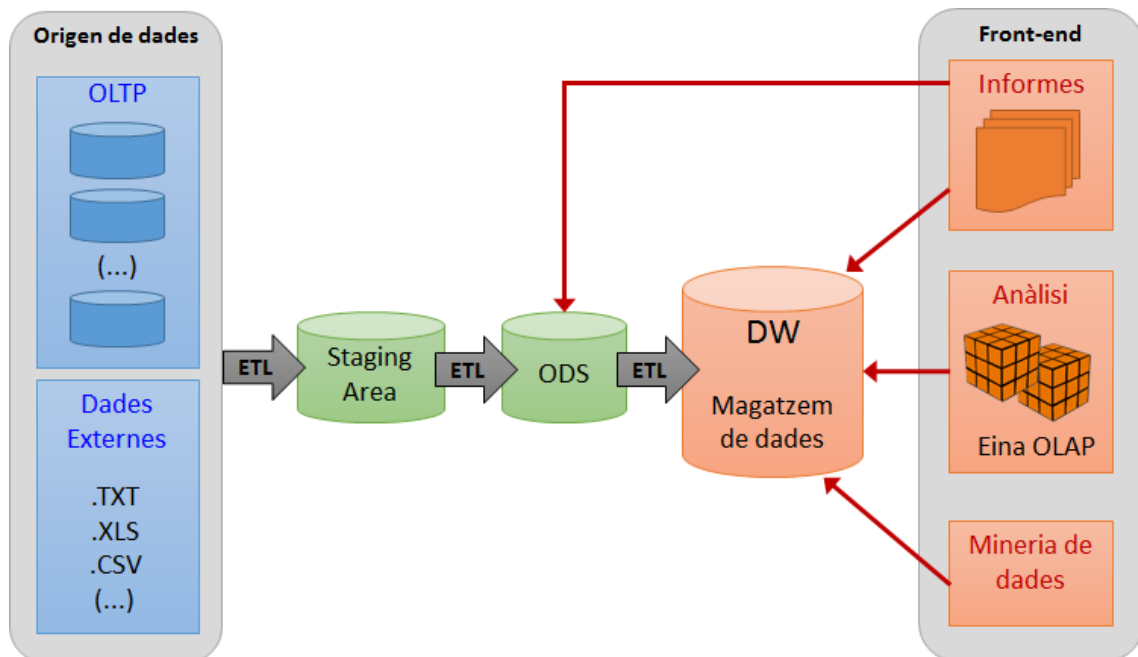


Figura 1. Components d'un magatzem de dades

a) Origen de dades (Data sources)

Sistemes origen des d'on el magatzem de dades o Data Warehouse carrega les dades mitjançant l'ús d'eines ETL (processos de càrrega). Bàsicament, hi ha de dos tipus:

- Sistema operacional corporatiu (transaccional): sobre una base de dades relacional, amb dades normalitzades (optimitzades per processos transaccionals).
- Fitxers: procedents de sistemes aliens a la corporació (com és el cas que ens ocupa) i/o de sistemes corporatius sense connexió directa al sistema operacional. Es tracta de fitxers quins continguts mantenen una estructura tractable per un procés de càrrega de dades (ex: text pla amb camps delimitats per tabulador, XML, XLS, etc.)

b) BD temporal o Staging Area

Emmagatzematge temporal on els usuaris no hi tenen accés. Les dades extretes mitjançant eines ETL des dels sistemes origen es guarden amb la seva estructura original.

És un component opcional, però la seva presència facilita l'eficiència dels processos ETL.

c) Operational Data Store (ODS)

Emmagatzematge on es troben les dades normalitzades (base de dades relacional) recollides des de la *staging area*, i que després es "desnormalitzen" (les dades s'organitzen per tema, no per funció) quan es transfereixen al Magatzem de dades.

Aquest component és opcional i sovint s'utilitza per l'extracció d'informes, així com per tenir una còpia normalitzada de les dades de la *staging area* que permet l'ampliació/modificació de les estructures multidimensionals des de taules estructurades.

d) Magatzem de dades o Data Warehouse (DW)

El magatzem de dades es construeix sobre una base de dades en la que operen les eines OLAP. Es tracta de la informació a analitzar, organitzada en taules de dimensió i de taules de fet.

Alhora, aquestes taules adopten una estructura multidimensional (estrella i/o floc de neu) proporcionada per les eines OLAP.

- Quan la estructura multidimensional es guarda físicament a la base de dades, s'empra la tecnologia MOLAP (*Multidimensional OLAP*), i la informació s'organitza en cubs.
- Quan l'estructura no es guarda enlloc (es basa en vistes) diem que s'utilitza la tecnologia ROLAP (*Relational OLAP*), i la informació s'organitza en taules.

També val anomenar les metadades (Metadata) com a component del magatzem de dades.

e) Accés usuaris (front-end)

No es ben bé un component del magatzem de dades, però forma part de la seva lògica, ja que és el resultat visible dels components anteriorment esmentats.

- Anàlisi (*Analysis*): els usuaris accedeixen a la informació multidimensional mitjançant eines adequades per aquesta tasca (ex: MS Excel amb Power Pivot).
- Informes (*Reporting*): accés a la informació en forma de llista.
- Minería de dades (*Data mining*): accés a la informació per "aprendre" de les dades emmagatzemades mitjançant l'ús d'eines especialitzades basades en estadística.

1.2. Justificació del TFC

Des del punt de vista purament acadèmic, el projecte permet aprofundir els coneixements adquirits a les assignatures relacionades amb Bases de Dades i Enginyeria del Programari tot i aplicant-los al camp dels magatzems de dades.

Tanmateix, i ja centrant-nos en els magatzems de dades, ens permet iniciar-nos en el món de la intel·ligència de negoci (essent els magatzems de dades un component essencial), quin ús a qualsevol empresa és un factor estratègic per a la presa de decisions.

Evidentment, aquest factor encara cobra més importància davant la situació de crisi que estem vivint, on l'ús d'aquestes tecnologies poden marcar la diferència en un mercat altament competitiu, i en molts casos, en una situació de saturació de producte.

En el meu cas particular, fa un temps que estic en contacte amb la intel·ligència de negoci, mitjançant l'ús d'eines professionals. Aquestes eines, però, no em permeten conèixer a fons el que hi ha darrera de les tecnologies emprades, per la qual cosa l'aprenentatge derivat d'aquest TFC serà molt útil pel meu desenvolupament professional.

Des del punt de vista del client (FECRES, representat al TFC per la UOC), aquest treball representarà una solució per millorar la seguretat i la densitat de tràfic a Catalunya mitjançant l'anàlisi de les dades aportades per diferents entitats, consolidades en un magatzem de dades centralitzat.

Aquesta tasca d'anàlisi és realitzada per un analista de negoci, per la qual cosa es justifica l'ús d'un magatzem de dades basat en eines OLAP ja que:

- Els analistes de negoci no són informàtics → han de ser les eines les que s'adaptin a l'usuari i no a l'inrevés.
- Els usuaris no han de dependre del departament d'Informàtica, les eines OLAP han de presentar les dades com els analistes estan acostumats a veure-les (en termes de fets i dimensions en comptes de taules, atributs i claus foranes).

1.3. Descripció del projecte

Aquest TFC es dividirà en quatre fases:

- a) Pla de treball i anàlisi preliminar de requeriments
 - Planificació
 - Anàlisi preliminar
 - Anàlisi de les fonts de dades operacionals
- b) Anàlisi de requeriments i disseny conceptual i tècnic
 - Anàlisi detallat de requeriments
 - Descripció del model dimensional
 - Disseny dels procediments d'extracció de dades a alt nivell
- c) Implementació
 - Construcció del magatzem de dades
 - Configuració de l'eina d'explotació
 - Construcció dels informes i anàlisi de la informació
- d) Memòria i presentació

1.4. Objectius del projecte

1.4.1. Generals

Adquirir experiència en el disseny, construcció i explotació d'un magatzem de dades a partir de la informació disponible en una base de dades transaccional.

- Aprendre tots els conceptes relacionats amb magatzems de dades.
- Entendre la seva utilitat per saber on poden ser necessaris.
- Entendre tot allò relacionat amb bases de dades multidimensionals, com ara estructures OLAP, consultes MDX, ...

1.4.2. Específics

Entendre el funcionament i ús de la base de dades relacional Oracle 11g, incloent-hi PL/SQL (nivell bàsic). L'aprenentatge portat a terme ens ha de permetre comparar amb altres motors de bases de dades relacionals existents al mercat, dels quals ja es tenia alguna experiència prèvia per tal de tenir un millor criteri en cas de dependre de nosaltres la tria d'aquest producte.

Aprenentatge de les eines que Microsoft, en la seva versió 2012 per servidors, té disponibles per intel·ligència de negoci com MS Analysis Services 2012 (SSAS) i MS Integration Services (SSIS).

Tanmateix, es vol aprofundir en el coneixement de MS Excel © (en les seves versions 201x) com a eina d'anàlisi mitjançant l'ús de taules dinàmiques i el component *Power Pivot*. Es tracta d'una eina quin ús es troba molt estès com a eina d'anàlisi per la seva facilitat d'ús i flexibilitat.

1.5. Enfocament i mètode seguit

En l'elaboració d'aquest TFC s'han considerat els següents punts:

- Cobrir tots els requeriments exposats pel client (FECRES) en les seves especificacions.
- Deixar el projecte obert a característiques que es consideren interessants per a futures implementacions, com ara la geo-localització.
- Centrar la funcionalitat del producte resultant en l'analista de negoci, usuari principal. Les decisions preses en quant a les eines utilitzades tenen el seu origen en aquest fet.
- S'intenta, en mida que es pugui, aprofitar les eines que ens proveeix la màquina de desenvolupament per estalviar costos al client i no desestabilitzar la infraestructura proporcionada amb components de programari que puguin afectar al seu bon funcionament → es valora la homogeneïtat en l'entorn de programari (Microsoft, en aquest cas).
- La taxa de recursos emprats en aquest TFC ha sigut dedicada, majoritàriament, a la implementació dels processos ETL per garantir una bona qualitat de dades, ja que es considera la part més important del projecte.

El mètode seguit en l'elaboració d'aquest TFC per cadascuna de les seves fases s'ha basat en :

- Lectura de materials proporcionats a l'aula del TFC, així com d'altres materials disponibles a la Biblioteca virtual del campus relacionats amb Magatzems de Dades.
- Estudi de les opcions disponibles per desenvolupar el projecte (eines OLAP, eina informes, eina ETL, ...) mitjançant bibliografia disponible o consultes a fonts fiables d'Internet. Addicionalment es fan proves de funcionament per valorar la seva idoneïtat.
- Lectura i aplicació, en mida que ha sigut possible, dels suggeriments rebuts pel consultor de les PACs lliurades.
- S'ha intentat, en tot moment, fer un paral·lelisme amb el món empresarial (què faria a la meua empresa en aquest cas?) per intentar apropar-se el màxim possible a la realitat.

1.6. Planificació del projecte

1.6.1. Fites a complir

Les fites del projecte venen determinades per les dates de presentació de cadascuna de les PAC, coincidint amb el final de cada fase del projecte, i detallat al Pla Docent de l'assignatura.

<u>Títol</u>	<u>Inici / Enunciat</u>	<u>Lliurament</u>
PAC1 - Pla de treball	19/09/2013	01/10/2013
PAC2 - Anàlisi i Disseny	02/10/2013	05/11/2013
PAC3 - Implementació	06/11/2013	18/12/2013
PAC4 - Lliurament Final i Defensa	19/12/2013	06/01/2014

1.6.2. Planificació inicial vs. planificació final (detall)

Nom de la tasca	Inici	Fi	Ajustos
Inici	17/09/13	18/09/13	
Inici del semestre	17/09/13	17/09/13	
Lectura del Pla docent	17/09/13	17/09/13	
Lectura i comprensió enunciat TFC	18/09/13	18/09/13	

Taula 1. Ajustos planificació fase d'inici

Nom de la tasca	Inici	Fi	Ajustos
Fase 1: Pla de Treball	19/09/13	01/10/13	
Inici PAC1	19/09/13	19/09/13	
Lectura materials sobre Magatzems de dades	19/09/13	23/09/13	
Elaboració del pla de treball	24/09/13	28/09/13	
Anàlisi preliminar de requeriments	29/09/13	30/09/13	
Comprovació funcionament màquina virtual	27/09/13	30/09/13	
Lliurament PAC1	01/10/13	01/10/13	

Taula 2. Ajustos planificació fase 1: Pla de Treball

Nom de la tasca	Inici	Fi	Ajustos
Fase 2: Anàlisi i Disseny	02/10/13	05/11/13	
Inici PAC2	02/10/13	02/10/13	
Anàlisi de requeriments	02/10/13	06/10/13	
Estudi materials OLAP	07/10/13	10/10/13	
Model conceptual	11/10/13	13/10/13	
Disseny BD / Diagrama ER	14/10/13	16/10/13	
Disseny model multidimensional	17/10/13	21/10/13	
(*) Disseny processos ETL	22/10/13	28/10/13	22/10 – 30/10
(*) Disseny informes	29/10/13	01/11/13	31/10 – 01/11
Control de qualitat de les dades	02/11/13	04/11/13	
Lliurament PAC2	05/11/13	05/11/13	

Taula 3. Ajustos planificació fase 2: Anàlisi i disseny

(*) Variacions de la fase "Anàlisi i Disseny" en les tasques indicades

Nom de la tasca	Inici	Fi	Ajustaments
Fase 3: Implementació	06/11/13	18/12/13	
Inici PAC3	06/11/13	06/11/13	
Generació BD relacional	07/11/13	09/11/13	
(*) Implementació processos ETL	10/11/13	23/11/13	10/11 – 07/12
(*) Implementació estructures OLAP	24/11/13	07/12/13	08/12 – 11/12
(*) Implementació quadre de comandament	08/12/13	11/12/13	12/12 – 15/12
(*) Implementació informes	12/12/13	14/12/13	16/12 – 16/12
(*) Documentació	15/12/13	16/12/13	17/12 – 18/12
(**) Generació màquina virtual comprimida	17/12/13	18/12/13	
Lliurament PAC3	18/12/13	18/12/13	

Taula 4. Ajustos planificació fase 3: Implementació

(*) Variacions de la fase "Implementació" en les tasques indicades

(**) Finalment, a les indicacions de lliurament de la fase 3 no s'indica la necessitat de lliurar la màquina virtual comprimida, atès que s'està treballant amb un entorn accessible online.

Nom de la tasca	Inici	Fi	Comentaris
Fase 4: Memòria i Presentació virtual	19/12/13	06/01/14	
Memòria	20/12/13	25/12/13	
Presentació virtual	26/12/13	05/01/14	
Lliurament final	06/01/14	06/01/14	

Taula 5. Ajustos planificació fase 4: Memòria i Presentació virtual

1.6.3. Anàlisi de Riscos

Riscos associats al desenvolupament d'aquest projecte

Incidència	Pla de contingència
Reducció de disponibilitat de l'estudiant causada per malaltia, obligacions laborals, viatges imprevistos, etc.	1) Si la manca de disponibilitat no és massa crítica, l'estudiant incrementarà el nombre d'hores de dedicació al projecte en quant torni al nivell de disponibilitat habitual. 2) Si la manca de disponibilitat és crítica, es comunicarà al consultor del TFC, atès que es tracta d'un treball personal i intransferible subjecte a avaluació.
Avaria de l'equipament de desenvolupament	1) Equipament del consultor: es disposarà d'una altra màquina de característiques similars. 2) Màquina virtual d'Amazon: es demanarà suport al consultor del TFC perquè el personal tècnic de la UOC es posi en contacte amb aquest proveïdor, que tornarà a generar l'entorn.
Pèrdua d'informació	1) Equipament del consultor: s'efectuarà una còpia diària de la informació en un dispositiu d'emmagatzemament USB. 2) Màquina virtual d'Amazon: s'efectuarà una còpia diària les dades relacionades amb el projecte en una carpeta comprimida que posteriorment s'enviarà per FTP a un servidor accessible pel consultor.

Taula 6. Riscos associats al desenvolupament del projecte

Riscos associats al funcionament del producte final

Incidència	Pla de contingència
Canvien les especificacions de les fonts de dades operacionals procedents de les entitats col·laboradores.	S'establirà un protocol de comunicació entre les entitats i FECRES per tal d'informar adequadament qualsevol canvi o imprevist en l'enviament de les dades. Tanmateix, el sistema permetrà la càrrega reiterada de la informació en cas de rebre versions addicionals dels fitxers que contenen les dades esmentades.

Taula 7. Riscos associats al funcionament del producte final

Riscos inherents al desenvolupament del projecte

Hi ha riscos no previsibles relacionats amb la presa de decisions, com ara la selecció d'una eina que, per les seves característiques, cobreix els requisits de la solució i després no compleix les expectatives o, simplement no es troba integrada en l'entorn de desenvolupament de la manera esperada.

Aquest ha sigut el cas pel que fa al TFC que ens ocupa respecte a l'eina *MS Reporting Services 2012 (SSRS)*, inicialment planificada per cobrir les funcionalitats relacionades amb informes. Un cop es va iniciar la fase d'implementació, es va comprovar que el producte no estava completament instal·lat a l'entorn de desenvolupament (màquina Amazon), i que la versió de SQL Server Business Intelligence 2012 instal·lada al servidor (*Enterprise*) no permetia afegir més productes de la mateixa *suite*, atès que els fitxers originals d'instal·lació no es trobaven al servidor.

L'únic pla de contingència possible és trobar una altra eina que pugui aportar la mateixa funcionalitat en el mínim temps possible per no afectar a la planificació del projecte, amb una corba d'aprenentatge mínima i cost controlat. Es proposa la creació d'una taula amb les possibles alternatives i una valoració d'acceptació/rebuig per cadascuna d'elles.

A mode d'exemple, s'indica el procediment seguit per aquest projecte:

Producte	OK	Motius acceptació / rebuig
PENTAHO reporting	NO	- Requeria la instal·lació de més components de la suite PENTAHO per funcionar de manera òptima. - El consultor a càrrec del projecte (estudiant TFC) no tenia coneixements previs d'aquest producte
JASPER reports	NO	- Necessita habilitar la connexió a MS Analysis Services 2012 via HTTP → un cop habilitat, s'observa inestabilitat en la connexió. - L'eina no ofereix la possibilitat de dissenyar quadres de comandament (funcionalitat restringida a la versió Enterprise de pagament).
MS Excel 201x	SÍ	- Disponible actualment al client (FECRES rebia informació en aquest format de les entitats externes). - Ja hi ha un coneixement previ de l'eina entre els usuaris. - Proporciona totes les funcionalitats requerides.

Taula 8. Anàlisi per acceptació/rebuig d'eines per confecció d'informes

1.7. Productes obtinguts

- Bases de dades relacionals (Oracle Express 11g):
 - **DW_staging**: emmagatzematge de la staging area. Taules que respecten l'estructura dels fitxers externs al sistema amb totes les dades originals, excepte les redundants.
 - **DW_ODS**: emmagatzematge de la ODS. Taules amb estructura normalitzada que guarden les dades procedents de la staging area.
 - **DW**: magatzem de dades. Taules de dimensió i taules de fet.
- Base de dades multidimensional **TFC**, construïda sobre SSAS 2012 (*Microsoft Analysis Services*).
- Processos **ETL**: basats en SSIS 2012 (*Microsoft Integration Services*). A continuació s'indiquen els processos principals (es pot trobar el detall a l'apartat [3.4. Disseny processos de càrrega](#) d'aquest document):
 - **ETL_DWstaging**: càrrega de la *staging area* des de fonts de dades externes.
 - **ETL_BDrel**: càrrega de la ODS des de la *staging area*.
 - **ETL_DW**: càrrega del magatzem de dades des de l'ODS.
 - **SSAS**: generació de cubs sense connexió (arxius .CUB, permeten l'anàlisi de dades des de MS Excel) des del magatzem multidimensional TFC.
 - **RPT**: generació d'informes des de l'ODS (arxius XLS).
- Quadre de comandament basat en MS Excel 201x, amb connexió al servidor OLAP (C:\FECRES\QC.xlsx): permet l'anàlisi de la informació segons les especificacions proporcionades pel client.
- Informes unificats en un arxiu MS Excel 201x (C:\FECRES\Informes_local.xlsx): aquest arxiu conté vincles a d'altres arxius .XLS generats per l'aplicació (procés ETL RPT).
- Cubs sense connexió (.CUB): permeten l'anàlisi mitjançant taules dinàmiques de tota la informació gestionada pel projecte des de MS Excel 201x sense necessitat d'estar connectat al servidor. Els cubs són generats pel procés ETL anomenat **SSAS**.
- Portal online al servidor de FECRES que dona accés a:
 - Informació pública: estadístiques generals disponibles per a qualsevol usuari.
 - Informació privada FECRES: accés restringit al personal de FECRES per a descàrregues d'informació actualitzada (informes, cubs *offline* i arxius base).

