

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

Jordi Alvarez Pitarque

Enginyeria Tècnica d'informàtica de Gestió
Universitat Oberta de Catalunya
Treball Fi de Carrera - Magatzems de dades
Curs 2013-2014

Consultor: Carles Llorach Rius

Data Lliurament: 06 de gener del 2014

Resum

Aquest document defineix la construcció i explotació d'un magatzem de dades per a la Fundació d'Estudis per a la Conducció Responsable. Aquesta feina és el resultat del treball de fi de carrera dels estudis d'Enginyeria Tècnica d'Informàtica de Gestió de la Universitat Oberta de Catalunya.

L'objectiu del projecte és homogeneïtzar la informació que rep la Fundació, des de diverses fonts i en diferents formats, consolidar-la en un únic magatzem de dades, i habilitar eines que facilitin la seva explotació i anàlisi. La consecució d'aquestes fites és determinant per què la direcció conegui l'evolució del trànsit rodat de vehicles a Catalunya, i minimitzi el riscs en cas de qualsevol presa de decisions.

Com indica el títol, el projecte té dues fases clarament diferenciades: crear el magatzem de dades i poblar-ho amb dades prèviament normalitzades; i, desenvolupar utilitats per a analitzar la informació.

Per a assolir la primera fase, el sistema segueix un procés estàndard que consisteix a extraure les dades de diversos fitxers que es lliuren a la Fundació, transformar-les per què segueixin els mateixos criteris, i finalment lliurar-les en una única base de dades.

La segona fase permet als usuaris del producte consultar la informació consolidada al magatzem. Hi ha un tipus de consulta lliure que ofereix a l'usuari resultats des de diferents perspectives; i un altre, que es lliura amb dades prefixades per mitjà d'informes a l'usuari.

Per concloure, el producte desenvolupat consolida i unifica, en un únic repositori, la informació heterogènia enviada periòdicament per els diferents organismes oficials. Es pretén minimitzar al màxim qualsevol intervenció manual en el procés de càrrega de la informació. Addicionalment, hi ha disponible un quadre de comandament que agrupa els darrers indicadors actualitzats. Finalment, a banda d'una eina de consulta lliure de la informació en forma de cub, hi ha un mínim d'informes parametrizats que es poden ampliar en un futur si son de la satisfacció dels usuaris.

Paraules clau

BI (*Business Intelligence*), magatzem de dades, *Data Warehouse*, ETL (*Extract, transformation and load*), Indicador, Quadre de comandament, model multi-dimensional, reporting

Construction and operation of a data warehouse for analysis of information about traffic vehicle

Author: Jordi Alvarez Pitarque
January 6th, 2014

Summary

This document explains the construction and operation of a data warehouse for the *Fundació d'Estudis per a la Conducció Responsable*. This memory is the result of the Degree Project of the Technical Engineering in Computer Management of the *Universitat Oberta de Catalunya*.

The project aim is to standardize the information received by the Foundation from different sources and formats, consolidate them into a single data store, and provide tools to accommodate their use and analysis. The achievement of these actions is crucial to know the evolution of vehicle traffic in Catalonia, and minimize risk in the event of any decision-making of the Foundation's Chairman.

As the title suggests, the project has two distinct phases: creating the data store and populate it with pre-normalized data; and, developing utilities to analyze information.

To achieve the first stage, the system follows a standard process of extracting data from multiple files that are delivered to the Foundation, transforming them applying the same criteria, and finally, loads them in a single database.

The second phase allows, to system's users, consulting the information loaded in the data warehouse. There is a type of open query that gives users results from different perspectives, and another one, that is delivered with predefined data through of reports to the user.

To sum up, the developed tool consolidates and unifies in a single repository, heterogeneous information periodically sent by the different government agencies. It also minimizes any manual intervention in the process of uploading information. Additionally, there is a control panel that shows grouped the latest updated indicators. Finally, besides of an open consultation tool through a cube, there is a minimum of parameterized reports that might be expanded in the future if they are beneficial to users.

Key words

BI (*Business Intelligence*), *Data Warehouse*, ETL (*Extract, transformation and load*), indicator, Dash board, multidimensional model, reporting

Construcción y explotación de un almacén de datos para el análisis de información sobre el tráfico rodado de vehículos

Autor: Jordi Alvarez Pitarque
6 de enero de 2014

Resumen:

Este documento define la construcción y explotación de un almacén de datos para la *Fundació d'Estudis per a la Conducció Responsable*. Este trabajo es el resultado del trabajo de fin de carrera de los estudios de Ingeniería Técnica de Informática de Gestión de la *Universitat Oberta de Catalunya*.

El objetivo del proyecto es homogeneizar la información que recibe la Fundación, desde diversas fuentes y en diferentes formatos, consolidarla en un único almacén de datos, y habilitar herramientas que faciliten su explotación y análisis. La consecución de estas acciones es determinante para que la dirección conozca la evolución del tráfico rodado de vehículos en Catalunya, y minimice los riesgos en caso de cualquier toma de decisiones.

Como indica el título, el proyecto tiene dos fases claramente diferenciadas: crear el almacén de datos y poblar-lo con datos previamente normalizadas y, desarrollar utilidades para analizar la información.

Para alcanzar la primera fase, el sistema sigue un proceso estándar que consiste en extraer los datos de varios archivos que se entregan a la Fundación, transformarlos para que sigan los mismos criterios, y finalmente gravarlos en una única base de datos.

La segunda fase permite a los usuarios del producto consultar la información consolidada en el almacén. Existe un tipo de consulta libre que ofrece al usuario resultados desde diferentes perspectivas, y otro, que se entrega con datos predefinidos a través de informes al usuario.

Para concluir, el producto desarrollado consolida y unifica en un único repositorio, la información heterogénea enviada periódicamente por los diferentes organismos oficiales. También, se minimiza al máximo cualquier intervención manual en el proceso de carga de la información. Adicionalmente, existe un cuadro de mando que agrupa los últimos indicadores actualizados. Finalmente, aparte de una herramienta de consulta libre de la información en forma de cubo, existe un mínimo de informes parametrizados que se pueden ampliar en un futuro si resultan provechosos para los usuarios.

Palabras clave

BI (Business Intelligence), almacén de datos, Data Warehouse, ETL (Extract, transformation and load), Indicador, Cuadro de mando, modelo multi-dimensional, reporting

Índex de continguts

1.	Capítol1, Introducció	7
1.1.	Justificació del TFC	7
1.2.	Objectius del TFC.....	7
1.3.	Enfocament i mètode seguit	8
1.4.	Planificació inicial vs planificació final.....	8
1.5.	Anàlisi de riscos. Resultats esperats.....	10
1.6.	Productes obtinguts	12
2.	Capítol 2, Anàlisi	13
2.1.	Anàlisi de requeriments.....	13
2.2.	Diagrama de casos d'ús amb una explicació de cada cas d'ús.....	16
2.3.	Diagrama del model conceptual.....	18
3.	Capítol 3, Disseny	23
3.1.	Diagrama de l'arquitectura del software.....	23
3.2.	Diagrama de l'arquitectura hardware	23
3.3.	Disseny de la base de dades i diagrama del model físic	24
3.4.	Disseny i descripció dels informes creats	27
3.5.	Disseny i descripció del cub creats.....	34
4.	Captures de pantalla.....	37
4.1.	Accessos als productes	37
4.2.	Detall de les ETLs. Funcionament dels processos de càrrega:	38
4.2.1.	Job principal:: Actualització del DW	38
4.2.2.	Job Càrrega de fitxers.....	39
4.2.3.	Transformació de consolidació dels indicadors	45
4.2.4.	Job de transformació al model dimensional.....	46
5.	Conclusions	50
6.	Línies d'evolució futur	50

Índex de figures

Il·lustració 1: Diagrama de Gantt.....	9
Il·lustració 2: Cas d'ús Consultar Informació	17
Il·lustració 3: Cas d'ús Realitzar Manteniment	17
Il·lustració 4: Model Entitat-Relació	18
Il·lustració 5: cub de dades	20
Il·lustració 6: Exemple de Cub	21
Il·lustració 7: Model en estrella	22
Il·lustració 8: Diagrama de l'arquitectura del software	23
Il·lustració 9: Arquitectura del hardware	24
Il·lustració 10: Quadre de comandament de l'estat del trànsit	29
Il·lustració 11: Quadre de comandament de l'evolució del trànsit.....	29
Il·lustració 12: Informe vehicles filtrat.....	30
Il·lustració 13: Informe dels radars	31
Il·lustració 14: Informe dels radars usant navegació drill-down	32
Il·lustració 15: Informe permís	33
Il·lustració 16: Informe permís filtrat	34
Il·lustració 17: Disseny del cub de trànsit	35
Il·lustració 18: Exemple Consulta a cub 1.....	36
Il·lustració 19: Exemple Consulta a cub 2.....	36
Il·lustració 20: Connexió al repositori.....	37
Il·lustració 21: Catàleg de jobs i transformacions del projecte	38
Il·lustració 22: job principal > job jActualitzarDW	38
Il·lustració 23: Connexió a Oracle des de Pentaho Data Integration	39
Il·lustració 24: job jCarregarFitxersOrigen	39
Il·lustració 25: Transformació de municipi	41
Il·lustració 26: Transformació de la població	42
Il·lustració 27: Transformació dels radars.....	43
Il·lustració 28: Transformació dels conductors	44
Il·lustració 29: Transformació dels vehicles	45
Il·lustració 30: Transformació Trans_Consolidat.....	46
Il·lustració 31: job jModelDimensional.....	46
Il·lustració 32: Dimensió Any	47
Il·lustració 33: Dimensió Territori.....	48
Il·lustració 34: Dimensió Vehicle	48
Il·lustració 35: Dimensió via	49
Il·lustració 36: Transformació dels fets	49

1. Capítol1, Introducció

El treball de fi de carrera, del que forma part aquesta memòria, queda delimitat a l'àrea de Magatzem de Dades de la UOC.

El cas pràctic resol un problema habitual de les organitzacions: analitzar un gran volum de dades relacionades amb la seva activitat el més simple possible per assolir amb èxit la seva missió.

Aquests darrers temps, de menor inversió en sistemes d'informació, les empreses i organitzacions orienten els seus esforços a treure el màxim profit dels sistemes existents, tant propis com aliens; normalment gestionen un ventall d'aplicacions, de tecnologies diferents, que tenen poca cohesió.

En aquest context, afegir una capa de negoci que recopili i normalitzi tota la informació esdevé una tasca imprescindible.

Per aquest motiu, la meua orientació va dirigida cap al camp del BI (*Business Intelligence*). Malauradament, a mida que conec la multitud d'eines i productes que el formen, me n'adono, que és una tasca que requereix molt d'anàlisi i molta pràctica per treure profit tant del negoci del client com dels productes relacionats existents de BI.

1.1. Justificació del TFC

Actualment, hi ha accés a grans quantitats de dades, i la tendència és augmentar-la vertiginosament els propers anys. Termes com el *Big Data*, *Open Data*, *Cloud computing*, geolocalització, etc, són nomenclatures o conceptes utilitzats per apropar, compartir i aprofitar-se d'informació provinent d'arreu. Òbviament, les organitzacions volen fer-ne un ús per millorar els seus processos i arribar als seus objectius dedicant els mínims recursos possibles.

Un magatzem de dades, o més conegut pel terme en anglès, *Data Warehouse*, es pot veure com una mena de diamant que s'obté després del laboriós procés d'extraure'l de la roca i tallar-lo en brut. Extrapolant el símil al context que ens ocupa, l'aportació del treball serà el d'extraure les dades, filtrar-les com ho faria un embut, i dipositar-les en un repositori d'informació per què s'analitzin amb més facilitat.

1.2. Objectius del TFC

a. Generals

L'objectiu principal del projecte és la construcció d'un magatzem de dades, a partir de diferents orígens, incloent l'explotació de la informació.

A més a més, es pretén demostrar la correcta gestió d'un projecte en totes les seves fases: planificació, anàlisi, disseny, implementació, proves i lliurament.

Adicionalment, ofereix l'oportunitat als alumnes, d'aprendre i practicar amb entorns virtuals i amb diferents solucions disponibles al mercat, relacionades amb el *Business Intelligence*, de forma gratuïta o sota llicència educativa adquirida per la UOC

b. Específics.

De forma específica, els objectius del projecte són els següents:

- Adquirir habilitats en l'anàlisi de les dades.
- Conèixer els processos ETL: extracció, transformació i carrega de les dades en un magatzem de dades des de diferents orígens d'informació.
- Dissenyar i implementar un magatzem de dades.
- Disseny de cubs i utilització eines OLAP.
- Dissenyar informes ad hoc.
- Crear una memòria per a documentar el projecte d'una forma tècnica.

