



Sistema d'intel·ligència de negoci per a l'anàlisi dels tractaments de reducció del colesterol

Raimon Silvestre Puig

Màster Universitari en Enginyeria Informàtica
Business Intelligence

David Amorós Alcaraz
Ferran Prados Carrasco

Gener de 2019



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

FITXA DEL TREBALL FINAL

Títol del treball:	<i>Sistema d'intel·ligència de negoci per a l'anàlisi dels tractaments de reducció del colesterol</i>
Nom de l'autor:	<i>Raimon Silvestre Puig</i>
Nom del consultor/a:	<i>David Amorós Alcaraz</i>
Nom del PRA:	<i>Ferran Prados Carrasco</i>
Data de lliurament (mm/aaaa):	<i>01/2019</i>
Titulació o programa:	Màster Universitari en Enginyeria Informàtica
Àrea del Treball Final:	Business Intelligence
Idioma del treball:	<i>Català</i>
Paraules clau	<i>LDL Business Intelligence</i>

Resum del Treball:

La finalitat del treball és crear una plataforma *Business Intelligence* (BI) implementada amb tecnologies *Open Source* que possibiliti l'anàlisi de la informació que ha estat generada durant un experiment que té com a objectiu comprovar l'eficàcia dels diferents tractaments en la reducció dels nivells de colesterol.

Per tal de realitzar aquest anàlisi en l'entorn de *Business Intelligence* es requerirà dissenyar i implementar un magatzem de dades (*Data Warehouse*) que permeti emmagatzemar la informació adquirida. Així mateix, s'hauran de programar els processos d'extracció, transformació i càrrega (ETL) per tal d'alimentar el magatzem de dades amb els fitxers base facilitats. Finalment s'analitzen i implementen les eines necessàries, entre les diferents plataformes *BI Open Source*, per disposar de la capa de programari necessària que permeti l'anàlisi multidimensional de la informació.

La metodologia utilitzada ha estat la descomposició del treball en fases diferenciades que corresponen a unitats funcionals. Aquestes etapes han estat realitzades de forma seqüencial, tot i això, cada una de les fases implica la revisió iterativa d'una o de diverses etapes dels nivells anteriors.

El projecte respon amb eficàcia a les diferents preguntes analítiques que s'estableixen com a objectiu inicials. D'altra banda, conclou amb un entorn d'Intel·ligència de Negoci que executa consultes complexes i retorna dades precises, en un temps de resposta baix, que pot ser utilitzat per analistes experts en el domini i uns coneixements mínims d'ofimàtica.

Abstract:

The purpose of the work is to create a Business Intelligence platform (BI) implemented with Open Source technologies that enables the analysis of the information generated during an experiment whose objective is to check the effectiveness of the different treatments in the reduction of the levels of cholesterol

In order to carry out this analysis in the Business Intelligence environment, it will be necessary to design and implement a Data Warehouse that allows you to store the information acquired. Likewise, the extraction and processing and loading (ETL) processes must be programmed to load the data warehouse with the source file. Finally, the necessary tools are analyzed and implemented between the different Open Source BI platforms in order to have the software layer that allows the multidimensional analysis of the information.

The methodology used was the decomposition of work in differentiated phases that correspond to functional units. These stages have been carried out sequentially, however, each phase involves the iterative revision of one or several stages of the previous levels.

The project responds effectively to the different analytical questions that are set as the initial objective. On the other hand, concludes with a Business Intelligence environment that performs complex queries and returns precise data in a low response time, which can be used by expert analysts in the domain and not need advanced computer skills.

Índex

Índex	iii
1. Introducció.....	1
1.1 Context i justificació del Treball	1
1.2 Objectius del Treball.....	2
1.3 Enfocament i mètode seguit	3
1.4 Planificació del Treball.....	3
1.5 Breu sumari de productes obtinguts	6
1.6 Breu descripció dels altres capítols de la memòria	7
2. Anàlisi i selecció de les eines BI.....	9
2.1 <i>Bussines Intelligence</i>	9
2.2 Anàlisi i d'eines BI	9
2.3 Selecció entorn BI	13
2.4 Anàlisi i selecció eines SGBD	13
3. Model dimensional	15
3.1 El Data Warehouse	15
3.2 Anàlisi de les dades.....	16
3.3 Model Conceptual	17
3.4 Implementació física del model de dades.....	24
4. Processos ETL	25
4.1 Processos de càrrega transformació i extracció.....	25
4.2 Dimensió activity.	25
4.3 Dimensió diet.....	26
4.4 Dimensió date.	27
4.5 Dimensió patient.....	28
4.6 Taula de fets registres.....	29
5. Implementació eines OLAP	31
5.1 Eines OLAP i multidimensionalitat.....	31
5.2 Disseny i implementació.....	32
5.3 Publicació i proves.	33
6. Anàlisis de dades	35
6.1 Abast de l'anàlisi.	35
6.2 Mètriques.....	35
6.3 Quina és la relació entre els diferents tractaments i l'evolució dels pacients?	36
6.4 Existeixen teràpies més eficaces?	37
6.5 Ha influït en el resultat, els hàbits dels pacients?	38
6.6 L'evolució al llarg del temps, per un mateix tractament , depenen d'algun factor com els hàbits?	40
6.7 Hi ha diferències en el resultat d'un tractament segons el lloc geogràfic del pacient?	43

6.8 Hi ha algun període de l'any on el tractament sigui més o menys efectiu?	43
7. Conclusions.....	45
8. Glossari	47
9. Bibliografia.....	49
10. Annexos	51
10.1 Creació Schema de MySQL	51
10.2 Cub OLAP XML.....	53

Llista de figures

- Figura 1. Diagrama de Gantt amb la planificació del projecte**
- Figura 2. Breu esquema de la plataforma BI obtinguda**
- Figura 3. Representació de les dades disponibles**
- Figura 4. Diagrama del model d'alt nivell**
- Figura 5. Diagrama físic del model de dades**
- Figura 6. Seqüència de Processos ETL**
- Figura 7. Processos ETL de la dimensió activity**
- Figura 8. Dades normalitzades de la dimensió activity**
- Figura 9. Processos ETL de la dimensió diet**
- Figura 10. Dades normalitzades de la dimensió diet**
- Figura 11. Processos ETL de la dimensió date**
- Figura 12. Dades normalitzades de la dimensió date**
- Figura 13. Processos ETL de la dimensió patient**
- Figura 14. Dades normalitzades de la dimensió patient**
- Figura 15. Processos ETL de la taula de fets**
- Figura 16. Dades normalitzades de la taula de fets**
- Figura 17. Cub OLAP**
- Figura 18. Dimensions, jerarquies i granularitat**
- Figura 19. Visualització amb Jpivot del cub OLAP**
- Figura 20. Visualització amb Saiku del cub OLAP**
- Figura 21. Taula d'anàlisi del SBP i el DBP en relació al tractament**
- Figura 22. Gràfic d'anàlisi del SBP i el DBP en relació al tractament**
- Figura 23. Taula d'anàlisi entre els tractaments i l'evolució dels pacients**
- Figura 24. Gràfic d'anàlisi entre els tractaments i l'evolució dels pacients**
- Figura 25. Taula d'anàlisi de l'evolució dels diferents tractaments**
- Figura 26. Gràfic d'anàlisi de l'evolució dels diferents tractaments**
- Figura 27. Taula d'anàlisi dels resultats en relació a l'activitat**
- Figura 28. Gràfic d'anàlisi dels resultats en relació a l'activitat**
- Figura 29. Gràfic d'anàlisi dels resultats en relació a l'activitat**
- Figura 30. Taula d'anàlisi dels resultats en relació a l'activitat de cada pacient**
- Figura 31. Gràfic d'anàlisi dels resultats en relació a l'activitat de cada pacient**
- Figura 32. Taula d'anàlisi dels resultats en relació a la dieta**
- Figura 33. Gràfic d'anàlisi dels resultats en relació a la dieta**

Figura 34. Taula d'anàlisi dels resultats en relació a la dieta de cada pacient

Figura 35. Gràfic d'anàlisi dels resultats en relació a la dieta de cada pacient

Figura 36. Taula resum de l'evolució dels diferents tractaments relacionats amb els hàbits de cada pacient.

Figura 37. Gràfic d'anàlisi de l'evolució dels diferents tractaments relacionats amb els hàbits de cada pacient.

Figura 38. Gràfic d'anàlisi de l'evolució dels diferents tractaments relacionats amb els hàbits.

Figura 39. Taula d'anàlisi de l'evolució dels diferents tractaments relacionats amb els lloc geogràfic de cada pacient

Figura 40. Gràfic d'anàlisi de l'evolució dels diferents tractaments relacionats amb els lloc geogràfic de cada pacient

Figura 41. Gràfic d'anàlisi trimestral de l'evolució dels diferents tractaments

Figura 42. Gràfic d'anàlisi mensual de l'evolució dels diferents tractament

Figura 43. Gràfic d'anàlisi setmanal de l'evolució dels diferents tractaments

1. Introducció

1.1 Context i justificació del Treball

El colesterol és un lípid (grassa). Es forma en el fetge a partir d'aliments grassos i és necessari per al funcionament normal de l'organisme. El colesterol està present en la membrana plasmàtica (capa exterior) de totes les cèl·lules de l'organisme.

El colesterol es desplaça per la sang mitjançant unes molècules denominades lipoproteïnes. Els tres tipus principals són:

Les lipoproteïnes de baixa densitat (LDL) o "colesterol dolent"; es creu que causen malalties arterials. Les LDL transporten el colesterol des del fetge a les cèl·lules i poden causar una acumulació nociva si hi ha més del que les cèl·lules poden usar.

Les lipoproteïnes d'alta densitat (HDL) o "colesterol bo"; es creu que prevenen les malalties arterials. Les HDL es porten el colesterol de les cèl·lules i ho retornen al fetge on es descompon i s'elimina com a residu corporal.

Els triglicèrids es formen en el fetge i estan presents en productes làctics, carn i olis culinaris. L'obesitat i l'alimentació rica en greixos augmenten el risc de tenir nivells alts de triglicèrids.

El colesterol alt està relacionat amb malalties greus, com les cardiopaties, l'angina de pit i els accidents cerebrovasculars. La causa de la cardiopatia coronària és l'estrenyiment de les artèries (aterosclerosi) que subministren l'aportació de sang al cor. Els dipòsits grassos, com el colesterol o els productes residuals s'acumulen a l'interior de les artèries. Aquesta acumulació es diu placa i impedeix el flux de sang per les artèries.

Si té símptomes de aterosclerosi, també pot tenir un nivell alt de colesterol. Els símptomes inclouen l'angina de pit (dolor en el pit causat per una reducció de l'aportació sanguínia al cor), dolor de cama (a causa de l'estrenyiment de les artèries que porten sang a les extremitats) i coàguls sanguinis en les artèries que transporten sang al cor (trombosi coronària). Els coàguls sanguinis poden portar a una deficiència cardíaca.

Les taques espesses de color groc (xantomes) al voltant dels ulls o en alguna altra zona de la pell es formen pels dipòsits de colesterol. Sovint, es poden apreciar en persones amb colesterol alt hereditari (per familiars amb antecedents de colesterol alt).

El treball concret que es portarà a terme és crear un entorn *Business Intelligència* (BI) que possibiliti l'anàlisi de la informació generada durant un experiment que té com objectiu comprovar la eficàcia dels diferents tractament que tenen com finalitat la reducció dels nivells de colesterol. Tanmateix la informació subministrada podrà ser analitzada en forma de cubs de dades el que permetrà un anàlisi molt més acurat i l'extracció de conclusions molt més elaborades.

1.2 Objectius del Treball

L'objectiu d'aquest treball és el disseny i implementació d'un sistema de Business Intelligència que faciliti l'adquisició, l'emmagatzemament i l'explotació de dades associades a pacients als que se'ls ha diagnosticat els nivells de colesterol (LDL). Per tant, el treball tindrà els següents objectius:

1. Dissenyar un magatzem de dades (*Data Warehouse*) que permeti emmagatzemar la informació adquirida dels diferents orígens de dades. Tenint en compte que tindrem un conjunt de pacients que han estat sotmesos, per grups, a diferents tractaments.
2. Implementar aquest magatzem de dades i programar els processos ETL (sigles en anglès d'extracció, transformació i càrrega) que permetin alimentar el *Data Warehouse* a partir dels fitxers base facilitats.
3. Analitzar les diferents plataformes BI *Open Source* disponibles al mercat que ens permetrien explotar la informació emmagatzemada.
4. Seleccionar i implantar una d'aquestes eines *Open Source* de tal forma que es disposi d'una capa de programari per l'anàlisi de la informació.

Les preguntes analítiques que es volen resoldre són les següents:

- Quina és la relació entre els diferents tractaments i l'evolució dels pacients?
- Existeixen teràpies més eficaces?
- Ha influït en el resultat, els hàbits dels pacients?
- L'evolució al llarg del temps, per un mateix tractament, depenen d'algun factor com els hàbits.
- Hi ha diferències en el resultat d'un tractament segons el lloc geogràfic del pacient?
- Hi ha algun període de l'any on el tractament sigui més o menys efectiu?

1.3 Enfocament i mètode seguit

La dimensió i els objectius del treball orienten la solució a una estratègia centrada a seleccionar i implementar una plataforma BI *Open Source* de les disponibles al mercat de tal forma que es disposi d'una capa de programari per l'anàlisi de la informació disponible i assolir així els objectius del projecte.

En la gestió del projecte, un cop analitzat el problema i avaluats els recursos, es decideix una descomposició del treball en 7 fases que corresponen a unitats funcionals que seran abordades de forma seqüencial. Tot i això, cada una de les fases implica la revisió iterativa d'una o de diverses de les fases anteriors.

Aquestes fases, definides com a tasques principals del pla de treball, son:

- Planificació.
- Analitzar i seleccionar entre les diferents plataformes BI Open Source i SGBD.
- Dissenyar i implementar d'un magatzem de dades.
- Programar els processos ETL.
- Implementar i fer proves de l'entorn BI Open Source.
- Explotar i analitzar la informació.
- Finalitzar el projecte.

Així mateix, en el transcurs del projecte s'utilitzen estratègies i pràctiques diverses, moltes d'elles utilitzant com a referències d'una banda el PMBOK, com a marc conceptual i col·lecció del que es pot considerar "bones pràctiques generalment acceptades en gestió de projectes" i d'altra banda The Data Warehouse Toolkit de Kimball, R. i Ross, M. en la seva tercera edició com a biblioteca completa de tècniques de modelització digital dimensionada.

1.4 Planificació del Treball

La descomposició del treball en les següents tasques s'ha fonamentat en l'anàlisi comparatiu de projectes centrats en el disseny i implementació de sistemes de *Business Intelligence* per l'adquisició, l'emmagatzemament i l'explotació de dades.

Una part important del projecte és l'anàlisi i selecció de les diverses plataformes així com les eines a utilitzar, aquest nivell d'incertesa

s'haurà de corregir amb una revisió continuada de la planificació on es tingui en compte la possible necessitat de formació en les tecnologies escollides així com la modificació de tasques o activitats relacionades.

La següent taula mostra les principals tasques del treball amb les seves activitats relacionades i els recursos assignats en una primera planificació.

Tasques	Activitats	Dies
		12
Planificació	Estudi dels mòduls i de l'enunciat del treball	3
	Cerca bibliogràfica (biblioteca, publicacions, cercadors, etc.)	5
	Elaboració del pla de treball segons les especificacions PAC1	4
		12
Analitzar i seleccionar entre les diferents plataformes BI Open Source i SGBD	Anàlisi i selecció entre les diferents eines BI <i>Open Source</i>	6
	Anàlisi i selecció entre els diferents SGBD compatibles	3
	Documentació (memòria del projecte)	3
		23
Dissenyar i implementació d'un magatzem de dades (Data Warehouse)	Anàlisi de les dades i les seves interrelacions	3
	Disseny lògic i físic del repositori	10
	Instal·lació del software i implementació del model	5
	Documentació (memòria del projecte)	5
		21
Programar els processos ETL	Disseny dels processos d'extracció, transformació i càrrega	4
	Comprovació de dades i proves	8
	Càrrega de dades al <i>Data Warehouse</i>	4
	Documentació (memòria del projecte)	5
		7
Implementació i proves de l'entorn BI Open Source	Implementació de la capa d'anàlisi de dades	4
	Proves i ajustos de la solució BI	2
	Documentació (memòria del projecte) PAC3	1
		24
Explotació i anàlisi de la informació.	Explotació de les dades mitjançant la plataforma BI	10
	Dissenyar i implementar els informes d'anàlisi.	6
	Extreure conclusions del projecte	2
	Documentació (memòria del projecte)	6
		11
Finalització del projecte	Memòria del projecte	7
	Elaboració de la presentació virtual	4

Els recursos previstos en aquest projecte estan definits per la càrrega de 12 crèdits ECTS que son concretats per l'equivalència de 25 hores de dedicació per cada crèdit.