1.8. Continguts de la memòria

La resta de capítols que formen el cos d'aquesta memòria, i que veurem a continuació, tracten sobre d'aquest document podem trobar els següents continguts:

- Anàlisi: inclou els requeriments, casos d'ús i model conceptual
- Disseny: inclou diagrames d'arquitectura de programari i maquinari, disseny de la base de dades, diagrama del model físic i descripció dels informes creats.
- Captures de pantalla, conclusions i línies d'evolució futures.

2. Anàlisi

2.1. Requeriments de la solució

2.1.1. Actors

Consultor/a UOC

Representat per l'estudiant que realitza aquest TFC. Les seves funcions principals seran:

- Creació de la base de dades relacional corporativa.
- Càrrega de dades inicial
- Disseny d'estructures OLAP necessàries per poder dur a terme l'anàlisi de dades.
- Disseny d'informes i consultes inicial.
- Formació al/s gestor/s de BD del sistema.

Gestor BD

Representat pel personal d'informàtica del client (o en el seu defecte algú amb la formació necessària), quines funcions principals seran:

- Càrrega – almenys – anual de les dades procedents de les entitats col·laboradores.
- Revisió i depuració de les dades carregades.
- Creació de nous informes/consultes a petició del client.

Analista de negoci

Personal de FECRES especialitzat en l'anàlisi de la informació del magatzem de dades.

- Anàlisi multidimensional (ús de consultes basades en motor OLAP).
- Generació d'informes estadístics en base a l'anàlisi efectuat.
- Establiment d'indicadors.

Entitat col·laboradora

Entitats que proveeixen les dades a carregar en el magatzem de dades:

- IDESCAT: Institut d'Estadística de Catalunya. Dades relatives a municipis i vehicles.
- DGT: Direcció General de Tràfic. Aporta el cens de conductors dels darrers 5 anys.
- SERVEI CATALÀ DE TRÀNSIT: dades relatives a radars fixos.

Aquestes entitats no interactuen directament amb el sistema, però són destinataris de notificacions que genera el sistema en relació a les dades rebudes (majorment incidències).

Usuari

Qualsevol usuari, ja que tindrà accés a la informació pública de la web corporativa de FECRES per consulta d'estadístiques bàsiques i públiques.

2.1.2. Requeriments Funcionals

- a) Creació de la base de dades relacional corporativa on s'emmagatzemaran les dades operacionals procedents de les entitats col·laboradores.
- b) Càrrega de dades operacionals en la base de dades corporativa, almenys un cop a l'any. S'empraran processos ETL i es generarà un informe de càrrega consultable.
- c) Anàlisi de la informació del magatzem de dades.
- d) Disseny de consultes i informes.

2.1.3. Requeriments no funcionals

- a) Requisits de la interfície
 1. Com a base per tot el projecte, es necessita una base de dades relacional on guardar la informació normalitzada.
→ *Oracle Server Express 11g* compleix amb les necessitats i es troba instal·lada a l'entorn.
 2. Atès el clar enfocament analític d'aquest projecte, es requerirà una eina que permeti tant l'emmagatzemament de les estructures multidimensionals com l'explotació de la informació que contenen.
→ *MS Analysis Services 2012 (SSAS)*, ja que es troba instal·lada a l'entorn proporcionat per aquest projecte, i permet el model MOLAP (més òptim en quant al rendiment de les consultes).
 3. Tanmateix, serà necessari disposar d'una eina per l'execució d'informes que permeti, almenys, l'ús de filtres. També seria recomanable que inclogués la possibilitat de fer *drill-down* sobre les dades d'un nivell de detall superior (en termes de llistats, plegar/desplegar grups d'informació)
→ *MS Excel 2010/2013*: el consumidor de la informació generada pel magatzem de dades serà l'analista de negoci, sense formació tècnica per utilitzar eines sofisticades d'anàlisi. Per aquest motiu optem per una eina senzilla que no necessita de la intervenció del departament de informàtica, i que pot treballar sense connexió al servidor (excepte quan actualitza dades).
 4. En rebre la informació del magatzem de dades des de fora del sistema en formats diversos, es necessita una eina ETL per carregar aquestes dades
→ de nou, aprofitem una de les eines de l'entorn: *MS Integration Services 2012 (SSIS)*, que a més proporciona mecanismes avançats de generació de processos de càrrega complexes.
 5. Pel que fa al front-end (usuari), bàsicament es treballarà amb:
 - Navegador Internet: per mostrar el portal corporatiu (públic i privat).
 - *MS Excel 2010/2013 + Power pivot*: informes, anàlisi multidimensional i quadre de comandament. Integració amb SSAS.

b) Requisits de prestacions

- El sistema haurà de disposar de suficient capacitat tant per el tractament de la informació multidimensional com pel que fa als requeriments de totes les eines instal·lades i esmentades al punt anterior.
- El nombre d'usuaris concurrents no sembla ser un punt crític, ja que els usuaris que accedeixen al sistema són força limitats. Per tant, l'única consideració serà disposar de les llicències d'ús del programari adients.

c) Requisits d'operació

- L'operació del sistema serà efectuada, en la seva posada en marxa pel Consultor i, posteriorment pel Gestor BD.
- Ambdós perfils hauran de comptar amb els coneixements i experiència necessaris per operar en aquest sistema, així com accés al sistema proporcionat pel propietari de l'entorn.

d) Requisits de documentació

Segons les especificacions a l'enunciat d'aquest TFC, seran necessaris – almenys – els següents documents.

- Memòria
- Presentació del producte
- Model de dades

e) Requisits de seguretat

- Accés a serveis i màquines: només el Gestor BD tindrà accés a aquests recursos.
- Autenticació i autorització d'usuaris: la seguretat treballarà en tres nivells:
 - i. Sistema operatiu: els usuaris autoritzats tindran accés a les carpetes del sistema amb informació relacionada amb aquest projecte (seguretat NTFS), així com les eines necessàries instal·lades al seu equip pel departament d'IT (els usuaris no són administradors de les seves màquines).
 - ii. Base de Dades: seguretat integrada amb el Sistema Operatiu, i definida a nivell d'accés (login) i a nivell d'objecte (permisos sobre taules, vistes, etc.)
 - iii. Aplicació: aquest projecte inclou una relació d'usuaris i els seus accessos a les diferents parts del producte.

f) Requisits de portabilitat

No hi ha requisits específics de portabilitat, per aquest motiu s'han escollit les eines en entorn Windows, prioritant la disponibilitat, qualitat i integració.

g) Requisits de qualitat

- Proves de sistema: com a qualsevol projecte de programari, es concertarà amb els usuaris un període de proves al sistema per comprovar el seu funcionament (correctesa, requeriments i rendiment).
- Qualitat de la informació al sistema: la qualitat de la informació és aliena a FECRES, ja que la informació ens arriba d'entitats externes → els processos de càrrega (ETL) hauran de gestionar el control de qualitat de les dades esmentades.

h) Requisits de fiabilitat

- L'entorn de treball el projecte inclou un contracte amb especificacions SLA sobre els *ratios* de fiabilitat del sistema i el temps de resposta en cas de fallida.
- Es programaran còpies de seguretat tant a nivell de base de dades com de sistema operatiu per garantir la seva recuperació en cas de fallida.

i) Requisits de permanència i actualització

- Tot el programari del servidor s'ha d'actualitzar regularment mitjançant la plataforma Windows Update de Microsoft sota la supervisió del gestor BD i el responsable d'infraestructura.

2.1.4. Origen de dades ¹

Les dades que s'han d'incorporar al sistema tenen les següents característiques:

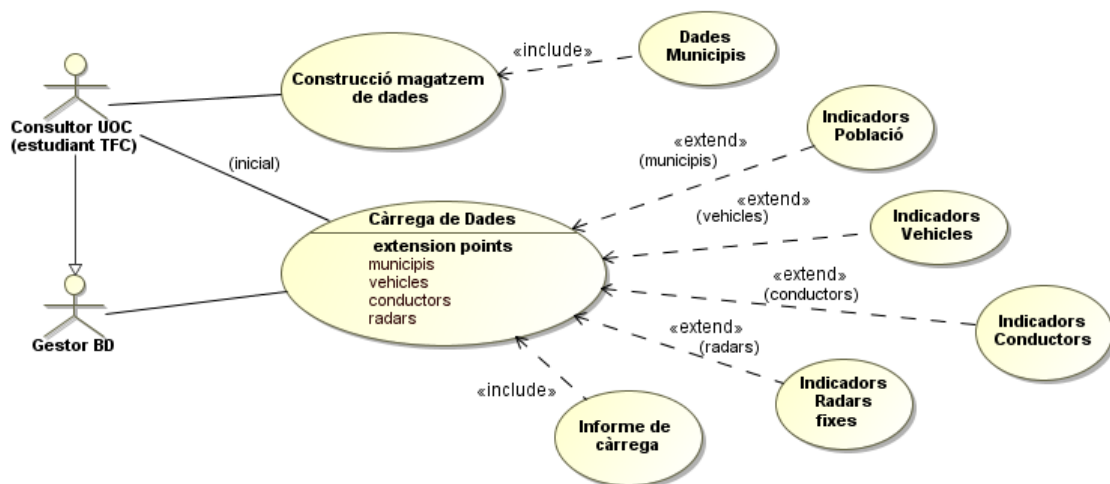
a) Dades procedents d'entitats col·laboradores (IDESCAT, DGT i SCT) :

- 1) Son fonts de dades alienes al sistema operacional de FECRES → necessitaran validació prèvia a la seva incorporació al magatzem de dades, així com control de qualitat dels continguts.
- 2) Es tracta d'arxius de text pla i/o *MS Excel* → necessitem processos ETL i una *staging area* per guardar la informació rebuda abans d'incorporar-la al sistema.
- 3) Els arxius rebuts només contenen els municipis on hi ha disponibles les dades corresponents a cada fitxer. Tanmateix, no hi ha coincidència entre els diferents arxius en quant als municipis inclosos, ni es disposa en tots els casos d'un codi/denominació únic per dits municipis que garanteixi la coherència exigida pel sistema.

1. Veure **Annex A**: Documents adjunts a la memòria (Processos de càrrega)

- 4) La informació obtinguda no s'incorporarà a la base de dades operacional, ja que no interactuen amb el sistema de FECRES ni formen part d'ell. Tanmateix no compleix amb els nivells mínims de normalització necessaris. S'optarà, doncs, per la seva càrrega al magatzem de dades corporatiu.
 - i. Abans de la càrrega al magatzem de dades corporatiu, es guardarà una còpia de les dades normalitzades per proporcionar una via de consulta per informes.
 - 5) No es disposa de les dades del cens de conductors per l'any 2012 → no es contemplaran, per tant, aquestes dades ni es buscaran mètodes alternatius d'incorporar-les en no ser crítiques pel sistema.
 - 6) La informació relativa a permisos de conducció i llicències no inclou el detall relatiu al tipus de permís de conducció (A, A1, B1, etc.), per la qual cosa només es consideraran dos tipus de permís: permís i llicència.
- b) Dades obtingudes per FECRES per complementar / donar coherència al sistema d'anàlisi d'informació.
- Dades "mestres" de comarques i municipis obtingudes des de l'IDESCAT: es decideix incorporar aquesta informació atès que les dades procedents d'entitats col·laboradores no inclouen una informació unificada i completa, i és vital per fer l'anàlisi posterior.

2.2. Casos d'ús



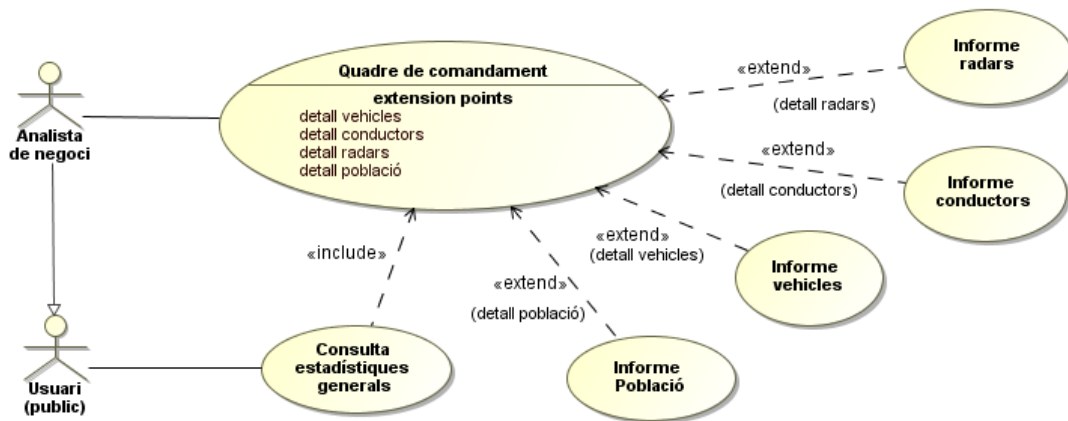


Figura 2. Diagrama de casos d'ús

Cas d'ús Construcció magatzem de dades

Resum de la funcionalitat: dur a terme la creació d'un magatzem de dades i tots els elements relacionats.

Actors: **Consultor UOC**

Pre-condició: el sistema disposa de totes les eines necessàries (base de dades relacional, motor per construir magatzem de dades multidimensional, etc.).

Post-condició: existència d'un magatzem de dades amb les estructures multidimensionals necessàries i les dades mestres de municipis carregades.

Procés:

1. El **consultor UOC** genera les taules necessàries a la base de dades relacional corporativa per allotjar les dades mestres normalitzades.
2. El **consultor UOC** genera les taules i estructures en estrella necessàries a la base de dades multidimensional.

Cas d'ús Dades municipis

Resum de la funcionalitat: càrrega de dades mestres d'ubicació (Comarca, Municipi) per donar coherència a les dades externes proporcionades per les entitats col·laboradores. S'entén com una funcionalitat exclusiva de la posada en marxa del sistema o necessària en cas de canviar alguna dada de la xarxa de municipis/comarques de Catalunya. S'obtenen des de la web de l'IDESCAT

Actors: **Consultor UOC**

Pre-condició: existència de les taules necessàries per carregar aquesta informació a la base de dades relacional.

Post-condició: dades mestres de Municipis carregades a la base de dades relacional.

Procés principal / alternatives de procés i excepcions: veure descripció del procés de càrrega [ETL Municipis](#) a l'apartat 3.4.1 d'aquest document.

Casos d'ús relacionats: Construcció magatzem de dades.

Cas d'ús Càrrega de Dades

Resum de la funcionalitat: Carregar la informació proporcionada per les entitats col·laboradores des de orígens de dades externs al sistema a la base de dades relacional mitjançant processos ETL (el disseny detallat d'aquests processos es pot trobar a [l'apartat 3.4](#) d'aquest document per als processos principals o als documents adjunts ¹).

Actors: **Consultor UOC** (cas càrrega inicial), **Gestor BD**

Pre-condició:

- El sistema disposa d'una base de dades relacional amb les taules necessàries per carregar la informació.
- El sistema disposa d'una *staging area* on deixar les dades "en brut" recopilades per a la seva validació com a pas previ a la seva incorporació a la base de dades relacional.

Post-condició: dades carregades a la base de dades relacional.

Procés:

1. El **consultor UOC** ó el **Gestor BD** obté el fitxer enviat per l'entitat col·laboradora externa i ho des a la ubicació prevista (carpeta) pel procés de càrrega.
2. El **consultor UOC** ó el **Gestor BD** accedeix a l'eina ETL i selecciona el procés de càrrega de dades (procés ETL) i l'executa.
3. El procés ETL carrega les dades seleccionades a la *staging area*.
4. El procés ETL valida les dades carregades complint amb les especificacions del sistema.
5. El procés ETL transfereix les dades des de la *staging area* a la base de dades relacional.
6. El procés ETL genera un informe de càrrega amb els resultats.

Alternatives de procés i excepcions:

- 1a. El fitxer extern no es troba a la ubicació esperada → el sistema informa del problema i finalitza el cas d'ús.
- 3a. El fitxer extern no conté dades o les dades tenen una estructura incorrecta → el sistema informa del problema i finalitza el cas d'ús.
- 4a. Es troben errors durant el procés de càrrega.
 - 4a1. Tots els registres són erronis → el sistema informa del problema i finalitza el cas d'ús.
 - 4a2. Alguns registres són erronis → el sistema genera un informe detallat dels problemes i finalitza el cas d'ús sense incorporar cap dada al sistema.

Casos d'ús relacionats: Indicadors Població, Indicadors Vehicles, Indicadors Conductors, Indicadors Radars, Informe de càrrega.

1. Veure **Annex A**: Documents adjunts a la memòria (Model Multidimensional)

Cas d'ús Indicadors Població

Resum de la funcionalitat: càrrega d'indicadors relacionats amb població, procedents de l'arxiu proporcionat per l'IDESCAT (*Dades_municipis.xls*).

Actors: **Consultor UOC** (cas càrrega inicial), **Gestor BD**

Pre-condició: existència de les taules necessàries per carregar aquesta informació a la base de dades relacional.

Post-condició: indicadors relacionats amb la població carregats a la base de dades relacional.

Procés principal / alternatives de procés i excepcions: veure descripció del procés de càrrega ETL_IndPoblacio als documents adjunts.

Casos d'ús relacionats: Càrrega de dades.

Cas d'ús Indicadors Vehicles

Resum de la funcionalitat: càrrega d'indicadors relacionats amb el parc de vehicles, procedents de l'arxiu proporcionat per l'IDESCAT (*Dades_vehicles.xls*).

Actors: **Consultor UOC** (cas càrrega inicial), **Gestor BD**

Pre-condició: existència de les taules necessàries per carregar aquesta informació a la base de dades relacional.

Post-condició: indicadors relacionats amb el parc de vehicles carregats a la base de dades relacional.

Procés principal / alternatives de procés i excepcions: veure descripció del procés de càrrega ETL_IndVehicles als documents adjunts.

Casos d'ús relacionats: Càrrega de dades.

Cas d'ús Indicadors Conductors

Resum de la funcionalitat: càrrega d'indicadors relacionats amb cens de conductors (permisos i llicències), procedents de l'arxiu proporcionat per la DGT (*Dades_conductorsYYYY.txt*).

Actors: **Consultor UOC** (cas càrrega inicial), **Gestor BD**

Pre-condició: existència de les taules necessàries per carregar aquesta informació a la base de dades relacional.

Post-condició: indicadors relacionats amb conductors carregats a la base de dades relacional.

Procés principal / alternatives de procés i excepcions: veure descripció del procés de càrrega ETL_IndConductors als documents adjunts.

Casos d'ús relacionats: Càrrega de dades.

Cas d'ús [Indicadors Radars fixes](#)

Resum de la funcionalitat: càrrega d'indicadors relacionats amb els radars fixes (ubicació), procedents de l'arxiu proporcionat pel Servei Català de Trànsit (*Radars_SCT.txt*).

Actors: **Consultor UOC** (cas càrrega inicial), **Gestor BD**

Pre-condició: existència de les taules necessàries per carregar aquesta informació a la base de dades relacional.

Post-condició: indicadors relacionats amb el cens de conductors carregats a la base de dades relacional.

Procés principal / alternatives de procés i excepcions: veure descripció del procés de càrrega ETL_IndRadars als documents adjunts.

Casos d'ús relacionats: [Càrrega de dades](#).

Cas d'ús [Quadre de Comandament](#)

Resum de la funcionalitat: Proporcionar un quadre de comandament o Dashboard per facilitar l'anàlisi de la informació del trànsit rodat de vehicles.

Actors: **Analista de negoci**

Pre-condició: El sistema disposa de les eines OLAP necessàries per efectuar l'anàlisi, i de les eines de *reporting* per generar el quadre de comandament.

Post-condició: visualització del quadre de comandament.

Procés principal / alternatives de procés i excepcions: veure descripció detallada a l'apartat 3.5 ([Quadre de comandament](#)) d'aquest document.

Casos d'ús relacionats: [Informe Radars](#), [Informe Conductors](#), [Informe Vehicles](#).