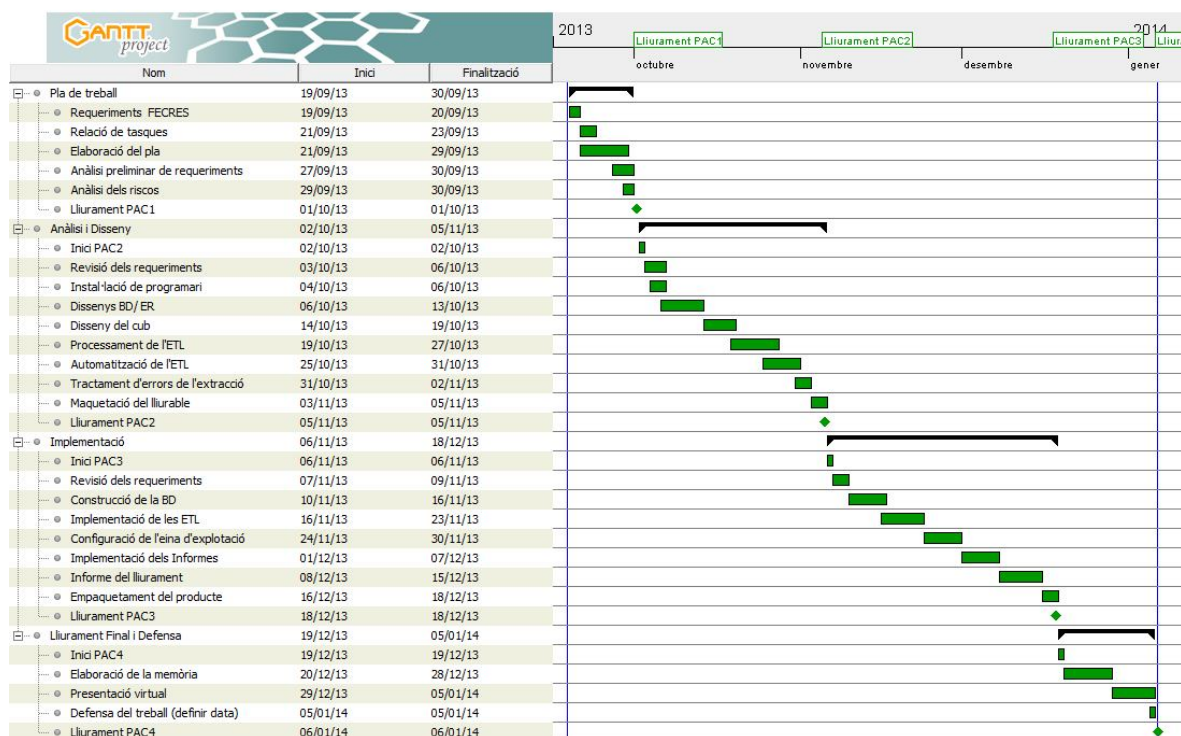
1.3. Enfocament i mètode seguit

El projecte s'ha desenvolupat, en la mida de lo possible, segons els lliuraments planificats per l'avaluació continuada de la UOC. Atès que cada lliurament està associat a una etapa del cicle de vida clàssic del programari, és el mètode seguit per a la realització del projecte:

- Fase 1. Pla de treball i anàlisi preliminar (recollida de requeriments): correspon als inicis d'un projecte, és a dir, comunicació amb el client, revisió del material disponible, tecnologies a utilitzar, relació d'activitats a realitzar, planificació del projecte, anàlisi dels requeriments i l'avaluació dels riscos, principalment.
- Fase 2. Anàlisi de requeriments i disseny conceptual (anàlisi i disseny). Anàlisi en profunditat dels requeriments i de les dades, disseny dels models conceptual i lògic i esbós del model multi-dimensional i descripció de les ETLs són les tasques més destacades.
- Fase 3. Implementació (Realització): Aquesta etapa és la de construcció del producte:
 - o Construcció del magatzem de dades
 - o Càrregues de informació
 - o Configuració de l'eina d'explotació
 - o Construcció dels informes i anàlisi de la informació.
- Fase 4. Lliurament de la solució (test, manteniment i lliurament): revisió de les càrregues, ajustaments de les configuracions, proves d'anàlisi, etc..

1.4. Planificació inicial vs planificació final

La planificació inicial es mostra en el següent diagrama de Gantt:



II-lustració 1: Diagrama de Gantt

La planificació final s'ha desviat a mida que avançava el projecte. El principals motius dels endarreriments han estat: un risc no identificat com el desconeixement de les eines a utilitzar; i d'altres identificats com la configuració de les eines com el report server (que no està en producció), el funcionament del *Pentaho Data Integration* a l'entorn virtual (java i drivers a Oracle), i copiar fitxers a un entorn virtual,

A continuació es mostra la desviació segons la relació de tasques fins a la data del lliurament:

Tasca	Estimació dies	Data límit planificada	Desviació
PLA DE TREBALL	13 dies	01/10/2013	0%
Revisió dels requeriments	2 dies	20/09/13	
Relació de tasques	3 dies	23/09/2013	
Elaboració del pla	9 dies	29/09/2013	
Anàlisi preliminar de requeriments	4 dies	30/09/2013	
Anàlisi de riscos	2 dies	30/09/2013	
Lliurament	1 dia	01/10/2013	
ANÀLISI I DISSENY	35 dies	05/11/2013	20%
Revisió dels requeriments	4 dies	06/10/2013	0%
Instal·lació del programari	3 dies	06/10/2013	0%
Dissenys BD/ ER	8 dies	13/10/2013	0%
Disseny del cub	6 dies	19/10/2013	0%

Processament de l'ETL	9 dies	27/10/2013	90%
Automatització de l'ETL	9 dies	31/10/2013	10%
Tractament d'errors de l'extracció	3 dies	02/11/2013	100%
Maquetació del lliurable	3 dies	05/11/2013	10%
Lliurament	1 dia	05/11/2013	80%
IMPLEMENTACIÓ	43 dies	18/12/2013	66%
Revisió dels requeriments	3 dies	09/11/2013	0%
Construcció de la BD	7 dies	16/11/2013	20%
Implementació de les ETL	8 dies	23/11/2013	20%
Configuració de l'eina d'exploració	7 dies	30/11/2013	50%
Implementació dels Informes	7 dies	07/12/2013	100%
Informe del lliurament	7 dies	15/12/2013	100%
Empaquetament del producte	3 dies	18/12/2013	100%
Lliurament	1 dia	18/12/2013	66%

1.5. Anàlisi de riscos. Resultats esperats.

Risc 01	Connexió als fitxers d'origen
Descripció	Sempre hi ha risc quan s'estableix connexions entre entorns.
Tractament	<ul style="list-style-type: none"> - Evitar espais als fitxers de d'origen. - Canviar el format o aconseguir <i>drivers</i> si l'eina no permet establir connexió.
Alternativa	Demanar els fitxers als sistemes d'informació per què canviïn els lliurin compatibles amb la nostra solució.
Observacions	En les ETLs programades cal evitar qualsevol tasca manual

Resultat: no hi ha hagut problemes per a connectar-se als diferents formats d'origen al marge d'endarreriment en la previsió inicial.

Risc 02	Extracció de les dades
Descripció	Els fitxers d'origen contenen dades no estructurades a l'inici i al final de l'arxiu en alguns casos. També poden donar problemes els delimitadors.
Tractament	<ul style="list-style-type: none"> - Eliminar manualment les files sobrants sinó es poden treure de forma automàtica. - Procurar que el caràcter (o tabulador) no formi part del contingut del text. P.ex. les dobles cometes.
Alternativa	Demanar els fitxers als sistemes d'informació per què l'enviïn estructurat.
Observacions	En les ETLs programades cal evitar qualsevol tasca manual

Resultat: no hi ha excessius problemes per a extraure les dades sense intervenció encara que un endarreriment en la previsió inicial.

Risc 03	Connexió al DW des de l'ETL
Descripció	Sempre hi ha risc quan s'estableix connexions entre entorns. En aquest cas des de l'eina on s'implementi l'ETL cap al DW de l'Oracle XE
Tractament	- Utilitzar el suport del laboratori.
Alternativa	- Canviar l'eina per a implementar les ETLs
Observacions	---

Resultat: aquest risc ha suposat el principal endarreriment de la tercera PAC. No s'ha usat el suport del laboratori per què l'eina amb problemes era el Pentaho. No s'ha canviat d'eina per què la majoria de tasques relacionades amb les ETLs s'havien desenvolupat localment sense problemes. El problema era amb la localització del driver que utilitza Pentaho per a connectar-se a l'Oracle.

Risc 04	Carrega de les dades
Descripció	Incompatibilitats entre les dades transformades i el disseny de les taules de destí
Tractament	- Revisar els tipus de dades - Revisar la longitud de cada camp - Revisar els nom dels objectes (taula, atributs...) - Revisar les restriccions del destí (<i>constraints</i>)
Alternativa	---
Observacions	---

Resultat: cap problema al marge de restriccions de la pròpia base de dades.

Risc 05	Instal·lació de programari
Descripció	Tot i que es disposa d'un entorn virtual amb les aplicacions principals instal·lades, hi ha un risc amb la resta de programari (Java, sistemes de 32 o 64 bits...) o amb la disponibilitat de llicències.
Tractament	- Utilitzar el suport del laboratori - Crea una VM amb la Virtual Box d'Oracle (Sun).
Alternativa	Buscar alternatives de programari de lliure distribució. P.ex. <i>Pentaho</i> en comptes del BI de l'Sql Server; o l' <i>Online Diagram Software</i> en comptes del <i>Magic Draw</i> .

Resultat: Problemes amb la instal·lació de Pentaho atès que no reconeixia el Java de la màquina virtual. La solució era executar un procés anomenat Kitchen.bat que defineix la ubicació del Java.

Risc 06	Comunicacions Client - Servidor
Descripció	Pot ser que no hi hagi visibilitat entre els serveis del servidor i el client.
Tractament	- Testejar un servei del servidor des del client. P.ex. l'accés a l'Oracle
Alternativa	---
Observacions	Pot ser necessari establir un servei d'FTP per a traspasar fitxers entre servidor i client.

Resultat: problemes alhora de copiar fitxers al servidor. Després de modificar permisos, revisar FTPs, etc. la solució era tan senzilla com arrossegar els fitxers. Manca d'experiència en entorns virtuals.

Un risc important no tingut en compte, és el desconeixement de les eines a utilitzar que no hi havia destinat dates per a l'aprenentatge.

1.6. Productes obtinguts

Els productes obtinguts per ordre cronològic son els següents

- **Processos ETL:** transformacions i treballs dissenyats amb el *Pentaho Data Integration* i disponibles al seu catàleg. Les transformacions extrauen la informació, la manipulen i la carreguen al magatzem de dades; els treballs o *jobs*, executen lots de transformacions i també els propis *jobs*.
- **El magatzem de dades:** Construït amb l'*Oracle Database 11g XE*, dins l'esquema anomenat Pentaho.
- **Cub:** model dimensional de la informació dissenyat amb l'*SQL Server 2012*. Mostra les dades des de diferents perspectives com el tipus de vehicles, la província o l'any. Està disponible dins el *Visual Studio* amb el nom *pjt_transit_molap*.
- **Informes:** s'elaboren els següents informes amb l'*SQL Server 2012*:
 - *rp01_QC*: Quadre de comandament que mostra la darrera actualització dels indicadors requerits
 - *rp02_Vehicles*: llistat parametrizat que mostra informació del parc de vehicles de Catalunya.
 - *rp03_QC_Evolucio*: Quadre de comandament que mostra l'evolució del trànsit.
 - *rp04_Radars*: llistat parametrizat que mostra la informació disponible dels radars.
 - *rp05_Permis*: llistat parametrizat que mostra la informació disponible dels conductors.Estan disponibles dins el projecte quadreComandament del *Visual Studio*.

2. Capítol 2, Anàlisi

2.1. Anàlisi de requeriments

2.1.1. Requeriments

a. Funcionals

Les funcionalitats que requereix FECRES per a la nova solució, son:

- La creació d'un magatzem de dades corporatiu amb la següent informació:
Informació sobre municipis i vehicles de motor, sobre els conductors dels darrers cinc anys, i per últim, la localització dels radars fixos.
- La disponibilitat dels següents indicadors per un any concret, segons la província, la comarca, el tipus de vehicle i el tipus de permís de circulació:
 1. Total de vehicles
 2. Total de conductors
 3. % de vehicles respecte població
 4. Densitat de població (habitants/km²) i densitat de trànsit (vehicles/km²)
 5. Nombre de vehicles/Nombre de radars
 6. % de conductors per radar
 7. Indicador de conductors vs habitants per gènere
 8. Indicador de radars vs vehicles
 9. Rati de vehicles x conductor
 10. Quantitat de vehicles / superfície del territori
- Elaborar un catàleg d'informes prefixats que mostrin la informació anterior.

En resum, FECRES demana integrar en un nou magatzem de dades la informació obtinguda de diverses entitats, i la creació d'un quadre de comandament que els ajudi a aprofundir en l'evolució del trànsit rodat de vehicle.

b. No funcionals

Els requeriments no funcionals, son aquells aspectes que no s'indiquen de forma específica però es pressuposen a la consecució del projecte:

- Robust: s'espera que sigui un sistema sense errades, o amb errades controlades, que siguin intel·ligibles per als usuaris de l'aplicació.
- Segur: atès que accedim a diferents utilitats i gestors de bases de dades, serà necessari crear diferents perfils, del tipus administrador i consultor.
- Escalable: es possible que es requereixi més informació una vegada l'usuari veu les possibilitats que ofereixi el sistema, així, que hem de preveure el creixement dels cubs o l'ampliació de les dimensions i les mètriques.
- Compatible: els accessos a l'eines per navegador, és a dir, que no requereixin instal·lació, han d'estar alineats amb les darreres versions i en especial, amb les versions actuals del client.
- Usable: Atès que les característiques dels sistemes analítics que implementem estan orientats a l'usuari i al seu model de negoci,

l'aplicació hauria de ser fàcilment usable per l'usuari, és a dir, que no requereixi un esforç extraordinari per els usuaris de l'aplicació.

2.1.2. Estudi de les dades d'origen

2.1.2.1. Relació de fitxers

La relació dels fitxers proporcionats pel client es repassen a continuació de forma esquemàtica:

Fitxer	Radars_SCT.txt
Font	Servei Català de Trànsit http://www.gencat.cat/transit/radars.htm
Descripció	localització dels radars fixos
Primera fila que conté dades	13
Nombre de registres	217
LayOut	Via Text Municipi Text Comarca Text Demarcació Text
Observacions	No existeix un codi identificatiu per als municipis.

Fitxer	Dades_municipis.xls
Font	IDESCAT http://www.idescat.cat/
Descripció	Informació relativa a població i extensió dels municipis dels darrers sis anys
Primera fila que conté dades	2
Nombre de registres	466
LayOut	Municipi Text Codi INE Text(5) Població 2012 Numèric Població 2011 Numèric Població 2010 Numèric Població 2009 Numèric Població 2008 Numèric Població 2007 Numèric Extensió (Km2) Numèric
Observacions	- Al web d'IDESCAT hi ha disponible una API per a fer peticions de les dades dels municipis: http://www.idescat.cat/api/emex/ - La província es pot deduir del codi_ine - No hi ha informació de la comarca.