En la temporització de les diferents tasques s'ha mirat d'assignar una durada realista segons la seva càrrega i complexitat així com dels coneixements i formació prèvia. A més, les activitats crítiques contemplen uns coixins de dedicació entre el 10 i el 20% en la seva planificació.

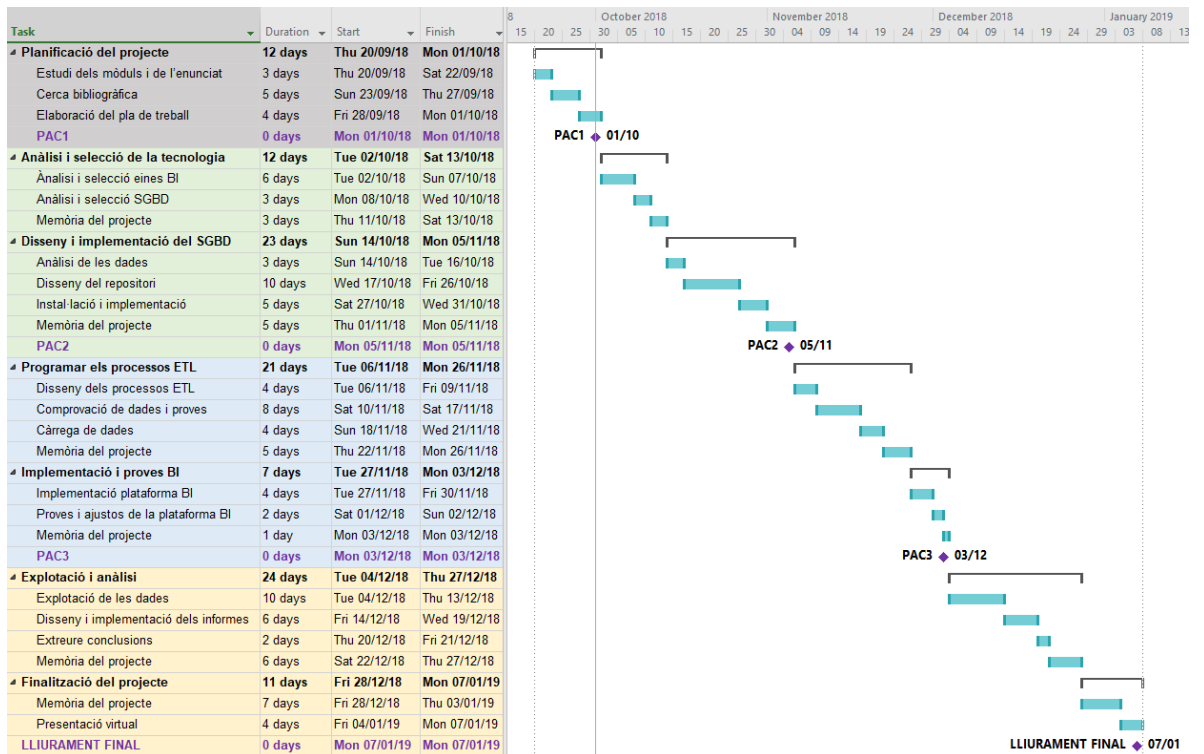
El temps previst per a la documentació serà aproximadament un 35% del global, això és lleugerament inferior a la previsió que es realitza en un projecte d'aquesta tipologia. La resta dels recursos, un 65%, es destinaran a la construcció del producte en base als objectius definits.

En tot cas, per tal de garantir la qualitat del projecte, s'ha previst la possibilitat d'utilitzar recursos addicionals que haurien de permetre fer front a possibles problemes així com a la necessitat de més dedicació. La següent taula mostra les fites del projecte amb les seves corresponents dates.

Fita	Data
Inici del projecte	20-09-2018
Planificació del projecte	01-10-2018
Anàlisi i selecció de la tecnologia	13-10-2018
Disseny i implementació del SGBD	05-11-2018
Programar els processos ETL	26-11-2018
Implementació i proves BI	03-12-2018
Explotació i anàlisi de la informació.	27-02-2018
Finalització del projecte	07-01-2018
Lliurament final	07-01-2018

El diagrama de Gantt de la figura 1 mostra les fites del projecte amb les seves tasques i les dates de planificació.

Figura 1. Diagrama de Gantt amb la planificació del projecte

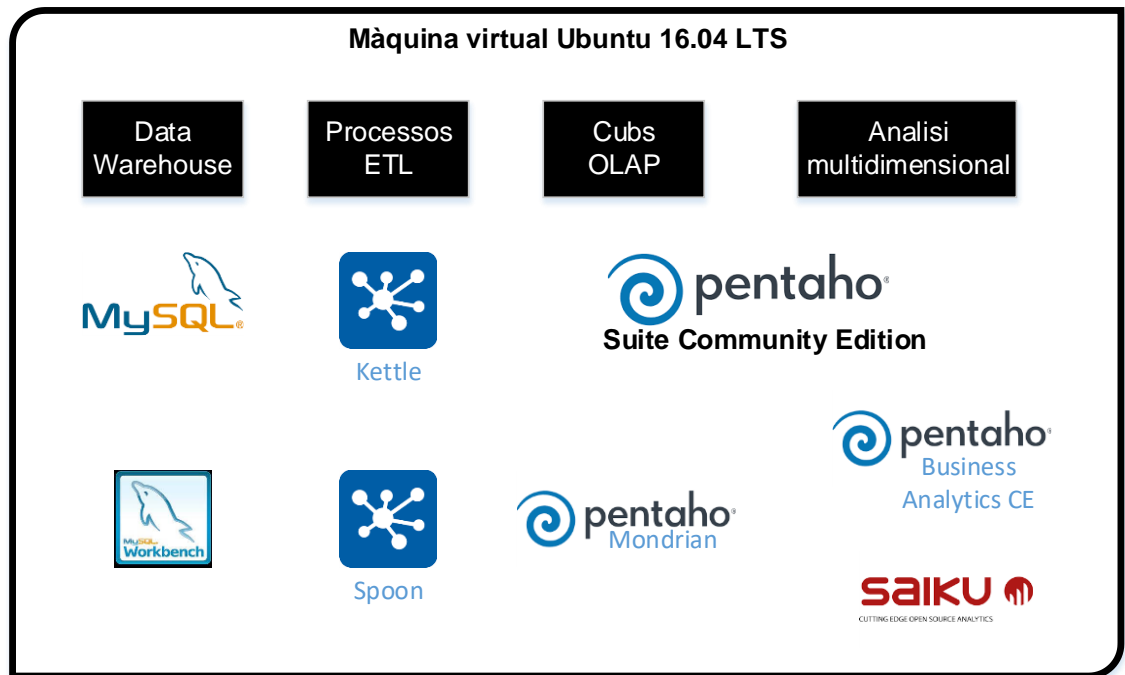


1.5 Breu resum de productes obtinguts

La realització d'aquest treball ha generat un seguit de fitxers que han permès la creació del *Data Warehouse*, la càrrega de dades per mitjà dels corresponents processos ETL, la creació del cub OLAP i finalment també s'han obtingut els informes analítics per respondre les preguntes definides com a objectius del treball.

Amb tot, el principal producte és un sistema de *Business Intelligence* que facilita l'adquisició, l'emmagatzemament i l'explotació de dades associades a pacients als que se'ls ha diagnosticat els nivells de colesterol (LDL). Aquesta plataforma BI *Open Source* està implementada en una màquina virtual Ubuntu 16.04 LTS on hi ha instal·lades i configurades totes les eines necessàries per realitzar el projecte i generar els fitxers descrits. La figura 2 mostra un resum gràfic de l'entorn BI obtingut.

Figura 2. Breu esquema de la plataforma BI obtinguda



1.6 Breu descripció dels altres capítols de la memòria

Anàlisi i selecció de les eines BI

L'elecció de les eines d'intel·ligència de negoci *Open Source* és un dels factors més determinants del resultat del projecte. Es realitzarà un anàlisi comparatiu de les quatre eines *Business Intelligence Open Source* més utilitzades i més reconegudes del moment. Així mateix, es compararà i seleccionarà entre els diferents Sistemes de Bases de Dades per tal de crear el Data Warehouse. Les plataformes BI analitzades han estat Pentaho, SpagoBI, Jasper i BIRT

Model dimensional (DWH)

Un anàlisi de les dades disponibles agrupades en diferents fulles d'un llibre d'Excel permetrà iniciar la definició del model conceptual utilitzant eines especificades en el llibre "The Data Warehouse Toolkit" de Kimball i Ross, en primer lloc és definirà un "Bus Matrix", seguit d'un model dimensional d'alt nivell i finalment del model dimensional detallat. El capítol conclou amb la Implementació física del model de dades en un magatzem de dades multidimensional.

Processos ETL

Per tal d'alimentar el DWH i entregar una col·lecció d'informació dimensional per la seva consulta i anàlisi es programen els diferents processos d'extracció (extract), transformació (transform) i càrrega (load) de les dades. Aquests processos han estat definits i portats a terme iterativament fins a obtenir un procediment automàtic i senzill de modificar que permet la seva execució en qualsevol moment. En el capítol es descriuen les principals fases de cadascun dels processos necessaris per implementar la càrrega de dades així com les transformacions modelades en l'etapa de disseny.

Implementació eines OLAP

Disseny i implementació amb eines OLAP (*Online Analytical Processing*) a partir del model físic de dades multidimensionals. Es definirà un esquema OLAP amb un sol cub que tindrà una taula de fets i diverses dimensions, cadascuna amb una o més jerarquies. Els fitxers resultants són models de metadades XML amb esquemes definits en una estructura específica que posteriorment implementarà el servidor d'anàlisi BI.

Anàlisi de dades

L'anàlisi es concentrarà en resoldre les diferents preguntes analítiques plantejades com a objectius del projecte. La plataforma d'intel·ligència de negoci i anàlisi de dades en un entorn multidimensional permet disposar de la informació que requereix i aquesta es mostrarà en taules de dades i representacions gràfiques per tal d'ajudar a una millor interpretació dels resultats.

Conclusions

Es realitza una descripció de les conclusions del treball així com una reflexió crítica tant de l'assoliment dels objectius inicials com del seguiment de la planificació i la metodologia al llarg de projecte. També s'apunten possibles línies de treball futures.

2. Anàlisi i selecció de les eines BI

2.1 *Business Intelligence*

Per tal de definir la intel·ligència de negoci o més usualment BI trobem que la reconeguda consultora *Gartner Inc.* la defineix en el seu glossari de l'any 2018 com a “un terme que inclou les aplicacions, la infraestructura i les eines i les millors pràctiques que permeten l'accés i l'anàlisi de la informació per millorar i optimitzar les decisions i el rendiment”.

Així mateix, *Gartner Inc.* especifica que les plataformes de *Business Intelligence* “permeten a les empreses crear aplicacions de BI aportant capacitats en tres categories: anàlisi, com ara processament analític en línia (OLAP); lliurament d'informació, com informes i taulers de comandament; i la integració de plataformes, com ara la gestió de metadades de BI i un entorn de desenvolupament”.

2.2 Anàlisi i d'eines BI

L'elecció de la plataforma d'intel·ligència de negoci utilitzada per aquest projecte és un dels factors més determinants del seu resultat. Aquesta plataforma ens ha de permetre transformar les dades per tal d'adaptar-les al model conceptual per medi del procés, anomenat ETL de càrrega transformació i extracció (*Extraction, Transformation, Load*). Posteriorment s'haurà d'extreure de les dades la informació requerida utilitzant informes amb l'objectiu de respondre a les preguntes analítiques que afronta el projecte.

Existeix una ampla varietat d'eines en el mercat de *Business Intelligence* que ens permeten realitzar les tasques descrites, ara bé, els requisits d'aquest treball estableixen la necessitat que les eines utilitzades siguin *Open Source*. Amb tot, un anàlisi comparatiu de totes les opcions disponibles sobrepassaria l'abast d'aquest projecte, així, s'han analitzat les quatre eines *Business Intelligence Open Source* més utilitzades i més reconegudes del moment.

D'altra banda l'elecció de la plataforma BI ens permetrà comparar i seleccionar entre els diferents Sistemes de Bases de Dades compatibles per tal de crear el *Data Warehouse*. Les plataformes BI analitzades han estat Pentaho, SpagoBI, Jasper i BIRT

Pentaho

Empresa nord americana especialitzada en programari d'intel·ligència de negoci, va se adquirida per Hitachi Data Systems l'any 2015 i actualment forma part d'Hitachi Vantara, una nova empresa que unifica les operacions de Pentaho, Hitachi Data Systems i Hitachi Grup Insight. Tot i això, manté el nom de Pentaho i continua oferint una versió sense cost de llicència anomenada Community Edition. La versió Enterprise Edition està disponible a través d'una subscripció anual i conté funcions addicionals i suport que no es troba a l'edició de la comunitat.

Pentaho es defineix a si mateix com una plataforma de BI "orientada a la solució" i "centrada en processos". Proporciona integració de dades, serveis OLAP, informes, panells d'informació, mineria de dades i capacitats per extreure, transformar i càrrega r orígens de dades diversos (ETL).

Aquestes solucions es componen principalment d'una infraestructura d'eines d'anàlisi i informes integrada amb un motor de flux de treball dels diferents processos que disposen de la capacitat d'executar les regles de negoci necessàries en forma de processos i activitats per tal de mostrar la informació requerida. La plataforma és millorada pels productes complementaris, generalment en forma de plug-ins desenvolupats per l'empresa o per la seva ampla comunitat d'usuaris.

La consultora Gartner ha valorat Pentaho en els seus informes anuals com a programari *Open Source* de provada eficàcia en projectes de tractament de dades que requereixen de la construcció de processos ETL.

Gartner ha subratllar la importància que un programari ETL disposi de les següents característiques: connectivitat, capacitat d'entrega de dades, de metadades i modelatge de dades, de disseny i entorn de desenvolupament, de gestió de dades d'administració, a més de capacitats SOA i un cert grau de compactació, consistència i interoperabilitat.

La instal·lació del servidor requereix 8GB de memòria RAM amb 4 dedicades a Pentaho servers així com 20GB d'espai lliure per fitxers de sistema després de la instal·lació. És compatible amb diverses bases de dades Open Source incloent MySql i PostgreSQL. Entre d'altres, permet realitzar la instal·lació en sistemes operatius Open Source com CentOS o Ubuntu.

SpagoBI

A partir de la seva versió 6.0 SpagoBI ha passat a anomenar-se Knowage actualment en la versió 6.2. Aquesta plataforma BI pertany a la iniciativa Free and open-source (FOSS) de SpagoWorld, fundada i recolzada per Engineering Group. És la única analitzada que disposa de tot el programari Open Source.

Disposa d'una arquitectura modular on els diferents mòduls es relacionen amb el nucli del sistema oferint estabilitat i capacitat d'evolució.

En la seva instal·lació completa incorpora mòduls per extreure, transformar i càrrega r (ETL) orígens de dades diversos, permet fer anàlisi de dades multidimensional OLAP, mineria de dades (*Data Mining*) així com realitzar consultes i visualització de dades. Cal destacar que presenta una ampla varietat de funcions d'anàlisi i un ampli ventall de funcions avançades de visualització de dades.

Es reconeguda com l'eina BI *Open Source* més potent i completa però alhora disposa d'una interfície complexa i una corba d'aprenentatge alta, tant per la configuració com pel seu ús. D'altra banda, no existeix una documentació molt extensa accessible ni disposa d'un fòrum de suport oficial.