Cas d'ús [Informe Població](#)

Resum de la funcionalitat: generació d'un informe amb els detalls de les dades de població per municipis (el disseny detallat dels informes del sistema es troba a l'apartat 3.6 d'aquest document).

Actors: **Analista de negoci**

Pre-condició: el sistema disposa de les eines de *reporting* necessàries per generar els informes

Post-condició: visualització de l'informe.

Procés principal / alternatives de procés i excepcions: veure descripció de l'informe [RPT Poblacio](#) a l'apartat 3.6.1 d'aquest document.

Casos d'ús relacionats: [Quadre de comandament](#).

Cas d'ús [Informe Radars](#)

Resum de la funcionalitat: generació d'un informe amb els detalls dels radars per municipis (el disseny detallat dels informes del sistema es troba a l'apartat 3.6 d'aquest document).

Actors: Analista de negoci

Pre-condició: el sistema disposa de les eines de *reporting* necessàries per generar els informes

Post-condició: visualització de l'informe.

Procés principal / alternatives de procés i excepcions: veure descripció de l'informe [RPT Radars](#) a l'apartat 3.6.4 d'aquest document.

Casos d'ús relacionats: [Quadre de comandament](#).

Cas d'ús [Informe Conductors](#)

Resum de la funcionalitat: generació d'un informe amb els detalls del cens de conductors per municipis (el disseny detallat dels informes del sistema es troba a l'apartat 3.6 d'aquest document).

Actors: Analista de negoci

Pre-condició: el sistema disposa de les eines de *reporting* necessàries per generar els informes

Post-condició: visualització de l'informe.

Procés principal / alternatives de procés i excepcions: veure descripció de l'informe [RPT Conductors](#) a l'apartat 3.6.3 d'aquest document.

Casos d'ús relacionats: [Quadre de comandament](#).

Cas d'ús [Informe Vehicles](#)

Resum de la funcionalitat: generació d'un informe amb els detalls del parc de vehicles per municipis (el disseny detallat dels informes del sistema es troba a l'apartat 3.6 d'aquest document).

Actors: Analista de negoci

Pre-condició: el sistema disposa de les eines de *reporting* necessàries per generar els informes

Post-condició: visualització de l'informe.

Procés principal / alternatives de procés i excepcions: veure descripció de l'informe [RPT Vehicles](#) a l'apartat 3.6.2 d'aquest document.

Casos d'ús relacionats: [Quadre de comandament](#).

Cas d'ús **Consulta Estadístiques Generals**

Resum de la funcionalitat: generació d'un informe generalista amb accés públic des de la web corporativa de FECRES amb el resultat de l'anàlisi efectuat per la corporació.

Actors: **Usuari (públic)**

Pre-condició: l'analista de negoci ha generat l'informe des del quadre de comandament

Post-condició: visualització de l'informe.

Procés principal / alternatives de procés i excepcions: veure descripció de l'informe [RPT EstadístiquesWeb](#) a l'apartat 3.6.5 d'aquest document.

Casos d'ús relacionats: [Quadre de comandament](#).

2.3. Model conceptual

2.3.1. Model relacional

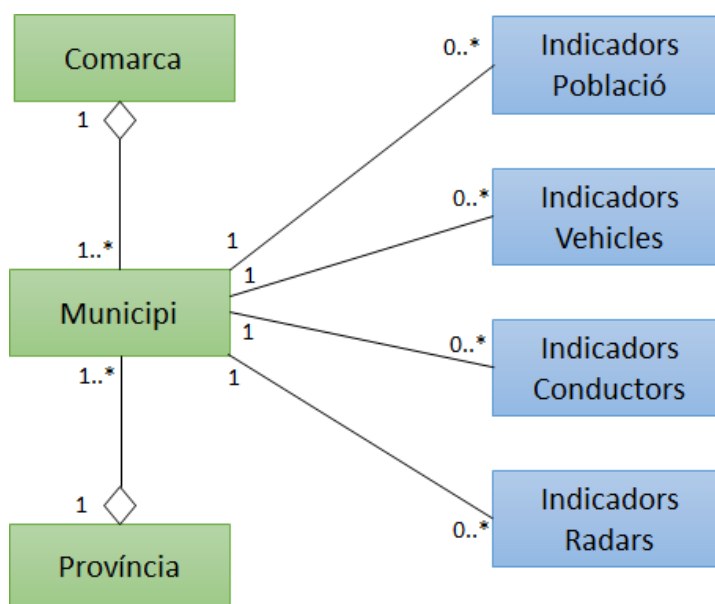


Figura 3. Model conceptual

El model relacional inclou dos tipus d'entitats:

- Entitats pròpies del sistema (en fons verd): les dades són pròpies del sistema, les obté FECRES per garantir l'unicitat i integritat de les dades dels municipis.
- Entitats alienes al sistema (en fons blau): les dades provenen de l'exterior, la seva estructura i continguts no poden ser controlats per FECRES, que disposarà d'un procés ETL que permetrà incorporar-les al sistema.

- Població: dades relacionades amb la població (nombre d'habitants, extensió, etc.).
- Vehicles: nombre de vehicles del parc classificat per tipus de vehicle.
- Conductors: nombre de permisos de conducció classificat per gènere i tipus.
- Radars: relació de radars per via.

Podríem parlar encara d'un tercer tipus d'entitats no relacionades amb les dades que formen part del nucli de la solució, les que donen suport a la interfície d'usuari (integrades a l'eina OLAP del servidor: *MS Analysis Services*).

- Metadades: atributs que descriuen les dades. La manera en què es presenta la informació a l'usuari no és mitjançant l'ús de noms tècnics (com ara noms de columna a les taules, sinó noms intel·ligibles que pugui entendre un usuari no informàtic (com ara l'analista de negoci).
- Seguretat: marca l'accés dels usuaris a les entitats de dades.
- KPI (*Key Performance Indicators*): permeten fer realment útils les dades analitzades dotant-les d'unes xifres objectiu i un significat.

2.3.2. Model multidimensional ¹

Seguint recomanacions dels mòduls didàctics de l'assignatura Magatzems de dades i disseny multidimensional (Mòdul 4. Disseny multidimensional, pàg.59), optem per l'ús del model d'**estrella** desnormalitzada en detriment del **floc de neu** (estrella normalitzada): *"El floc de neu és un error de disseny que genera un guany inapreciable d'espai i una gran pèrdua de rendiment. Per a millorar el rendiment de les consultes, hem d'evitar normalitzar els esquemes, llevat de casos extrems en què la grandària de la Dimensió sigui comparable a la del Fet."*

Dimensions

- Temps: jerarquia d'agregació (*all, any): no es poden obtenir més nivells atès que el nivell de dades més detallat correspon a l'any.
- Ubicació: jerarquia d'agregació (*all, província / comarca / municipi)
- Tipus de vehicle: jerarquia d'agregació (*all, tipus vehicle)
- Tipus de permís de conducció: jerarquia d'agregació (*all, tipus permís)
- Gènere: jerarquia d'agregació (*all, gènere)
- Via: jerarquia d'agregació (*all, via)

Taules de fet

Hi haurà una taula de fet per cada conjunt de mesures amb la mateixa granularitat. La taula de fet determina el punt d'estudi, allò que es vol analitzar. Les mesures que podem trobar en aquest projecte s'especifiquen a continuació (ressaltades en blau les mesures calculades):

- Taula de fet de Població
 - Nombre d'habitants
 - Extensió del municipi (Km2)

1. Veure **Annex A**: Documents adjunts a la memòria (Model Multidimensional)

- Taula de fet de Vehicles
 - Nombre de vehicles
- Taula de fet de Conductors
 - Nombre de conductors
- Taula de fet de Radars
 - Nombre de radars

Indicadors (KPI)

No s'especificaven a l'enunciat, però entenem que el valor real de l'anàlisi de dades s'obté en quan es defineixen els indicadors adequats (no hi ha mida útil sense objectiu ni comparació).

Els indicadors es defineixen en "mode semàfor":

- Verd: objectiu complert
- Groc: no hi ha canvis
- Vermell: objectiu no complert

A continuació passem a anomenar els indicadors per aquest projecte:

- Indicadors per Població
 - Població vs. any anterior
 - Densitat de població vs. any anterior
- Indicadors per Vehicles
 - Vehicles vs. any anterior
 - Vehicles per Habitant vs. any anterior
 - Vehicles per Radar vs. any anterior

Solució OLAP

Atès l'escàs volum d'informació que presenta el projecte i la freqüència amb que s'actualitzen les dades del magatzem de dades, es decideix que un model MOLAP seria més adequat. Ara bé, les especificacions donades al TFC indiquen "...El disseny haurà d'estar orientat a un magatzem de dades físic ROLAP, creant les dimensions, atributs i fets necessaris...", per la qual cosa s'opta per la implementació d'aquesta solució, tenint en compte les limitacions que ROLAP implica en un producte com *MS Analysis Services* ¹.

1. Microsoft, *Partition Storage Modes and Processing*
(<http://technet.microsoft.com/en-us/library/ms174915.aspx>)

ROLAP	MOLAP
<p>Les dades es guarden a una base de dades relacional</p> <p>→ temps de càrrega menor en utilitzar eines ETL</p> <p>→ pitjor rendiment (connexió addicional a la BD relacional).</p> <p>→ no necessita sincronització, sempre accedeix a les dades reals</p>	<p>Les dades es guarden en una BD pròpia</p> <p>→ continua necessitant una BD relacional per la càrrega de dades de fonts externes.</p> <p>→ millor rendiment</p> <p>→ necessita sincronització amb el DW per mantenir les dades actualitzades</p>
Més escalable per grans volums de dades (dimensions amb milions de registres)	No adequat per grans volums de dades
Permet un model de seguretat a nivell de files / columnes (el que proporciona l'SGDB)	Model de seguretat a més alt nivell i basat en l'eina OLAP
Per dades agregades, dóna un rendiment pobre a no ser que es generin taules amb dites dades amb les eines ETL	Les dades agregades es generen automàticament a la BD multidimensional

Taula 9. Comparativa solucions ROLAP/MOLAP

A continuació es detallen les estructures multidimensionals del projecte, seguint les recomanacions dels mòduls didàctics de l'assignatura .

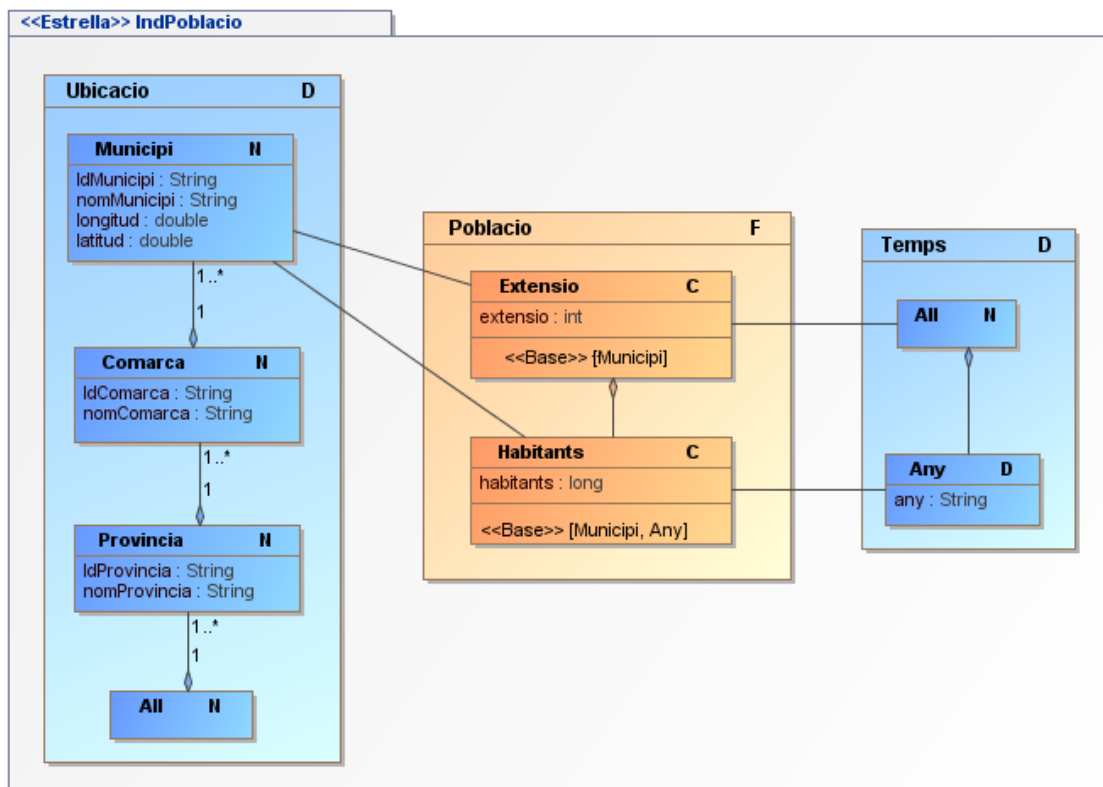


Figura 4. Esquema en estrella per dades Població (cub IndPoblacio)

Estrella: IndPoblacio

Taula de fet: Població

Granularitat: Municipi, Any

Mesures calculades:

- Densitat de població (habitants per Km2)
- % Població a la ubicació seleccionada (Província, Comarca o Municipi) vs. total
- TOTAL Habitants
- TOTAL Extensió

Observacions: es distingeixen dues Cel·les (IndHabitants i IndSuperfície) amb diferent granularitat (l'extensió no canvia en funció de l'any, no utilitza la dimensió de temps). Per simplificar l'aplicació, i atès que el nombre de cel·les resultant no és massa elevat, es duplicarà el valor de l'extensió de cada municipi per tots els anys.

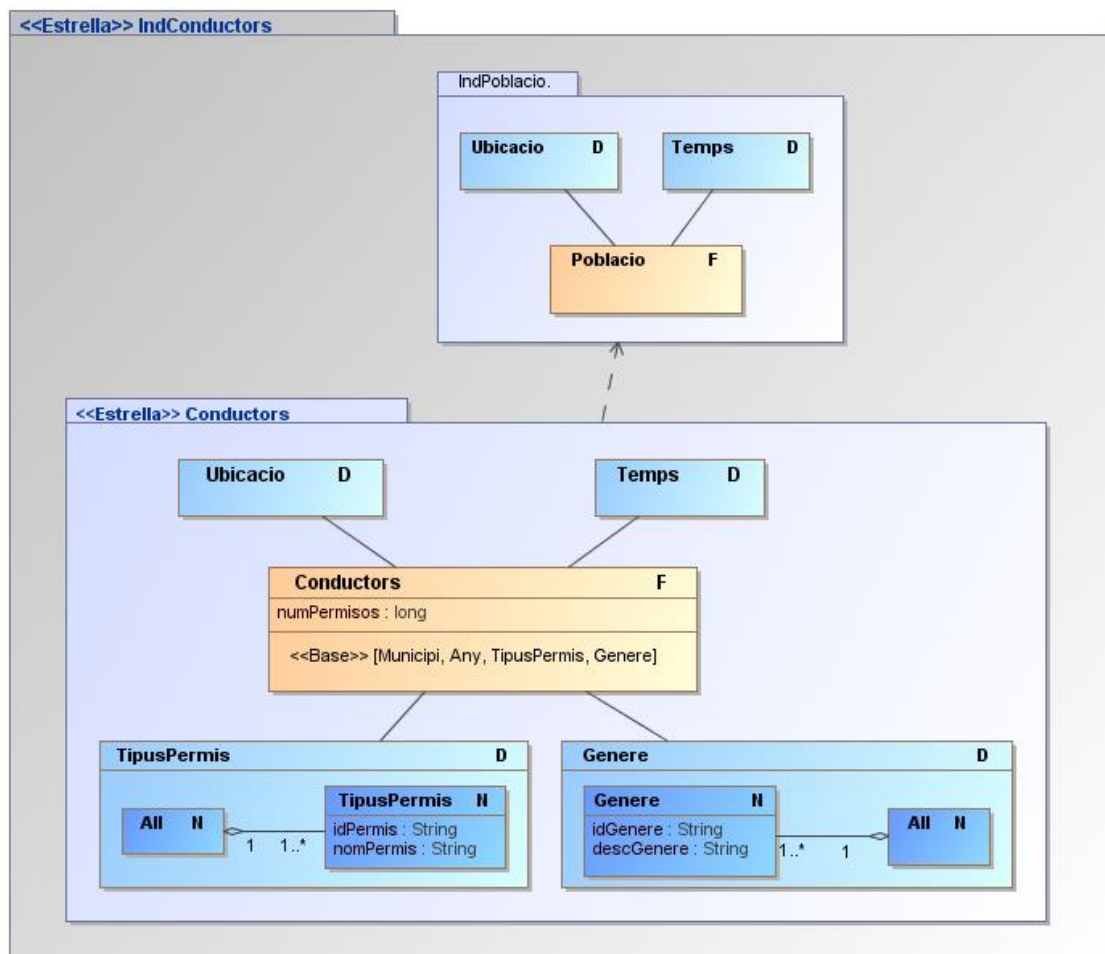


Figura 5. Esquema en estrella per dades Conductors (cub IndConductors)

Estrella: IndConductors

Taula de fet: Conductors.

Granularitat: Municipi, Any, Tipus de Permís (permís, llicència), Gènere (dona, home).

Mesures calculades (dependència d'estrella IndPoblacio):

- % Conductors del gènere seleccionat vs. total
- % Conductors amb permís del tipus seleccionat vs. total
- Conductors vs. Habitants per gènere
- TOTAL Conductors
- TOTAL Habitants
- TOTAL Extensió

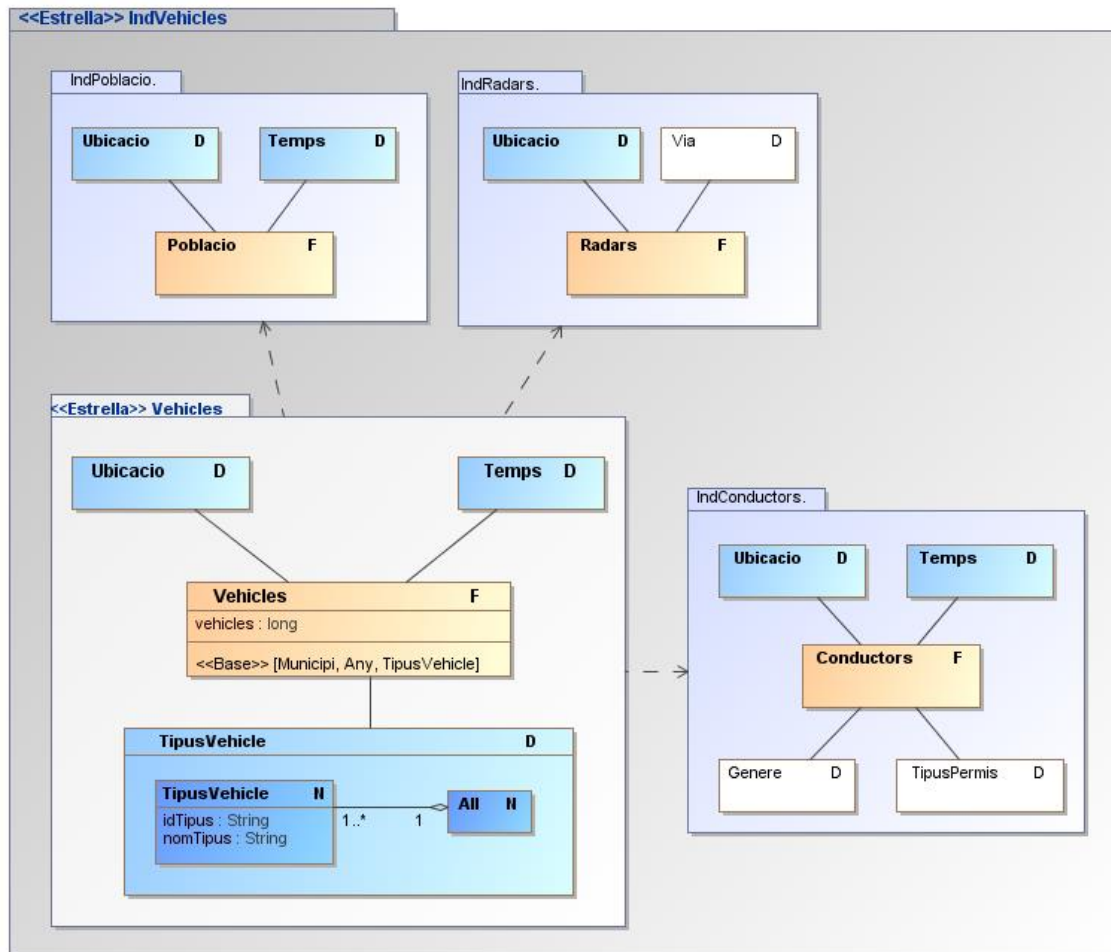


Figura 6. Esquema en estrella per dades Vehicles (cub IndVehicles)