Fitxer	Dades_vehicles.xls
Font	IDESCAT http://www.idescat.cat/
Descripció	Informació dels vehicles de motor per municipi i tipus de vehicle dels darrers sis anys
Primera fila que conté dades	2
Nombre de registres	466
LayOut	Municipi Text

d'aquesta perspectiva atès que no es pot perdre informació vital com els radars o algun tipus de vehicle.

- Tots els fitxers tenen en comú la informació referent al municipi.

2.1.2.2. Tractament dels fitxers per a extraure la informació:

Es considera imprescindible minimitzar la intervenció manual per a l'extracció de la informació dels fitxers d'origen. Aquest objectiu té varis propòsits:

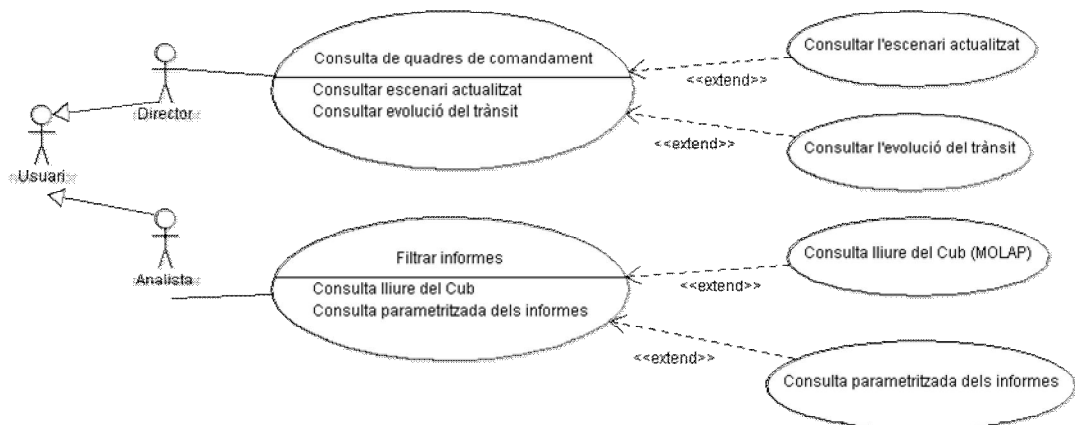
- **Repetir tantes càrregues** d'informació com necessiti el client. No es pot descartar que de forma asíncrona, cada organització envii dades actualitzades.
- **Reduir temps de presentació** de la informació, atès que preparar la informació de forma ideal per a l'extracció de la informació, requereix un temps inadmissible per la rapidesa amb la que actualment circula la informació.
- **Evitar que la manipulació manual** dels fitxers danyï l'estructura dels mateixos, i dificulti detectar l'error en els processos d'extracció.

Aquests avantatges també generen dificultat per als processos ETL i un manteniment anual dels mateixos:

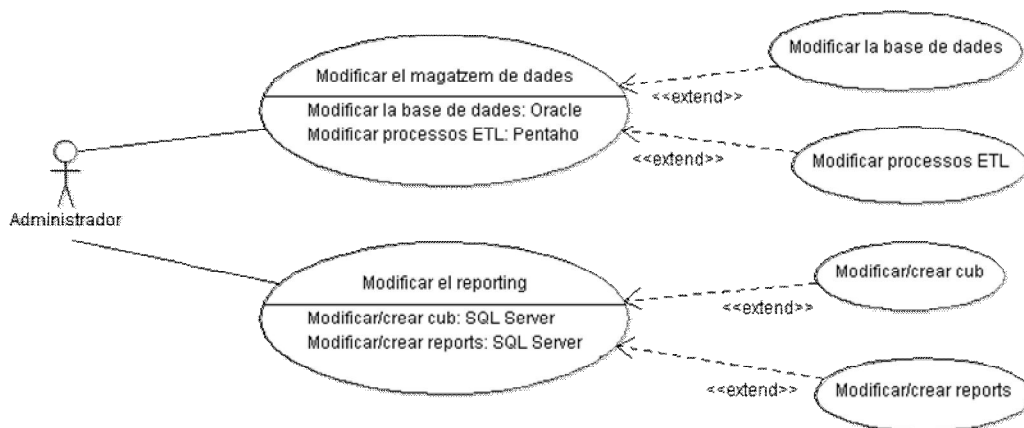
- **Fitxer radars.** No presenta cap dificultat per què l'enviament de nous fitxers substituïria els anteriors.
- **Dades municipis i dades vehicles.** L'enviament de nous fitxers amb la informació de nous anys implicaria modificar les ETLs d'extracció d'ambdós fitxers. El motiu principal, és que es necessita traslladar a nivell de fila les columnes etiquetades amb els anys, de forma que les columnes per exemple referents a 2007, 2008 o 2009 quedarien sota una única columna etiquetada com a any. Es tracta d'un manteniment laboriós.
- **Dades conductors.** L'enviament de nous fitxers anuals, implica afegir nous canals d'entrada al procés d'extracció. Es tracta d'un manteniment anual i molt senzill.

2.2. Diagrama de casos d'ús amb una explicació de cada cas d'ús.

S'identifiquen dos casos d'ús: cas d'ús Consultar Informació i el cas d'ús Realitzar Manteniment.



II-lustració 2: Cas d'ús Consultar Informació



II-lustració 3: Cas d'ús Realitzar Manteniment

Cas d'ús Consultar Informació

Resum de la funcionalitat: tracta la consulta d'informació per mitjà dels informes existents.

Actors: Analista, Director.

Explicació: Els usuaris del sistema poden ser de tipus analista, que requereix filtrar la informació fins a obtenir la que necessita; o amb el rol de director, qui simplement executa algun dels quadres de comandament disponibles.

Cas d'ús Realitzar Manteniment

Resum de la funcionalitat: tracta del manteniment del sistema

Actor: administrador

Explicació: Un usuari del tipus administrador que tingui coneixements de BI, preferiblement dels equips de IT/SI del client, realitza el manteniment del sistema.

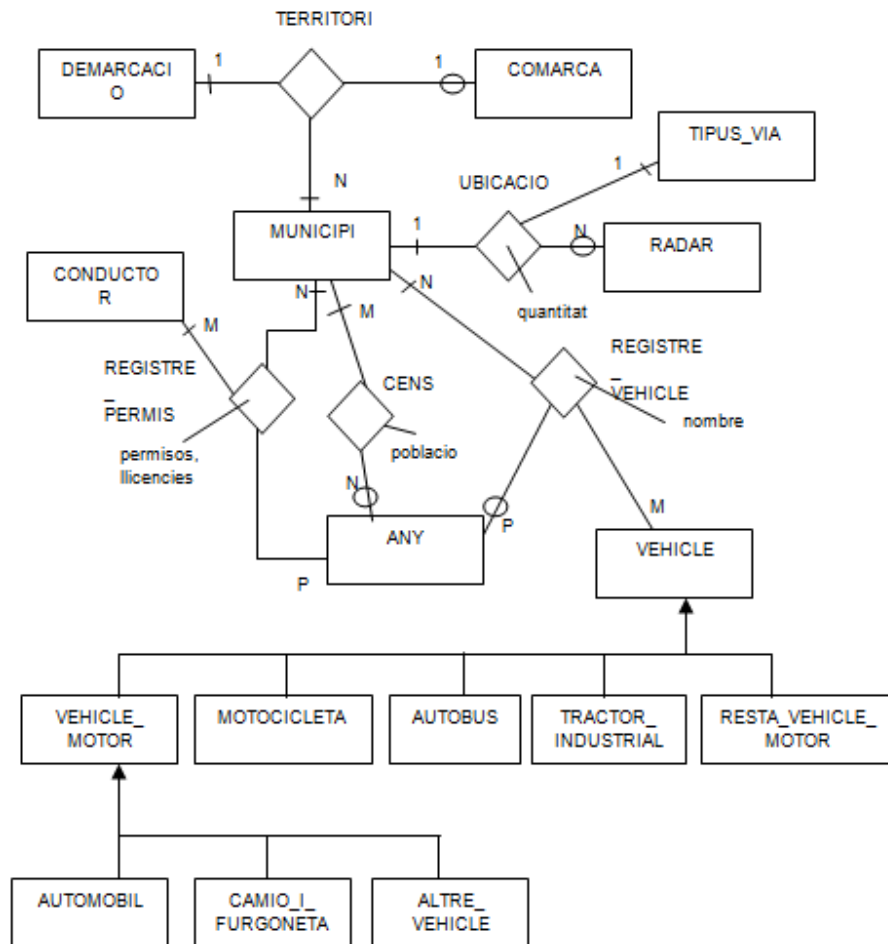
Serà necessari modificar les ETLs quan arribin nous fitxers sinó substitueixen els actuals; addicionalment, si es requereix nova informació, caldrà modificar la base de dades i les ETLs.

Possiblement l'administrador també crearà nous informes o modificarà els existents a mida que se'n faci us de la solució.

2.3. Diagrama del model conceptual

2.3.1. El model ER.

El model conceptual reflecteix un model de base de dades independent a la tecnologia a emprar. El resultat del model conceptual s'expressa en un model de dades d'alt nivell, en concret, en un model d'Entitat-Relació, abreviat com ER segons la nomenclatura anglesa *Entity-Relationship*. El model ER complet és el següent:



II-lustració 4: Model Entitat-Relació

A continuació s'estructura la informació segons els diferents conceptes:

a. Dades territori.

Segons l'anàlisi de la informació, el territori es compon de municipi, comarca i demarcació. Atès que la comarca només consta al fitxer de radars, hi ha municipis que no la tenen informada. Finalment, un municipi només pertany a

una comarca i a una demarcació concreta, en canvi tant la comarca, com la demarcació poden contenir N municipis.

Els atributs de les entitats que figuren al diagrama son els següents (claus primàries subratllades):

DEMARCACIO
nom-demarcacio
COMARCA
nom-comarca
MUNICIPI
nom-municipi, codi-ine, extensio

No és una situació ideal per què el codi_ine per si sol pot ser clau primària, però hi ha al voltant de 500 municipis (p.ex. Capolat) que no el tenen informat. Son municipis dels fitxers de les dades dels conductors i dels radars que no hi figuren a les dades del municipi.

b. Dades radars

Els radars estan ubicats a les vies a diferents municipis. Hi pot haver municipis sense radars, altres municipis amb més d'un radar, o fins i tot, tipus de vies (p.ex. autopistes) que travessen varis municipis on s'ubiquen diversos radars:

MUNICIPI
nom-municipi, codi-ine, extensio
RADAR
via
TIPUS_VIA
tipus

Cens població

De les dades existents al fitxer de municipis, hi ha alguna població com la Canonja, que no disposa del cens de la població. Aquesta excepció es reflecteix al model per mitjà d'un cercle a la banda de l'any. Tampoc hi ha disponible dades de municipis dels fitxers de conductors i dels radars que no hi figuren al fitxer de municipis:

MUNICIPI
nom-municipi, codi-ine, extensio
ANY
any

c. Dades conductors

Les dades dels conductors permeten conèixer el nombre de permisos i llicències, segons el gènere del conductor, que s'atorguen cada any.

CONDUCTOR
genere
MUNICIPI
nom-municipi, codi-ine, extensio
ANY
any

d. Dades vehicles

A la interrelació REGISTRE_VEHICLE s'indica el nombre de vehicles registrats als diferents municipis catalans anualment segons el tipus de vehicle. Hi ha vehicles que no tenen informat l'any. Al model s'indica la generalització/especialització dels vehicles només per entendre el disseny conceptual per què cap de les especialitzacions tenen atributs propis en aquest context.

MUNICIPI

nom-municipi, codi-ine, extensio

ANY

any

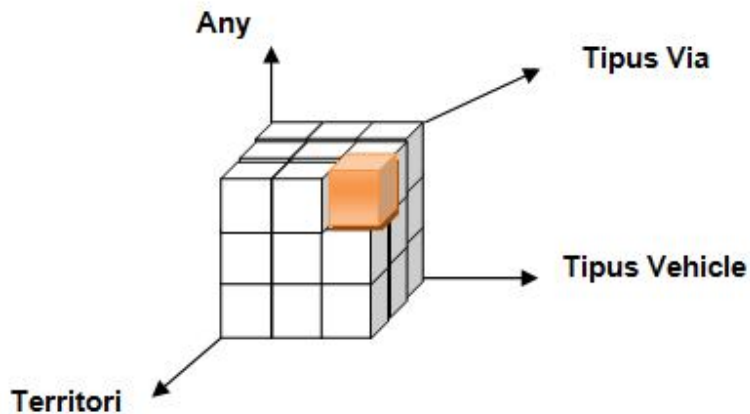
VEHICLE

tipus

2.3.2. Model multi-dimensional

En els models de dades dimensionals la informació es representa com un cub de dades on les cel·les contenen valors que anomenem fets i les arestes dimensions d'exploració que ens ofereixen diferents perspectives per a l'anàlisi de la informació.