La instal·lació del servidor requereix 3GB de memòria RAM així com 2GB d'espai per fitxers de sistema. La base de dades externa ha de ser MySql o MariaDB. Entre d'altres, permet realitzar la instal·lació en sistemes operatius Open Source com CentOS o Ubuntu.

Jaspersoft Community

Jaspersoft adquirida per TIBCO Software l'any 2014 proporciona programari comercial al voltant del producte JasperReports. Aquesta plataforma està disponible sota una llicència de codi obert per al seu ús juntament amb una infraestructura de codi obert. També existeixen diverses tipologies de llicències comercials per a implementacions empresarials que impliquin bases de dades comercials i servidors d'aplicacions.

Incorpora eines gràfiques per extreure, transformar i càrrega r (ETL) orígens de dades diversos, també a partir del núvol. Suporta la majoria de bases de dades relacionals i diversos orígens NoSql, també disposa d'algunes connexions natives amb aplicacions ERP i CRM. La plataforma també incorpora tècniques d'anàlisi multidimensional (OLAP) i destaca per ser l'eina *Open Source* més utilitzada per tal d'integrar informes dins d'aplicacions, tot i que aquesta funcionalitat no està

disponible en la versió Community els informes es poden utilitzar de forma autònoma.

Tot i que no és vital per aquest projecte, cal destacar que la versió Community tampoc disposa de la possibilitat d'incorporar un quadre de comandament (*Dashboard*) ni la realització informes interactius.

La instal·lació del servidor requereix 4/8GB de memòria RAM així com 10/40GB d'espai per fitxers de sistema. És compatible amb diverses bases de dades Open Source incloent MySql i MariaDB de forma nativa. Entre d'altres, permet realitzar la instal·lació en sistemes operatius *Open Source* com CentOS o Ubuntu.

BIRT

Business Intelligence and Reporting Tools (BIRT) és un projecte de programari de codi obert dins de la Fundació Eclipse, un consorci independent sense ànim de lucre de proveïdors de la indústria del programari i una comunitat de codi obert. El projecte està patrocinat per Actuate juntament amb les contribucions d'IBM i Innovent Solutions.

Eclipse defineix BIRT com a una plataforma tecnològica per crear visualitzacions de dades i informes que es poden integrar *rich client platform* (RCP) i aplicacions web, especialment aquells entorns basats en Java i Java EE.

Disposa de dos components principals: d'una banda dissenyador d'informes visuals dins de l' Eclipse IDE per crear informes que inclou un motor gràfic i anàlisi multidimensional (OLAP), i d'altra banda un component en temps d'execució per generar informes que es poden implementar en qualsevol entorn Java. Els informes extreuen dades d'un origen de dades o fonts, realitzen manipulacions i càlculs sobre les dades per respondre a preguntes sobre empreses i presenten els resultats com a informació en un formulari format i convenient per a l'usuari empresarial. Tot i això, no disposa de forma nativa de mòdul (ETL).

La instal·lació del servidor no especifica uns requeriments mínims de manera que s'assimila a eclipse i dependrà en cada moment de la magnitud del projecte. És compatible amb diverses bases de dades Open Source incloent MySql i MariaDB. Entre d'altres, permet realitzar la instal·lació en sistemes operatius Open Source com CentOS o Ubuntu.

2.3 Selecció entorn BI

Els requeriments d'instal·lació tant de maquinari com del SGBD necessaris en qualsevol de les quatre plataformes son assumibles en aquest projecte i no condicionen l'elecció de l'eina.

En una primera selecció es trien les dos plataformes que incorporen totes les eines necessaris per totes les fases d'un projecte d'intel·ligència de negoci per tal de reduir la complexitat i evitar problemes d'interoperabilitat no previstos.

Així, un cop l'elecció es focalitza entre SpagoBI i Pentaho Suite Community Edition es selecciona aquesta darrera plataforma per realitzar el projecte. Els aspectes més rellevants per prendre la decisió han estat:

- És la plataforma de *Bussines Intelligencie Open Source* més utilitzada.
- Disposa d'un complet motor ETL senzill d'utilitzar.
- Proporciona d'una corba d'aprenentatge baixa.
- Facilita la instal·lació amb un procediment senzill.
- Disposa d'un ampli suport per medi de la seva comunitat i els fòrums existents.
- Disposa de molta informació i manuals online.

2.4 Anàlisi i selecció eines SGBD

EL projecte necessita un repositori central preferiblement relacional on integrar les dades obtingudes de diferents fulls d'Excel.

Aquest magatzem de dades o *Data Warehouse* (DWH) és definit per Hugh j. Watson com "una col·lecció d'informació creada per donar suport a les aplicacions per la pressa de decisions". Per la seva banda, Ralph Kimball el va definir com "un magatzem de dades que, neteja, conforma i entrega una font de dades dimensional para la seva consulta i anàlisis".

Així, es fa necessari un sistema gestor de base de dades (SGBD) per tal d'emmagatzemar el *Warehouse*. Per l'elecció s'han analitzat les dos eines Open Source especialitzades en la implementació de magatzems de dades més utilitzades del mercat: PostgreSQL i MySQL i així mateix s'ha considerat mariaDB, eina que deriva de MySQL i està projectada per tal de mantenir la llicència de programari lliure independentment de l'evolució futura que pogués fer MySQL.

Els tres SGBD son compatibles amb Pentaho i cobreixen sense dubtes els requisits d'aquest projecte. PostgreSQL presenta un conjunt de funcionalitats més extensa destacant en la fiabilitat, integritat de dades i un seguit de característiques enfocades al desenvolupador, és apreciat el seu rendiment treballant amb grans volums de dades. Per la seva banda MySQL i conseqüentment mariaDB esdevenen sistemes molt ràpids tot i que amb menys funcionalitats.

Donat el volum de dades a treballar i l'abast del projecte no han estat analitzades altres consideracions que podrien condicionar al selecció del producte en escenari diferent. Així, es selecciona MySql en base a ser el SGBD més utilitzat dels tres i, que a més disposa d'una amplia documentació i comunitat de suport.

3. Model dimensional

3.1 El Data Warehouse

Un *Data Warehouse* (DWH) és un magatzem de dades que, neteja, conforma i entrega una col·lecció d'informació dimensional per la seva consulta i anàlisi per tal de donar suport a les aplicacions per la presa de decisions.

Una característica fonamental d'aquest tipus de magatzem de dades el trobem en les estructures que utilitza per guardar la informació, els models de taules en estrella o floc de neu, entre d'altres, generen dades homogènies i fiables que permeten òptimes consultes i el tractament jerarquitzat de les mateixes.

Aquest modelatge amb múltiples dimensions tracta les dades com si fossin un cub i és àmpliament acceptat com la tècnica per presentar dades analítiques, ja que d'una banda proporciona dades comprensibles per als usuaris empresarials i d'altra banda permet realitzar les consultes amb un alt rendiment.

Fets

La taula de fets en un model dimensional emmagatzema les dades que es volen analitzar. La idea que un esdeveniment mesurat en el món físic tingui una relació un a un amb una sola fila en la taula de fets corresponent. Això és un principi fonamental per a la modelització dimensional. La granularitat és el nivell de detall que es vol mantenir aquest concepte permet reduir l'espai de les dades reduint alhora el detall de les consultes. Les claus foranies, son claus dels registres a les taules de dimensió

Dimensions

Les taules de dimensió contenen el context textual associat als fets. Les dimensions s'associen amb l'esdeveniments i descriuen qui, què, on, quan, com i per què. Són les dades que s'utilitzaran per fer les cerques i permetran que els fets siguin agrupats, etiquetats i filtrats. Una taula de dimensions disposa d'una clau pròpia i un conjunt d'atributs que el descriuen. Així mateix, les jerarquies, són els diferents nivell de granularitat duna dimensió.

3.2 Anàlisi de les dades

Les dades que s'utilitzen en el projecte estan disponible en un llibre d'Excel. Aquesta informació està agrupada en diferents fulles que contenen la següent informació:

Patiens

Hi ha les dades relatives als pacients com la informació sobre la ciutat a la que pertanyen i a la comunitat autònoma, el rang d'edat i el gènere al que pertanyen així com la tipologia de tractament que està rebent.

Tractament

Full on hi ha 3 grups diferents de teràpies:

Teràpies farmacològiques

Són aquelles que es tracten amb medicaments farmacèutics que han superat totes fases clíniques establertes.

Teràpies naturals

Són aquelles teràpies que fan servir productes naturals (arrels, llavors, etc.).

Teràpies homeopàtiques

Aquelles que fan servir productes homeopàtics.

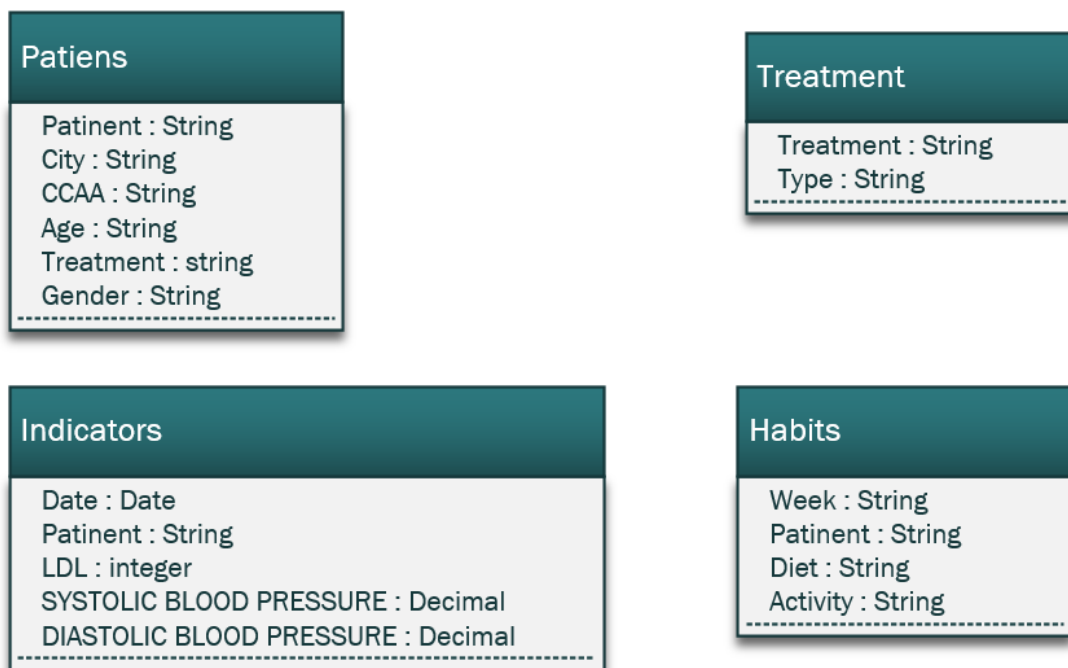
Habits

Recull de registres setmanals amb informació dels hàbits que ha seguit un pacient durant el tractament indicant l'activitat i dieta realitzada en la setmana de referència. Podran variar al llarg del temps per cada pacient.

Indicadors

El full hi ha els registres del control dels pacients on hi ha el nivell de colesterol LDL així com la informació de la tensió arterial.

Figura 3. Representació de les dades disponibles



3.3 Model Conceptual

Per tal de definir el model conceptual s'utilitzaran algunes de les eines especificades en el llibre *The Data Warehouse Toolkit* de Kimball i Ross. En primer lloc és definirà un "Bus Matrix", seguit d'un model dimensional d'alt nivell i finalment un model dimensional detallat. Aquest model serà revisat i validat en l'execució del projecte de manera iterativa.

Bus Matrix

El propòsit d'aquesta matriu és obtenir una alta abstracció que permet una visió general de la integració representant les diferents dimensions de consulta i les àrees (objectius) en les que intervenen, al mateix temps permet dividir el problema en parts més manejables.

Per tal de definir les diferents dimensions i fets s'analitzen les diferents preguntes analítiques que es volen resoldre com a objectiu d'aquest projecte:

Quina és la relació entre els diferents tractaments i l'evolució dels pacients?

En aquesta pregunta s'ha d'analitzar la teràpia de cada pacient i els indicadors, principalment el nivell de colesterol LDL, però també la tensió arterial del pacient. Així definim la dimensió **pacient** on hi tindrem el tractament associat. Els indicadors son registres que formaran part de la taula fets

Existeixen teràpies més eficaces?

La consulta que permet donar resposta a aquesta qüestió també implica les dimensions ja definides, donat que amb una comparativa dels diferents tractaments per cada pacient es pot avaluar la seva eficàcia en base als indicadors.

Ha influït en el resultat, els hàbits dels pacients?

Afegit a les dimensions ja identificades és necessitaran dos dimensions per avaluar la influència dels hàbits dels pacients, **diet** i **activity**.

L'evolució al llarg del temps, per un mateix tractament , depenen d'algun factor com els hàbits?

A més de les dimensions teràpia, pacient i hàbits serà necessària una dimensió **date** que permetí fer consultes. Es vol agrupar tota la informació a una granularitat mínima d'una setmana.

Hi ha diferències en el resultat d'un tractament segons el lloc geogràfic del pacient?

Es farà us de la dimensió pacient per medi de jerarquies per poder comparar els resultats segons la ciutat on resideix el pacient. Aquesta dimensió disposarà del nivell CCAA per tal de definir consultes amb una abstracció més elevada.

Hi ha algun període de l'any on el tractament sigui més o menys efectiu?

En aquest cas l'anàlisi de l'eficàcia dels tractaments en períodes d'any necessita de les dimensions ja detectades, pacient i data que permeti saber l'evolució dels indicadors en els períodes de l'any que es defineixin.

Tot i no aparèixer en les preguntes analítiques els pacients es poden agrupar per rang d'edat i/o per sexe. Aquesta informació estarà recollida en la pròpia dimensió pacient

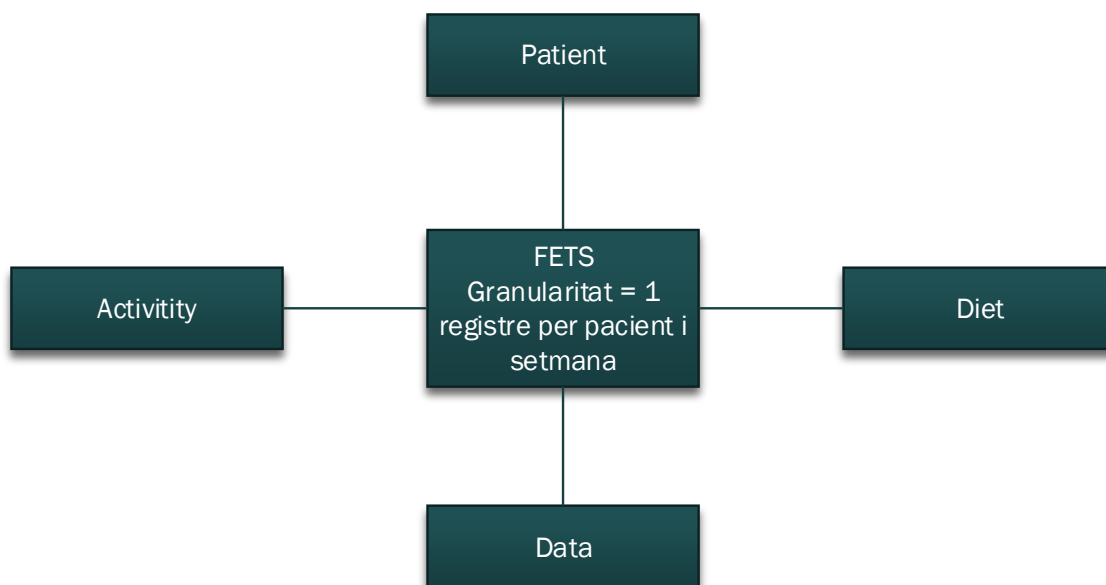
En la següent matriu "*Bus Matrix*" es pot observar cada pregunta analítica i la seva relació amb les dimensions detectades.