Estrella: IndVehicles

Taula de fet: Vehicles

Granularitat: Municipi, Any, Tipus de Vehicle

Mesures calculades (dependència d'estrelles IndPoblacio, IndConductors i IndRadars):

- Vehicles per habitant
- Densitat de trànsit (vehicles per Km2)
- Vehicles per Radar
- % Vehicles a la ubicació seleccionada vs. Total
- % Vehicles del tipus seleccionat vs. total
- Radars per Km2
- Radars per Habitant
- Conductors per Radar
- TOTAL Vehicles
- TOTAL Habitants
- TOTAL Extensió
- TOTAL Radars

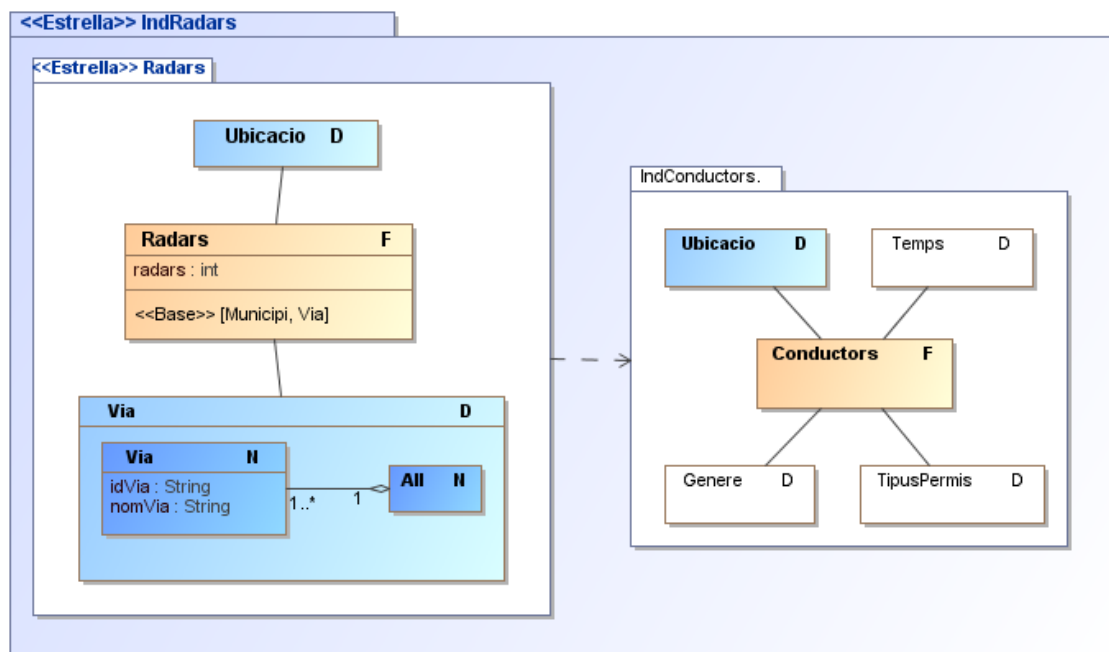


Figura 7. Esquema en estrella per dades Radars (cub IndRadars)

Estrella: IndRadars

Taula de fet: Radars, amb mesures relacionades amb els radars (nombre de radars).

Granularitat: Municipi, Via.

Mesures calculades (dependència d'estrella IndConductors):

- % Radars a la ubicació seleccionada vs. total
- Conductors per radar
- TOTAL Conductors
- TOTAL Radars

Comentaris sobre el model

Inicialment s'ha realitzat un model basat únicament en les taules de fet i les dimensions associades en funció de la granularitat requerida (una estrella simple).

A continuació, s'ha anat modificant el model en funció de les mesures calculades (càlculs sobre les mesures de les taules de fet) que es requerien pel projecte. Ha sigut llavors quan hem tingut la necessitat de dissenyar "superestrelles" (agrupació d'una o més estrelles simples).

En aquest últim cas, observem que algunes de les dimensions de les estrelles bàsiques (en fons blanc), no són utilitzades a les superestrelles generades.

3. Disseny

3.1. Arquitectura del projecte

3.1.1. Arquitectura del programari

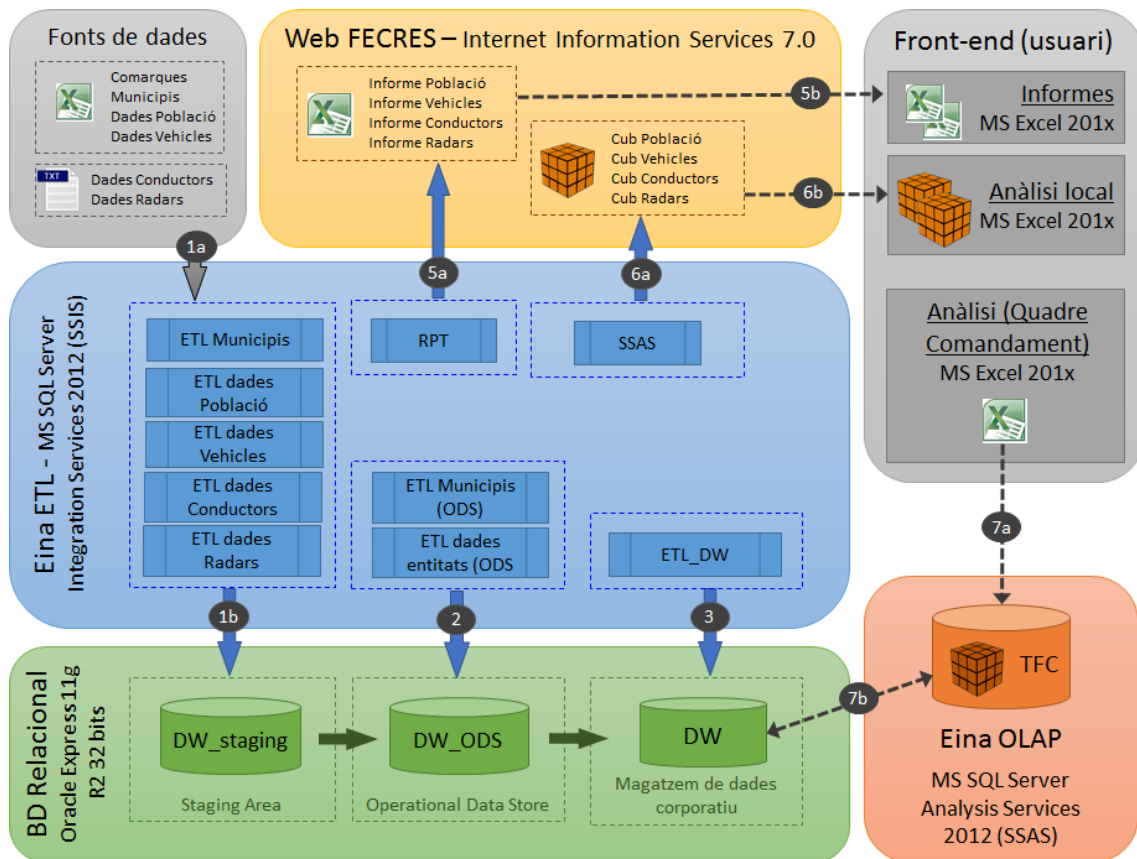


Figura 8. Diagrama arquitectura del programari

- 1) Càrrega de dades des de fonts de dades externes
 - a. Càrrega des dels arxius plans (XLS, XLSX, TXT) obtinguts des de les entitats col·laboradores (IDESCAT, DGT, Servei Català de Trànsit) o directament per FECRES.
 - b. Càrrega de dades a la *Staging Area* (DW_staging).
- 2) Càrrega de dades des de la *Staging Area* a l'*Operational Data Store* (DW_ODS).
- 3) Càrrega de dades des de l'*Operational Data Store* al Magatzem de Dades corporatiu (DW).
- 4) Generació d'informes.
 - a. Generació d'informes (format XLS) des de l'*Operational Data Store* (DW_ODS).
 - b. Publicació en la web corporativa de FECRES per a la seva descàrrega.
- 5) Generació de cubs *offline*.
 - a. Generació de cubs *offline* (format .CUB, gestionats des de MS Excel).
 - b. Publicació en la web corporativa de FECRES per a la seva descàrrega.

- 6) Informació per anàlisi al Quadre de Comandament (QC).
- El quadre de comandament s'ha elaborat en format *MS Excel 201x*, amb connexions cap al servidor OLAP de FECRES (és necessari disposar de connexió a Internet i el servidor OLAP ha d'estar funcionant).
 - El servidor OLAP consulta directament les dades proporcionades des del magatzem de dades corporatiu (DW), atès que el model OLAP utilitzat és ROLAP.

Justificació en la selecció del programari

Eina ETL – *MS Integration Services 2012 (SSIS)*

- Eina de reconeguda fiabilitat, que proporciona un ampli ventall de controls per realitzar quasi qualsevol tasca de càrrega.
- Eina ja instal·lada a l'entorn de desenvolupament / producció (màquina Amazon)
- El consultor de la UOC (estudiant TFC) té uns mínims coneixements de l'eina
- Entorn homogeni: el sistema operatiu i l'eina OLAP triada són del mateix fabricant

Eina OLAP – *MS Analysis Services 2012 (SSAS)*

- Eina de reconeguda fiabilitat i rendiment.
- Eina ja instal·lada a l'entorn de desenvolupament / producció (màquina Amazon)
- Permet el model d'emmagatzemament MOLAP (la informació s'emmagatzema en la base de dades multidimensional, en comptes de fer-ho en una BD relacional), més adequat en quant a rendiment. En el projecte que ens ocupa s'ha optat per una solució ROLAP atès que era un requisit de l'enunciat, però les agregacions s'han configurat per emmagatzematge MOLAP per guanyar en rendiment.
- Interfície d'usuari senzill i prou extens com per poder generar consultes MDX de forma visual i senzilla.

Eina d'Informes i Anàlisi – *MS Excel 201x (preferentment 2013, més estable)*

S'ha pres la decisió d'utilitzar les eines de Microsoft pels processos de càrrega (SSIS) i OLAP (SSAS) pels motius descrits amb anterioritat, i preferim mantenir l'homogeneïtat del sistema i apostar també per eines de la mateixa família per la realització d'informes.

En el plantejament inicial del projecte es va decidir utilitzar SSRS (*MS Reporting Services*) pels informes, però vàrem observar que la llicència del producte Microsoft instal·lada a la màquina d'Amazon no ens permetia afegir més components al sistema si no disposàvem del disc original. Davant aquesta situació, ens vàrem plantejar tres possibilitats:

- 1) Instal·lar una altra versió d'SSRS, a més de la base de dades a *MS SQL Server* necessària per allotjar la base de dades de *Reporting Services*.
 - CON: no teníem la seguretat de que la instal·lació d'una altra versió pogués afectar a la ja existent.
 - PRO: permet quadres de comandament bàsics i publicació automàtica/programada d'informes.

- 2) Instal·lar productes *open source*, com *Jasper Reports* o *Pentaho Reporting*.
- CON: manca de *know-how* d'aquests productes.
 - CON: excessiva complexitat en una gama de programari que, a més, si no es disposa de versions Enterprise, no permeten la realització de quadres de comandament amb les especificacions requerides pel projecte.
 - PRO: eines gratuïtes força esteses a nivell professional.
- 3) Utilitzar també *MS Excel* com a eina d'informes, ja que atesa l'escassa cadència d'actualització de la informació històrica del magatzem de dades (anual), no resulta excessivament necessari disposar d'una versió al servidor.
- CON: limitem l'explotació del sistema a un fabricant concret.
 - PRO: Integració avançada amb servidor *MS Analysis Services*
 - PRO: permet l'anàlisi de cubs sense connexió (vista multidimensional sense connexió al servidor).

Tanmateix, el perfil d'usuari consumidor d'informes i executor de l'anàlisi (analista de negoci), sovint pertanyent a l'alta direcció, necessita:

- Informació disponible *offline* (sense connexió). Accés durant viatges de negocis, visites a clients, etc.
- Independència del departament d'informàtica: necessita una eina d'ús extens a l'organització, fàcil d'utilitzar i que permeti organitzar la informació en funció de les necessitats (diferents "vistes" de dades).

, i aquestes dues característiques són proporcionades per l'eina seleccionada.

Base de Dades Relacional – Oracle Express 11g

- Eina de reconeguda fiabilitat i rendiment. Oracle és un referent en bases de dades orientades a l'empresa.
- Eina ja instal·lada a l'entorn de desenvolupament / producció (màquina Amazon).
- Addicionalment s'utilitzarà l'eina gratuïta SQL Developer, també proporcionada per Oracle. Es necessita la instal·lació de la màquina virtual Java per al seu funcionament.

Servidor Web – Internet Information Services 7.0

- IIS7 és un servidor web estable que manté l'homogeneïtat de l'entorn vist fins ara.
- S'instal·la com a component disponible al Servidor proporcionat per la màquina virtual Amazon (Servidor d'aplicacions).

Observacions addicionals

- Tot el programari especificat (excepte el *Frontend*, que correspon a la màquina de l'usuari) està instal·lat al servidor de FECRES.
- El programari de Microsoft (SSIS, SSAS) correspon a la versió *SQL Server 2012 SP1*.
- Es disposa de la suite *Microsoft Visual Studio Express 2010* pel disseny dels processos ETL (SSIS), informes (SSRS) i modelització de les estructures multidimensionals (SSAS).
- Addicional, com a eina de suport per al consultor UOC, s'utilitzarà l'eina *Oracle SQL Developer* per al disseny

3.1.2. Arquitectura del maquinari

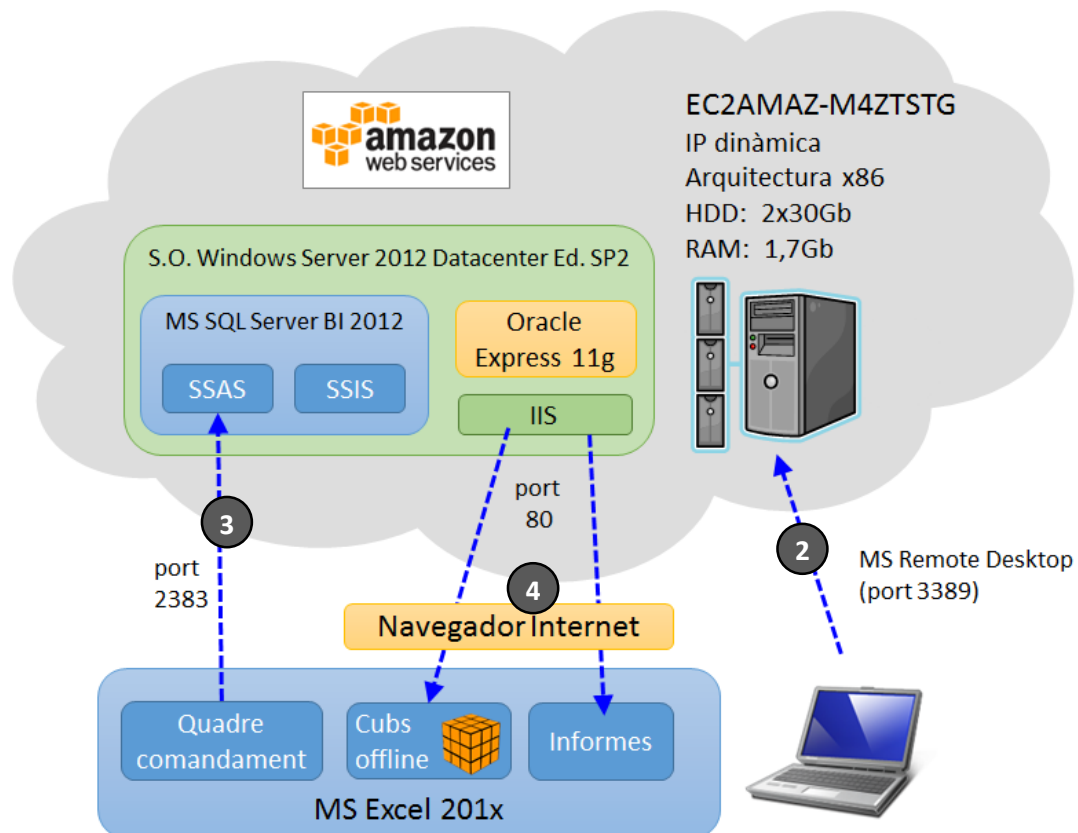


Figura 9. Diagrama d'arquitectura del maquinari

El servidor de la solució és en realitat una màquina virtual ubicada a *Amazon Web Services*, una entorn *cloud computing* que ofereix IAAS (*Infrastructure As A Service*), és a dir, una infraestructura de maquinari i programari adaptada a les nostres necessitats.

- 1) Obtenció d'una direcció IP per la màquina virtual Amazon. En el projecte que ens ocupa aquesta funció l'ofereix el campus de la UOC.

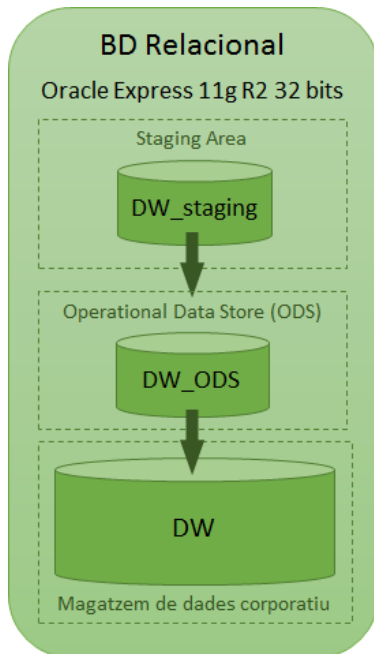
Rol administrador / gestor BD:

- 2) El client inicia la connexió amb la IP obtinguda anteriorment amb l'aplicació MS Remote Desktop o similar. Per facilitar i simplificar el treball amb aquesta màquina, s'ha establert un sistema d'assignació de nom fixe (AWSTFC) a la IP activa mitjançant l'addició d'una entrada a l'arxiu `C:\Windows\System32\drivers\etc\hosts`, tal com s'explica al document adjunt corresponent ¹.

Rol analista de negoci:

- 3) Des del Quadre de Comandament (arxiu `C:\FECRES\QC.xlsx`) l'analista inicia una connexió online amb la màquina per tal de treballar amb dades actualitzades.
- 4) Des del navegador, s'accedeix a la pàgina privada de FECRES: <http://AWSTFC/FECRES/> i es poden descarregar tant els cubs *offline* com els informes.

3.2. Disseny de la Base de Dades relacional



- Decidim la creació d'una *Staging Area*, ja que es guarda la informació amb la mateixa estructura que en les fonts de dades originals (excepte en casos d'informació redundant), no normalitzada de tal manera que actua com una còpia de recolzament d'aquestes dades, que a més no han de ser accedides per usuaris consumidors de dades.

- En el primer plantejament que es va fer d'aquest projecte, no s'havia considerat la possibilitat de construir un *Operational Data Store*, però després d'analitzar-lo amb més deteniment arribem a la conclusió de què és recomanable disposar d'un magatzem intermediari amb dades normalitzades no accessibles des del sistema transaccional, però sí per algunes aplicacions que requereixen d'informació normalitzada i consultable (ex: informes).