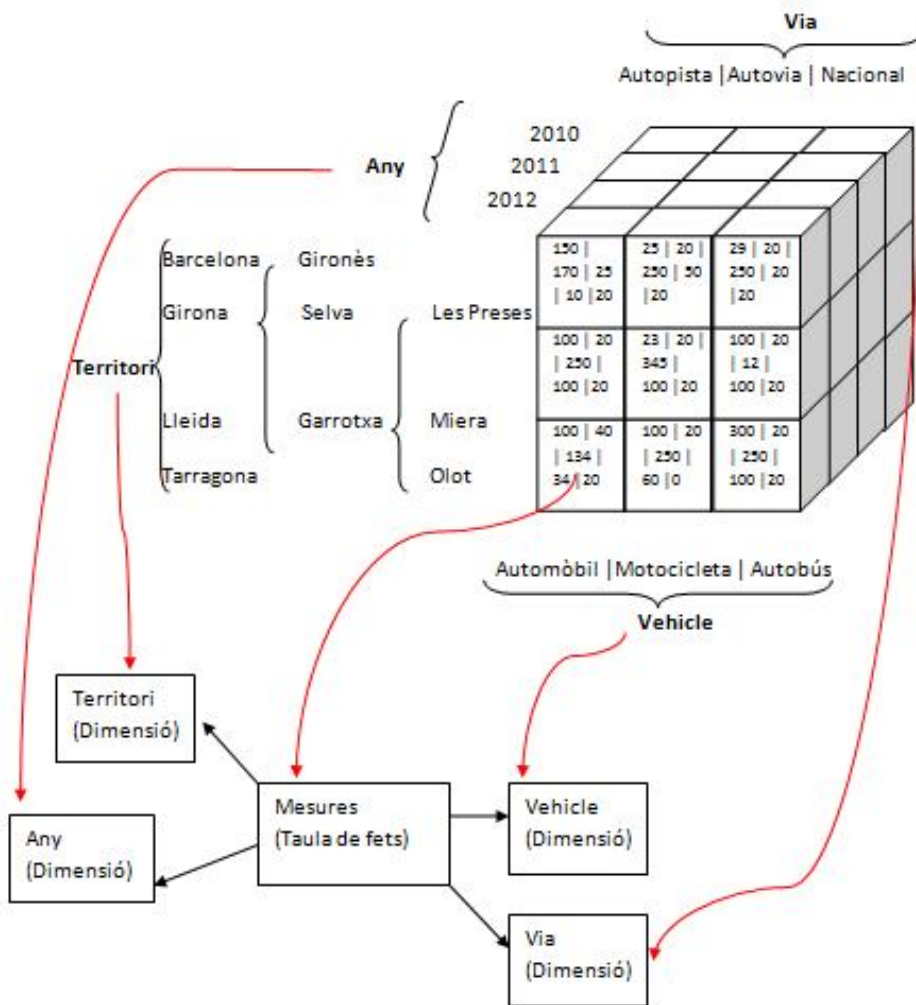
Les dimensions utilitzades en el model permeten consultar la informació de forma agregada per territori, per tipus de vehicle, per tipus de via i per any:



Il·lustració 5: cub de dades

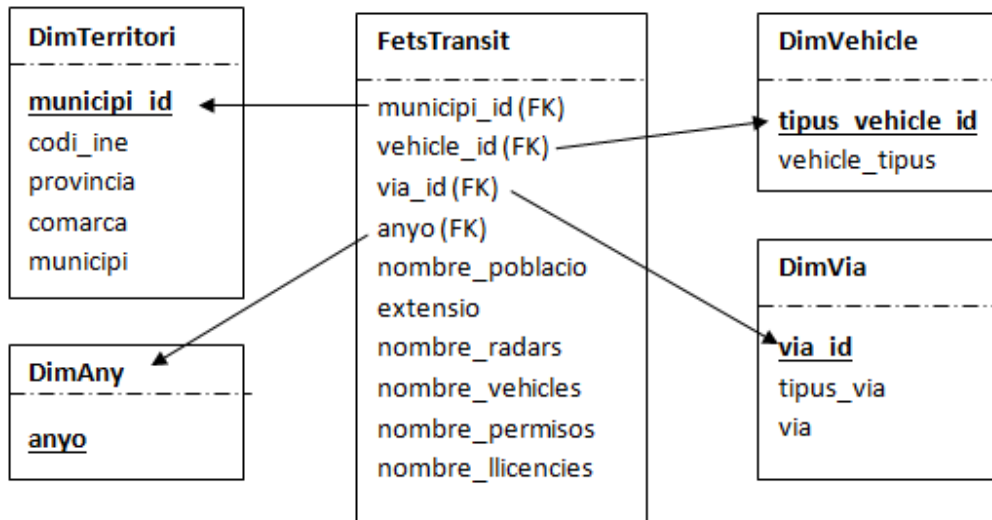
Dins d'una dimensió podem distingir varis atributs organitzats de forma jeràrquica. La profunditat d'aquesta jerarquia determina el nivell de granularitat de la informació. En el model presentat, només la dimensió territori presenta varis nivells d'agregació: Província > Comarca > Municipi.

Represento a continuació el cub amb diferents atributs de les dimensions, i de quina forma es combinen per formar taules amb relacions lògiques:



II-lustració 6: Exemple de Cub

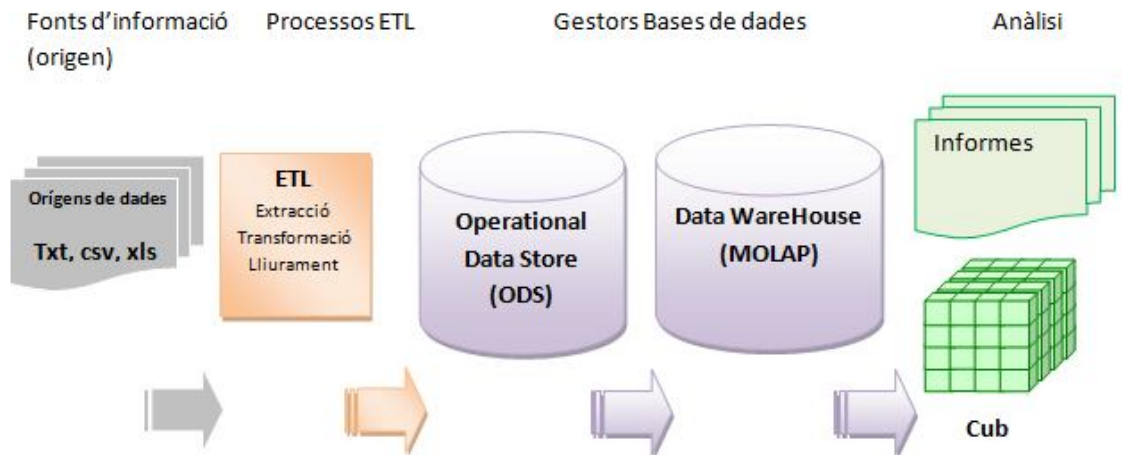
Per l'elaboració de les taules de dimensions s'utilitza el model en estrella. Es tracta de l'estructura més senzilla, on cada dimensió es basa en una taula de dimensions vinculada directament a la taula de fets mitjançant una relació de clau primària a clau forana:



Il·lustració 7: Model en estrella

3. Capítol 3, Disseny

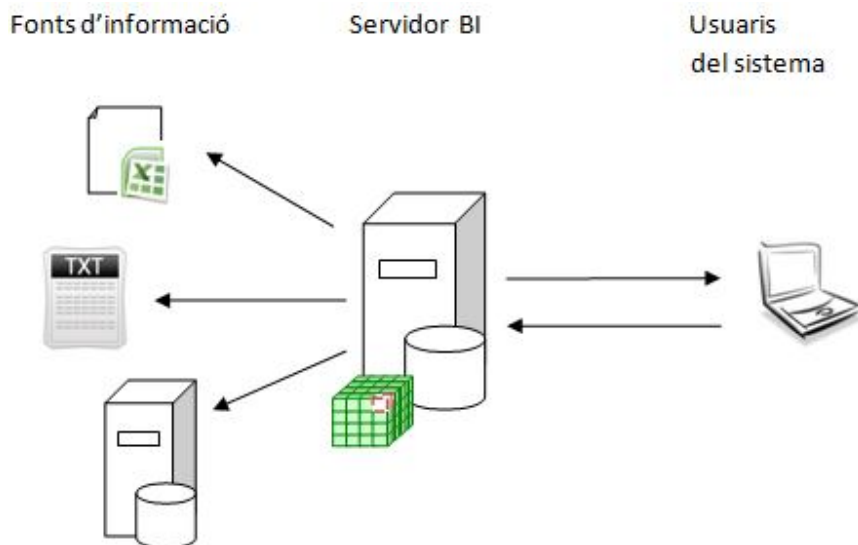
3.1. Diagrama de l'arquitectura del software



Il·lustració 8: Diagrama de l'arquitectura del software

Des del *Pentaho Data Integration* es creen connexions als diferents orígens de dades, que en aquest cas són fitxers amb diferents formats, i a continuació, s'utilitzen processos ETL per a extraure les dades, transformar-les de forma homogènia, i finalment carregar-les en una base de dades d'*Oracle XE*. Aquesta base de dades és el pas previ per a desnormalitzar-la en un model analític multi-dimensional que formarà el magatzem de dades del client. Amb l'*SQL Server Analysis Services*, modelarem la informació amb cubs i usarem eines OLAP per què siguin consultables per l'usuari. Atès que l'*SQL Server* inclou solucions de BI, s'utilitza per crear els informes (*Reporting*) i els indicadors agrupats en un quadre de comandament (*Dash Board*)

3.2. Diagrama de l'arquitectura hardware



Il·lustració 9: Arquitectura del hardware

Els components hardware necessaris per a desenvolupar el projecte de forma molt simplificada son:

- Un servidor virtual, amb el programari descrit a l'arquitectura del programari, per a ubicar el magatzem de dades i per què realitzi les funcions de servidor analític.
- Un client com a mínim per a fer les consultes dels informes i del cub al servidor.

3.3. Disseny de la base de dades i diagrama del model físic

3.3.1. Disseny lògic

El disseny lògic, és la transformació del model ER al model relacional. Es tracta de transformar totes les entitats en relacions definits a l'apartat de diagrama del model conceptual, i totes les interrelacions 1:1 i 1:N en claus foranies d'aquestes relacions:

DEMARCACIO (codi-demarcacio, nom-demarcacio)

COMARCA (codi-comarca, nom-comarca, codi-demarcacio)
on {codi-demarcacio} referencia DEMARCACIO

MUNICIPI (nom-municipi, codi-ine, extensio, codi-demarcacio, codi-comarca)
on {codi-demarcacio} referencia DEMARCACIO,
{codi-comarca} referencia COMARCA
i codi-comarca pot prendre valors nuls

CONDUCTOR (codi-gener, gener)

ANY (any)

RADAR (codi-via, via)

VEHICLE (codi-tipus, tipus)

UBICACIO(codi-ine, codi-via, quantitat)
on {codi-ine} referencia MUNICIPI
i {codi-via} referencia RADAR

CENS (codi-ine, any, poblacio)
on {codi-ine} referencia MUNICIPI
i {any} referencia ANY

REGISTRE_PERMIS (codi-ine, any, codi-genero, permisos, llicencies)
on {codi-ine} referencia MUNICIPI,
{any} referencia ANY
i {codi-genero} referencia CONDUCTOR

REGISTRE_VEHICLE (codi-ine, any, codi-tipus, nombre)
on {codi-ine} referencia MUNICIPI,
{any} referencia ANY
i {codi-tipus} referencia VEHICLE

3.3.2.Disseny físic

Hi ha implementat un ODS (*Operational Data Store*) que s'utilitza per a carregar la informació dels fitxers en un sistema més estable i protegit. Es compon de les següents taules sense relacionar:

Taula	Camps
TR_CENS	CODI_INE VARCHAR2(10) NOT NULL , ANYO VARCHAR2(4) NOT NULL , POBLACIO NUMBER NOT NULL
TR_CONDUCTOR	CODI_INE VARCHAR2(10) , MUNICIPI VARCHAR2(150) NOT NULL , DEMARCACIO VARCHAR2(100) , PERMISOS_DONA NUMBER(6,0) DEFAULT 0 , PERMISOS_HOME NUMBER(6,0) DDEFAULT 0 , LLICENCIES_DONA NUMBER(6,0) DDEFAULT 0 , LLICENCIES_HOME NUMBER(6,0) DDEFAULT 0 , ANYO VARCHAR2(4)
TR_MUNICIPI	CODI_INE VARCHAR2(10) NOT NULL , MUNICIPI VARCHAR2(150) NOT NULL , COMARCA VARCHAR2(100) , DEMARCACIO VARCHAR2(100) , EXTENSIO NUMBER DEFAULT 0 , CONSTRAINT TR_MUNICIPI_PK PRIMARY KEY (CODI_INE)
TR_RADAR	VIA VARCHAR2(100) NOT NULL , TIPUS_VIA VARCHAR2(100) , MUNICIPI VARCHAR2(150) NOT NULL , COMARCA VARCHAR2(100) , DEMARCACIO VARCHAR2(100) , CODI_INE VARCHAR2(10)
TR_VEHICLE	CODI_INE VARCHAR2(10) NOT NULL , MUNICIPI VARCHAR2(150) NOT NULL , TIPUS VARCHAR2(100) , NOMBRE NUMBER DEFAULT 0 , ANYO VARCHAR2(4)

A continuació definim les taules que formen el DataWareHouse pròpiament dit. Una taula unificada que serveix de base per als informes, i les taules que alimenten el cub MOLAP:

Taula	Camps
TR_CONSOLIDAT	INDICADOR VARCHAR2(100) NOT NULL , DIM_CODI_INE VARCHAR2(10) NOT NULL , DIM_MUNICIPI VARCHAR2(150) NOT NULL , DIM_COMARCA VARCHAR2(100) , DIM_DEMARCACIO VARCHAR2(100) , DIM_VIA VARCHAR2(100) , DIM_TIPUS_VIA VARCHAR2(100) , DIM_TIPUS_VEHICLE VARCHAR2(20) , YYYY VARCHAR2(4) , MET_EXTENSIO NUMBER , MET_POBLACIO NUMBER , MET_PERMISOS NUMBER , MET_LLICENCIES NUMBER , MET_NOMBRE_VEHICLES NUMBER , MET_NOMBRE_RADARS NUMBER CONSTRAINT TR_CONSOLIDAT_UK1 UNIQUE (INDICADOR , DIM_MUNICIPI , YYYY)
TR_DIM_VEHICLE	TIPUS_VEHICLE_ID NUMBER NOT NULL , VEHICLE_TIPUS VARCHAR2(100 BYTE) , CONSTRAINT TR_DIM_VEHICLE_PK PRIMARY KEY (TIPUS_VEHICLE_ID)
TR_DIM_VIA	VIA_ID NUMBER NOT NULL , TIPUS_VIA VARCHAR2(100 BYTE) , VIA VARCHAR2(100 BYTE) , CONSTRAINT TR_DIM_VIA_PK PRIMARY KEY (VIA_ID)
TR_DIM_ANY	ANYO VARCHAR2(4 BYTE) NOT NULL , CONSTRAINT TR_DIM_ANY_PK PRIMARY KEY (ANYO)
TR_DIM_TERRITORI	MUNICIPI_ID NUMBER(6, 0) NOT NULL , CODI_INE VARCHAR2(10 BYTE) NOT NULL , MUNICIPI VARCHAR2(150 BYTE) NOT NULL , COMARCA VARCHAR2(100 BYTE) , PROVINCIA VARCHAR2(100 BYTE) , CONSTRAINT MUNICIPI_PK PRIMARY KEY (MUNICIPI_ID)
TR_FETS	MUNICIPI_ID NUMBER(6, 0)

	, VEHICLE_ID NUMBER , VIA_ID NUMBER , ANYO VARCHAR2(20 BYTE) , EXTENSIO NUMBER DEFAULT 0 , POBLACIO NUMBER DEFAULT 0 , RADARS NUMBER DEFAULT 0 , VEHICLES NUMBER DEFAULT 0 , LLICENCIES NUMBER DEFAULT 0 , PERMISOS NUMBER DEFAULT 0
--	--

3.4. Disseny i descripció dels informes creats

Hi ha dos tipus d'informes. Un per què la direcció disposi del darrera actualització dels indicadors requerits en forma de quadre de comandament, i l'altre tipus, per què es pugui navegar i consultar dades més detallades de qualsevol any.