	Data	Pacient	Activity	Diet
Quina és la relació entre els diferents tractaments i l'evolució dels pacients?	X	X		
Existeixen teràpies més eficaces?		X		
Ha influït en el resultat, els hàbits dels pacients?		X	X	X
L'evolució al llarg del temps, per un mateix tractament, depenen d'algun factor com els hàbits?	X	X	X	X
Hi ha diferències en el resultat d'un tractament segons el lloc geogràfic del pacient?		X		
Hi ha algun període de l'any on el tractament sigui més o menys efectiu?	X	X		

Diagrama del model d'alt nivell

Utilitzant el “*Bus Matrix*” es crea un model d'alt nivell conegut com a bombolla que representa les diferents dimensions i taules de fets.

Figura 4. Diagrama del model d'alt nivell



Les dimensions i els fets es modelitzen com a taules en el magatzem de dades i habitualment corresponen a esquemes d'estel o de floc de neu. L'esquema en estel té alguns desavantatges, com una certa redundància de dades, que fa que els canvis siguin més costosos i augmenta la mida física del magatzem de dades, per la seva banda les consultes d'aquest model són més senzilles i ràpides.

En el projecte es podrien fer refinaments normalitzant alguna de les taules de dimensió per tal d'evolucionar el model d'esquema presentat en un model de floc de neu. Ara bé, els avantatges proporcionats no serien significatius alhora que faria el sistema més complex. Per tant, s'utilitzarà l'esquema d'estel en *el Data Warehouse*.

Model dimensional detallat

Aquest model es centra en els detalls i defineix les dimensions de manera detallada per cada taula i per cada columna. El detall identifica també els atributs necessaris dins de cada dimensió per a cada taula de fets.

Nom de taula	DimRegistres							
Tipus de taula	Fets							
Nom de la vista	Registres							
Descripció	Taula on hi ha els registres de cada control del pacient							
	Destí			Origen				
Nom	Descripció	Type	exemple	Sistema d'entrada	Full origen	Nom origen	Type	Regles ETL
registresKey	Clau primària	Int	1,2,3...	ETL				
PatientKey	Clau foranea de la taula patient	int	1,2,3...	ETL				
DateKey	Clau foranea de la taula Data	Int	1,2,3...	ETL				
TreatmentKey	Clau foranea de la taula Treatment	Int	1,2,3...	ETL				
CityKey	Clau foranea de la taula City	Int	1,2,3...	ETL				
ActivityKey	Clau foranea de la taula Activity	int	1,2,3...	ETL				
DietKey	Clau foranea de la taula Diet	Int	1,2,3...	ETL				
LDL	Nivell de colesterol	Int	1,2,3...	Excel	Indicators	LDL	number	
SBP	SYSTOLIC BLOOD PRESSURE	Dec	6,82	Excel	Indicators	SBP	number	
DBP	DIASTOLIC BLOOD PRESSURE	Dec	6,82	Excel	Indicators	DBP	number	

Nom de taula	DimDate							
Tipus de taula	Dimensió							
Nom de la vista	Date							
Descripció	Dates dels diferents anàlisi							
Mida	6 registres							
	Destí			Origen				
Nom	Descripció	DataType	exemple	Sistema d'entrada	Full origen	Nom origen	Dataype	Regles ETL
DateKey	Clau primària	Int	1,2,3...	ETL				
Date	Data obtinguda de la font	date	30/10/2017, 03/01/2018	Excel	Indicators	Date	Data	
Day	Número del dia	Int	1, 31 ...	ETL				
Week	Número setmana	Int	1, 54 ...	ETL				
Month	Número del mes	Int	1, 12 ...	ETL				
Trimestrer	Número del trimestre	Int	1, 4 ...	ETL				
Year	Número de l'any	int	1, 12 ...	ETL				
Jerarquia	Nivell 0	Day						
	Nivell 1	Week						
	Nivell 2	Month						
	Nivell 3	Trimestrer						
	Nivell 4	Year						

Nom de taula	DimPatient							
Tipus de taula	Dimensió							
Nom de la vista	Patient							
Descripció	Registres dels pacients amb la seva informació personal							
Mida	6 registres							
	Destí			Origen				
Nom	Descripció	DataType	exemple	Sistema d'entrada	Full origen	Nom origen	Dataype	Regles ETL
PatientKey	Clau primària	int	1,2,3...	ETL	Patinets	Patinet	text	
Patient	Nom del pacient	string	P1,P2...	Excel	Patinets	Patinet	text	
Age	Grup d'edat del pacient	string	60-99, 26-45...	Excel	Patinets	age	text	
Gender	Edat del pacient	caracter	M, F	Excel	Patients	gender	text	
Treatment	Tipologia del grup de tractament	string	PHARMA, NATURAL, HOMEOPATHIC	Excel	Treatment	Type	text	
City	Nom de la ciutat	string	N1, N2...	Excel	Patinets	City	text	
CCAA	Nom de la comunitat	String	C1, C2...	Excel	Patinets	CCAA	text	
Jerarquia 1		Nivell 0	Patient					
		Nivell 1	Treatment					
Jerarquia 2		Nivell 0	Patient					
		Nivell 1	City					
		Nevell 2	CCAA					

Nom de taula	DimActivity							
Tipus de taula	Dimensió							
Nom de la vista	Activity							
Descripció	informació de l'activitat setmanal que ha seguit un pacient durant el tractament.							
Mida	3 registres							
	Destí				Origen			
Nom	Descripció	DataType	exemple	Sistema d'entrada	Full origen	Nom origen	Dataype	Regles ETL
ActivityKey	Clau primària	int	1,2,3...	ETL				
Activity	Tipus de dieta	string	Healty ; Normal...	Excel	Hàbits	Activity	text	
Jerarquia	Nivell 0			Activity				

Nom de taula	DimDiet							
Tipus de taula	Dimensió							
Nom de la vista	Diet							
Descripció	informació de la dieta setmanal que ha seguit un pacient durant el tractament.							
Mida	2 registres							
0								
	Destí				Origen			
Nom	Descripció	DataType	exemple	Sistema d'entrada	Full origen	Nom origen	Dataype	Regles ETL
DietKey	Clau primària	int	1,2,3...	ETL				
Diet	Tipus de dieta	string	FAT MEDITERRANEAN...	Excel	Hàbits	Diet	text	
Jerarquia	Nivell 0			Diet				

3.4 Implementació física del model de dades

Un cop definit el model lògic de dades s'implementa el model físic utilitzant l'eina MySQL Workbench. Aquesta aplicació permet el desenvolupament, l'administració, el disseny, la creació i el manteniment en un únic entorn de desenvolupament per al sistema de bases de dades MySQL.

El resultat és un *schema* de MySQL que contindrà les taules definides així com els camps i relacions establertes. En primer lloc es defineix el model amb les eines visuals incorporades amb MySQL Workbench. Això permet exportar l'*script* SQL de creació del *schema*. Finalment, un cop connectada la instància de MySQL amb el MySQL Workbench s'executa l'*script* i és pot verificar que la creació física del model ha estat la prevista.

La següent imatge ens mostra el disseny del model físic que ha permès l'automatització de les sentències SQL de creació.

Figura 5. Diagrama físic del model de dades



4. Processos ETL

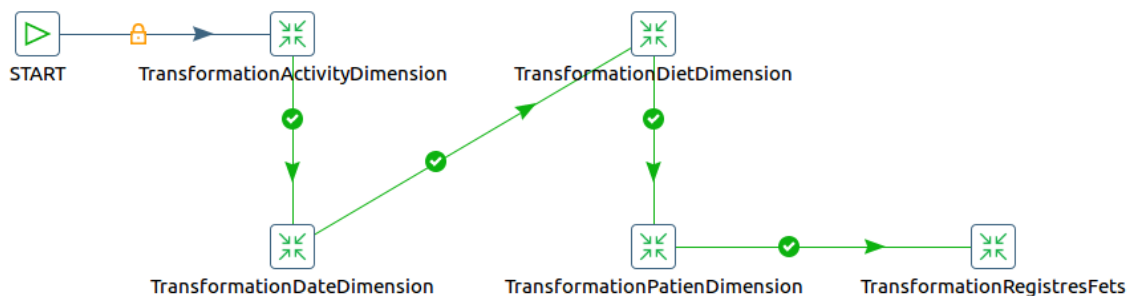
4.1 Processos de càrrega transformació i extracció.

Un cop s'ha definit i creat el magatzem de dades és realitzen els diferents processos coneguts com a ETL. Així, utilitzant com a base els fitxers de dades en format Excel, ja estudiats, es processarà l'extracció (*extract*) la transformació (*transform*) i la càrrega (*load*) de les dades per tal d'alimentar el DWH i entregar una col·lecció d'informació dimensional para la seva consulta i anàlisis per tal de donar suport a les aplicacions per la presa de decisions.

Tal i com es va definir en l'etapa en que es va analitzar i seleccionar entre les diferents plataformes *BI Open Source* i SGBD s'utilitzarà Pentaho Data Integration, amb nom de codi Kettle *Kettle ETL Environment*. Aquesta aplicació consisteix en un motor d'integració de dades bàsics (ETL) i aplicacions GUI que permeten a definir treballs i transformacions d'integració, que inclouen l'extracció, transformació, transport i càrrega de dades. Spoon és una interfície gràfica d'usuari que permet dissenyar transformacions i treballs que es poden executar amb les eines Kettle.

La següent imatge representa gràficament la seqüència dels diferents processos ETL realitzats.

Figura 6. Seqüència de Processos ETL



Aquets processos han estat definits i portats a terme iterativament fins a obtenir un procediment automàtic i senzill de modificar que permet la seva execució en qualsevol moment. Els següents apartats d'aquest capítol descriuen les principals fases de cadascun dels processos ETL necessaris per implementar la càrrega de dades així com les transformacions modelades en l'etapa de disseny.

4.2 Dimensió activity.

Els diferents processos han estat llegir les dades requerides del full d'origen Excel HABITS, ordenar les dades i extreure possibles caràcters

buits a l'inici o finalització de les dades, eliminar registres repetits, afegir una clau autonumèrica i finalment inserir les dades al DWH. Les següents imatges mostren d'una banda la representació gràfica de la transformació i d'altra banda les dades ja normalitzades en el repositori.

Figura 7. Processos ETL de la dimensió activity



Figura 8. Dades normalitzades de la dimensió activity

1 • `SELECT * FROM TFM2018.dimActivity;`

#	ActivityKey	Activity
1		HEALTHY
2		NORMAL
3		SEDENTARIAN
	NULL	NULL

4.3 Dimensió diet.

Els diferents processos han estat llegir les dades requerides del full d'origen Excel HABITS, ordenar les dades i extreure possibles caràcters buits a l'inici o finalització de les dades, eliminar registres repetits, afegir una clau autonumèrica i finalment inserir les dades al DWH. Les següents imatges mostren d'una banda la representació gràfica de la transformació i d'altra banda les dades ja normalitzades en el repositori.

Figura 9. Processos ETL de la dimensió diet



Figura 10. Dades normalitzades de la dimensió diet

```
1 • SELECT * FROM TFM2018.dimDiet;
```

#	DietKey	Diet
1	1	FAT
2	2	MEDITERRANEAN
3	3	VEGETARIAN
*	NULL	NULL

4.4 Dimensió date.

Els diferents processos han estat llegir les dades requerides del full d'origen Excel INDICATORS, ordenar les dades i extreure possibles caràcters buits a l'inici o finalització de les dades, eliminar registres repetits, utilitzar la transformació "calculator" per crear els nous camps day, week, month, year i trimestrer, afegir una clau autonumèrica i finalment inserir les dades al DWH. Les següents imatges mostren d'una banda la representació gràfica de la transformació i d'altra banda una mostra de les dades ja normalitzades en el repositori

Figura 11. Processos ETL de la dimensió date

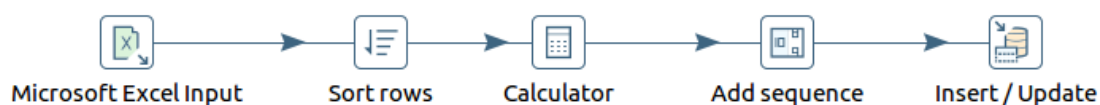


Figura 12. Dades normalitzades de la dimensió date

```
1 • SELECT * FROM TFM2018.dimDate;
```

dateKey	Date	Day	Week	Month	Trimestrer	Year
1	2017-01-02	2	1	1	1	2017
2	2017-01-09	9	2	1	1	2017
3	2017-01-16	16	3	1	1	2017
4	2017-01-23	23	4	1	1	2017
5	2017-01-30	30	5	1	1	2017
6	2017-02-06	6	6	2	1	2017
7	2017-02-13	13	7	2	1	2017
8	2017-02-20	20	8	2	1	2017
9	2017-02-27	27	9	2	1	2017
10	2017-03-06	6	10	3	1	2017
11	2017-03-13	13	11	3	1	2017
12	2017-03-20	20	12	3	1	2017
13	2017-03-27	27	13	3	1	2017
14	2017-04-03	3	14	4	1	2017
15	2017-04-10	10	15	4	1	2017
16	2017-04-17	17	16	4	1	2017
17	2017-04-24	24	17	4	1	2017
18	2017-05-01	1	18	5	1	2017
19	2017-05-08	8	19	5	1	2017
20	2017-05-15	15	20	5	1	2017
21	2017-05-22	22	21	5	1	2017
22	2017-05-29	29	22	5	1	2017
23	2017-06-05	5	23	6	1	2017
24	2017-06-12	12	24	6	1	2017
25	2017-06-19	19	25	6	1	2017
26	2017-06-26	26	26	6	1	2017
27	2017-07-03	3	27	7	1	2017
28	2017-07-10	10	28	7	1	2017
29	2017-07-17	17	29	7	1	2017
30	2017-07-24	24	30	7	1	2017
31	2017-07-31	31	31	7	1	2017
32	2017-08-07	7	1	8	1	2017
33	2017-08-14	14	2	8	1	2017
34	2017-08-21	21	3	8	1	2017
35	2017-08-28	28	4	8	1	2017
36	2017-09-04	4	5	9	1	2017
37	2017-09-11	11	6	9	1	2017
38	2017-09-18	18	7	9	1	2017
39	2017-09-25	25	8	9	1	2017
40	2017-10-02	2	9	10	1	2017
41	2017-10-09	9	10	10	1	2017
42	2017-10-16	16	11	10	1	2017
43	2017-10-23	23	12	10	1	2017
44	2017-10-30	30	13	10	1	2017
45	2017-11-06	6	14	11	1	2017
46	2017-11-13	13	15	11	1	2017
47	2017-11-20	20	16	11	1	2017
48	2017-11-27	27	17	11	1	2017
49	2017-12-04	4	18	12	1	2017
50	2017-12-11	11	19	12	1	2017
51	2017-12-18	18	20	12	1	2017
52	2017-12-25	25	21	12	1	2017

4.5 Dimensió patient.

Els diferents processos han estat llegir les dades requerides de forma separada dels fulls d'origen Excel PATIENTS I TREATMENT, ordenar cada una de les dades i extreure possibles caràcters buits a l'inici o finalització de les dades, eliminar registres repetits, unificar les dades necessàries en una sola consulta, afegir una clau autonumèrica i finalment inserir les dades al DWH. Les següents imatges mostren d'una banda la representació gràfica de la transformació i d'altra banda les dades ja normalitzades en el repositori.