Figura 10. Base de Dades Relacional

3.2.1. Diagrama d'Entitat-Relació (DER)

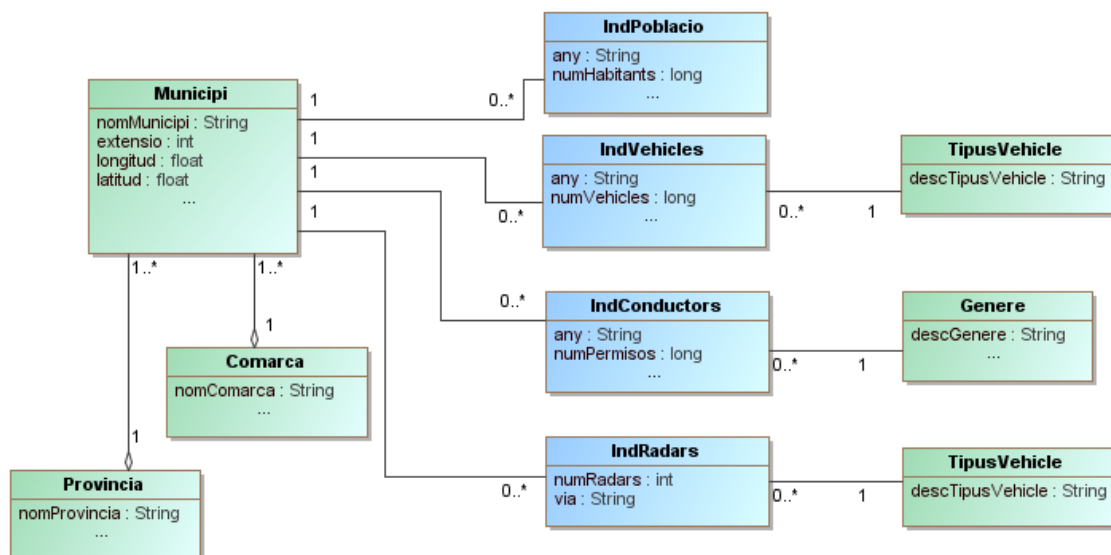


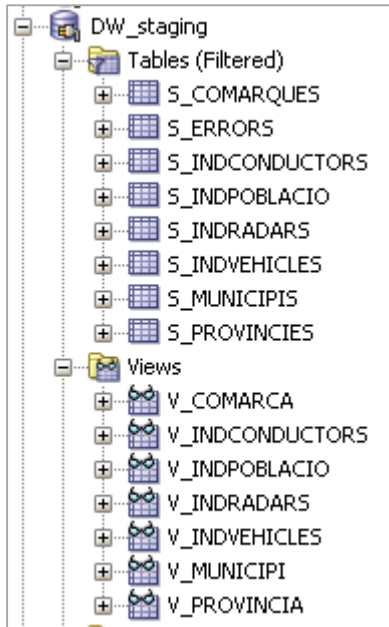
Figura 11. Diagrama d'entitat-relació (DER)

Aquest diagrama està basat en l'ODS, atès que és l'única BD normalitzada.

Les entitats amb fons verd són gestionades directament per FECRES, mentre que les de fons blau depenen completament de les dades proporcionades per les entitats col·laboradores.

3.2.2. Descripció detallada

DW_Staging



Usuari (*schema*) associat: **dw_stg**

En aquesta base de dades no es defineixen claus, atès que es tracta d'una còpia de les dades transferides des dels orígens de dades externs, amb totes les inconsistències corresponents (incloent-hi duplicitats), que en aquest nivell no controlem.

Tanmateix, no té sentit definir restriccions com ara claus foranes, ja que totes les taules són inicialitzades per complet cada vegada que es fa la càrrega de dades.

Es defineixen vistes que transformen la informació per adaptar-la a les estructures normalitzades de la base de dades DW_ODS, i d'aquesta manera simplificar la tasca de l'eina ETL que utilitzarà aquest model com a referència d'origen.

Figura 12. Objectes DW_staging

S_Municipis: municipis de Catalunya				
Clau	Columna	Descripció	Tipus	Observacions
	codiComarca	Codi comarca	varchar(2)	
	codiMunicipi	Codi municipi	varchar(6)	
	nomMunicipi	Nom municipi	varchar(45)	

S_Comarques: comarques de Catalunya				
Clau	Columna	Descripció	Tipus	Observacions
	codiComarca	Codi comarca	varchar(2)	
	nomComarca	Nom comarca	varchar(20)	

S_Provincies: províncies de Catalunya				
Clau	Columna	Descripció	Tipus	Observacions
	codiProvíncia	Codi província	varchar(2)	
	nomProvíncia	Nom província	varchar(20)	

S_Errors: registre d'errors dels processos de càrrega				
Clau	Columna	Descripció	Tipus	Observacions
	taula	Nom de la taula origen	varchar(50)	
	idError	Codi error eina ETL	number(38)	
	idColumna	Codi columna	number(38)	
	dades	Dades del registre amb error	clob	texte

S_IndPoblacio: dades Població				
Clau	Columna	Descripció	Tipus	Observacions
	codMunicipi	Nom municipi	varchar(45)	
	nomINE	Codi INE	varchar(5)	
	pob2012	Població 2012	number(7,0)	
	pob2011	Població 2011	number(7,0)	
	pob2010	Població 2010	number(7,0)	
	pob2009	Població 2009	number(7,0)	
	pob2008	Població 2008	number(7,0)	
	pob2007	Població 2007	number(7,0)	
	extensio	Extensió (Km2)	number(3,0)	

S_IndConductors: dades Conductors				
Clau	Columna	Descripció	Tipus	Observacions
	anny	Any	varchar(4)	
	codMunicipi	Nom província + municipi	varchar(65)	
	permisosDona	Nombre permisos dona	number(6,0)	
	permisosHome	Nombre permisos home	number(6,0)	
	llicenciesDona	Nombre llicències dona	number(5,0)	
	llicenciesHome	Nombre llicències home	number(5,0)	

S_IndVehicles: dades Vehicles				
Clau	Columna	Descripció	Tipus	Observacions
	nomMunicipi	Nom municipi	varchar(45)	
	codINE	Nom província + municipi	varchar(65)	
	automòbils_2012	Nombre automòbils any 2012	number(6,0)	
	automòbils_2011	Nombre automòbils any 2011	number(6,0)	
	automòbils_2010	Nombre automòbils any 2010	number(6,0)	
	automòbils_2009	Nombre automòbils any 2009	number(6,0)	
	automòbils_2008	Nombre automòbils any 2008	number(6,0)	
	automòbils_2007	Nombre automòbils any 2007	number(6,0)	
	camions_furgonetes_2012	N.camions i furgonetes any 2012	number(6,0)	
	camions_furgonetes_2011	N.camions i furgonetes any 2011	number(6,0)	
	camions_furgonetes_2010	N.camions i furgonetes any 2010	number(6,0)	
	camions_furgonetes_2009	N.camions i furgonetes any 2009	number(6,0)	
	camions_furgonetes_2008	N.camions i furgonetes any 2008	number(6,0)	
	camions_furgonetes_2007	N.camions i furgonetes any 2007	number(6,0)	
	altres_2012	Nombre altres vehicles any 2012	number(6,0)	
	altres_2011	Nombre altres vehicles any 2011	number(6,0)	
	altres_2010	Nombre altres vehicles any 2010	number(6,0)	
	altres_2009	Nombre altres vehicles any 2009	number(6,0)	
	altres_2008	Nombre altres vehicles any 2008	number(6,0)	
	altres_2007	Nombre altres vehicles any 2007	number(6,0)	

S_Radars: dades Radars				
Clau	Columna	Descripció	Tipus	Observacions
	via	Via	varchar(10)	
	nomMunicipi	Nom municipi	varchar(45)	
	nomComarca	Nom comarca	varchar(20)	
	nomProvíncia	Nom província	varchar(10)	

DW_ODS

Usuari (schema) associat: **dw_ods**



Figura 13. Objectes DW_ODS

Aquesta base de dades es troba completament normalitzada, i les dades que conté provenen de la *Staging Area*, prèviament extretes i transformades pels processos ETL corresponents.

Tanmateix, s'han dissenyat dos tipus de vistes, per tal de facilitar processos posteriors:

- Vistes per càrrega del Magatzem de dades corporatiu: es corresponen amb les vistes anomenades com *v_dim_xxx* (dimensions) i *v_fact_xxx* (taules de fet), on trobem la informació organitzada per tema (optimitzada pel seu tractament multidimensional).
- Vistes per generació d'informes: es corresponen amb els informes en format XLS generats per l'eina ETL, i que posteriorment seran descarregat per l'usuari corresponent.

Municipi: municipis de Catalunya				
Clau	Columna	Descripció	Tipus	Observacions
PK	idMunicipi	Codi municipi	varchar(6)	
FK	idComarca	Codi comarca	varchar(2)	
FK	idProvíncia	Codi província	varchar(2)	
	nomMunicipi	Nom municipi	varchar(45)	
	extensio	Extensió municipi (Km2)	number(3,0)	
	longitud	Coordinada longitud	varchar(9)	
	latitud	Coordinada latitud	varchar(9)	

Municipi (idMunicipi, idComarca, idProvíncia, nomMunicipi, extensió, longitud, latitud)
 {idComarca} clau forana cap a Comarca (idComarca)
 {idProvíncia} clau forana cap a Província (idProvíncia)

Comarca: comarques de Catalunya				
Clau	Columna	Descripció	Tipus	Observacions
PK	idComarca	Codi comarca	varchar(2)	
	nomComarca	Nom comarca	varchar(20)	

Comarca (idComarca, nomComarca)

Província: províncies de Catalunya				
Clau	Columna	Descripció	Tipus	Observacions
PK	idProvíncia	Codi província	varchar(2)	
	nomProvíncia	Nom província	varchar(10)	

Província (idProvíncia, nomProvíncia)

IndPoblacio: dades de població				
Clau	Columna	Descripció	Tipus	Observacions
PK	anny	Any	varchar(4)	
PK, FK	idMunicipi	Codi Municipi	varchar(6)	
	numHabitants	Nombre d'habitants	number(7,0)	

IndPoblacio (anny, idMunicipi, numHabitants)
{idMunicipi} clau forana cap a Municipi (idMunicipi)

IndVehicles: dades del parc de vehicles				
Clau	Columna	Descripció	Tipus	Observacions
PK	anny	Any	varchar(4)	
PK, FK1	idMunicipi	Codi Municipi	varchar(6)	
PK, FK2	tipusVehicle	Tipus de Vehicle	varchar(2)	
	numVehicles	Nombre de vehicles	number(7,0)	

IndVehicles (anny, idMunicipi, tipusVehicle, numVehicles)
{idMunicipi} clau forana cap a Municipi (idMunicipi)
{tipusVehicle} clau forana cap a TipusVehicle (tipusVehicle)

IndConductors: dades cens de conductors (permisos/llicències)				
Clau	Columna	Descripció	Tipus	Observacions
PK	anny	Any	varchar(4)	
PK, FK	idMunicipi	Codi Municipi	varchar(6)	
PK	idTipusPermis	Tipus de Permis	varchar(2)	
PK	idGenere	Gènere (home, dona)	char(1)	
	numPermisos	Nombre de permisos	umber(6,0)	

IndConductors (anny, idMunicipi, idTipusPermis, tipusGenere, numPermisos)
{idMunicipi} clau forana cap a Municipi (idMunicipi)
{idTipusPermis} clau forana cap a TipusPermis (idTipusPermis)
{idGenere} clau forana cap a Genere (idGenere)

IndRadars: dades de radars fixes				
Clau	Columna	Descripció	Tipus	Observacions
PK, FK	idMunicipi	Codi municipi	varchar(6)	
PK	via	Via	varchar(10)	
	numRadars	Nombre de radars	number(2,0)	

IndRadars (idMunicipi, via, numRadars)
{idMunicipi} clau forana cap a Municipi (idMunicipi)

TipusVehicle: relació de tipus de vehicle				
Clau	Columna	Descripció	Tipus	Observacions
PK	tipusVehicle	ID tipus vehicle	varchar(2)	
	descTipusVehicle	Descripció	varchar(25)	

TipusVehicle (tipusVehicle, descTipusVehicle)

TipusPermis: relació de tipus de permís de conducció				
Clau	Columna	Descripció	Tipus	Observacions
PK	tipusPermis	ID tipus permís	varchar(2)	
	descTipusPermis	Descripció	varchar(10)	

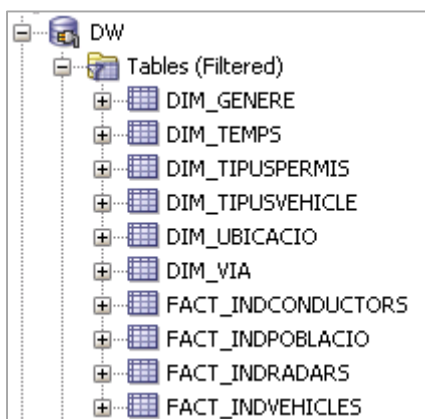
TipusPermis (tipusPermis, descTipusPermis)

Genere: relació gèneres				
Clau	Columna	Descripció	Tipus	Observacions
PK	idGenere	ID gènere	char(1)	
	descGenere	Descripció	varchar(10)	

Genere (idGenere, descGenere)

DW

Usuari (*schema*) associat: **dw**



Les taules d'aquesta base de dades tenen l'estructura necessària per treballar amb les estructures multidimensionals que es veuran a continuació.

Figura 14. Objectes DW

3.3. Disseny multidimensional detallat

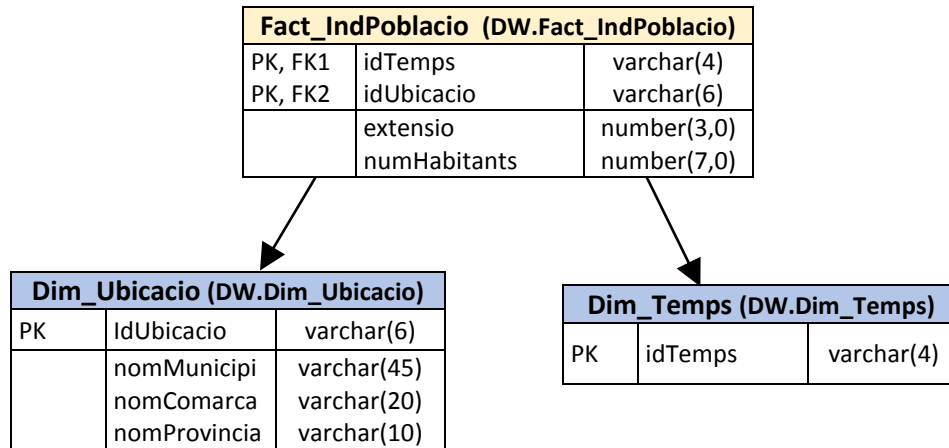


Figura 15. Estrella detallada dades de Població

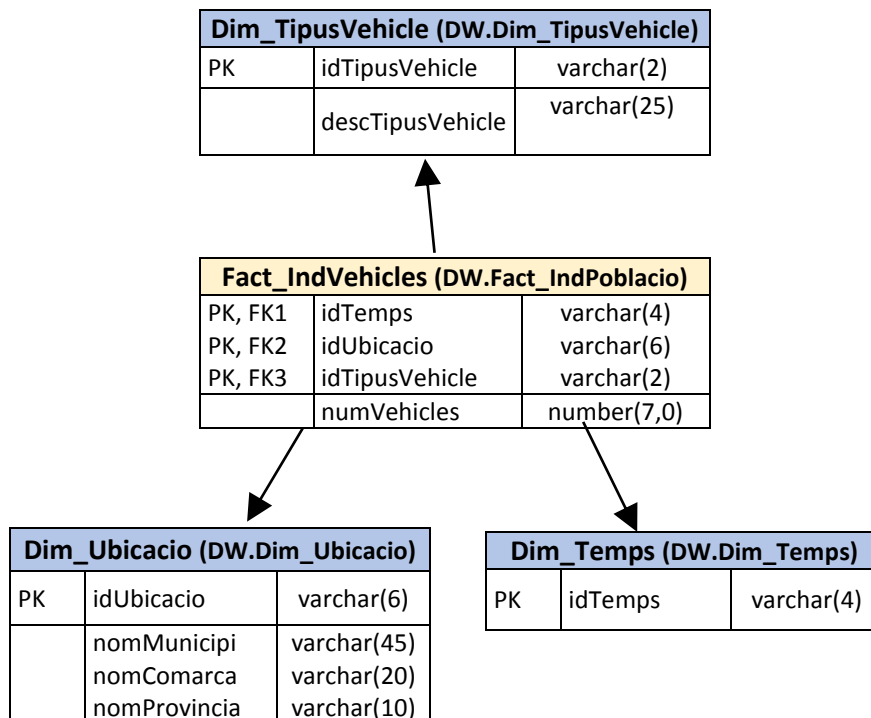


Figura 16. Estrella detallada dades de Vehicles

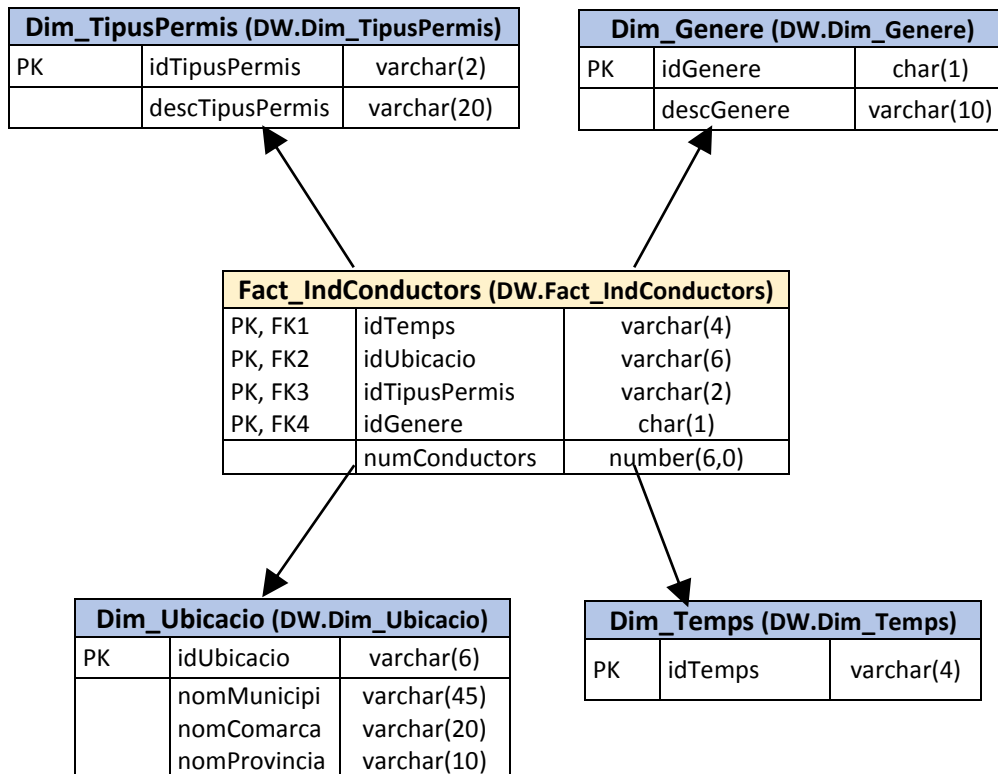


Figura 17. Estrella detallada dades de Conductors

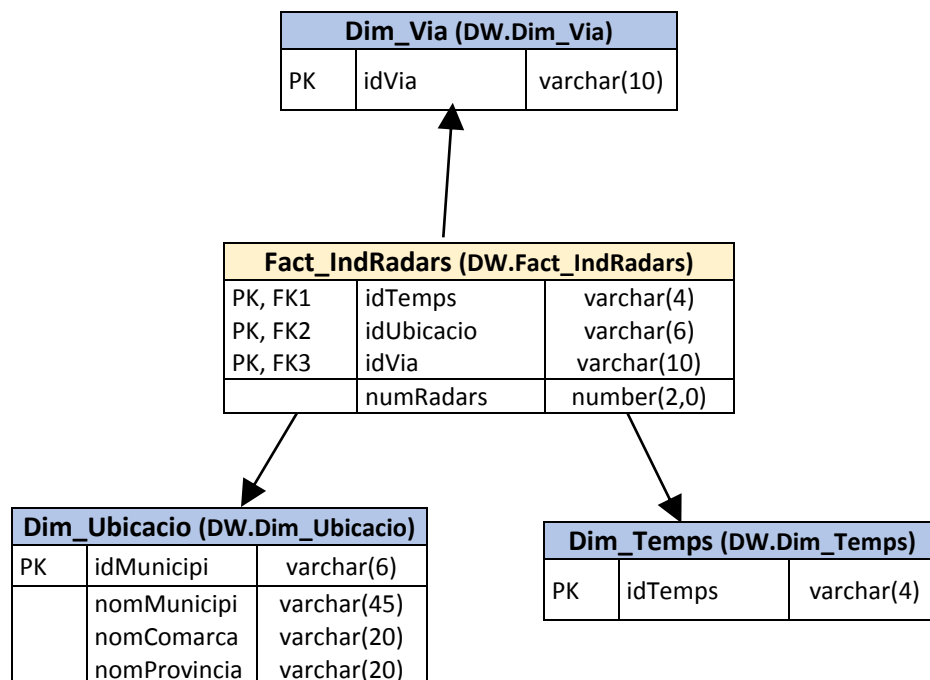


Figura 18. Estrella detallada dades de Radars

3.4. Disseny dels processos de càrrega

Tal com es comenta a l'apartat 3.1.1. ([Arquitectura del programari](#)), l'eina utilitzada per implementar els processos de càrrega serà Microsoft Integration Services (SSIS).