Tots els informes estan dins el projecte QuadreComandament de Visaul Studio

3.4.1. Quadres de comandament

3.4.1.1. Quadre de comandament de la situació del trànsit (rp01_QC)

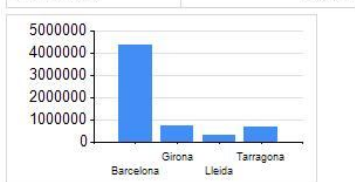
Aquest és l'informe principal del producte desenvolupat atès que presenta en una única vista, tots els darrers indicadors disponibles a la base de dades. Per a cada indicador, es presenta el darrer any disponible. És el tipus d'informe orientat a la direcció de l'organització per què no requereix cap acció addicional, és a dir, de forma dinàmica, busca els registres que contenen la darrera data o els que no tenen data informada:

FECRES:: QUADRE DE COMANDAMENT DE LA SITUACIÓ DEL TRÀNSIT

IND01. Total de vehicles

Actualització: 2012

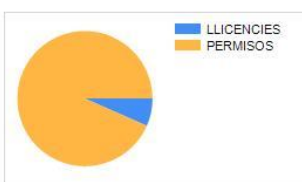
TIPUS	VEHICLES
Altres vehicles de motor	1072940
Autobusos	8519
Automòbils	3179182
Camions i furgonetes	720280
Motocicletes	665233
Reste vehicles de motor	379975
Tractors industrials	24442
Total vehicles	6050571



IND02. Total de conductors

Actualització: 2011

	LLICENCIES	PERMISOS
Barcelona	198986	2999685
Girona	40559	446926
Lleida	17091	248006
Tarragona	37214	457825
Total	293850	4152442



IND03. % de vehicles respecte la població

Actualització dades població: 2012

Actualització dades vehicles: 2012

Província	% Vehicles
Barcelona	79.03
Girona	99.75
Lleida	83.65
Tarragona	86.42



IND04. Densitat

Actualització dades població: 2012

Actualització dades vehicles: 2012

Província	Hab / Km2	Província	Veh / Km2
Barcelona	1116	Barcelona	882
Girona	237	Girona	237
Lleida	93	Lleida	78
Tarragona	193	Tarragona	167

IND05. No. vehicles / No. radars

Actualització dades vehicles: 2012

Província	Vehicles per radar
Barcelona	43154
Girona	19037
Lleida	8388
Tarragona	16314

IND06. % Conductors per radar

Actualització dades conductor: 2011

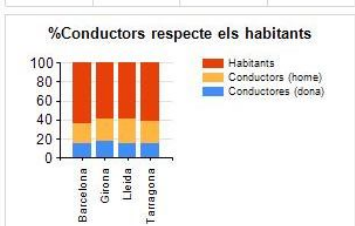
No. permisos	No. radars	% cond. per radar
4446292	217	0.004880%

IND07. Conductors vs habitants per gènere

Actualització dades conductors: 2011

Actualització dades població: 2012

	No. Cond. (Dona)	No. Cond. (Home)	Habitants
Barcelona	1307068	1891603	5515458
Girona	204895	282590	706141
Lleida	101997	163100	381017
Tarragona	198650	296389	773966
Total	1812610	2633682	7376582



IND08. Radars vs vehicles

Actualització dades vehicles: 2012

	No. Radars	No. Vehicles	% radars vs vehicles
Barcelona	101	4358597	0.0020
Girona	37	704378	0.0050
Lleida	38	318737	0.0120
Tarragona	41	668859	0.0060
Total	217	6050571	0.0250

IND09. Rati de vehicles per conductor

Actualització dades vehicles: 2012

Actualització dades conductor: 2011

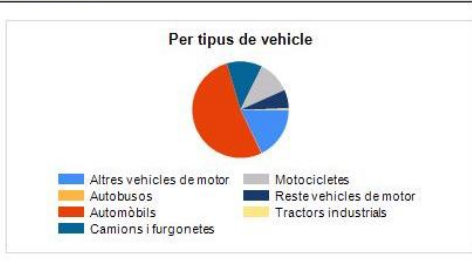
	No. Cond.	No. Vehicles	Rati veh. per conductor
Barcelona	3198671	4358597	136.26%
Girona	487485	704378	144.49%
Lleida	265097	318737	120.23%
Tarragona	495039	668859	135.11%
Total	4446292	6050571	136.08%

IND10. Quantitat vehicles / superfície territori

Actualització dades vehicles: 2012

Extensió del territori: 16031

	No. Vehicles	veh.km2
Altres vehicles de motor	1072940	66.93
Autobusos	8519	0.53
Automòbils	3179182	198.31
Camions i furgonetes	720280	44.93
Motocicletes	665233	41.50
Reste vehicles de motor	379975	23.70



II-lustració 10: Quadre de comandament de l'estat del trànsit

3.4.1.2. Quadre de comandament de l'evolució del trànsit (rp03_QC_Evolucio)

A diferència de l'anterior, el següent quadre de comandament es centra en l'evolució del trànsit, per tant, presenta únicament aquella informació que disposa de la dimensió temps:



FECRES:: QUADRE DE COMANDAMENT DE L'EVOLUCIÓ DEL TRÀNSIT

Evolució dels vehicles

	2007	2008	2009	2010	2011	2012	
Altres vehicles de motor	944349	1021005	1055804	1065978	1072940	1072940	
Automòbils	3036841	3155465	3175857	3165529	3177032	3179182	
Camions i furgonetes	708077	743945	742306	734200	727937	720280	
Vehicles de motor	4689267	4920415	4973967	4965707	4977909		

Evolució dels conductors *

LLicències	2008	2009	2010	2011	
Barcelona	59	57	61	59	
Girona	29	29	30	28	
Lleida	158	167	144	116	
Tarragona	314	313	267	244	

Permisos	2008	2009	2010	2011	
Barcelona	2955575	2965658	2984171	2999685	
Girona	437393	439571	444162	446926	
Lleida	243192	244102	245963	248006	
Tarragona	441434	444105	452161	457825	

* Es descarta el 2007 per què hi figura l'acumulat de les llicències i destorsionaria l'evolució.

II-lustració 11: Quadre de comandament de l'evolució del trànsit

En aquests tipus de presentació s'afegeix gràfics a nivell de registre per què ràpidament, la direcció conegui la tendència de cada indicador.

3.4.2. Informes parametritzats

Els següents informes, es presenten amb paràmetres per què l'usuari filtri la informació segons la necessiti. A més, s'habilita l'opció *drill-down* que li permet anar a buscar el detall de la informació atès que els informes es mostren inicialment agregats.

3.4.2.1. Informe de vehicles (rp02_Vehicles)

L'informe es mostra inicialment amb tota la informació disponible al magatzem de forma agregada per any i província. La mètrica que es mostra, és el nombre de vehicles.

L'usuari te habilitats els paràmetres de l'any, la província i el tipus de vehicle per cercar la informació que més li convingui.

Any	Província	Municipi	Tipus	Nombre
2012				448461
	Tarragona			448461
		Albinyana	Automòbils	1327
		Albinyana	Camions i furgonetes	360
		Alcanar	Camions i furgonetes	2115
		Alcanar	Automòbils	4988
		Alcover	Camions i furgonetes	923
		Alcover	Automòbils	2664
		Aldea (L')	Automòbils	2152
		Aldea (L')	Camions i furgonetes	1041
		Alforja	Automòbils	950
		Alforja	Camions i furgonetes	363
		Almoster	Camions i furgonetes	209

II-lustració 12: Informe vehicles filtrat

3.4.2.2. Informe de radars

El següent informe presenta el nombre de radars de les vies catalanes de forma agregada per província i comarca. Per defecte es mostra amb tota la informació disponible al magatzem de dades.

Els paràmetres que pot utilitzar l'usuari son la província i el tipus de via. P.ex. pot conèixer el nombre de radars ubicats a les autopistes de la província de Tarragona .

rp05_Permis.rdl [Design] rp04_Radars.rdl [Design] rp03_QC_Evolucio.rdl [Design] rp02_Vehicles.rdl [Des

Design Preview

Província Tipus de via

1 of 1 100% Find

Informe Radars

Província	Comarca	Municipi	Tipus via	via	Nombre
<input type="checkbox"/> Barcelona					101
<input type="checkbox"/> Girona					37
<input type="checkbox"/> Lleida					38
<input type="checkbox"/> Tarragona					41

II-lustració 13: Informe dels radars

La següent il·lustració mostra el mateix informe després de navegar a través dels diferents nivells d'informació. En aquest exemple interessava veure els radars ubicats a les poblacions de la comarca del Ripollès.


rp05_Permis.rdl [Design] rp04_Radars.rdl [Design] rp03_QC_Evolucio.rdl [Design] rp02_Vehicles.rdl [De

Design Preview

Província: Barcelona, Girona, Lleida, Tarr. Tipus de via: Autopista, Autovia, Autovia Be

1 of 1 100% Find

Informe Radars



Província	Comarca	Municipi	Tipus via	via	Nombre
+	Barcelona				101
-	Girona				37
	+	Alt Empordà			8
	+	Baix Empordà			5
	+	Garrotxa			4
	+	Gironès			6
	+	Pla de l'Estany			2
	-	Ripollès			2
		Sant Joan de les Abadesses	Autovia autonòmica	C-38	1
		Campdevàno	Carretera nacional	N-260	1
	+	Selva			10
+	Lleida				38
+	Tarragona				41

Il·lustració 14: Informe dels radars usant navegació drill-down

3.4.2.3. Informe de permisos

L'informe de permisos permet veure el nombre de permisos i llicències atorgades a Catalunya els darrers anys. En el cas dels permisos es presenta la dada acumulada; en canvi, les llicències es mostra únicament les lliurades un any concret.

Inicialment es carrega tota la informació del magatzem, però l'usuari pot filtrar la informació per tipus de permís, any i província.

rp05_Permis.rdl [Design] X rp04_Radars.rdl [Design] rp03_QC_Evolucio.rdl [Design] rp02_Veh


Design Preview

Tipus de permís LLICÈNCIA, PERMÍS Any 2007, 2008, 2009, 2010, 2011

Província Barcelona, Girona, Lleida, Tarr.

1 of 1 100%

Informe Permís



Any	Província	Municipi	Tipus	Nombre
2007				4058601
2008				4078154
2009				4094002
2010				4126959
2011				4152889

Il·lustració 15: Informe permís

A continuació es mostra la informació filtrada per els tipus de permisos que no son llicències de la província de Barcelona actualitzades al 2011:

rp05_Permis.rdl [Design] × rp04_Radars.rdl [Design] rp03_QC_Evolucio.rdl [Design] rp02