Figura 13. Processos ETL de la dimensió patient

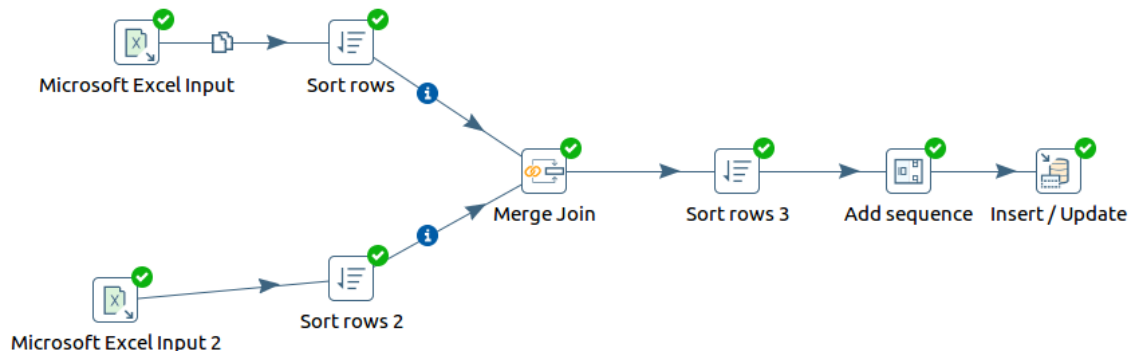


Figura 14. Dades normalitzades de la dimensió patient

```
1 • SELECT * FROM TFM2018.dimPatient;
```

#	PatienKey	Age	Patient	Gender	Treatment	City	CCAA
1	1	60-99	P1	M	PHARMA	BARCELONA	CATALUNYA
2	2	26-45	P2	M	NATURAL	VILAFRANCA P.	CATALUNYA
3	3	46-60	P3	M	HOMEOPATHIC	ZUMAIA	EUSKADI
4	4	60-99	P4	F	NATURAL	VITORIA	EUSKADI
5	5	26-45	P5	F	PHARMA	GETAFE	MADRID
6	6	46-60	P6	F	HOMEOPATHIC	MADRID	MADRID

4.6 Taula de fets registres.

Els diferents processos han estat llegir les dades de forma separada dels fulls d'origen Excel INDICATORS i HABITS, ordenar cada una de les dades i extreure possibles caràcters buits a l'inici o finalització de les dades, eliminar registres repetits, unificar les dades necessàries en una sola consulta, afegir una clau autonumèrica, utilitzar les dades obtingudes per cercar en les taules del DWH les claus de les diferents dimensions per tal de garantir que conflueixen en la taula de fets els nivells més baixos de cada dimensió. Finalment es fa un insert de les dades obtingudes al DWH. Les següents imatges mostren d'una banda la representació gràfica de la transformació i d'altra banda una mostra de les dades ja normalitzades en el repositori.

Figura 15. Processos ETL de la taula de fets

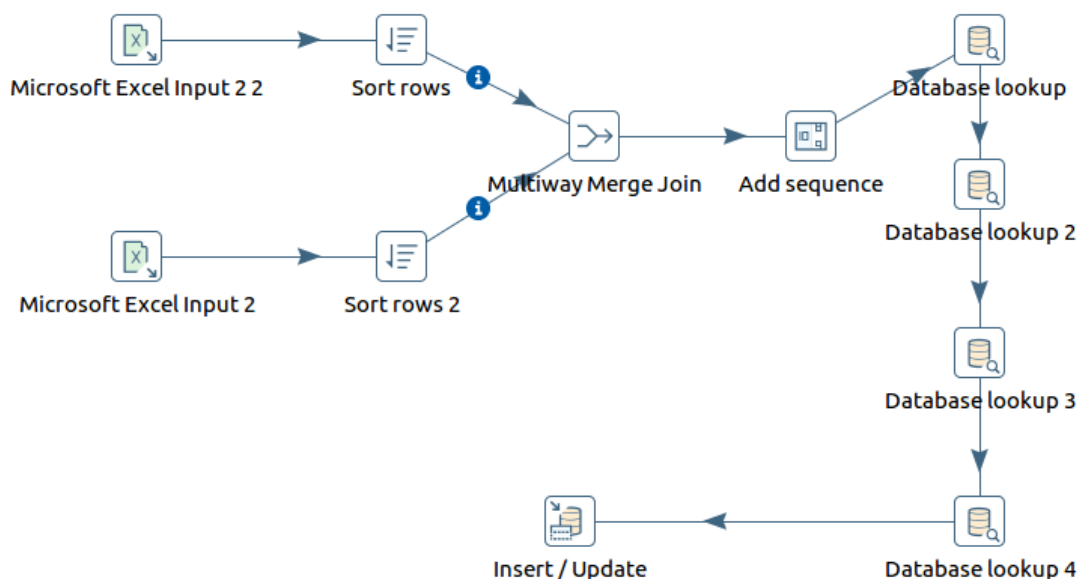


Figura 16. Dades normalitzades de la taula de fets

```
1 • SELECT * FROM TFM2018.registres;
```

RegistresKey	dimPatinent_	dimDate_dateKey	dimActivity_activit	dimDiet_dietKey	ldl	sbp	dbp
1	1	1	3	1	247	7.15	11.75
2	1	2	3	2	241	7.24	11.75
3	1	3	3	2	243	6.8	11.75
4	1	4	3	1	238	6.93	11.75
5	1	5	3	2	250	7.23	11.75
6	1	6	3	2	247	7.02	11.75
7	1	7	3	2	236	7.1	11.75
8	1	8	3	1	245	6.82	11.75
9	1	9	2	2	247	7.22	11.75
10	1	10	3	2	247	7.23	11.75

5. Implementació eines OLAP

5.1 Eines OLAP i multidimensionalitat.

Un cop realitzats els processos ETL per tal de transformar i emmagatzemar les dades de tal manera que presentin la dimensionalitat necessària per respondre les preguntes plantejades, es fa necessari el disseny i implementació d'eines OLAP (*Online Analytical Processing*).

OLAP es un terme introduït per E.F. Codd l'any 1993, per definir una categoria d'eines de programari que tenen la finalitat de permetre el coneixement de les dades d'una manera ràpida, consistent i interactiva presentant les dades que es volen analitzar en termes de fets i dimensions, de tal manera que es poden situar en un espai n-dimensional anomenant cub. Aquest cub, generalment de més de tres dimensions, hi ha d'una banda els fets, dades mètriques, que es volen analitzar i les seves dimensions que es poden entendre com les dades descriptives que contextualitzen i defineixen els fets (què, a qui, on, quan, com, etc.).

El conjunt d'eines OLAP es poden definir utilitzant el test presentat per Nigel Pendse l'any 2005, FASMI (*Fast analysis of Shared multidimensional information.*):

Fast

De forma habitual les consultes han de donar el seu resultat en uns cinc segons. Ara bé, consultes excepcionals poden disposar d'una durada inferior a 20 segons.

Analysis

Les eines OLAP han d'oferir a usuaris no especialitzats en programació o sistemes de gestió un ventall de possibilitats que els permetin l'anàlisi estadístic i la generació d'informes en relació als fets que es pretenen estudiar.

Shared

Aquesta categoria de programari ha d'implementar mecanismes de concurrència així com de control d'accés i confidencialitat de les dades. Això ha de permetre emmagatzemar i compartir la informació resultant.

Information

El sistema ha de tenir la capacitat d'emmagatzemar tota la informació necessària independentment del seu volum alhora que permet inserir metadades per tal d'enregistrar el seu significat.

5.2 Disseny i implementació.

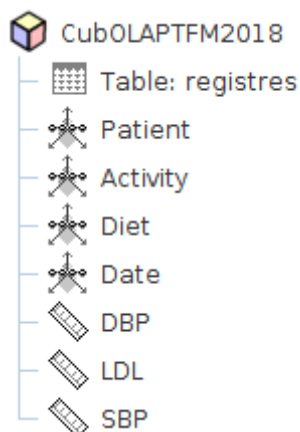
La definició prèvia del model multidimensional així com l'execució dels diferents processos ETL permet ara dissenyar i implementar els cubs per tal de publicar-los en el servidor BI i el seu posterior anàlisi.

L'eina seleccionada de la solució Pentaho Suite Community Edition per la definició i el disseny del cub és Penhaho Schema Workbench anomenada Mondrian, aplicació que permet crear i provar visualment cubs OLAP. A partir del model físic de dades multidimensionals, crearà un esquema amb el model lògic d'una estructura de bases de dades també multidimensional. En el cas d'aquest projecte es definirà un esquema Mondrian amb un sol cub que tindrà una taula de fets i diverses dimensions, cadascuna amb una o diverses jerarquies que consisteix en un nombre diferents de nivells.

Els fitxers resultants són models de metadades XML amb esquemes ROLAP (relacional OLAP) que es creen en una estructura específica que no requereix que es construeixi o mantingui un cub físic real; només es crea el model de metadades que posteriorment implementarà el servidor d'anàlisi BI.

La següent imatge mostra el Cub OLAP ja definit a Mondrian Schema Workbench amb les diferents dimensions, mesures i la taula de fets.

Figura 17. Cub OLAP

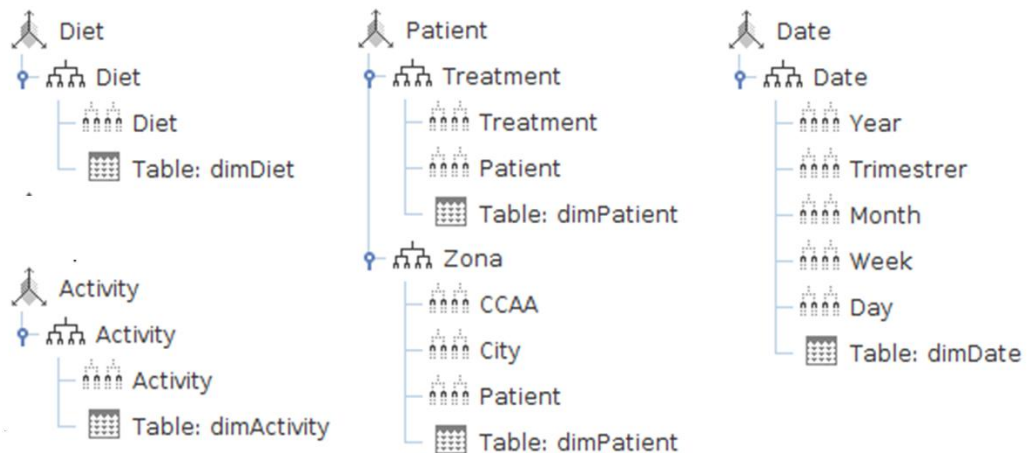


Aquest nivell superior de visualització s'observa el cub OLAP conformat per la taula de fets anomenada registres vinculada a les diferents dimensions per les seves claus principals com a claus foranies.

D'altra banda, hi ha les mètriques necessàries per l'anàlisi de les dades que corresponen a les lipoproteïnes de baixa densitat (LDL) i les mètriques que informen de la tensió arterial del pacient *systolic blood pressure* (SBP) i *diastolic blood pressure* (DBP).

Les dimensions han estat creades a nivell d'esquema, tot i no ser necessari per aquest projecte es defineixen d'aquesta manera per si en un futur es volgués genera un cub adicional poder-les reutilitzar.

Figura 18. Dimensions, jerarquies i granularitat



En aquesta figura hi ha les diferents dimensions amb les seves jerarquies i nivells. Cada dimensió vincula la seva taula corresponent amb la taula de fets del cub.

Les diferents jerarquies de les dimensions ens han de permetre manipular del cub de manera fàcil i flexible en el servidor BI sense perdre capacitat de càlcul. Així, a més de les operacions *Slice* i *Dice* aquesta granularitat ens permetrà canviar el nivell de detall i l'objecte d'anàlisi per medi d'operacions del tipus *Roll-up*, *Drill-down* i *Drill-across*.

5.3 Publicació i proves.

Com es va definir en etapes anteriors del projecte s'ha implementat el servidor *Open Source Pentaho Business Analytics CE* per tal de disposar d'una capa de programari per l'anàlisi de la informació. Un cop realitzada la publicació del cub OLP al servidor per medi de les funcionalitats pròpies de *Mondrian Schema Workbench* es disposa d'accés al cub per realitzar les primeres comprovacions i proves.

En una primera instància es visualitza el cub i es realitzen les modificacions necessàries en el disseny en l'entorn *Mondrian* per tal d'aconseguir un refinament en la configuració del cub que permeti validar els resultats per medi d'exemples aleatoris realitzats de manera empírica.

Per tal de poder navegar els cubs OLAP i mostrar la informació en forma de taules i gràfics *Pentaho Server CE* integra en la seva instal·lació per defecte *Jpivot* que permet realitzar les operacions típiques dels entorns OLAP. El seu anàlisi per defecte permet seleccionar entre les diferents dimensions, nivells i jerarquies del cub i visualitzar les dades referenciades en la nostra selecció. En la següent imatge es visualitza el cub implementat utilitzant *Jpivot*.

Figura 19. Visualització amb Jpivot del cub OLAP

				Measures
Treatment	Activity	Diet	Date	DBP
All Patient. Treatments	All Activitys	All Diets	All Dates	13,114
HOMEOPATHIC	All Activitys	All Diets	All Dates	13,253
		HEALTHY	All Dates	13,283
		FAT	All Dates	14,005
		MEDITERRANEAN	All Dates	13,07
			2017	13,07
			1	12,867
			2	13,245
			3	13,265
			4	13,084
			VEGETARIAN	All Dates
	NORMAL	All Diets	All Dates	13,119
NATURAL	All Activitys	All Diets	All Dates	13,212
PHARMA	All Activitys	All Diets	All Dates	12,875

Tot i això després de consultar la documentació de referència s'ha decidit utilitzar Saiku en la seva versió Saiku Analytics CE una suite modular d'anàlisi BI de codi obert que disposa d'una versió gratuïta en el Market Place de Pentaho instal·lable com a un Plugin del servidor Pentaho.

Saiku és considerada com a una solució potent i flexible que permet l'anàlisi OLAP de manera fàcil i intuïtiva. La següent imatge mostra una aproximació a l'entorn d'usuari de Saiku.

Figura 20. Visualització amb Saiku del cub OLAP

The screenshot shows the Saiku Analytics user interface. On the left, there is a 'Cubes' dropdown set to 'CubOLAPTFM2018'. Below it, the 'Measures' section lists 'DBP', 'LDL', and 'SBP'. The 'Dimensions' section shows a tree structure with 'Activity', 'Date' (expanded to show 'Date', '(All)', 'Year', 'Trimestrer', 'Month', 'Week', 'Day'), 'Diet' (expanded to show '(All)', 'Diet'), and 'Patient'. On the right, a pivot table is displayed for the year 2017. The table has columns for 'Year', 'Trimestrer', and four quarters (1, 2, 3, 4). The rows represent different diet types: 'FAT', 'MEDITERRANEAN', and 'VEGETARIAN'. The data values are numerical, representing measurements like DBP, SBP, and LDL.

Diet	2017											
	1		2		3		4		DBP		LDL	
Trimestrer	DBP	SBP	LDL	DBP	SBP	LDL	DBP	SBP	LDL	DBP	SBP	LDL
FAT	13,416	7,466	267,147	13,581	7,478	265	13,748	7,542	270,12	13,805	7,5	269,864
MEDITERRANEAN	12,99	7,531	285,139	12,902	7,471	273,162	12,934	7,458	262,707	12,892	7,453	260,4
VEGETARIAN	12,494	7,39	347,75	12,457	7,37	341,8	12,409	7,359	339,667	12,652	7,491	301,312

6. Anàlisi de dades

Arribat a aquest capítol es disposa d'un sistema d'intel·ligència de negoci implementat amb eines Open Source, amb els diferents components: el magatzem de dades, els processos d'extracció i transformació, la creació del magatzem de dades, l'anàlisi multidimensional i la capacitat de realització d'informes.

L'anàlisi es centrarà en la resolució de les diferents preguntes analítiques que es volen respondre. En els següents apartats es realitzen una sèrie de consideracions inicials i s'aborden les sis preguntes mostrant taules de dades i representacions gràfiques per tal d'ajudar a una millor comprensió dels resultats.

6.1 Abast de l'anàlisi.

La plataforma implementada en aquest projecte proporciona un sistema d'intel·ligència de negoci i anàlisi de dades en un entorn multidimensional que permet als analistes poder disposar de la informació que requereixin, per tal de poder fer els estudis i extreure les conclusions que la seva expertesa en la matèria els hi permet.

Ara bé, en aquest treball hem de considerar que les dades proporcionades corresponen a les analítiques clíniques i el seu entorn en el cas de sis pacients durant el període d'un any. Per tant, les respostes seguidament proporcionades han de quedar clarament circumscrites a aquestes condicions, en cap cas es pretenen inferir conclusions més enllà de l'anàlisi directe de les dades disponibles.