El procés de càrrega com a tal s'ha organitzat per tipus d'informació en paquets de càrrega (SSIS Package), que pot involucrar més d'una càrrega en el mateix procés.

Es pot trobar informació més detallada sobre els processos de càrrega en el document adjunt corresponent ¹.

Mètode de desenvolupament

En tots els processos de càrrega, dels quals s'afegeix un exemple a la figura de sota), segueixen un diagrama de flux i compleixen les següents premisses:

- Gestió d'errors: els registres d'error es desvien del flux normal del procés (línies vermelles de la figura a sota) i es grava informació descriptiva del registre, així com la taula d'on procedeix en una taula d'errors (`DW_staging.S_ERRORS`).
- Es filtren les files amb informació en blanc (capçaleres/peus dels arxius externs, per norma general) i també es desvien a l'arxiu d'error.
- Tots els elements del flux de càrrega són descriptius de la tasca que realitzen (s'identifiquen fàcilment).

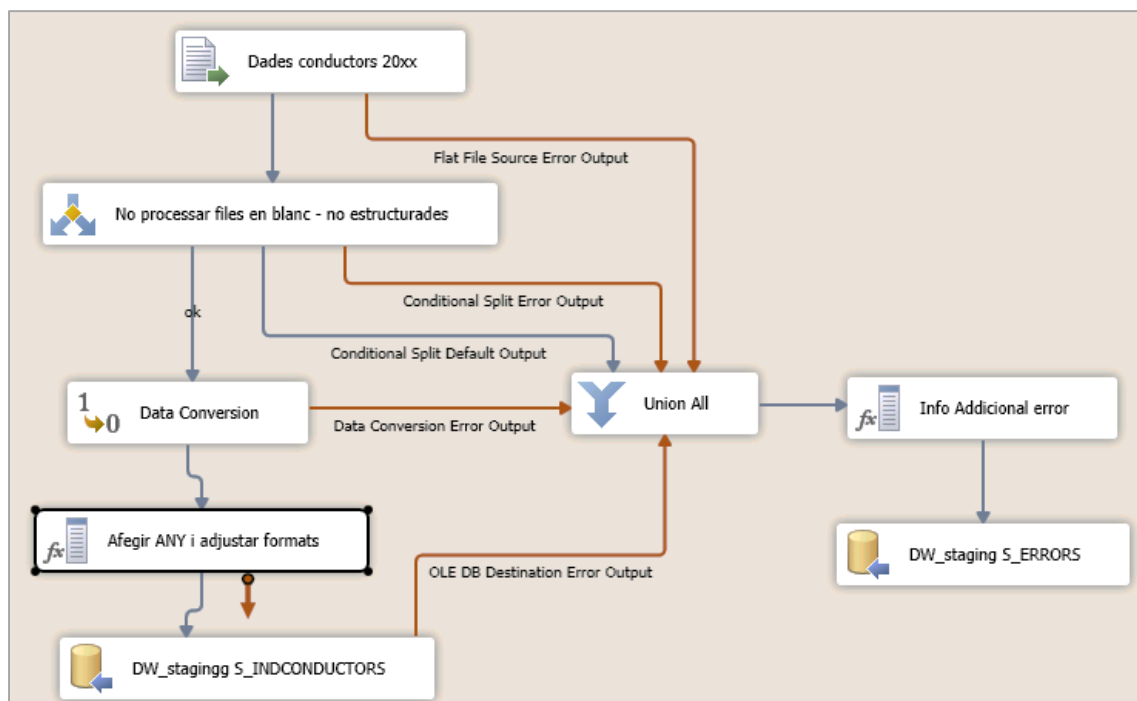


Figura 19. Exemple de diagrama de flux SSIS

1. Veure **Annex A**. Documents adjunts a la memòria (Processos de càrrega).

Processos de càrrega principals

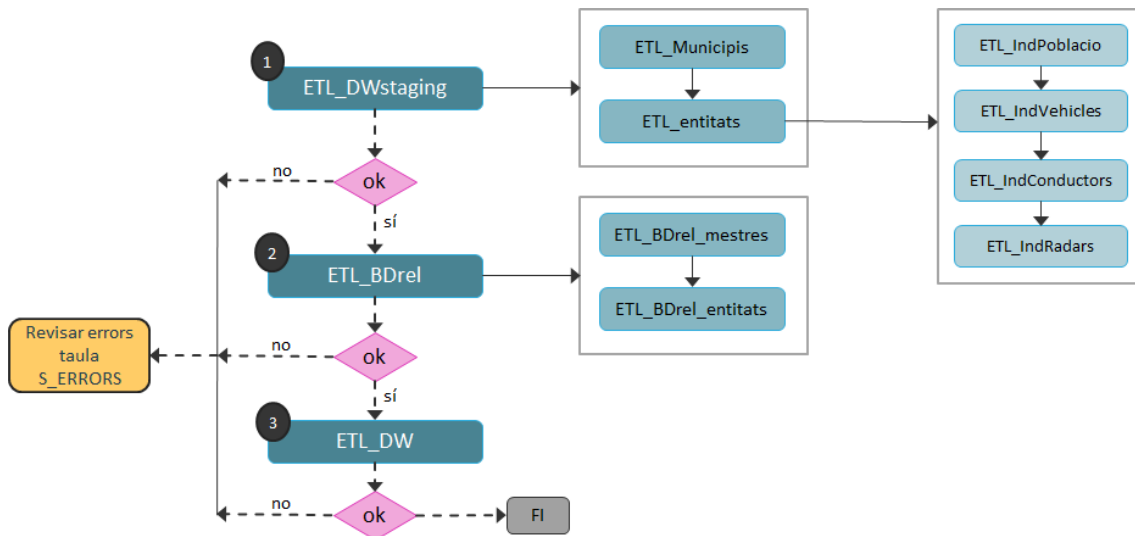


Figura 20. Diagrama de processos de càrrega (paquets SSIS) de la *Staging Area*

1) **ETL_DWstaging**: càrrega de la *Staging Area* des de fonts de dades externes

- a. **ETL_Municipis**: Dades de Municipis des d'arxius en format XLS/XML descarregats per FECRES al web de l'IDESCAT. No formava part de les especificacions, però ho considerem necessari com a referència per garantir la integritat de les dades, atès que no totes les estadístiques aportades per les entitats inclouen dades de tots els municipis ni la relació amb les comarques/províncies que són necessàries pel nivell d'agregació que es sol·licita a l'enunciat.
- b. **ETL_Entitats**: Dades enviades per les entitats col·laboradores (DGT, IDESCAT, SCT) amb la informació que formarà part del magatzem de dades per la seva anàlisi. Es tracta d'arxius plans en format XLS i TXT.

2) **ETL_BDrel**: càrrega de l'ODS des de la *Staging Area*

- a. **ETL_BDrel_mestres**: Càrrega de dades de municipis considerades dades "mestres", ja que la informació aportada per les entitats es recollirà en elles per organitzar-se adequadament. Provenen del procés de càrrega anterior **ETL_Municipis**.
- b. **ETL_BDrel_entitats**: Càrrega dades aportades per les entitats. Provenen del procés de càrrega anterior **ETL_entitats**.

3) **ETL_DW**: Càrrega del magatzem de dades corporatiu des de l'ODS

Processos de càrrega per generar cub sense connexió des del magatzem de dades

SSAS: En aquest procés:

- Es refresquen les dades del magatzem de dades des de la BD relacional: tenint en compte que la informació que es carrega s'actualitza anualment, no veiem la necessitat de programar una actualització automàtica, creiem que és més adequat efectuar les actualitzacions sota demanda.
- Es generen els cubs sense connexió (*offline cubes*) que poden ser oberts per Microsoft Excel (triada com a eina d'anàlisi) i no requereixen estar connectats amb el servidor. Els cubs sense connexió permeten la vista multidimensional de les dades mitjançant l'ús de taules dinàmiques (en el cas que ens ocupa, *MS Excel*).

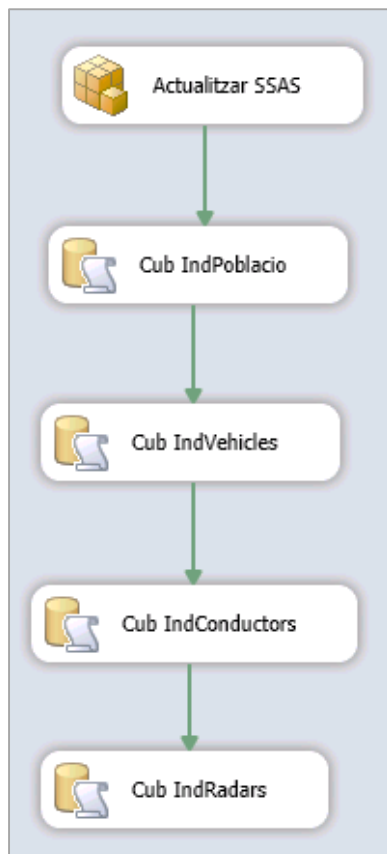


Figura 21. Procés de càrrega SSAS

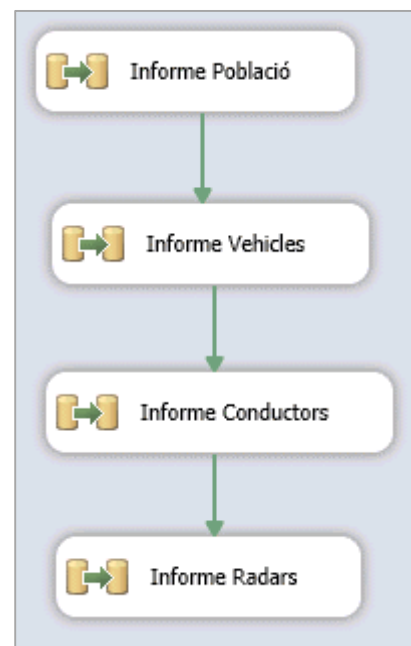


Figura 22. Procés de càrrega RPT

Processos de càrrega per informes

RPT: En aquest procés es generen arxius MS Excel (*RptPoblacio.xls*, *RptVehicles.xls*, *RptConductors.xls* i *RptRadars.xls*) amb les dades de l'informe corresponent. Els arxius generats seran referenciats posteriorment a l'arxiu base d'informes que té el client (*Informes.xlsx*), on es vinculen aquests arxius per unificar la seva aparença i facilitar la seva navegació mitjançant l'aplicació de filtres

3.4.1. Processos principals: Fonts de dades externes → *Staging Area*

Fulla de treball de les tables de la *staging area*¹ (DW_staging)

Paquet SSIS (D:\TFC\fonts_dades)	Taula (BD: DW_staging)	Freq.	Num. files	Tam. Fila bytes	Tam. Taula Kb	Mèt, Act.	Origen de dades
ETL_Municipis	S_Municipis	O/D	947	53	50,2	T/R	IDESCAT_municipis.xls
	S_Comarques		41	22	0,9	T/R	IDESCAT_comarques.xls
	S_Provincies		4	12	0,05	T/R	IDESCAT_provincies.xls
ETL_IndPoblacio	S_IndPoblacio	anual	466	96	44,7	T/R	Dades_municipis.xls
ETL_IndVehicles	S_IndVehicles	anual	466	212	98,8	T/R	Dades_vehicles.xls
ETL_IndConductors	S_IndConductors	anual	950	87	82,6	A	Dades_conductors2007.txt
			950				Dades_conductors2008.txt
			950				Dades_conductors2009.txt
			950				Dades_conductors2010.txt
			950				Dades_conductors2011.txt
ETL_IndRadars	S_IndRadars	O/D	217	85	18,4	T/R	Radars_SCT.txt

Taula 10. Staging tables volumètric worksheet

Freq.: freqüència amb que es produeix la càrrega de dades (o execució del procés ETL).

- **O/D:** *on demand*. La càrrega es produeix com a resposta a una petició del del sistema.

Mèt.Act.: mètode d'actualització de la taula (càrregues posteriors a la posada en marxa).

- **T/R:** *truncate and reload*. La taula receptora de la informació es buida completament i es tornen a carregar tots els continguts des del fitxer extern.
- **A:** *append*. La informació del fitxer extern s'afegeix a l'existent per l'any que s'està processant, esborrant prèviament les dades existents de l'any tractat.
- **U:** *Update*. La informació del fitxer extern es localitza a la taula corresponent per dada clau i s'actualitza la seva informació.

Consideracions

- A tots els processos de càrrega s'haurà de tenir en compte que per garantir la normalització de les taules de la base de dades relacional, en el moment de transferir la informació des de la taula de la *staging area*, s'hauran d'incloure consultes per resoldre els codis d'alguns atributs, atès que no sempre els orígens de dades inclouen aquesta informació (ex: les dades dels permisos dels conductors no inclouen el codi de municipi, sinó el seu nom en el format "Província – Municipi").
- S'han de tenir en compte les observacions efectuades a l'apartat 2.1.4 d'aquest document ([Orígens de dades](#)) a cadascun dels fitxers (formats, files a ignorar, etc.).
- Els arxius origen de dades es troben a la carpeta D:\TFC\fonts_dades\ del servidor de FECRES.
- Per tots els processos de càrrega especificats en aquest document, es poden trobar els detalls al document adjunt corresponent².

1. Basat en la proposta de R.KIMBALL, *The Data Warehouse ETL Toolkit*, pàg.33, figura 2.1: Staging tables volumetric worksheet.

2. Veure **Annex A**. Documents adjunts a la memòria (processos de càrrega).

ETL_Municipis

- **Descripció:** càrrega de dades relatives a municipis, comarques i províncies.
- **Observacions:** la càrrega de les coordenades dels municipis és opcional, només necessària en cas d'habilitar la geo-localització al projecte.
- **Procés:**
 - 1) Netejar la taula d'errors
 - 2) Si existeixen els arxius obtinguts per FECRES a la web de l'IDESCAT, continuar.
 - 3) Netejar taula de províncies (DW_staging.S_PROVINCIES).
 - 4) Carregar dades províncies des de l'arxiu IDESCAT_províncies.xls a la taula DW_staging.S_PROVINCIES.
 - 5) Netejar taula de comarques (DW_staging.S_COMARQUES).
 - 6) Carregar dades comarques des de l'arxiu IDESCAT_comarques.xls a la taula DW_staging.S_COMARQUES.
 - 7) Netejar taula de municipis (DW_staging.S_MUNICIPIS).
 - 8) Carregar dades comarques des de l'arxiu IDESCAT_municipis.xls a la taula DW_staging.S_MUNICIPIS.

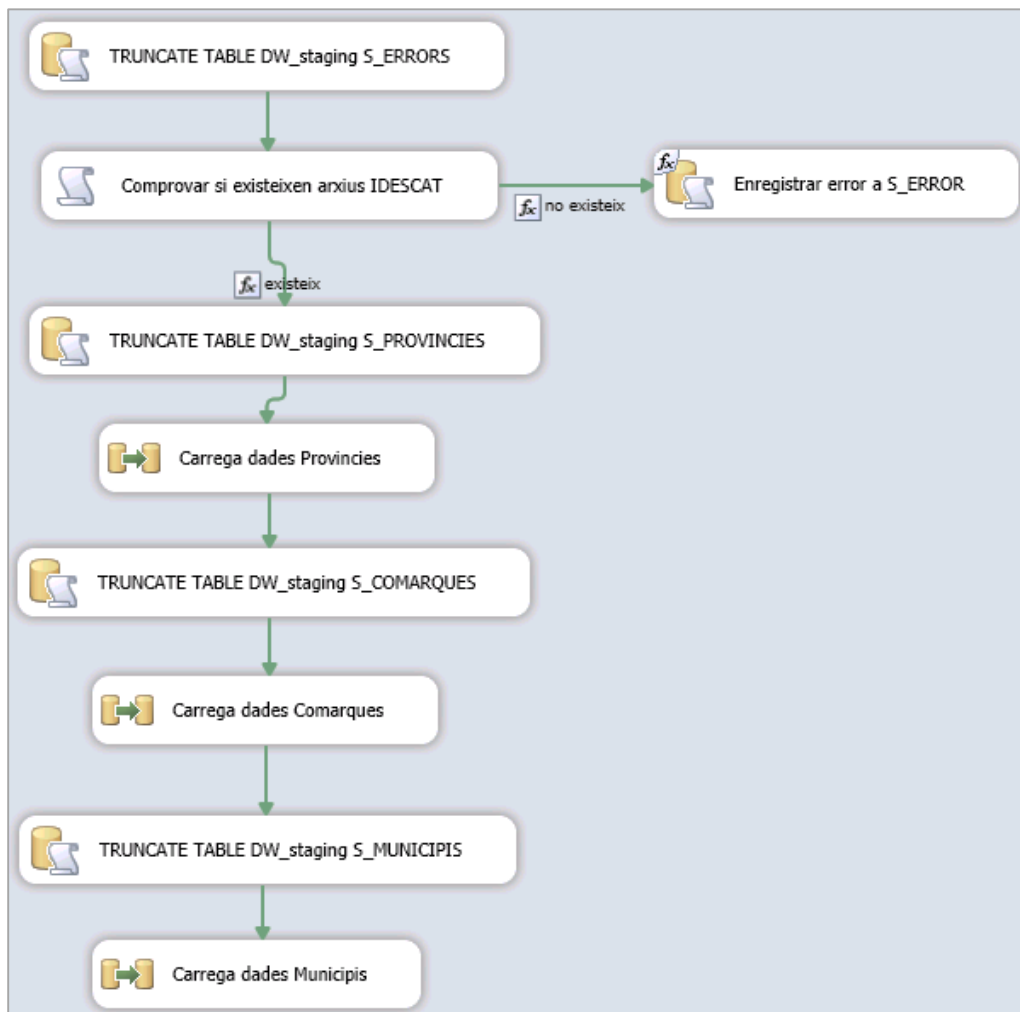


Figura 23. Procés de càrrega base per dades Ubicació: Província, Comarca i Municipi (ETL_Municipis)

ETL_Entitats

- **Descripció:** càrrega de dades relatives a municipis, comarques i províncies.
- **Observacions:** el format de l'arxiu de l'any 2011 es diferent a la resta (control·lat al procés ETL_IndConductors), i no s'han proporcionat les dades de l'any 2012..
- **Procés:**
 - 1) Netejar la taula d'errors per processos relacionats amb aquest paquet.
 - 2) Si existeix l'arxiu Dades_municipis.xls proporcionat per l'IDESCAT, continuar.
 - 3) Executar càrrega de dades de Població (procés **ETL_IndPoblacio**).
 - 4) Si existeix l'arxiu Dades_vehicles.xls proporcionat per l'IDESCAT, continuar; en cas contrari enregistrar l'error i continuar.
 - 5) Executar càrrega de dades de Vehicles (procés **ETL_IndVehicles**).
 - 6) Per cadascun dels arxius proporcionats per la DGT (anys 2007 al 2012):
 - Si existeix l'arxiu Dades_conductors20XX.txt, continuar; en cas contrari enregistrar l'error.
 - Executar càrrega de dades de conductors per l'any 20xx (procés **ETL_IndConductors**).
 - 7) Si existeix l'arxiu Radars_SCT.txt proporcionat per l'SCT, continuar; en cas contrari enregistrar l'error i continuar.
 - 8) Executar càrrega de dades de Radars (procés **ETL_IndRadars**).