Design Preview

Tipus de permís: PERMÍS Any: 2011

Província: Barcelona

1 of 1 100%

Informe Permís



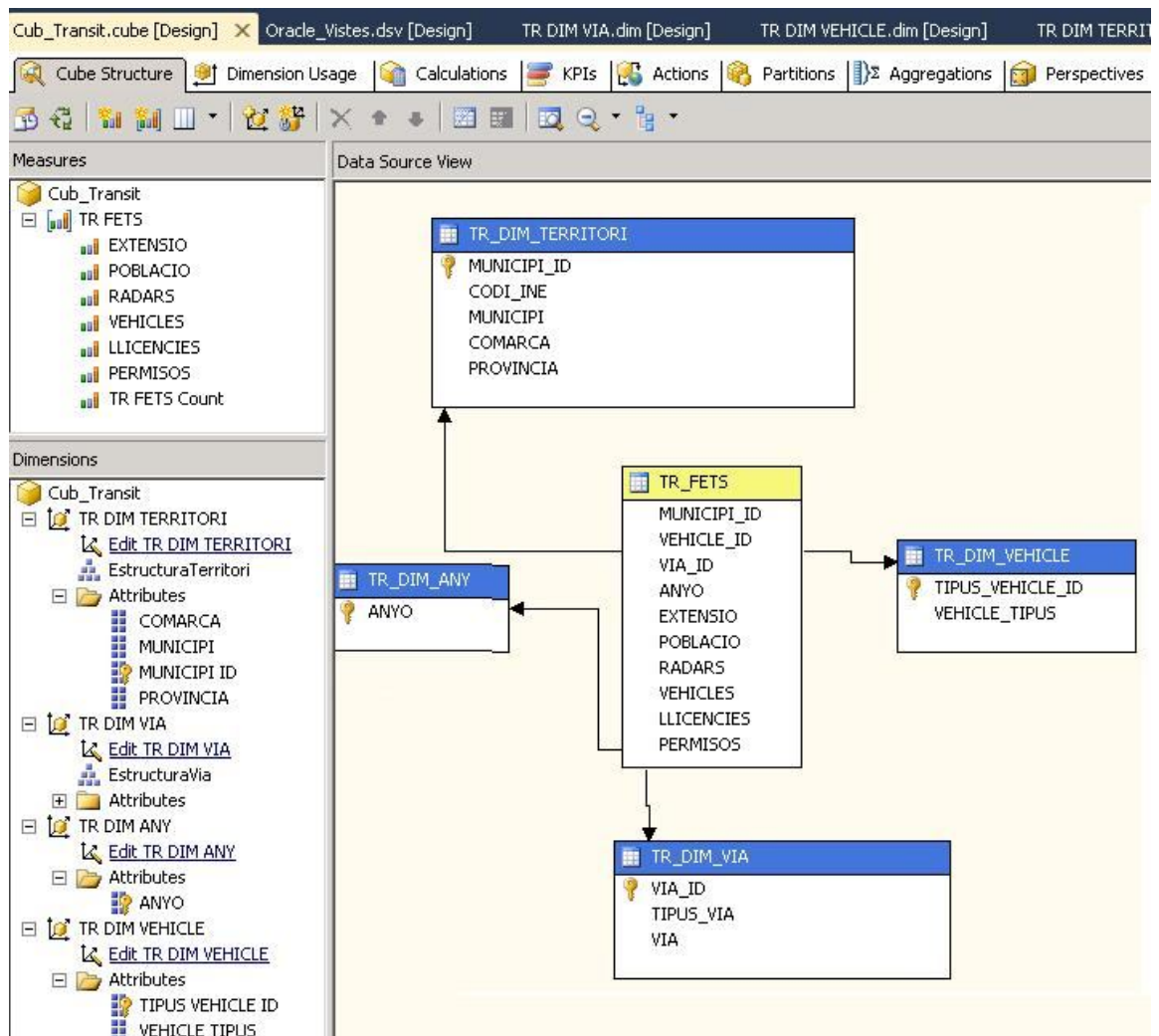
Any	Província	Municipi	Tipus	Nombre
2011				2999685
	Barcelona			2999685
		Abrera	PERMÍS	7225
		Aguilar de Segarra	PERMÍS	152
		Aiguafreda	PERMÍS	1514
		Alella	PERMÍS	6424
		Alpens	PERMÍS	188
		Ametlla del Vallès (L')	PERMÍS	5366
		Arenys de Mar	PERMÍS	8430
		Arenys de Munt	PERMÍS	5059
		Argentona	PERMÍS	7540
		Argençola	PERMÍS	118
		Artés	PERMÍS	3449
		Avinyonet del Penedès	PERMÍS	1057
		Avinyó	PERMÍS	1382

Il·lustració 16: Informe permís filtrat

3.5. Disseny i descripció del cub creats

A part dels informes, s'ha dissenyat un cub amb les dimensions i mètriques explicades a l'apartat del model conceptual, per donar als usuaris llibertat total per analitzar la informació disponible al magatzem de dades.

El disseny del cub anomenat Cub_Transit es mostra a continuació:



Il·lustració 17: Disseny del cub de trànsit

A continuació, es mostren uns exemples del seu ús a través de la pestanya del navegador del propi Visual Studio:

- En el primer exemple de la pàgina següent es veu la població de la comarca de la Selva censada al 2012 de cada municipi.
- En el segon, s'amplia les dades amb la informació de l'extensió de cada municipi i el nombre de vehicles també del 2012. Atès que l'extensió no té data d'actualització, s'afegeix al 2012, l'expressió *unknown*.

Dimension	Hierarchy	Operator	Filter Express
TR DIM TERRITORI	COMARCA	Equal	{ Selva }
TR DIM ANY	ANYO	Equal	{ 2012 }
<Select dimension>			

MUNICIPI	POBLACIO
Arbúcies	6741
Blanes	39785
Caldes de Malavella	7071
Hostalric	4018
Lloret de Mar	40837
Maçanet de la Selva	7175
Riudarenes	2180
Tossa de Mar	5952
Vidreres	7732
Vilobí d'Onyar	3053

II-lustració 18: Exemple Consulta a cub 1

Dimension	Hierarchy	Operator	Filter Expression
TR DIM TERRITORI	COMARCA	Equal	{ Selva }
TR DIM ANY	ANYO	Equal	{ 2012, Unknown, ... }
<Select dimension>			

MUNICIPI	POBLACIO	VEHICLES	EXTENSIO
Arbúcies	6741	5915	86
Blanes	39785	40892	18
Caldes de Malavella	7071	6319	57
Hostalric	4018	3793	3
Lloret de Mar	40837	36597	48
Maçanet de la Selva	7175	6961	46
Riudarenes	2180	2294	48
Tossa de Mar	5952	7604	38
Vidreres	7732	7625	48
Vilobí d'Onyar	3053	3670	32

II-lustració 19: Exemple Consulta a cub 2

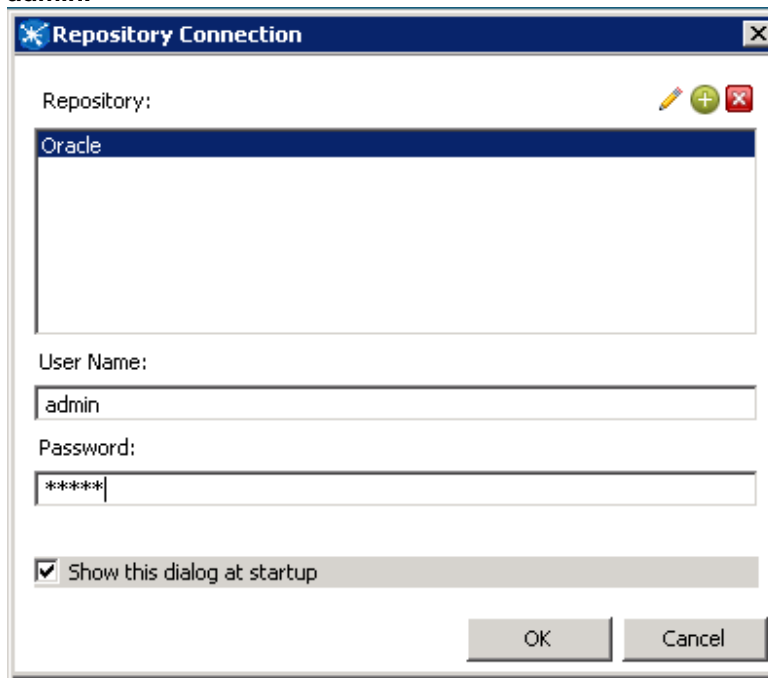
4. Captures de pantalla

4.1. Accessos als productes

a. PENTAHO

Per accedir a Pentaho hi ha un accés directe a l'escriptori a la carpeta D:\PDI. Primer s'executa el bat **Kitchen**, per mapejar els paths de JAVA, i a continuació el clica el bat **Spoon**.

A continuació ens connectem al repositori creat per al projecte: usuari **admin**, password **admin**.



Il·lustració 20: Connexió al repositori

Des de 'fitxer>>Obrir' podem accedir als jobs o les transformacions explicades en aquest document:

Nombre	Tipo
jActualitzarDW	job
jCarregarFitxersOrigen	job
jModelDimensional	job
Trabajo 1	job
Trans_cens	transformation
Trans_conductor	transformation
Trans_Consolidat	transformation
Trans_consolidat_cens	transformation
Trans_consolidat_municipi	transformation
Trans_Dim_Any	transformation
Trans_Dim_Territori	transformation
Trans_Dim_Vehicle	transformation
Trans_Dim_Via	transformation
Trans_Fets	transformation
Trans_municipi	transformation
Trans_radar	transformation
Trans_vehicle	transformation

II-lustració 21: Catàleg de jobs i transformacions del projecte

b. ORACLE

Base de dades/ SID: **XE**
 Esquema/usuari: **pentaho**
 Password: **source**
 Connexió al SQL Developer : **DW**
 Les taules i vistes utilitzades comencen amb el prefix **TR_**

c. SQL SERVER

Els informes i el cub es poden veure dins els projectes **QuadreComandament** i **pjt_transit_molap** del Visual Studio, i executar per mitjà de la pestanya *browser*. Està en mode integració per què la instància al servidor de reporting no està creada.

4.2. Detall de les ETLs. Funcionament dels processos de càrrega:

Els processos d'extracció, transformació i càrrega al DW estan desenvolupats amb el *Pentaho Data Integration*. Les càrregues es poden repetir tantes vegades com es necessitin per què es buiden les taules abans de cada càrrega.

4.2.1. Job principal:: Actualització del DW

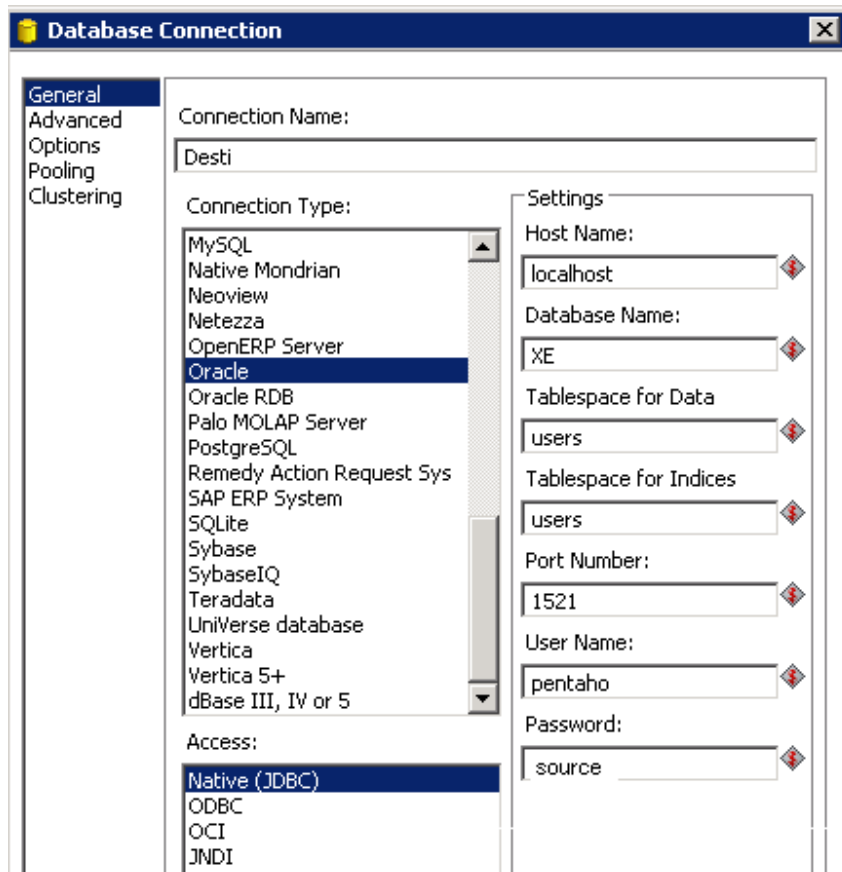
Tot el procés està unificat en el següent job que aniré desglossant:



II-lustració 22: job principal > job jActualitzarDW

Bàsicament, connecta amb els fitxers origen de les dades, els carrega a un ODS (Operational data store); consolida els indicadors en una única taula per unificar les dimensions comuns; i, finalment, transforma les dades consolidades de las taules a un model dimensional, és a dir, amb taules de dimensions i de fets en forma d'estrella il·lustrat a l'apartat del Model multi-dimensional

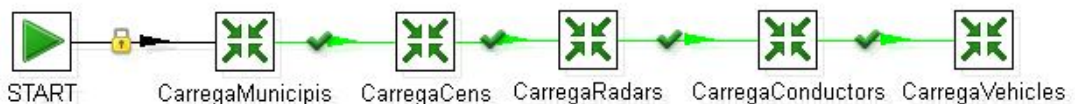
Comença el job amb el pas check Db connections que valida la connexió a la base de dades de destí:



Il·lustració 23: Connexió a Oracle des de Pentaho Data Integration

4.2.2. Job Càrrega de fitxers

Executa un conjunt de transformacions per fer la càrrega dels fitxers d'origen al model relacional creat a l'esquema de Pentaho (password source) dins l'Oracle XE:



Il·lustració 24: job jCarregarFitxersOrigen

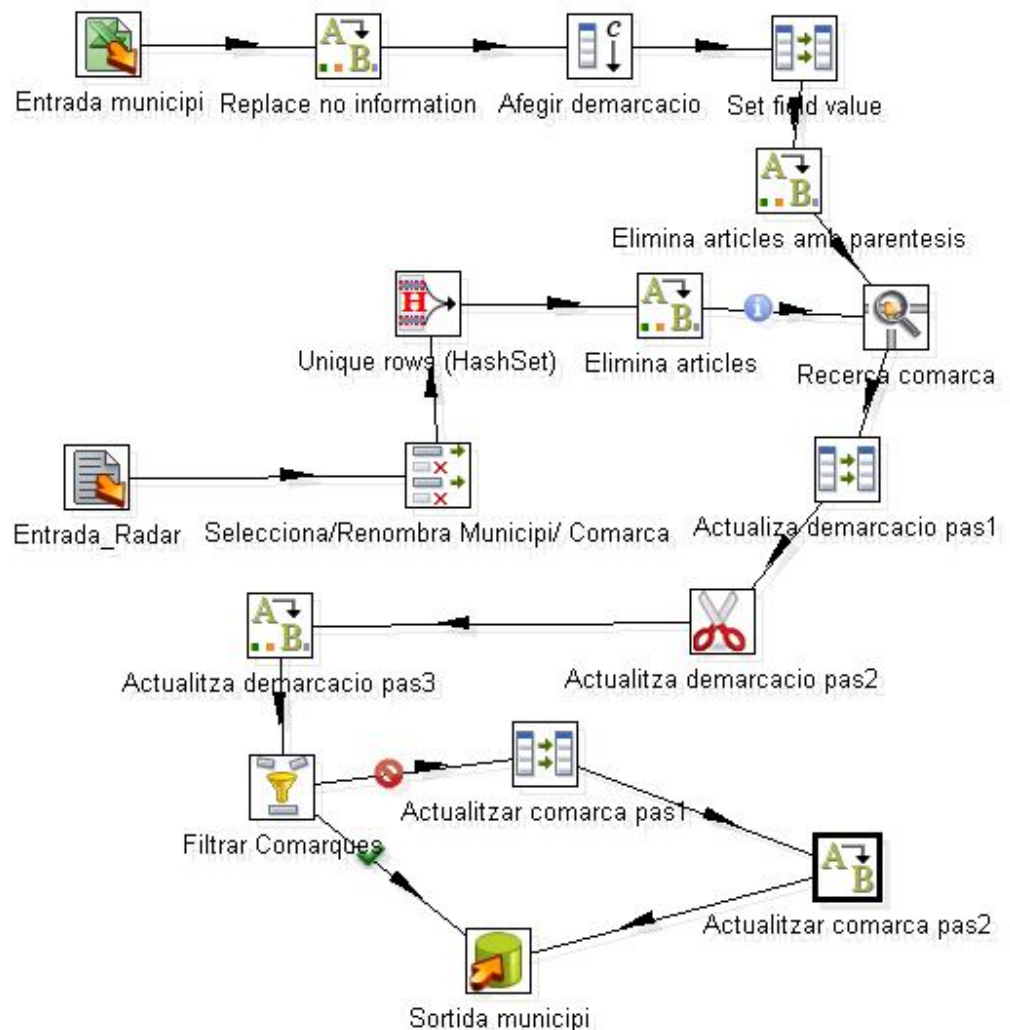
Resumeixo a la següent taula les passes que executa el job i les amplio a continuació:

Pas	Transformació	Fitxers d'origen	Taula destí
I. CarregaMunicipis	Trans_municipi	Dades_municipis.xls Radars_SCT.txt	TR_MUNICIPI
II. CarregaCens	Trans_cens	Dades_municipis.xls	TR_CENS
III. CarregaRadars	Trans_radar	Radars_SCT.txt	TR_RADAR
IV. CarregaConductors	Trans_conductor	Dades_conductors 2007.txt Dades_conductors 2008.txt Dades_conductors 2009.txt Dades_conductors 2010.txt Dades_conductors 2011.txt	TR_CONDUCTOR
V. CarregaVehicles	Trans_vehicle	Dades_vehicles.xls	TR_VEHICLE

I. CarregaMunicipis (Trans_municipi)

Transformació que extrau les dades del fitxer de municipis, complementa les dades de la comarca segons les dades obtingudes del fitxer de radars a partir de la canonització del municipi; afegeix la província segons els dos primers dígit del codi_ine; i finalment omple les comarques dels municipis que no apareixen al fitxer de radars amb informació del tipus "Altres comarques de Barcelona" per a la demarcació de Barcelona per exemple.

Per a cada municipi, es carrega a la taula **TR_MUNICIPI**: el **nom**, el **codi_ine**, la **comarca**, la **demarcació** i l'**extensió del territori**.

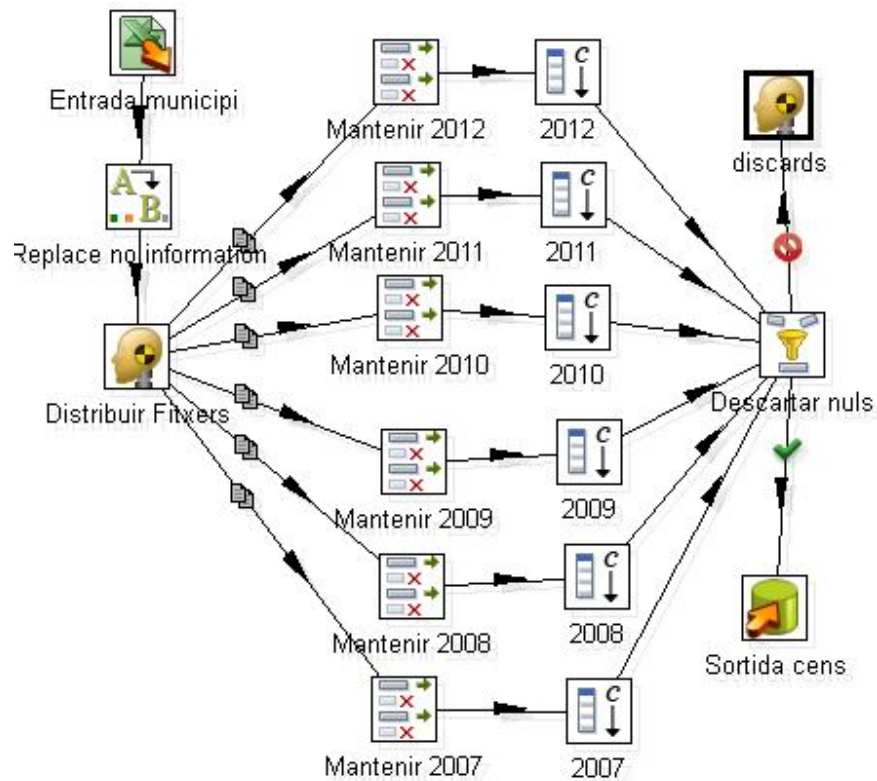


II-lustració 25: Transformació de municipi

II. CarregaCens (Trans_cens)

Transformació que extrau les dades referents al cens de cada municipi del fitxer de municipis. Cada columna etiquetada amb l'any es tracta per separat per unificar aquesta informació a nivell d'atribut (ANYO). El procés s'inicia reemplaçant els creuaments sense dades amb valors nuls que finalment es descarten al final de la transformació.

Per a cada any i cada municipi es carrega a la taula **TR_CENS** les següents dades: **l'any**, **el codi_ine** (la resta de la informació la tenim a la taula de municipis), i **el cens**.

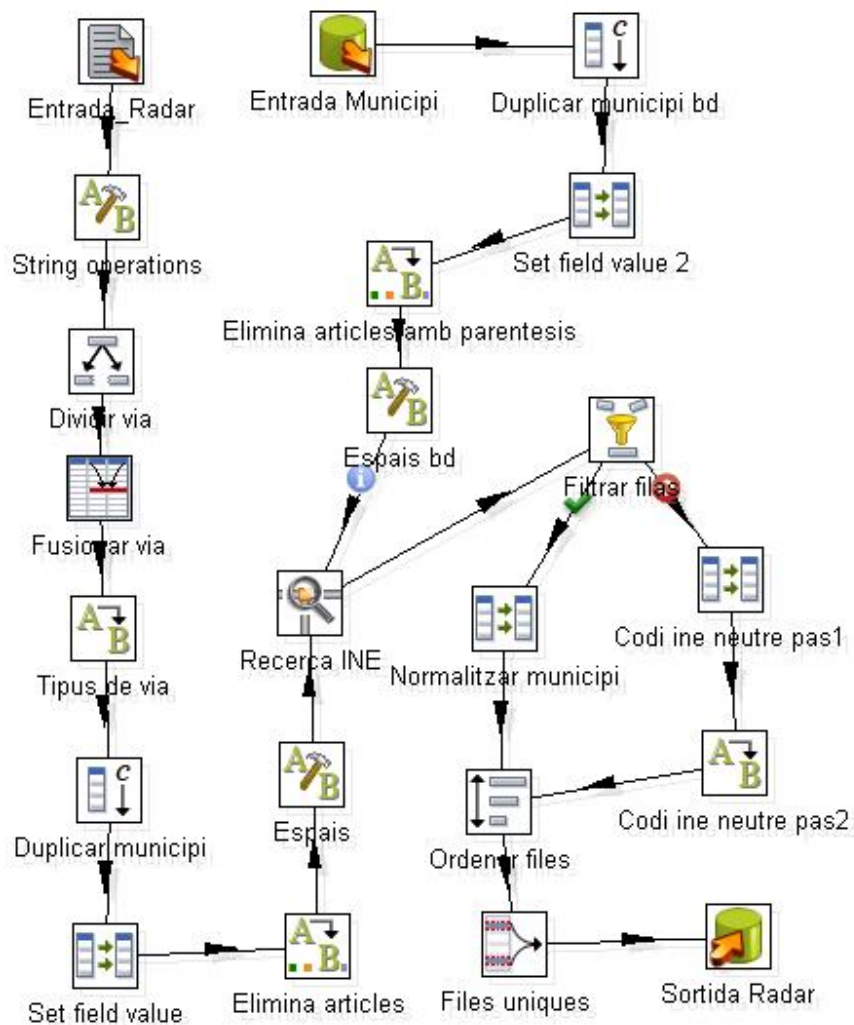


Il·lustració 26: Transformació de la població

III. CarregaRadars (Trans_radar)

Transformació que extrau les dades del fitxer de radars. El procés comença afegint informació del tipus de via (p.ex. A es autovia, AP es autopista...); després, consulta el codi ine de la taula de municipis després d'eliminar els articles per què tenen diferents posicions; i, finalment, s'informa dels municipis que no tenen codi ine a la bd amb codis neutre del tipus 25000 pels municipis de Lleida per exemple.

Els registres que es bolquen a la taula **TR_RADAR** contenen la següent informació: el **tipus de la via**, la **via**, el **municipi**, la **comarca**, la **demarcació**, el **codi_ine**, i el **nombre de radars**.

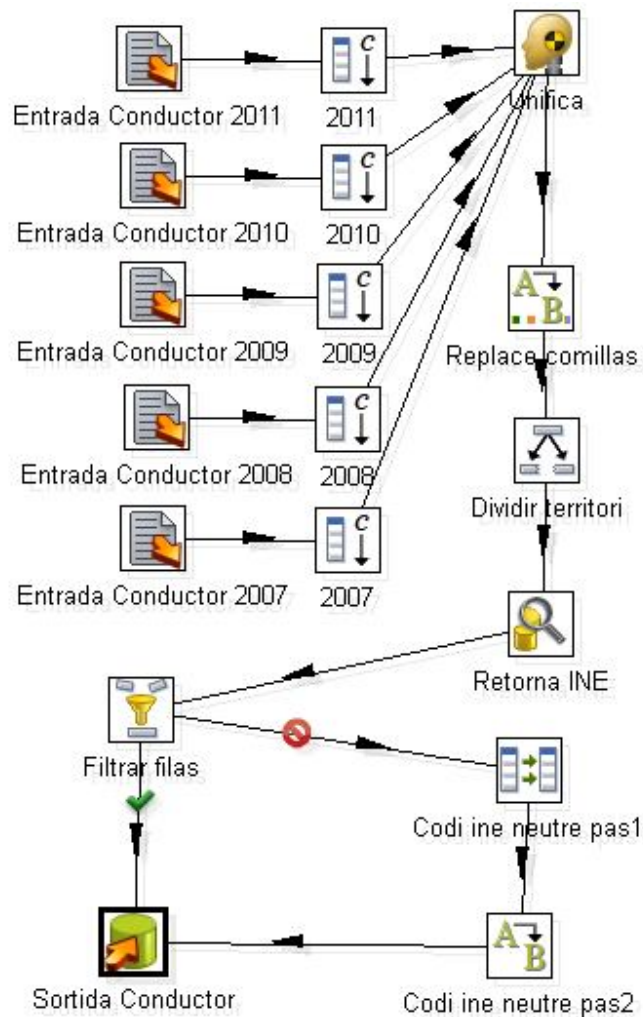


Il·lustració 27: Transformació dels radars

IV. CarregaConductors (Trans_conductor)

Transformació que extrau les dades dels permisos dels fitxers dels conductors. Comencem el procés unificant tots els fitxers afegint l'any del nom de cada fitxer a nivell d'atribut dels registres; eliminem les dobles cometes per tractar correctament la informació numèrica; i, a continuació recuperem si existeix el codi ine de la taula de municipis, altrament, li afegim un codi ine neutre del tipus 43000 per als municipis de Tarragona per exemple.

Els registres que es bolquen a la taula **TR_CONDUCTOR** contenen la següent informació: **el municipi, la comarca, la demarcació, el codi_ine, l'any i, els permisos i llicències tant de les dones com dels homes.**

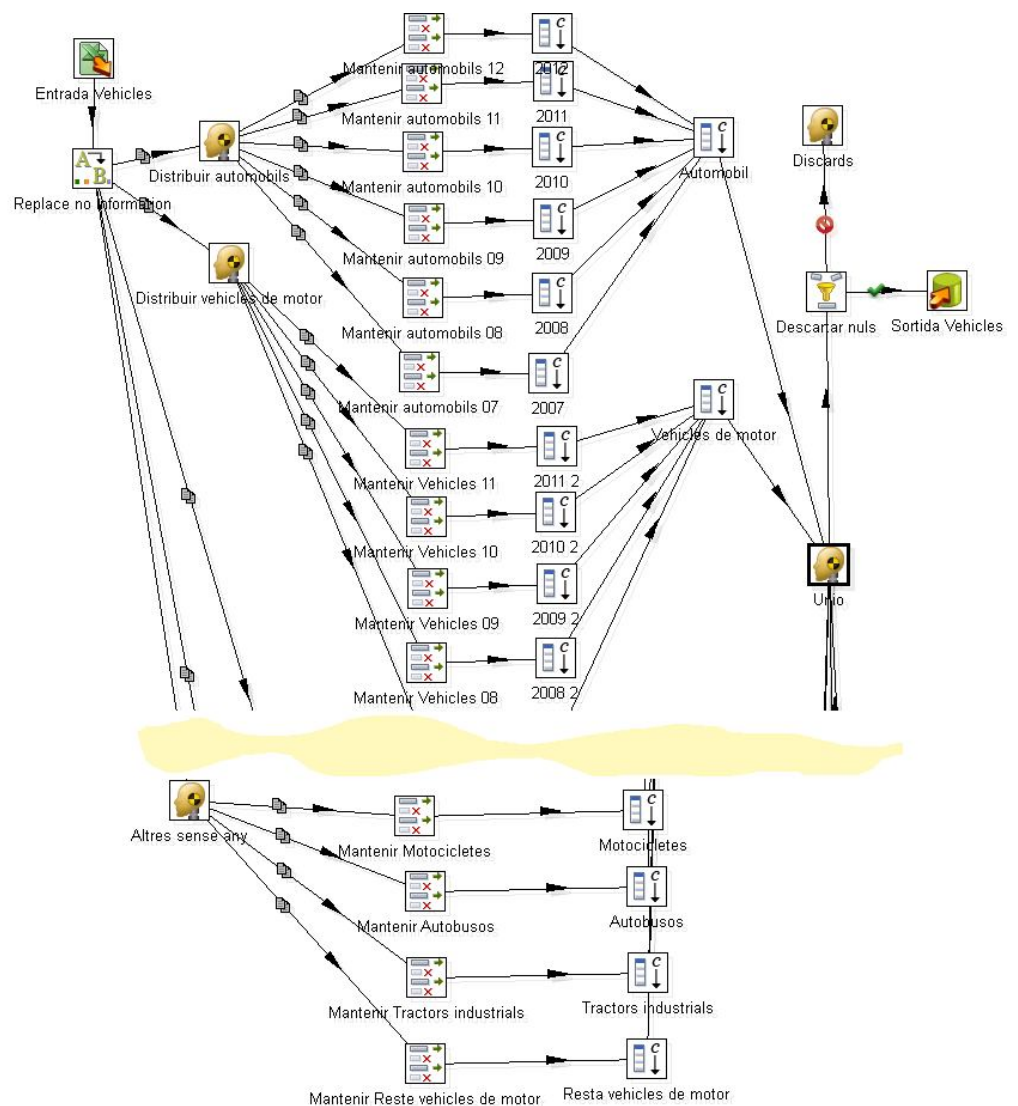


Il·lustració 28: Transformació dels conductors

V. CarregaVehicles (Trans_vehicle)

Transformació que extrau les dades del fitxer de vehicles. És un procés molt laboriós per què s'afegeix cada columna, que informa de l'any i del tipus de vehicle, a nivell d'atribut de cada registre. Conseqüentment, s'inicia un procés per tractar individualment cada porció del fitxer per normalitzar-los amb el següent denominador comú: municipi, codi_ine, tipus de vehicle, nombre de vehicles i l'any. Per els vehicles que no tenen informat l'any es desa el camp en blanc. Finalment, es descarten aquells registres que no tenen informat el nombre de vehicles.