6.2 Mètriques.

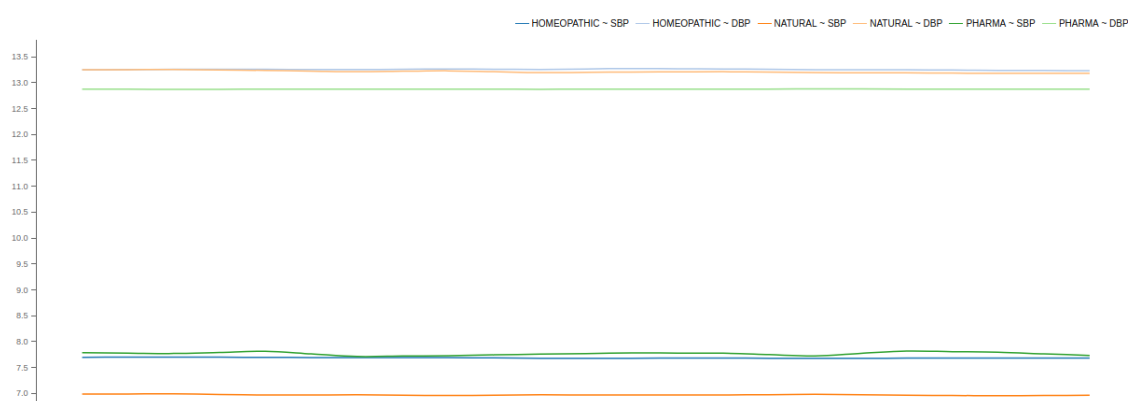
Les tres dades quantitatives de que es disposen per realitzar els diferents anàlisis, són d'una banda les lipoproteïnes de baixa densitat o LBD, conegudes també popularment com a "colesterol dolent" de les que utilitzarem el seu acrònim anglès LDL (de *Low Density Lipoprotein*) i d'altra banda els valors de la pressió arterial del pacient, tant del ventricle esquerra (pressió sistòlica o SBP) com del ventricle dret (pressió diastòlica o DBP).

Tanmateix, un anàlisi previ de les dades indica que la pressió arterial dels pacients no s'ha vist afectada al llarg dels diferents tractaments en uns valors que puguin ser significatius. Així, aquests valors no seran considerants per afrontar la resolució de les diferents qüestions, tot i això, les següents gràfiques deixen constància de la seva evolució en el transcurs de cada tractament en períodes agrupats d'un mes.

Figura 21. Taula d'anàlisi del SBP i el DBP en relació al tractament

Treatment - Treatment	HOMEOPATHIC		NATURAL		PHARMA	
Month	DBP	SBP	DBP	SBP	DBP	SBP
1	13,25	7,696	13,249	6,985	12,876	7,784
2	13,256	7,7	13,254	6,989	12,871	7,769
3	13,256	7,694	13,236	6,966	12,874	7,811
4	13,252	7,69	13,213	6,97	12,874	7,711
5	13,263	7,69	13,229	6,957	12,876	7,728
6	13,255	7,676	13,194	6,97	12,873	7,756
7	13,274	7,678	13,206	6,968	12,876	7,779
8	13,265	7,68	13,211	6,968	12,875	7,774
9	13,249	7,675	13,194	6,98	12,877	7,723
10	13,248	7,679	13,189	6,962	12,877	7,817
11	13,237	7,681	13,181	6,954	12,876	7,793
12	13,23	7,679	13,184	6,961	12,876	7,731

Figura 22. Gràfic d'anàlisi del SBP i el DBP en relació al tractament



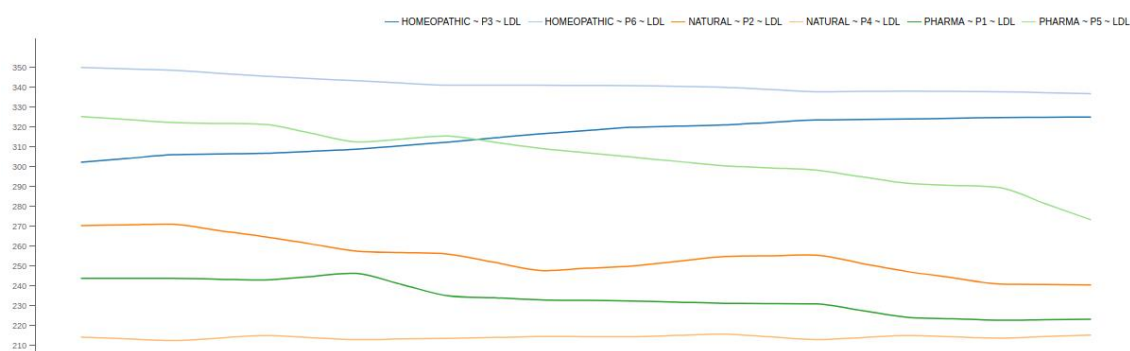
6.3 Quina és la relació entre els diferents tractaments i l'evolució dels pacients?

Per respondre aquesta pregunta s'utilitzen les dimensions treatment, patient i data per veure l'evolució dels pacients en els diferents tractaments, la dimensió data es granula a un mes per facilitar la visualització. En les figures 23 i 24 es pot observar que **el tractament farmacològic és l'únic que expressa una relació entre el tractament i l'evolució dels pacients** ja que redueix sensiblement l'LDL en els dos pacients. Per altra banda, els tractaments homeopàtics i naturals mostren resultats dispersos on hi ha pacients que redueixen els nivells i d'altres que els augmenten.

Figura 23. Taula d'anàlisi entre els tractaments i l'evolució dels pacients

Month		1	2	3	4	5	6	7	8	9	10	11	12
Treatment	Patient	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL
HOMEOPATHIC	P3	302,2	306	306,75	308,75	312,4	316,5	319,8	321	323,5	324	324,75	325
	P6	350	348,5	345,5	343,25	341	341	340,8	340	337,75	338	337,75	336,75
NATURAL	P2	270,4	271	264,75	257,5	256	247,75	250	254,75	255,5	247,2	241	240,5
	P4	214,2	212,5	215	213	213,6	214,5	214,4	215,75	213	215	213,75	215,25
PHARMA	P1	243,8	243,75	243	246,25	235	233	232,4	231,25	231	224,2	222,75	223,25
	P5	325,2	322,25	321,25	312,5	315,4	309,25	304,8	300,5	298,25	291,6	289,5	273,25

Figura 24. Gràfic d'anàlisi entre els tractaments i l'evolució dels pacients



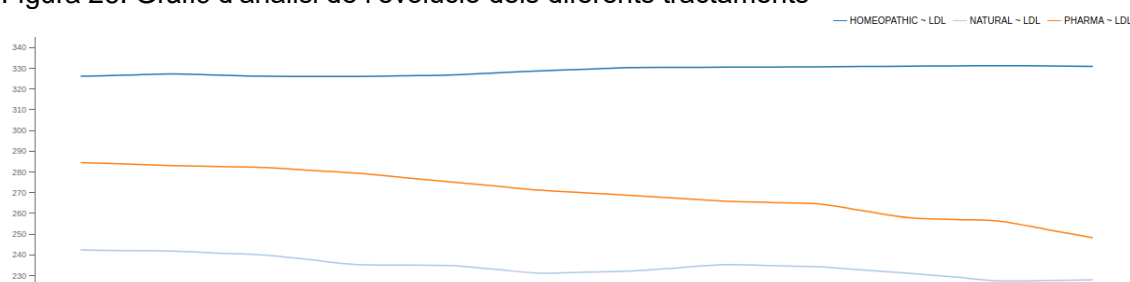
6.4 Existeixen teràpies més eficaces?

Sí, utilitzant la vista anterior i la representada en les següents dos figures, on s'agrupen els diferents tractaments, es pot observar com els valors medis dels **pacients amb tractament farmacològic han reduït el seu índex LDL en 12.74 %**, els pacients amb tractament natural han reduït en 5.95 % i finalment els pacients amb tractament homeopàtic han augmentat el seu índex LDL en 1.46 %. Tot i això, pel que fa al tractament natural cal fer notar els resultats amb tendències desiguals en l'evolució dels pacients, ja detectades en la pregunta anterior.

Figura 25. Taula d'anàlisi de l'evolució dels diferents tractaments

Month	1	2	3	4	5	6	7	8	9	10	11	12
Treatment	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL
HOMEOPATHIC	326,1	327,25	326,125	326	326,7	328,75	330,3	330,5	330,625	331	331,25	330,875
NATURAL	242,3	241,75	239,875	235,25	234,8	231,125	232,2	235,25	234,25	231,1	227,375	227,875
PHARMA	284,5	283	282,125	279,375	275,2	271,125	268,6	265,875	264,625	257,9	256,125	248,25

Figura 26. Gràfic d'anàlisi de l'evolució dels diferents tractaments



6.5 Ha influït en el resultat, els hàbits dels pacients?

En primer lloc s'estudia la relació entre l'índex LDL i l'activitat del pacient, en la figura 27 hi ha la taula resum de dotze mesos on és mostren diverses tendències que poden ser confirmades en la gràfica de la figura 28. Ara bé, en fer un detall setmanal (figura 29) es poden observar valors irregulars que corresponen a canvis d'hàbits dels pacients. Un anàlisi en detall de l'activitat de cada pacient en referència a la seva activitat física es mostren en la taula 30 i el gràfic 31.

Així, amb les mostres de que es disposen i amb l'estudi d'aquestes gràfiques **no es pot concloure que l'activitat física hagi tingut influència en l'evolució dels resultats** ja que en els índexs de cada pacient no es veu un patró de modificació en les setmanes en que hi ha una activitat diferent i pel que fa a la tendència anual en la comparativa entre pacients també hi ha resultats contradictoris entre activitat física i evolució del LDL que no permeten establir una influència directa.

Figura 27. Taula d'anàlisi dels resultats en relació a l'activitat

Month	1	2	3	4	5	6	7	8	9	10	11	12
Activity	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL
HEALTHY	309,143	304,333	305,667	307,364	292	299,417	302,214	300,182	310	299,6	301,167	298,692
NORMAL	285,133	285,091	260,286	255,625	290	273,286	262,5	276,75	252,5	258,5	248,333	236,667
SEDENTARIAN	261	235,25	259	259,8	230,833	228,4	232,5	227,4	233,667	224,2	223	221,5

Figura 28. Gràfic d'anàlisi dels resultats en relació a l'activitat

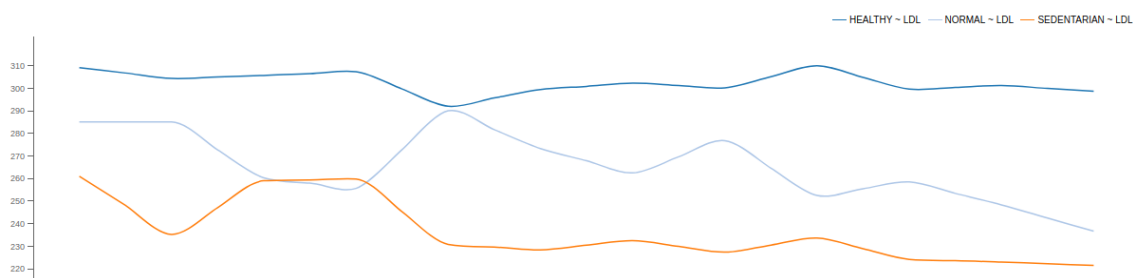


Figura 29. Gràfic d'anàlisi dels resultats en relació a l'activitat

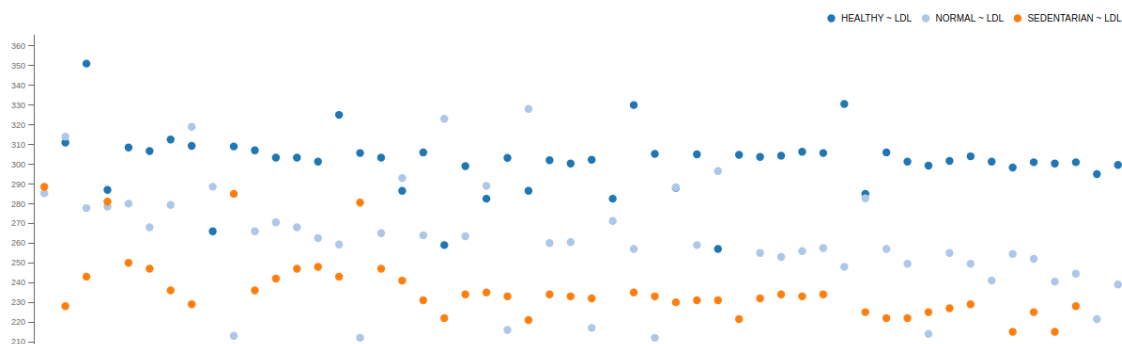
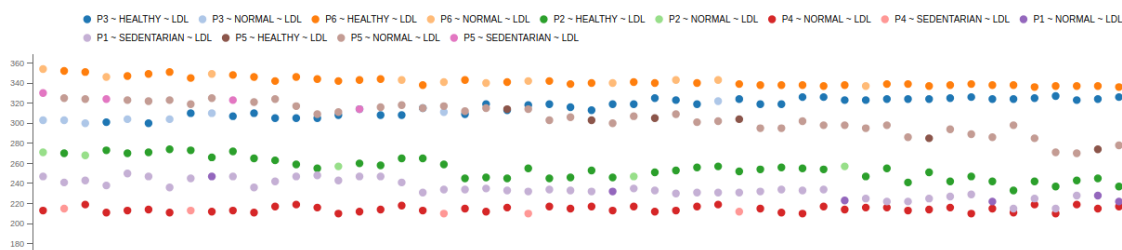


Figura 30. Taula d'anàlisi dels resultats en relació a l'activitat de cada pacient

Activity	HEALTHY				NORMAL						SEDENTARIAN		
	P3	P6	P2	P5	P3	P6	P2	P4	P1	P5	P4	P1	P5
Treatment - Patient	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL
1	301	350	271	-	302,5	350	269,5	214	-	324	215	243,8	327
2	305	348,333	271	-	307	349	-	212,333	247	322,25	213	242,667	-
3	306,75	345,5	264,75	-	-	-	-	215	-	320,667	-	243	323
4	308,75	343,25	257,667	-	-	-	257	213	-	312	-	246,25	314
5	312,75	340,5	256	-	311	341,333	-	214,5	-	315,4	210	235	-
6	316,5	340,667	247,75	314	-	342	-	216	-	307,667	210	233	-
7	319,8	340,333	250,75	304	-	341,5	247	214,4	232	305,333	-	232,5	-
8	320,667	339	254,75	304	322	343	-	217	-	299,333	212	231,25	-
9	323,5	337,75	255	-	-	-	257	213	223	298,25	-	233,667	-
10	324	338,25	247,2	285	-	337	-	215	-	293,25	-	224,2	-
11	324,75	337,75	241	-	-	-	-	213,75	222	289,5	-	223	-
12	325	336,75	240,5	274	-	-	-	215,25	225	273	-	221,5	-

Figura 31. Gràfic d'anàlisi dels resultats en relació a l'activitat de cada pacient



En segon lloc s'estudia la relació entre l'índex LDL i la dieta del pacient, es realitza l'anàlisi amb una graduació temporal mensual (figures 33 i 34) de l'evolució dels resultats per una mateixa dieta i un segon anàlisi amb el detall de la dieta habitual de cada pacient així com l'evolució dels resultats de forma setmanal (figures 34 i 35). En aquest cas, en els dos anàlisis, es pot apreciar una relació entre la dieta i les tendències de les mètriques.