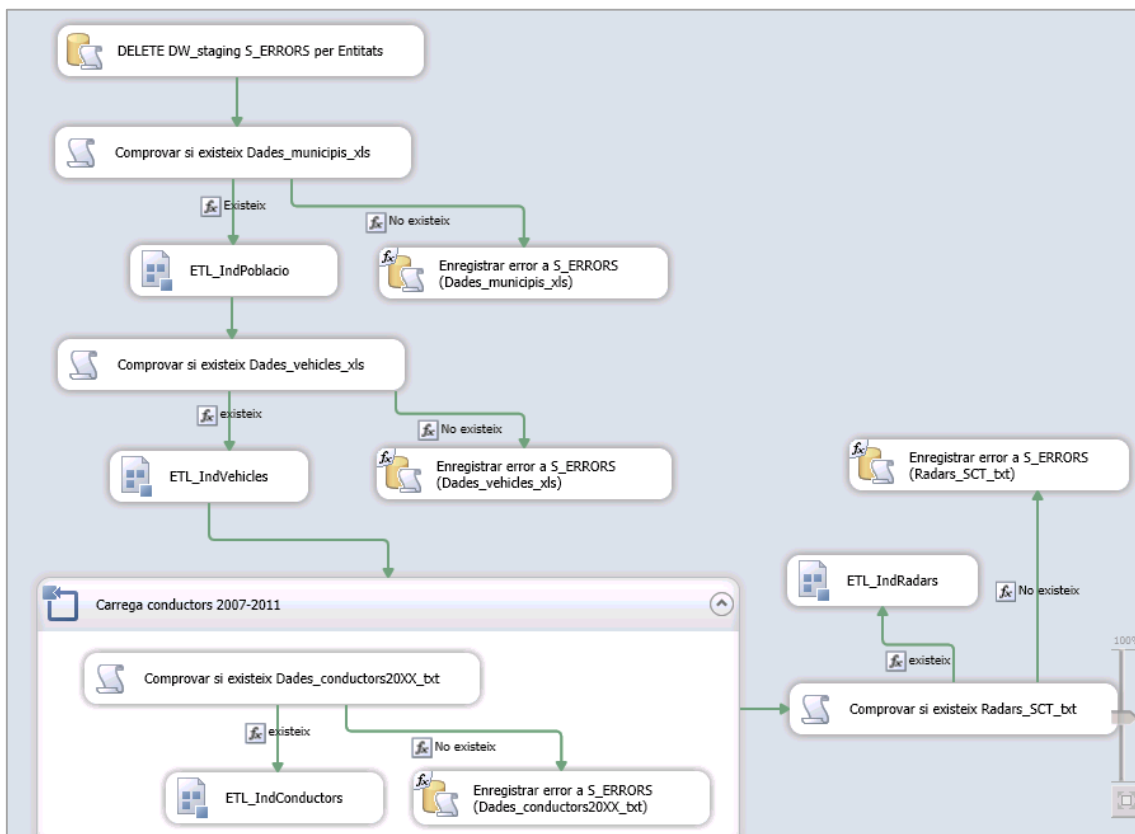


Figura 24. Procés de càrrega base per dades a analitzar (ETL_Entitats)

3.4.2. Processos principals: *Staging Area* → ODS

Els processos de càrrega es porten a terme en tot moment en l'àmbit de la base de dades relacional (no hi ha fonts d'origen extern). S'utilitza l'eina ETL per comoditat a l'hora d'unificar processos i per la capacitat d'utilitzar diagrames de flux que permeten controlar visualment les tasques efectuades.

ETL_BDrel_mestres

- **Descripció:** càrrega de dades relatives a municipis, comarques i províncies en taules normalitzades.
- **Procés:**
 - 1) Netejar la taula d'errors per processos relacionats amb aquest paquet.
 - 2) Netejar la taula DW_ODS.Provincia
 - 3) Transferir informació de províncies des de DW_staging.v_Provincia a DW_ODS.Provincia.
 - 4) Netejar la taula DW_ODS.Comarca (en cascada netejarà DW_ODS.Municipi).
 - 5) Transferir informació de províncies des de DW_staging.v_Comarca a DW_ODS.Comarca.
 - 6) Transferir informació de províncies des de DW_staging.v_Municipi a DW_ODS.Municipi.

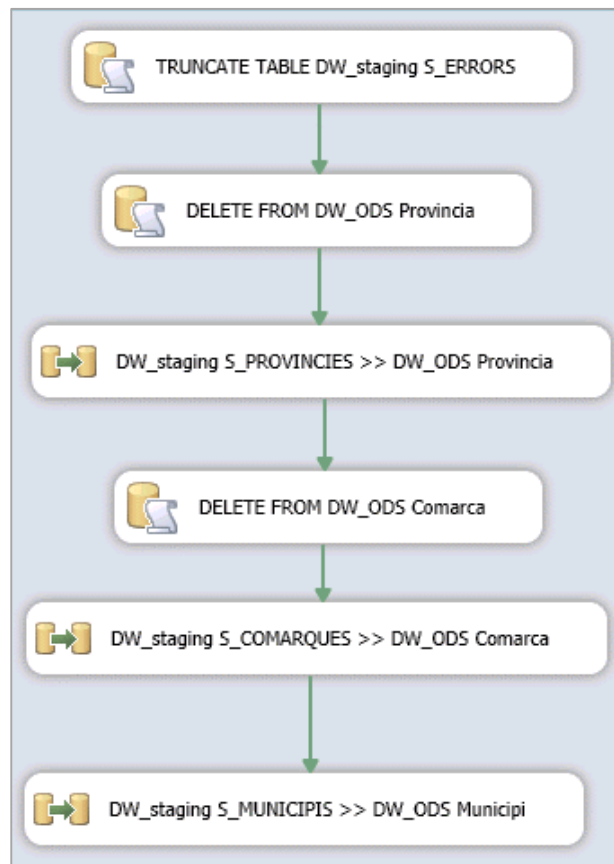


Figura 25. Procés de càrrega de l'ODS des de la *Staging Area* per dades base

ETL_BDrel_entitats

- **Descripció:** càrrega de dades per analitzar (població, vehicles, conductors i radars) en taules normalitzades.
- **Procés:**
 - 1) Netejar la taula DW_ODS.IndPoblacio
 - 2) Transferir dades de població des de DW_staging.v_IndPoblacio a DW_ODS.IndPoblacio.
 - 3) Netejar la taula DW_ODS.IndVehicles
 - 4) Transferir dades de població des de DW_staging.v_IndVehicles a la taula DW_ODS.IndPoblacio
 - 5) Netejar la taula DW_ODS.IndConductors
 - 6) Per cada any de dades del cens de conductors (2007 al 2011):
 - Trobar sentència SQL per recuperar les dades de l'any 20xx
 - Transferir dades de població des de DW_staging.v_IndConductors (filtre any 20xx) a DW_ODS.IndConductors.
 - 7) Netejar la taula DW_ODS.IndRadars
 - 8) Transferir dades de població des de DW_staging.v_IndRadars a la taula DW_ODS.IndRadars



Figura 26. Procés de càrrega de l'ODS des de la *Staging Area* per dades d'anàlisi

3.4.3. Processos principals: ODS → Magatzem de dades corporatiu (DW)

Els processos de càrrega es porten a terme en tot moment en l'àmbit de la base de dades relacional (no hi ha fonts d'origen extern). S'utilitza l'eina ETL per comoditat a l'hora d'unificar processos i per la capacitat d'utilitzar diagrames de flux que permeten controlar visualment les tasques efectuades.

En aquest procés es "desnormalitzen" les dades, que passen a estar organitzades per tema en comptes de per funció amb l'objectiu de ser utilitzades des de l'eina OLAP.

Les taules relacionals passen a ser:

- **Taules de fet** (nomenclatura fact_XXXX): objecte a analitzar, on trobem les mesures o valors.
- **Dimensions** (nomenclatura dim_XXX): atributs que proporcionen el context en què s'obtenen les mesures de fet.

Totes les tasques per cada taula segueixen la mateixa pauta:

- 1) Netejar la taula DW_ODS.XXX
- 2) Transferir la informació de la taula DW_ODS.v_XXX (la vista organitza les dades per la taula destí) a la taula DW.XXX

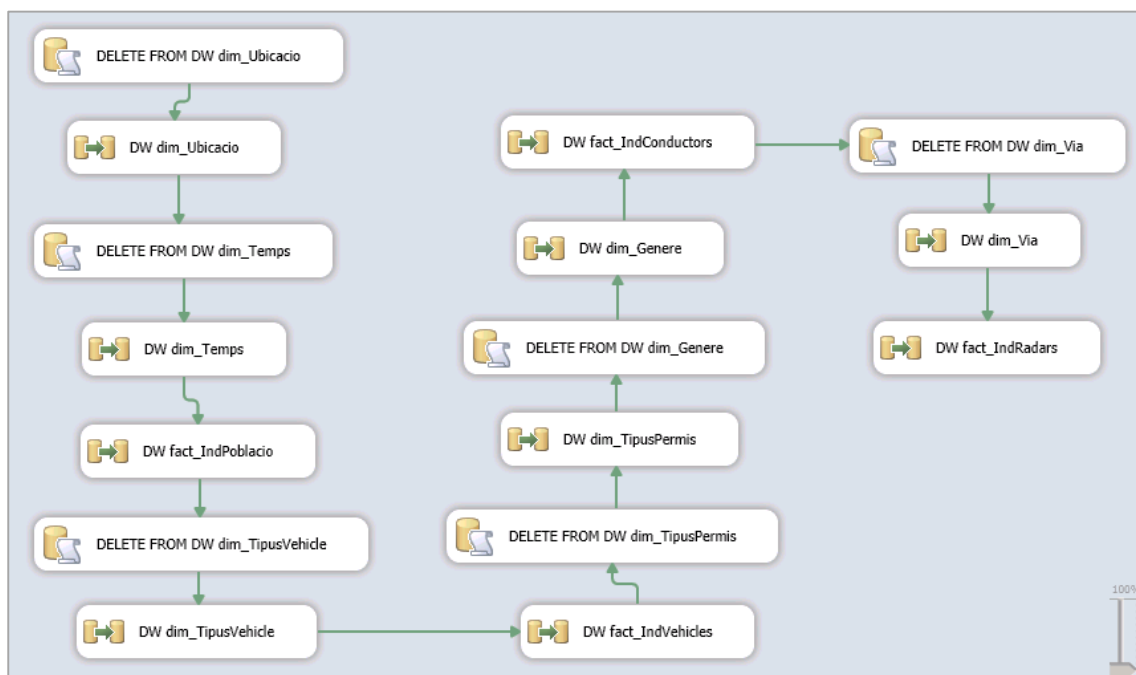


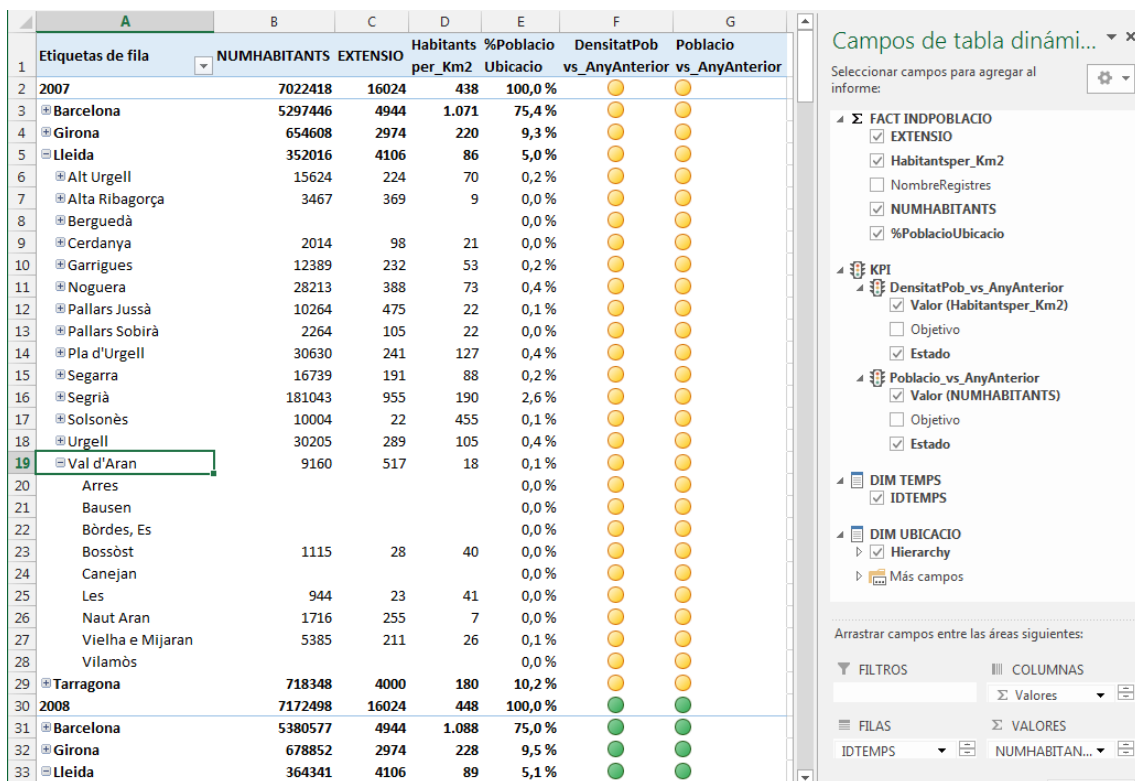
Figura 27. Procés de càrrega del Magatzem de Dades corporatiu des de l'ODS

3.5. Quadre de comandament (QC)

Cal destacar que el QC, basat en una eina versàtil i flexible com MS Excel 201x, no és una solució tancada, sinó ampliable i millorable pel propi usuari. Per tant, podem considerar que la solució que es presenta no és ni la millor ni l'única.

El QC es troba a l'arxiu C:\FECRES\QC.xlsx i treballa directament connectat amb el servidor de FECRES (MS Analysis Services), per la qual cosa requereix de connexió a Internet.

Es treballa amb **taules dinàmiques**, que permeten tractar la informació com una estructura multidimensional. A la figura de sota podem veure un exemple de taula dinàmica extreta d'un dels cubs *offline* generats per la solució.



Etiquetas de fila	NUMHABITANTS	EXTENSIO	Habitants per_Km2	%Poblacio Ubicacio	DensitatPob vs_AnyAnterior	Poblacio vs_AnyAnterior
2007	7022418	16024	438	100,0%		
Barcelona	5297446	4944	1.071	75,4%		
Girona	654608	2974	220	9,3%		
Lleida	352016	4106	86	5,0%		
Alt Urgell	15624	224	70	0,2%		
Alta Ribagorça	3467	369	9	0,0%		
Berguedà				0,0%		
Cerdanya	2014	98	21	0,0%		
Garrigues	12389	232	53	0,2%		
Noguera	28213	388	73	0,4%		
Pallars Jussà	10264	475	22	0,1%		
Pallars Sobirà	2264	105	22	0,0%		
Pla d'Urgell	30630	241	127	0,4%		
Segarra	16739	191	88	0,2%		
Segrià	181043	955	190	2,6%		
Solsonès	10004	22	455	0,1%		
Urgell	30205	289	105	0,4%		
Val d'Aran	9160	517	18	0,1%		
Arres				0,0%		
Bausen				0,0%		
Bòrdes, Es				0,0%		
Bossòst	1115	28	40	0,0%		
Canejan				0,0%		
Les	944	23	41	0,0%		
Naut Aran	1716	255	7	0,0%		
Vielha e Mijaran	5385	211	26	0,1%		
Vilamòs				0,0%		
Tarragona	718348	4000	180	10,2%		
2008	7172498	16024	448	100,0%		
Barcelona	5380577	4944	1.088	75,0%		
Girona	678852	2974	228	9,5%		
Lleida	364341	4106	89	5,1%		

Figura 28. Exemple de taules dinàmiques (cub *offline* IndPoblacio.cub)

Tanmateix, es poden definir **segmentacions de dades**, que permeten filtrar les dades de forma dinàmica en les taules de dimensió.

3.5.1. Dades Població

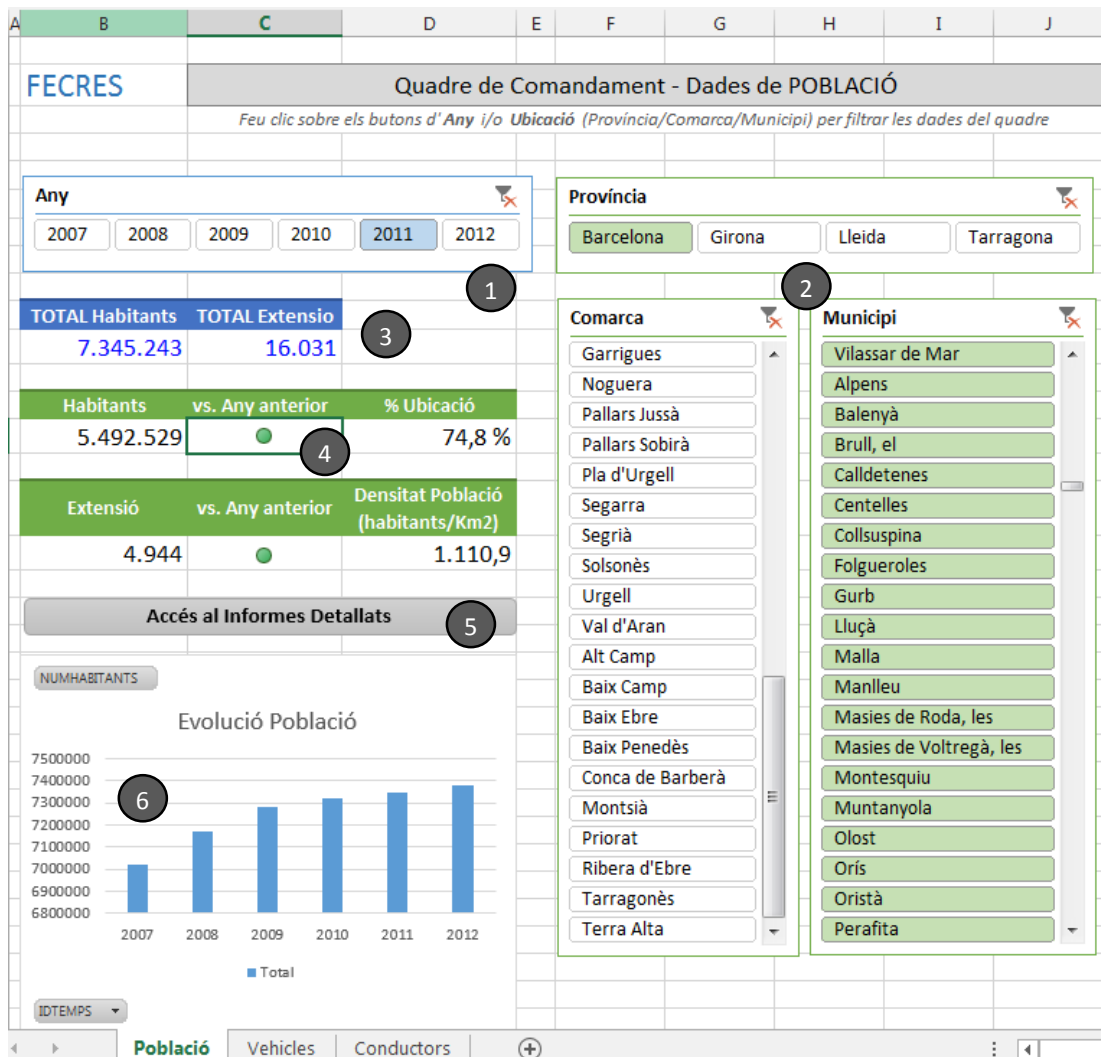


Figura 29. Quadre de Comandament dades Població

- 1) Segmentació de dades per Any: afecta a totes les taules dinàmiques.
- 2) Segmentació de dades per Ubicació (Província, Comarca, Municipi): afecten a totes les taules dinàmiques excepte les de totals (en blau).
- 3) Taula dinàmica de mesures de totals.
- 4) Taula dinàmica de mesures detallades segons segmentació. Els KPI anomenats vs.Any anterior comparen el nombre d'habitants amb el de l'any anterior tenint en compte tots els filtres aplicats (en verd: la població augmenta, en groc: la població es manté, en vermell: la població disminueix).
- 5) Accés a l'informe de dades de població (C:\FECRES\reports\RptPoblacio.xls).
- 6) Gràfic evolució de població: en la versió 2013 d'MS Excel es permet filtrar dinàmicament les dimensions que intervenen