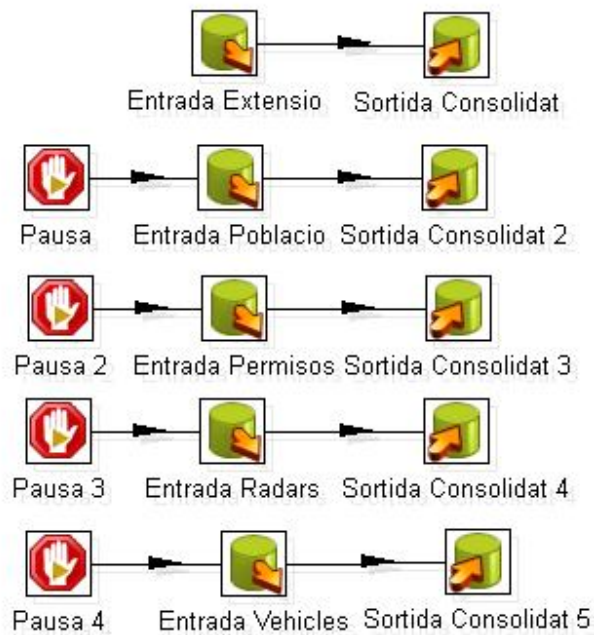
Es carreguen les següents dades a la taula **TR_VEHICLE**: **el municipi, el codi_ine, l'any, el tipus de vehicle, i el nombre de vehicles**. Atès que tots els municipis tenen informats el codi_ine recuperarem, si s'escau, la resta d'informació del territori.



Il·lustració 29: Transformació dels vehicles

4.2.3. Transformació de consolidació dels indicadors

ConsolidaIndicadors (Trans_consolidat) és una transformació que normalitza les taules del model relacional en una taula de consolidat, amb l'objectiu de normalitzar els atributs i etiquetar cada registre amb un indicador de l'origen (p.ex. "CENS" per a les dades que mostren la població de cada municipi i any):



Il·lustració 30: Transformació Trans_Consolidat

4.2.4. Job de transformació al model dimensional

jModelDimensional, anomenat TransformaModelDimensional al job principal, és un treball que s'ocupa de transformar la taula consolidada en un model dimensional (MOLAP) per què sigui intel·ligible per l'SQL Server i faciliti la creació del cub. Identifica amb claus úniques cada membre de la dimensió per què es puguin referenciar des de la taula de fets:



Il·lustració 31: job jModelDimensional

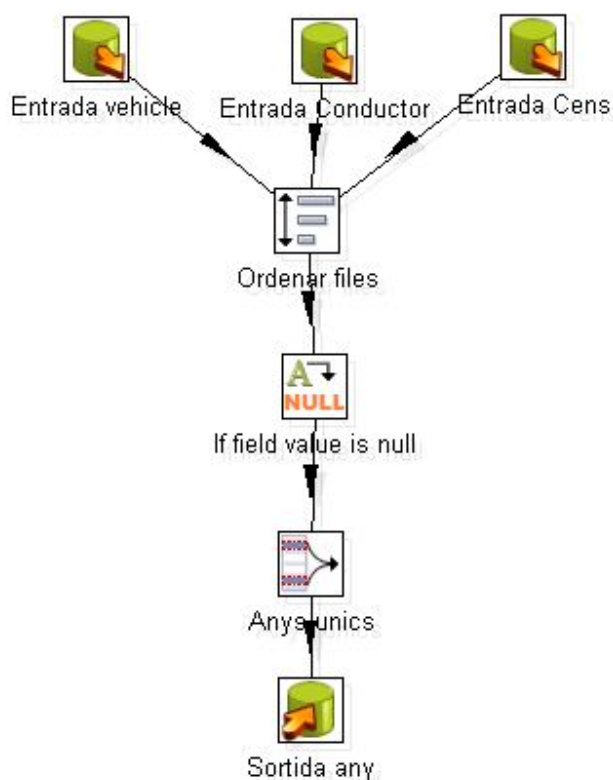
Resumeixo a la següent taula les fases que executa el job i les amplio a posteriorment:

Pas	Transformació	Taula origen	Taula destí
I. DimensioAny	Trans_Dim_Any	TR_VEHICLE TR_CONDUCTOR TR_CENS	TR_DIM_ANY
II. DimensioTerritori	Trans_Dim_Territori	TR_RADAR TR_CONDUCTOR TR_MUNICIPI	TR_DIM_TERRITORI
III. DimensioVehicle	Trans_Dim_Vehicle	TR_VEHICLE	TR_DIM_VEHICLE

IV. DimensioVia	Trans_Dim_Via	TR_RADAR	TR_DIM_VIA
V. Fets	Trans_Fets	TR_CONSOLIDAT	TR_FETS

I. DimensióAny (Trans_Dim_Any)

Transformació que extreu els anys de les taules que contenen aquesta informació i les unifica i ordena a la taula **TR_DIM_ANY**.



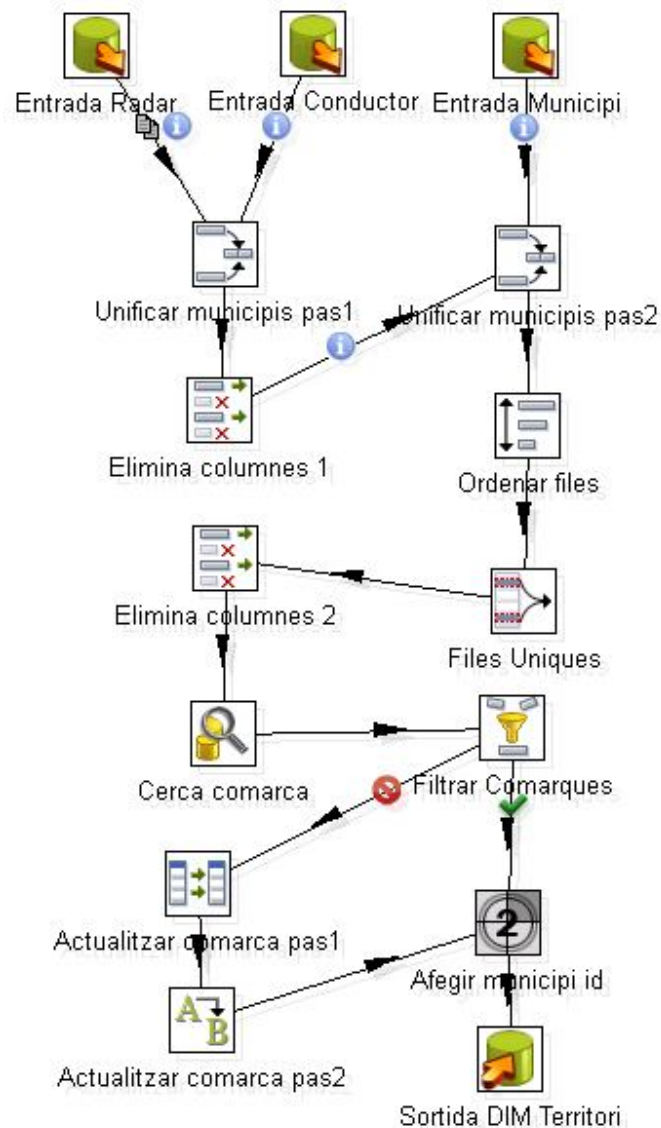
II-lustració 32: Dimensió Any

II. DimensioTerritori (Trans_Dim_Territori)

Transformació que extreu els municipis (codi ine, nom i demarcació) de les diferents taules; els complementa amb informació de la comarca, i se'ls assigna una de neutra sinó en tenen; finalment, s'afegeix un identificador únic a cada municipi.

La informació unificada es bolca a la taula **TR_DIM_TERRITORI**:

- municipi_id
- codi_ine
- municipi
- comarca
- provincia



II-lustració 33: Dimensió Territori

- III. DimensioVehicle (Trans_Dim_Vehicle)
 Transformació que extreu els tipus de vehicle de **TR_VEHICLE**, els identifica de forma inequívoca, i els bolca a la taula **TR_DIM_VEHICLE**:



II-lustració 34: Dimensió Vehicle

- IV. DimensioVia (Trans_Dim_Via)
 Transformació que extreu els tipus de via de **TR_VIA**, els identifica de forma inequívoca, i els bolca a la taula

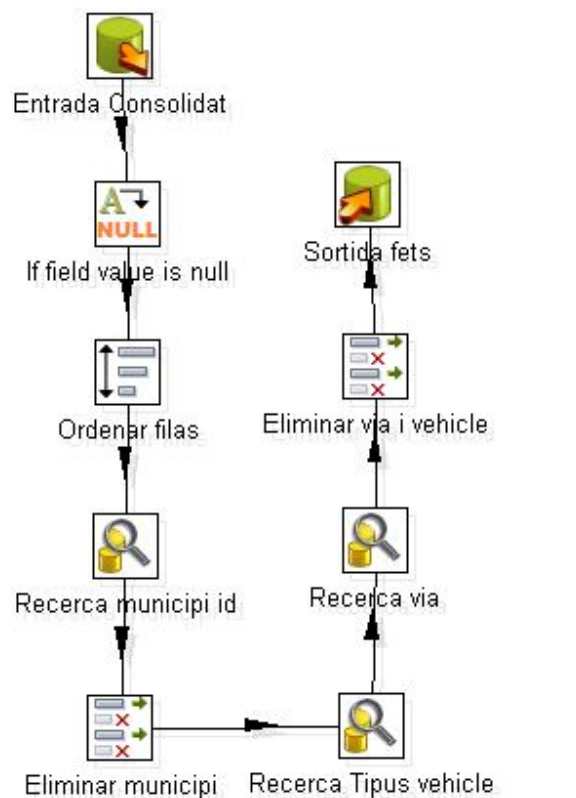
TR_DIM_VIA:



Il·lustració 35: Dimensió via

V. Fets (Trans_Fet)

Transformació que extreu les dades a TR_CONSOLIDAT i les normalitza amb els codis de les diferents dimensions. Els valors numèrics nuls, els hi assigna un 0, i als anys nuls, el text 'YYYY'. Es bolca la informació a la taula **TR_FETS: municipi_id, vehicle_id, via_id, anyo, extensio, poblacio, radars, vehicles, llicencies i permisos:**



Il·lustració 36: Transformació dels fets

5. Conclusions

Abans d'iniciar el projecte, havia participat en manteniments de reports de BI i d'alguns quadres de comandament però no havia tingut l'oportunitat de practicar tot el procés que requereixen aquests entorns des de l'inici.

Els desconeixements de les eines, la quantitat de funcionalitats repetides entre els productes (P.ex. Cub de Pentaho, cub de l'SQL, cub d'Oracle...), i la seva configuració, m'ha penalitzat en excés la consecució de les fites a mitjans del projecte. Aquesta experiència em servirà per a planificar millor els projectes i no infravalorar d'antuvi la dificultat de les eines de BI.

Al final del projecte, he après multitud de funcionalitats, i he après a superar diverses dificultats pròpies de la comunicació entre tants entorns. A part, he comprovat la potència d'aquet tipus de solucions.

Tot i no estar content amb el resultat de la memòria, ni del servidor de reports que no he aconseguit fer funcionar, si que valoro haver creat un quadre de comandament amb la actualització automàtica dels darrers indicadors sense cap tipus d'intervenció manual.

També estic satisfet dels processos ETLs atès que, en contra del que pensava inicialment, no hi ha hagut cap modificació manual.

Per concloure, estic satisfet del que he après, i espero repercutir-lo a la feina atès que tots els clients per els que he treballat, tenen curiositat en conèixer el que el Business Intelligence els hi pot oferir.

6. Línies d'evolució futur

Les línies de treball que haurien de desenvolupar-se son les següents:

- Crear un servidor de reports a producció.
- Crear un manual de manteniment de les ETLs per manipular les transformacions dels fitxers que contenen informació temporal.
- Cal associar a una tasca planificada el job principal de Pentaho que inicia totes les ETLs
- Per a la pròpia memòria, crear el glossari per a definir els termes tècnics