Per tant, **els pacients amb dietes principalment vegetarianes han millorat els seus valors un 18% i els pacients amb dietes mediterrànies ho fan en un 15% , per la seva banda el pacients amb dietes greixoses empitjoren en un 1%.**

Figura 32. Taula d'anàlisi dels resultats en relació a la dieta

Month	1	2	3	4	5	6	7	8	9	10	11	12
Diet	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL
FAT	268,133	276,636	252,25	271,636	255,2	267,5	263,222	268,375	279,625	269,75	269,25	270,833
MEDITERRANEAN	282,818	279,546	291,357	270,8	278,188	268	264,875	268,077	254	268,056	260,167	246,9
VEGETARIAN	349	349	344	343	341	341,667	340,8	340,333	337,75	304,25	310,5	295,25

Figura 33. Gràfic d'anàlisi dels resultats en relació a la dieta

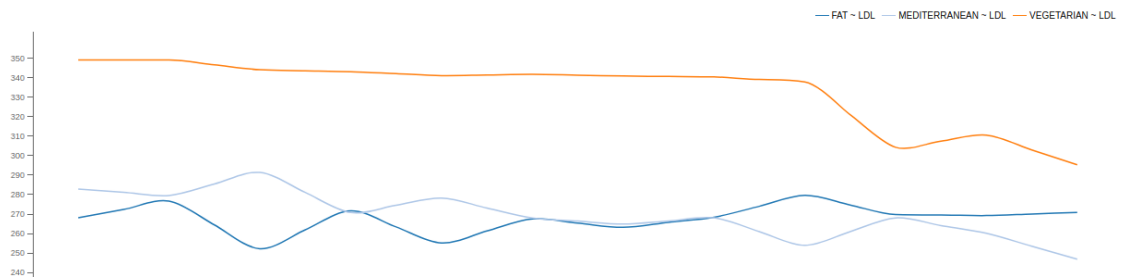
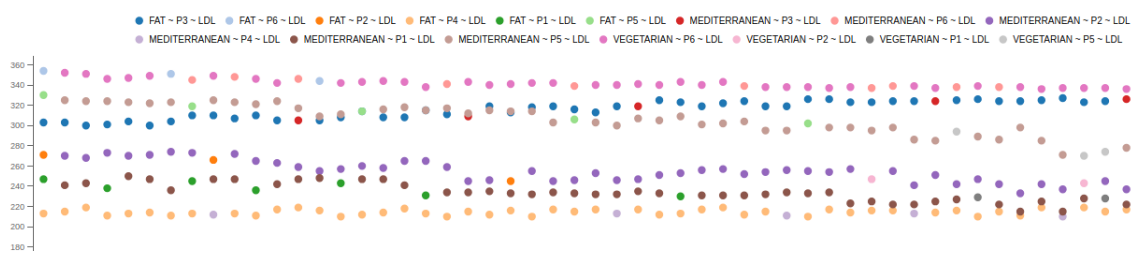


Figura 34. Taula d'anàlisi dels resultats en relació a la dieta de cada pacient

Diet	FAT						MEDITERRANEAN						VEGETARIAN			
	P3	P6	P2	P4	P1	P5	P3	P6	P2	P4	P1	P5	P6	P2	P1	P5
Treatment - Patient	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL	LDL
1	302,2	354	271	214,2	242,5	330	-	-	270,25	-	244,667	324	349	-	-	-
2	306	351	266	212,667	245	319	-	345	272,667	212	243,333	323,333	349	-	-	-
3	307,333	-	-	215	236	-	305	347	264,75	-	245,333	321,25	344	-	-	-
4	308,75	344	-	213	243	314	-	-	257,5	-	247,333	312	343	-	-	-
5	313,25	-	-	213,6	231	-	309	341	256	-	236	315,4	341	-	-	-
6	316,5	-	245	214,5	-	306	-	339	248,667	-	233	310,333	341,667	-	-	-
7	320	-	-	214,75	230	-	319	-	250	213	233	304,8	340,8	-	-	-
8	321	-	-	215,75	-	-	-	339	254,75	-	231,25	300,5	340,333	-	-	-
9	323,5	-	-	213,667	-	302	-	-	255,5	211	231	297	337,75	-	-	-
10	324	-	-	215,5	-	-	324	338	247,25	213	224,2	291	338	247	-	294
11	324,75	-	-	213,75	-	-	-	338	241	-	220,667	289,5	337,667	-	229	-
12	324,667	-	-	217	-	-	326	-	239,667	210	221,667	274,5	336,75	243	228	272

Figura 35. Gràfic d'anàlisi dels resultats en relació a la dieta de cada pacient



6.6 L'evolució al llarg del temps, per un mateix tractament, depenen d'algun factor com els hàbits?

En aquest cas s'utilitzaran les dimensions Activity i Diet per descobrir la seva possible incidència sobre l'evolució d'un tractament. És disposen de 6 teràpies que s'agrupen en 3 grups diferents. Cal tenir clar que cada pacient pot modificar els seus hàbits cada setmana, tot i això, en les

figures 37 i 38 s'observa que cada pacient disposa d'una sola teràpia i uns hàbits habituals que es mantenen sostinguts en el temps i permeten la seva comparació.

Les taula 37 permeten identificar aquests hàbits majoritaris en cada pacient conjuntament amb la teràpia que ha estat seguint, així mateix la gràfica 38 ens mostra aquest resultat així com les possible variacions setmanals. Amb aquesta informació s'elabora la taula resum 36 on es recull l'evolució en LDL de cada pacient conjuntament amb el tractament seguit i els hàbits habituals.

Finalment és pot inferir que els pacients amb dietes mediterrànies han obtingut millor resultats, cal destacar el cas del pacient P2 que tot i no estar amb tractament farmacològic mostra una millora considerable amb un tractament natural i una activitat saludable conjuntament amb la dieta mediterrània. D'altra banda **la comparació dels resultats de les sis teràpies agrupades en tres tractaments mostren que els pacients amb hàbits més saludables obtenen sempre millors resultats per un mateix tractament al llarg del temps.**

Figura 36. Taula resum de l'evolució dels diferents tractaments relacionats amb els hàbits de cada pacient.

	Treatment	Activity	Diet	Resultat
P5	FARMA	NORMAL	MEDITERRANEAN	19 %
P1	FARMA	SEDENTARIAN	MEDITERRANEAN	9 %
P2	NATURAL	HEALTHY	MEDITERRANEAN	12 %
P4	NATURAL	NORMAL	FAT	0 %
P6	HOMEOPATIC	HEALTHY	VEGETARIAN	4 %
P3	HOMEOPATIC	HEALTHY	FAT	-7 %

Figura 37. Gràfic d'anàlisi de l'evolució dels diferents tractaments relacionats amb els hàbits de cada pacient.

Patient	Treatment	Activity	Diet	Month												
				1	2	3	4	5	6	7	8	9	10	11	12	
P1	PHARMA	NORMAL	MEDITERRANEAN	-	247	-	-	-	-	-	232	-	223	-	222	222
			VEGETARIAN	-	-	-	-	-	-	-	-	-	-	-	-	228
		SEDENTARIAN	FAT	242,5	245	236	243	231	-	230	-	-	-	-	-	-
			MEDITERRANEAN	244,667	241,5	245,333	247,333	236	233	233,333	231,25	233,667	224,2	220	221,5	
P2	NATURAL	HEALTHY	FAT	-	266	-	-	-	-	245	-	-	-	-	-	
			MEDITERRANEAN	271	272,667	264,75	257,667	256	248,667	250,75	254,75	255	247,25	241	239,667	
			VEGETARIAN	-	-	-	-	-	-	-	-	-	247	-	243	
		NORMAL	FAT	271	-	-	-	-	-	-	-	-	-	-	-	
MEDITERRANEAN	268		-	-	257	-	-	247	-	257	-	-	-			
P4	NATURAL	NORMAL	FAT	214	212,5	215	213	214,5	216	214,75	217	213,667	215,5	213,75	217	
			MEDITERRANEAN	-	212	-	-	-	-	213	-	211	213	-	210	
		SEDENTARIAN	FAT	215	213	-	-	210	210	-	212	-	-	-	-	
P3	HOMEOPATHIC	HEALTHY	FAT	301	305	307,333	308,75	314	316,5	320	320,667	323,5	324	324,75	324,667	
			MEDITERRANEAN	-	-	305	-	309	-	319	-	-	-	324	-	326
		NORMAL	FAT	302,5	307	-	-	311	-	-	322	-	-	-	-	
P5	PHARMA	HEALTHY	MEDITERRANEAN	-	-	-	-	-	314	304	304	-	285	-	-	
			VEGETARIAN	-	-	-	-	-	-	-	-	-	-	-	274	
		NORMAL	FAT	-	319	-	-	-	306	-	-	-	302	-	-	-
			MEDITERRANEAN	324	323,333	320,667	312	315,4	308,5	305,333	299,333	297	293	289,5	274,5	
			VEGETARIAN	-	-	-	-	-	-	-	-	-	294	-	270	
		SEDENTARIAN	FAT	330	-	-	314	-	-	-	-	-	-	-	-	
MEDITERRANEAN	324		-	323	-	-	-	-	-	-	-	-	-			
P6	HOMEOPATHIC	HEALTHY	FAT	-	351	-	344	-	-	-	-	-	-	-	-	
			MEDITERRANEAN	-	345	347	-	-	339	-	339	-	338,5	338	-	
			VEGETARIAN	350	349	344	343	340,5	341,5	340,333	339	337,75	338	337,667	336,75	
		NORMAL	FAT	354	-	-	-	-	-	-	-	-	-	-	-	
			MEDITERRANEAN	-	-	-	-	341	-	-	-	-	-	337	-	
			VEGETARIAN	346	349	-	-	341,5	342	341,5	343	-	-	-	-	

Figura 38. Gràfic d'anàlisi de l'evolució dels diferents tractaments relacionats amb els hàbits.



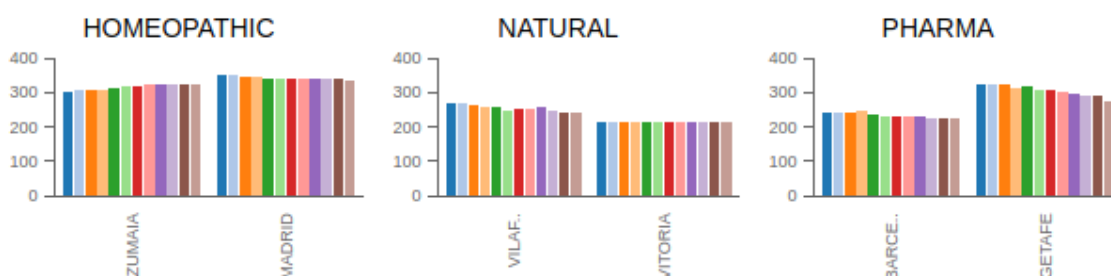
6.7 Hi ha diferències en el resultat d'un tractament segons el lloc geogràfic del pacient?

La següent taula 39 conjuntament amb la gràfica 40 estan elaborades utilitzant les dimensions de zona i els diferents tractaments. Es mostra que **per un mateix tractament, depenent del lloc geogràfic del pacient, els resultats tenen tendències que no permeten establir tendències, excepte en el tractament farmacològic, que sempre evoluciona de forma positiva.**

Figura 39. Taula d'anàlisi de l'evolució dels diferents tractaments relacionats amb els lloc geogràfic de cada pacient

Treatment	City	Month												
		1	2	3	4	5	6	7	8	9	10	11	12	
HOMEOPATHIC	ZUMAIA	302,2	306	306,75	308,75	312,4	316,5	319,8	321	323,5	324	324,75	325	
	MADRID	350	348,5	345,5	343,25	341	341	340,8	340	337,75	338	337,75	336,75	
NATURAL	VILAFRANCA P.	270,4	271	264,75	257,5	256	247,75	250	254,75	255,5	247,2	241	240,5	
	VITORIA	214,2	212,5	215	213	213,6	214,5	214,4	215,75	213	215	213,75	215,25	
PHARMA	BARCELONA	243,8	243,75	243	246,25	235	233	232,4	231,25	231	224,2	222,75	223,25	
	GETAFE	325,2	322,25	321,25	312,5	315,4	309,25	304,8	300,5	298,25	291,6	289,5	273,25	

Figura 40. Gràfic d'anàlisi de l'evolució dels diferents tractaments relacionats amb els lloc geogràfic de cada pacient



6.8 Hi ha algun període de l'any on el tractament sigui més o menys efectiu?

Es realitza una comparació de les 6 teràpies en graduacions temporals diferents obtenint una gràfica evolutiva per trimestres (figura 41), una per mesos (figura 42) i una darrera per setmanes (figura 43).

Comparant les diferents tendències en els períodes trimestrals i mensuals no s'aprecien etapes significativament diferenciades, ara bé, en el moment de fer la comparativa per setmanes **es pot observar una lleugera tendència negativa al voltant de les setmanes 34 i 35 on el tractament sembla menys efectiu.** Aquestes setmanes corresponent a l'equador del més d'agost.

Figura 41. Gràfic d'anàlisi trimestral de l'evolució dels diferents tractaments

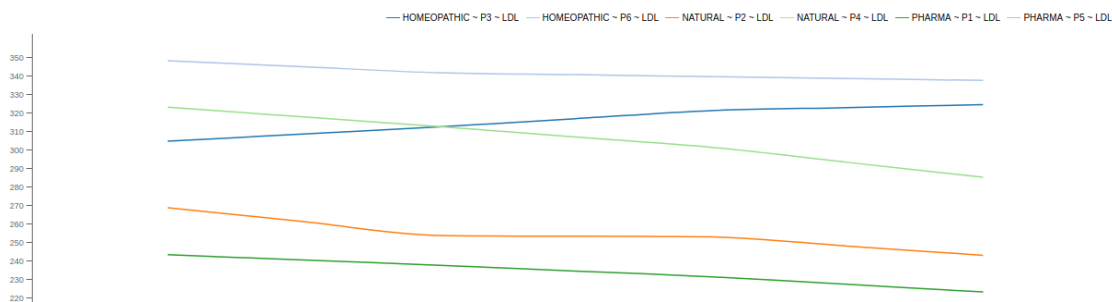


Figura 42. Gràfic d'anàlisi mensual de l'evolució dels diferents tractaments

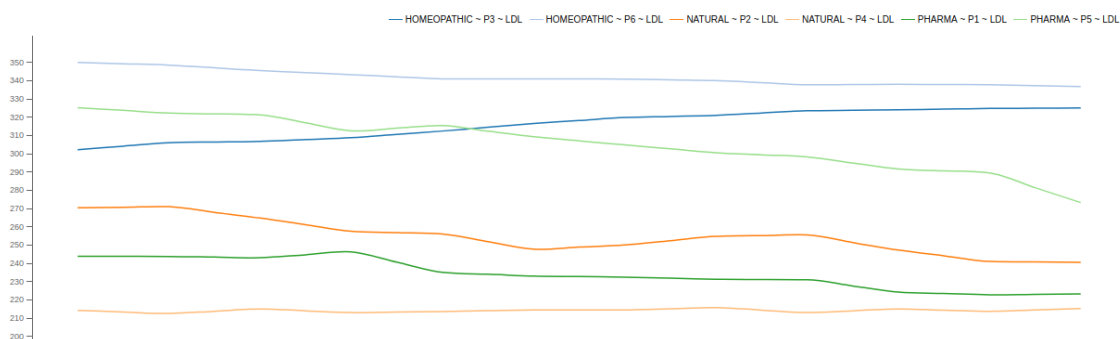
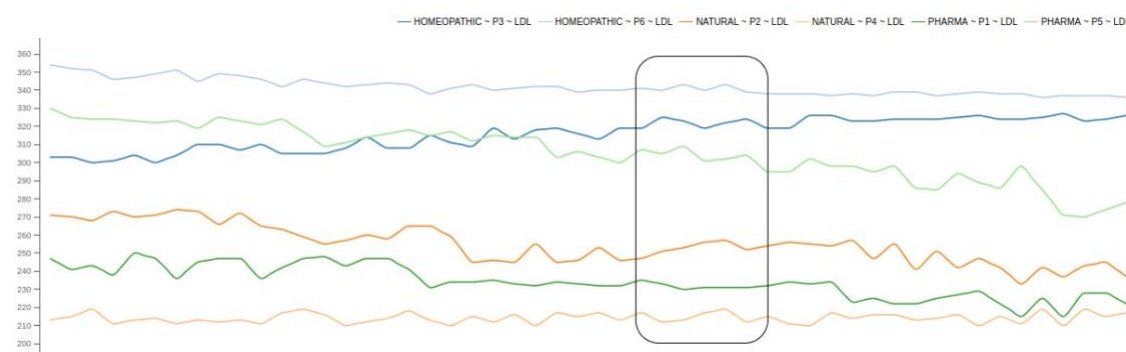


Figura 43. Gràfic d'anàlisi setmanal de l'evolució dels diferents tractaments



7. Conclusions

La realització d'aquest projecte ha permès aprendre i consolidar coneixements relacionats amb la intel·ligència de negoci com els necessaris per saber organitzar diferents estructures de dades per tal de modelar en múltiples dimensions un magatzem de dades (*Data Warehouse*) i implementar-ho amb els processos de càrrega transformació i extracció (ETL). Comprendre i saber aplicar els mètodes i eines d'anàlisi multidimensional (OLAP) i crear cubs d'anàlisi. Conèixer les eines *Open Source* en matèria d'intel·ligència de negoci i estar en condicions de desenvolupar i implantar un projecte *Business Intelligence* (BI) dins de l'empresa. Així mateix, una lliçó important, ha estat identificar la intel·ligència de negoci com a una oportunitat per tal de posar a disposició dels analistes de dades eines d'anàlisi en les principals funcions i processos empresarials que els permetin, amb uns requeriments mínims de coneixements d'ofimàtica, realitzar consultes de dades comprensibles amb un alt rendiment.