3.5.2. Dades Vehicles

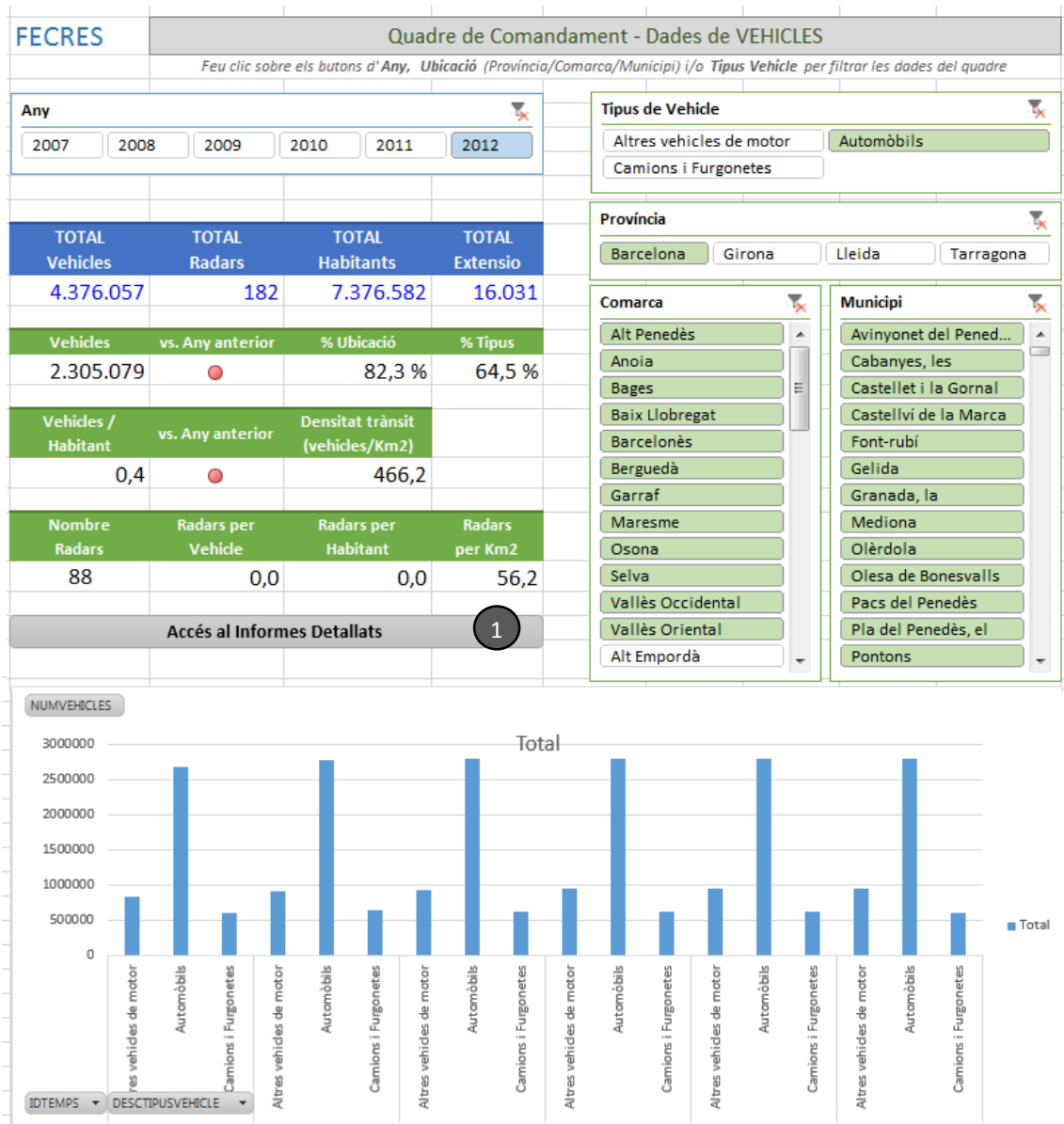


Figura 30. Quadre de Comandament dades Vehicles

1) Accés a l'informe de dades de població (C : \FECRES\reports\RptVehicles.xls).

3.5.3. Dades Conductors

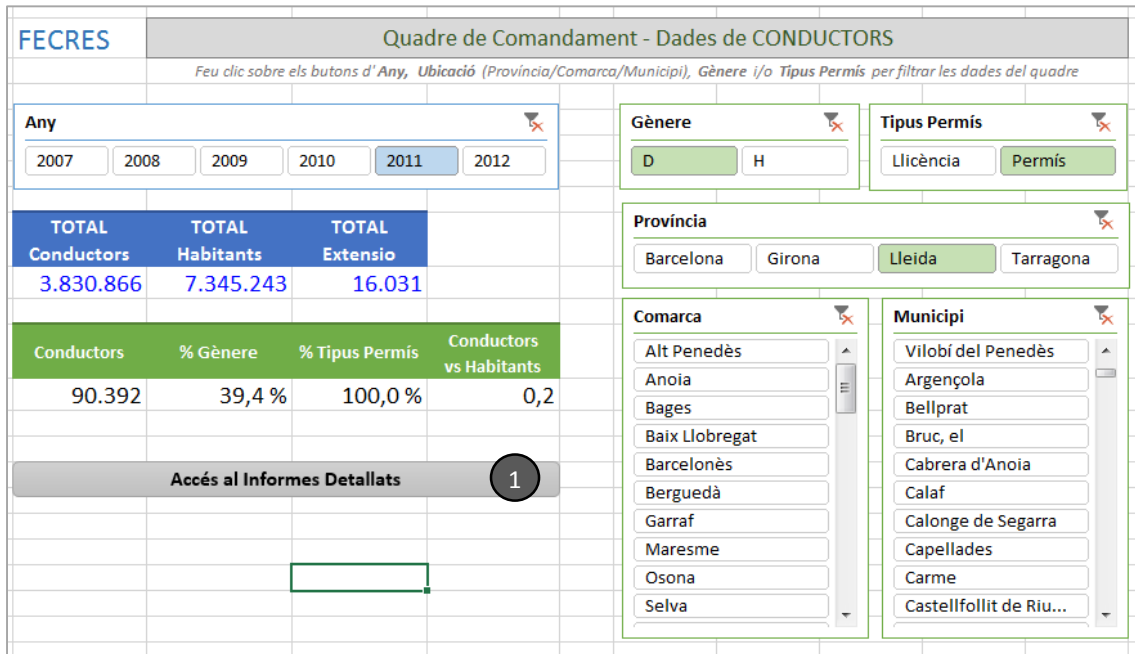


Figura 31. Quadre de Comandament dades Conductors

- 1) Accés a l'informe de dades de població (C:\FECRES\reports\RptConductors.xls).

3.6. Consultes i Informes

Tots els informes s'accedeixen des de l'arxiu C:\FECRES\Informes_local.xlsx, o individualment des de la carpeta C:\FECRES\reports\.

3.6.1. Informe dades població (RPT_Poblacio)

FECRES		Informe de dades de POBLACIÓ					
Any	Província	Comarca	Municipi	Nombre Habitants	Extensió	Longitud	Latitud
2007	Barcelona	Alt Penedès	Avinyonet del Penedès	1588	29		
2007	Barcelona	Alt Penedès	Castellet i la Gornal	2044	47		
2007	Barcelona	Alt Penedès	Castellví de la Marca	1596	29		
2007	Barcelona	Alt Penedès	Font-rubí	1430	37		
2007	Barcelona	Alt Penedès	Gelida	6151	27		
2007	Barcelona	Alt Penedès	Granada, la	1866	6		
2007	Barcelona	Alt Penedès	Mediona	2251	48		
2007	Barcelona	Alt Penedès	Olesa de Bonesvalls	1556	31		
2007	Barcelona	Alt Penedès	Olèrdola	3280	30		
2007	Barcelona	Alt Penedès	Pla del Penedès, el	891	9		

Figura 32. Informe dades Població

3.6.2. Informe dades vehicles (RPT_Vehicles)

FECRES		Informe de dades de VEHICLES			
Any	Província	Comarca	Municipi	Típus Vehicle	Nombre Vehicles
2007	Barcelona	Alt Penedès	Avinyonet del Penedès	Altres vehicles de motor	244
2007	Barcelona	Alt Penedès	Avinyonet del Penedès	Automòbils	877
2007	Barcelona	Alt Penedès	Avinyonet del Penedès	Camions i Furgonetes	303
2007	Barcelona	Alt Penedès	Castellet i la Gornal	Altres vehicles de motor	370
2007	Barcelona	Alt Penedès	Castellet i la Gornal	Automòbils	1023
2007	Barcelona	Alt Penedès	Castellet i la Gornal	Camions i Furgonetes	318
2007	Barcelona	Alt Penedès	Castellví de la Marca	Altres vehicles de motor	359
2007	Barcelona	Alt Penedès	Castellví de la Marca	Automòbils	860

Figura 33. Informe dades Vehicles

3.6.3. Informe dades conductors (RPT_Conductors)

FECRES		Informe de dades de CONDUCTORS				
Any	Província	Comarca	Municipi	Gènere	Típus Permís	Nombre Conductors
2007	Barcelona	Alt Penedès	Avinyonet del Penedès	Dona	Llicència	19
2007	Barcelona	Alt Penedès	Avinyonet del Penedès	Dona	Permís	370
2007	Barcelona	Alt Penedès	Avinyonet del Penedès	Home	Llicència	48
2007	Barcelona	Alt Penedès	Avinyonet del Penedès	Home	Permís	564
2007	Barcelona	Alt Penedès	Castellet i la Gornal	Dona	Llicència	57
2007	Barcelona	Alt Penedès	Castellet i la Gornal	Dona	Permís	466
2007	Barcelona	Alt Penedès	Castellet i la Gornal	Home	Llicència	66
2007	Barcelona	Alt Penedès	Castellet i la Gornal	Home	Permís	684

Figura 34. Informe dades Conductors

3.6.4. Informe dades radars (RPT_Radars)

FECRES		Informe de dades de RADARS		
Província	Comarca	Municipi	Via (Demarcació)	Nombre Radars
Barcelona	Alt Penedès	Sant Quintí de Mediona	C-244	1
Barcelona	Alt Penedès	Santa Margarida i els Monjos	N-340	1
Barcelona	Alt Penedès	Subirats	N-340	1
Barcelona	Alt Penedès	Torrelavit	C-15	1
Barcelona	Anoia	Calaf	C-1412a	1
Barcelona	Anoia	Castellolí	A-2	1
Barcelona	Anoia	Copons	C-1412a	1

Figura 35. Informe dades Radars

3.6.2. Web corporativa

Pàgina principal (pública): inclou un conjunt d'estadístiques bàsiques per accés públic. A data d'avui encara es disposa de les especificacions per implementar-ho.

Tanmateix cal definir la imatge corporativa de FECRES, no inclòs en l'àmbit d'aquest projecte.



Figura 36. Pàgina principal web corporativa

Pàgina d'accés restringit (personal de FECRES):

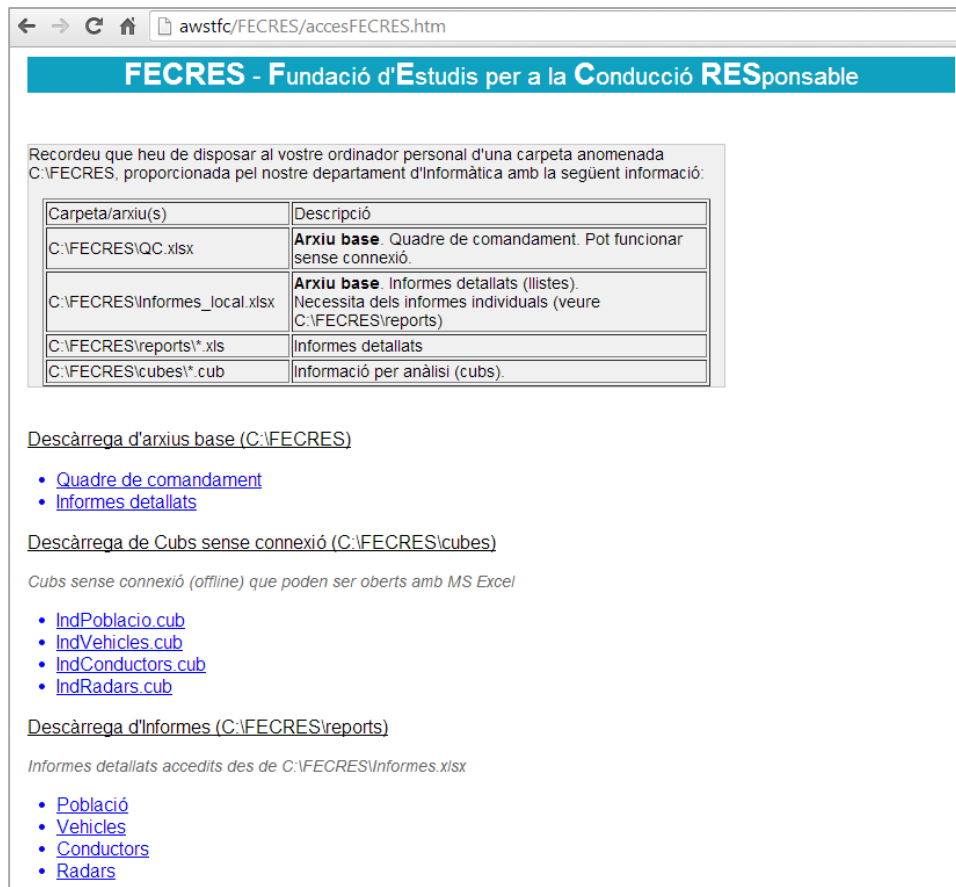


Figura 37. Pàgina accés privat a web corporativa

4. Conclusions

Aquest projecte ha permès aprofundir els coneixements sobre intel·ligència de negoci en general, i magatzems de dades en particular.

El que l'ha fet particularment interessant és el fet d'haver treballat sobre un cas eminentment pràctic i tenir la capacitat de decisió sobre les eines a emprar, així com els arguments necessaris per defensar les decisions preses.

Volem fer un incís sobre la importància estratègica que les eines de BI tenen, cada vegada més, en el món empresarial. L'activitat d'una empresa ha deixat de centrar-se només en les activitats clàssiques, com ara les vendes i la fabricació per considerar d'altres com l'estudi i aprenentatge de les dades històriques i les xarxes socials.

Ara ja no es tracta només d'estudiar l'evolució de les vendes, sinó de trobar els clients que són rentables, els productes que donaran un millor marge, i tot un seguit d'indicadors que la intel·ligència de negoci ens ofereix mitjançant l'ús de les eines adequades.

Considerem, doncs, que s'han assolit tant els objectius acadèmics del TFC d'aquesta assignatura com els requeriments de l'enunciat del projecte associat i els meus propis.

5. Línies d'evolució futura

Alguns punts que considerem importants per incloure a la solució en un futur serien:

- Incorporació de geo-localització: la solució ja contempla aquesta possibilitat amb la inclusió de camps específics per les coordenades geogràfiques a la base de dades.
- Implementació d'una versió del quadre de comandament que funcioni localment amb els cubs offline, per eliminar l'obligatorietat d'estar connectat.

6. Glossari

Cubs offline. *Offline cubes.*

Dice. Operació OLAP on es selecciona un subespai del cub original, sense reduir el nombre de dimensions.

Drill-across. Operació OLAP que canvia el tema de l'anàlisi. Continuem tenint el mateix espai n-dimensional però corresponen a un tipus de fet diferent (sobre les mateixes dimensions, canviem la taula de fets).

Drill-down. Operació OLAP que augmenta el detall amb què veiem les dades.

ETL. Acrònim per *Extract, Transform and Load*. Tècniques que s'apliquen per incorporar dades alienes al sistema des de fonts heterogènies. Sovint es confon amb el terme importació, tot i que en aquest últim cas només ens referim a la càrrega.

KPI. *Key Performance Indicator.* Mètriques utilitzades per quantificar objectius que reflexen el rendiment d'una organització. Aquests indicadors s'utilitzen en la Intel·ligència de Negoci per prescriure línies d'actuació futures.

MOLAP. *Multidimensional OLAP.* Solució OLAP que emmagatzema les dades en una base de dades multidimensional en la forma en què es volen analitzar.

ODS (*Operational Data Store*): base de dades dissenyada per a integrar dades de múltiples fonts per a fer anàlisi i informes més fàcilment.

OLAP. *Online analytic processing* o processament analític en línia. Solució utilitzada en el camp de la intel·ligència de negoci quin objectiu és agilitzar la consulta de grans quantitats de dades. Utilitza estructures multidimensionals o cubs OLAP que contenen dades resumides, calculades i no normalitzades (respecte a una base de dades relacional).

OLTP. *Online transaction processing.* Tipus de processament que facilita i administra aplicacions transaccionals. Els sistemes OLTP estan dissenyats per permetre la inserció ràpida d'informació (un dels seus objectius principals és l'entrada de dades) al nivell màxim de granularitat possible. A més, la informació es troba normalitzada per estalviar el màxim d'espai possible i evitar la redundància de dades. L'estructura bàsica d'emmagatzematge és la taula.

Pivot. Operació OLAP que consisteix en canviar l'ordre i posició de les dimensions.

Quadre de Comandament (*Dashboard*): Sistema que ens informa de l'evolució dels paràmetres fonamentals del negoci en forma d'indicadors.

ROLAP. *Relational OLAP.* És una solució OLAP que no emmagatzema dades, sinó que les accedeix a una base de dades relacional, que contindrà taules addicionals amb dades precalculades o agregades.

Roll-up. Operació OLAP que redueix el detall en què veiem les dades → agrupa les cel·les del cub seguint una certa jerarquia d'agregació. Per obtenir les dades d'una cel·la aplica una certa operació matemàtica (suma, mitjana) a les dades de les cel·les de cada grup.

Slice. Operació OLAP que consisteix a fer un tall al cub (estrella) de tal manera que es redueixi el nombre de dimensions → fixar el valor de les dimensions.

SSAS. *SQL Server Analysis Services.* Producte de Microsoft © que ofereix funcions de processament analític en línia (OLAP) i mineria de dades per aplicacions d'Intel·ligència de negoci. Analysis Services admet OLAP i permet dissenyar, crear i administrar estructures multidimensionals que contenen dades agregades des d'altres orígens de dades, com bases de dades relacionals.

SSRS. *SQL Server Reporting Services.* Producte de Microsoft © Reporting Services que ofereix una plataforma d'informes basada en servidor que proporciona la funcionalitat completa de generació d'informes per una gran varietat d'orígens de dades.

SSIS. *SQL Server Integration Services.* Plataforma de Microsoft © que ofereix processos d'extracció, transformació i càrrega de dades des de fonts de dades diverses.

7. Bibliografia

1KeyData, *Free online programming tutorials and information on data warehousing*
<http://www.1keydata.com/datawarehousing/data-warehouse-definition.html>

CURTO DÍAZ, JOSEP (2013). *Diseño de un Data Warehouse*. Barcelona: Universitat Oberta de Catalunya, Estudis d'Informàtica i Multimèdia.

RIUS GAVIDIA, ANGELS / SERRA VIZERN, MONTSE / ABELLÓ GAMAZO, ALBERTO / SAMOS JIMÉNEZ, JOSÉ / VIDAL PORTOLÉS, JOSEP / CURTO DÍAZ, JOSEP (2013). *Data warehouse: magatzems de dades i models multidimensionals* (2a ed), de l'assignatura Models multidimensionals i magatzems de dades (11.048). Barcelona: Universitat Oberta de Catalunya, Estudis d'Informàtica i Multimèdia.
URI campus UOC: https://cataleg.uoc.edu/record=b1047637~S1*cat

KIMBALL, RALPH / ROSS, MARGY (2002). *The Data Warehouse Toolkit: The complete guide to dimensional modeling*: John Wiley & Sons, Inc. (Second Edition).

KIMBALL, RALPH / JOE CASERTA (2004). *The Data Warehouse ETL Toolkit: Practical techniques for extracting, cleaning, conforming and delivering data*: Wiley Publishing, Inc.

KIMBALL, RALPH / MARGY ROSS / WARREN THORNTWHAITE / JOY MUNDY / BOB BECKER (2008). *The Data Warehouse Lifecycle Toolkit: Practical techniques for Building Data Warehouse and Business Intelligence Systems*: Wiley Publishing, Inc. (Second Edition).

8. Annexos

Annex A: Documents adjunts a la memòria

A1. Configuració i accés al sistema.pptx: informació d'accés a les diferents parts del sistema (sistema operatiu, bases de dades, web corporativa), així com la configuració necessària tant al client com al servidor.

A2. Processos de càrrega.pptx: anàlisi de les fonts de dades i explicació detallada dels processos de càrrega incloent-hi consultes SQL/MDX, scripts i configuracions.

A3. Model Multidimensional.pptx: detall del model multidimensional, incloent-hi particions, càlculs, scripts i configuracions.