L'objectiu del projecte ha estat assolit amb el disseny i implementació d'un sistema d'Intel·ligència de Negoci que facilita l'adquisició, l'emmagatzemament i l'explotació de dades associades a pacients als que se'ls ha diagnosticat alts nivells de colesterol. Per tant, ha estat possible donar resposta a les preguntes analítiques que es volien resoldre. Ara bé, l'extensió de l'experiment on s'han obtingut les dades base de forma conjunta al nombre de pacients existents, ha generat un espai d'informació limitat on l'anàlisi comparatiu entre les diferents dimensions ha resultat pobre per poder obtenir conclusions més enllà de l'anàlisi directe de les dades disponibles. Amb tot, el treball permet afirmar, que en el cas d'aquest experiment, els pacients amb tractament farmacològic i uns hàbits més saludables han reduït els seus nivells de lipoproteïnes de baixa densitat (LDL) de forma més eficient.

En el moment de realitzar un anàlisi del seguiment de la planificació i la metodologia escollida s'han de posar de relleu diferents dificultats, d'una banda la integració de les diferents eines *Open Source* escollides ha necessitat més temps del previst, d'altra banda l'ús de tecnologies que no es coneixien prèviament i la recerca de bones pràctiques tampoc es va dimensionar correctament. Així, tot i que les hores de dedicació estaven establertes per la càrrega en crèdits ECTS, per tal de garantir la qualitat del treball ha estat necessari una dedicació ha superior. Aquesta possibilitat ja estava contemplada en la programació i els canvis realitzats han esdevingut en variacions en la temporització que han pogut ser recuperades amb èxit gràcies a la disponibilitat de més recursos. En referència a la metodologia prevista, aquesta ha estat satisfactòria, de manera clara la descomposició del treball en fases gestionades com a petits projectes seqüencials però revisades de

manera iterativa. La definició correcta d'aquestes unitats ha resultat clau pel seguiment correcte del treball.

Tot i haver assolit els objectius del projecte, han quedat pendents línies de treball futur que no han pogut ser explorades com; l'ampliació dels anàlisis amb les dades dels pacients referents a la seva edat i sexe, la creació dels taulers de control (*Dashboard*) amb una representació gràfica que hauria de permetre una visualització ràpida de l'estat dels diferents indicadors clau. A més, a la vista dels resultats dels diferents anàlisis també s'obre una línia d'investigació per tal de determinar la idoneïtat dels tractaments homeopàtic i naturals, ja que, els resultats conviden a qüestionar-los.

8. Glossari

BI *Business Intelligence*. Intel·ligència empresarial o intel·ligència de negoci, així es denominen a les estratègies i tecnologies utilitzades per les empreses per a l'anàlisi de dades d'informació empresarial.

OLAP *On-line analytical processing*. Conjunt d'eines i tecnologies d'anàlisi multidimensional basades en la construcció d'estructures multidimensionals anomenades Cubes OLAP.

ETL *Extract, Transform, Load*. Procés que permet moure dades d'orígens diversos dins de l'organització realitzant les tasques d'extracció, transformació i càrrega. La funció principal es alimentar el *Data Warehouse* que ha de donar suport al posterior anàlisi de la informació.

DWH *Data Warehouse*. Bases de dades que integren diferents orígens d'informació històrica de l'empresa i amb l'objectiu principal d'ajudar a la presa de decisions en àrees d'interès de l'empresa.

SGBD Sistema de gestió de bases de dades. Les seves funcions principals són facilitar l'ús simultani a molts usuaris, independitzar l'usuari del món físic i mantenir la integritat de les dades.

Open Source El codi obert és mètode de desenvolupament de programari que aprofita el poder de permetre la revisió del codi font als usuaris per tal d'estudiar, canviar i distribuir el programari per a qualsevol propòsit.

Dashboard És una interfície gràfica centrada en l'usuari amb les principals mètriques o indicadors claus que intervenen en l'anàlisi de dades amb l'objectiu d'ajudar a l'organització en la presa de decisions.

LDL *Low Density Lipoprotein* És el nivell de les lipoproteïnes de baixa densitat o LBD, conegudes també popularment com a "colesterol dolent.

SBP *Systolic Blood Pressure* La pressió sistòlica és la pressió arterial del ventricle esquerra.

DBP *Diastolic Blood Pressure* La pressió diastòlica és la pressió arterial del ventricle dret.

SOA *Service Oriented Architecture* És un concepte d'arquitectura de programari que defineix la utilització de serveis per a donar suport als requisits del negoci.

RPC *Rich Client Platform* És una eina de programació que facilita la integració de components de programari independents

Bus Matrix Matriu que permet obtenir una alta abstracció al mateix temps permet dividir el problema en parts més manejables

Script Guió o conjunt d'instruccions informàtiques.

9. Bibliografia

Kimball, R. (2008). The data warehouse lifecycle toolkit. John Wiley & Sons.

E. F. Codd; S. B. Codd; C. T. Salley (1993). Providing OLAP to User-Analysts: An IT Mandate. Arbor Software, Technical Report.

Pendse, Nigel (2005), "What is OLAP?", in The BI Verdict, Business Application Research Center, 2009.

Project Management Institute (2008). A Guide to the Project Management Body of Knowledge (PMBOK Guide), 4th. Edition. Pennsylvania (EUA): PMI.

Data Modeling Techniques for Data Warehouse:
http://e-ammar.net/download/ibm_warehouse%20model.pdf
[Setembre-Desembre 2018]

Hugh J. Watson Terry College of Bussiness, University of Georgia Recent Developments in ata Warehouseing: A tutorial:
<https://slideplayer.com/slide/6095737/>
[Setembre-Desembre 2018]

Cano, J. L. (2007). Business intelligence: Competir con información. Banesto, Fundación Cultur [ie Cultural]
[Setembre-Desembre 2018]

OLAP Systems: <http://web.mit.edu/profit/PDFS/SlaughterA.pdf>
[Setembre-Desembre 2018]

Pentaho: <https://www.hitachivantara.com/go/pentaho.html>
[Setembre-Desembre 2018]

Pentaho: <https://help.pentaho.com/Documentation/8.1>
[Setembre-Desembre 2018]

Mondrian: <https://community.hitachivantara.com/docs/DOC-1009853>
[Setembre-Desembre 2018]

Mondrian Schema: <https://mondrian.pentaho.com/documentation/schema.php>
[Setembre-Desembre 2018]

MySQL: <https://www.mysql.com/>
[Setembre-Desembre 2018]

MySql Workbench : <https://www.mysql.com/products/workbench/>
[Setembre-Desembre 2018]

Saiku: <https://www.meteorite.bi/products/saiku>
[Setembre-Desembre 2018]

Kettle: <https://community.hitachivantara.com/docs/DOC-1009855-data-integration-kettle>

[Setembre-Desembre 2018]

Spoon: <https://help.pentaho.com/Documentation/6.0/0H0/070/010/020>

[Setembre-Desembre 2018]

Gartner: <https://www.gartner.com/it-glossary/b>

[Setembre-Desembre 2018]

Wikipedia: <https://en.wikipedia.org>

[Setembre-Desembre 2018]

10. Annexos

10.1 Creació Schema de MySQL

```
-- MySQL Script generated by MySQL Workbench  
-- Model: New Model   Version: 1.0  
-- MySQL Workbench Forward Engineering
```

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

```
-----  
-- Schema TFM2018  
-----
```

```
-----  
-- Schema TFM2018  
-----
```

```
CREATE SCHEMA IF NOT EXISTS `TFM2018` DEFAULT CHARACTER SET utf8 ;  
USE `TFM2018` ;
```

```
-----  
-- Table `TFM2018`.`dimPatient`  
-----
```

```
CREATE TABLE IF NOT EXISTS `TFM2018`.`dimPatient` (  
  `PatienKey` INT NOT NULL,  
  `Age` VARCHAR(15) NULL,  
  `Patient` VARCHAR(5) NULL,  
  `Gender` VARCHAR(5) NULL,  
  `Treatment` VARCHAR(45) NULL,  
  `City` VARCHAR(45) NULL,  
  `CCAA` VARCHAR(45) NULL,  
  PRIMARY KEY (`PatienKey`))  
ENGINE = InnoDB;
```

```
-----  
-- Table `TFM2018`.`dimDate`  
-----
```

```
CREATE TABLE IF NOT EXISTS `TFM2018`.`dimDate` (  
  `dateKey` INT NOT NULL,  
  `Date` DATE NULL,  
  `Day` INT NULL,  
  `Week` INT NULL,  
  `Month` INT NULL,  
  `Trimestrer` INT NULL,  
  `Year` INT NULL,  
  PRIMARY KEY (`dateKey`))  
ENGINE = InnoDB;
```

```
-----  
-- Table `TFM2018`.`dimDiet`  
-----
```

```
CREATE TABLE IF NOT EXISTS `TFM2018`.`dimDiet` (  
  `DietKey` INT NOT NULL,  
  `Diet` VARCHAR(45) NULL,  
  PRIMARY KEY (`DietKey`))
```

```
ENGINE = InnoDB;
```

```
-----  
-- Table `TFM2018`.`dimActivity`  
-----
```

```
CREATE TABLE IF NOT EXISTS `TFM2018`.`dimActivity` (  
  `ActivityKey` INT NOT NULL,  
  `Activity` VARCHAR(45) NULL,  
  PRIMARY KEY (`ActivityKey`))  
ENGINE = InnoDB;
```

```
-----  
-- Table `TFM2018`.`registres`  
-----
```

```
CREATE TABLE IF NOT EXISTS `TFM2018`.`registres` (  
  `RegistresKey` INT NOT NULL,  
  `dimPatinent_patienKey` INT NOT NULL,  
  `dimDate_dateKey` INT NOT NULL,  
  `dimActivity_activityKey` INT NOT NULL,  
  `dimDiet_dietKey` INT NOT NULL,  
  `Idl` INT NULL,  
  `sbp` DOUBLE NULL,  
  `dbp` DOUBLE NULL,  
  INDEX `fk_registres_dimPatinent1_idx` (`dimPatinent_patienKey` ASC),  
  INDEX `fk_registres_dimDate1_idx` (`dimDate_dateKey` ASC),  
  INDEX `fk_registres_dimActivity1_idx` (`dimActivity_activityKey` ASC),  
  INDEX `fk_registres_dimDiet1_idx` (`dimDiet_dietKey` ASC),  
  PRIMARY KEY (`RegistresKey`),  
  CONSTRAINT `fk_registres_dimPatinent1`  
    FOREIGN KEY (`dimPatinent_patienKey`)  
    REFERENCES `TFM2018`.`dimPatient` (`PatienKey`)  
    ON DELETE NO ACTION  
    ON UPDATE NO ACTION,  
  CONSTRAINT `fk_registres_dimDate1`  
    FOREIGN KEY (`dimDate_dateKey`)  
    REFERENCES `TFM2018`.`dimDate` (`dateKey`)  
    ON DELETE NO ACTION  
    ON UPDATE NO ACTION,  
  CONSTRAINT `fk_registres_dimActivity1`  
    FOREIGN KEY (`dimActivity_activityKey`)  
    REFERENCES `TFM2018`.`dimActivity` (`ActivityKey`)  
    ON DELETE NO ACTION  
    ON UPDATE NO ACTION,  
  CONSTRAINT `fk_registres_dimDiet1`  
    FOREIGN KEY (`dimDiet_dietKey`)  
    REFERENCES `TFM2018`.`dimDiet` (`DietKey`)  
    ON DELETE NO ACTION  
    ON UPDATE NO ACTION)  
ENGINE = InnoDB;
```

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

10.2 Cub OLAP XML

```

<Schema name="SchemaTFM2018">
  <Dimension type="StandardDimension" visible="true" highCardinality="false" name="Activity">
    <Hierarchy name="Activity" visible="true" hasAll="true" primaryKey="ActivityKey">
      <Table name="dimActivity">
      </Table>
      <Level name="Activity" visible="true" column="ActivityKey" nameColumn="Activity" type="String"
uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <Dimension type="StandardDimension" visible="true" highCardinality="false" name="Diet">
    <Hierarchy name="Diet" visible="true" hasAll="true" primaryKey="DietKey">
      <Table name="dimDiet">
      </Table>
      <Level name="Diet" visible="true" column="DietKey" nameColumn="Diet" type="String"
uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <Dimension type="StandardDimension" visible="true" highCardinality="false" name="Patient">
    <Hierarchy name="Treatment" visible="true" hasAll="true" primaryKey="PatienKey">
      <Table name="dimPatient">
      </Table>
      <Level name="Treatment" visible="true" column="Treatment" nameColumn="Treatment" type="String"
uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
      </Level>
      <Level name="Patient" visible="true" column="PatienKey" nameColumn="Patient" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
      </Level>
    </Hierarchy>
    <Hierarchy name="Zona" visible="true" hasAll="true" primaryKey="PatienKey">
      <Table name="dimPatient">
      </Table>
      <Level name="CCAA" visible="true" column="CCAA" nameColumn="CCAA" type="String"
uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
      </Level>
      <Level name="City" visible="true" column="City" nameColumn="City" type="String" uniqueMembers="true"
levelType="Regular" hideMemberIf="Never">
      </Level>
      <Level name="Patient" visible="true" column="PatienKey" nameColumn="Patient" type="String"
uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <Dimension type="TimeDimension" visible="true" highCardinality="false" name="Date">
    <Hierarchy name="Date" visible="true" hasAll="true" primaryKey="dateKey">
      <Table name="dimDate">
      </Table>
      <Level name="Year" visible="true" column="Year" nameColumn="Year" type="Integer"
uniqueMembers="true" levelType="TimeYears" hideMemberIf="Never">
      </Level>
      <Level name="Trimestrer" visible="true" column="Trimestrer" nameColumn="Trimestrer" type="Integer"
uniqueMembers="false" levelType="TimeQuarters" hideMemberIf="Never">
      </Level>
      <Level name="Month" visible="true" column="Month" nameColumn="Month" type="Integer"
uniqueMembers="false" levelType="TimeMonths" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>

```

```
<Level name="Week" visible="true" column="Week" nameColumn="Week" type="Integer"
uniqueMembers="false" levelType="TimeWeeks" hideMemberIf="Never">
</Level>
<Level name="Day" visible="true" column="Day" nameColumn="Day" type="Integer"
uniqueMembers="false" levelType="TimeDays" hideMemberIf="Never">
</Level>
</Hierarchy>
</Dimension>
<Cube name="CubOLAPTFM2018" visible="true" description=" Cub per a l&#8217;an&#224;lisi dels
tractaments de reducci&#243; del colesterol" cache="true" enabled="true">
<Table name="registres">
</Table>
<DimensionUsage source="Patient" name="Patient" visible="true" foreignKey="dimPatinent_patienKey"
highCardinality="false">
</DimensionUsage>
<DimensionUsage source="Activity" name="Activity" visible="true" foreignKey="dimActivity_activityKey"
highCardinality="false">
</DimensionUsage>
<DimensionUsage source="Diet" name="Diet" visible="true" foreignKey="dimDiet_dietKey"
highCardinality="false">
</DimensionUsage>
<DimensionUsage source="Date" name="Date" visible="true" foreignKey="dimDate_dateKey"
highCardinality="false">
</DimensionUsage>
<Measure name="DBP" column="dbp" datatype="Numeric" aggregator="avg" visible="true">
</Measure>
<Measure name="LDL" column="ldl" datatype="Integer" aggregator="avg" visible="true">
</Measure>
<Measure name="SBP" column="sbp" datatype="Numeric" aggregator="avg" visible="true">
</Measure>
</Cube>
</Schema>
```