



Construcció i explotació d'un magatzem de dades corporatiu per a l'organisme de Salut Pública de la Diputació de Girona

Astrid Victoria Desset Desset

Grau en Enginyeria Informàtica

Pere Juanola Juanola

16 de Juny de 2011

Agraïments

En el transcurs del Treball Final de Grau que presento a continuació, he rebut el suport i la comprensió, tant de companys de feina com d'amics i familiars.

Agraeixo especialment en David Ferrando -Cap de l'àrea d'Informació per a la Gestió i la Qualitat de Dipsalut- per confiar en mi malgrat saber que no havia implementat mai un magatzem de dades amb anterioritat, en Francesc Berjano -consultor extern estretament lligat a l'àrea- per recomanar-me les eines de codi obert de *Pentaho* i despertar en mi la curiositat pel món del *business intelligence*, i en Miquel Obrador -el millor company de feina que em podia tocar-. Tots tres m'han anat enviant missatges encoratjadors i han seguit l'avenç del treball, de vegades llegint-ne els esborranys en diumenges i festius. Vull agrair també a tota la resta de companys de les diferents àrees de l'ens, que han respost a consciència les meves preguntes referents a dades i usos de les mateixes.

Hi han hagut moments en que fins i tot els amics de sempre han “patit” el meu treball. Em refereixo a aquelles reunions de dissabtes vespre en les quals tothom s'esbargaix i que he mig compartit amb ells, asseguda en el sofà, amb l'ordinador sobre els genolls i intentant redactar frases tècniques.

Pel que fa al meu marit Joaquim i a la meva filla Laia -que diu que *“la UOC li ha robat la seva mama”*-, els hi vull atribuir tot el mèrit d'haver pogut concloure aquest Treball Final de Grau.

A tots, moltes gràcies!

Resum

El present Treball Final de Grau de l'àrea de Magatzem de Dades, consisteix en la **construcció i explotació d'un magatzem de dades corporatiu per a l'organisme de Salut Pública de la Diputació de Girona**, el qual ofereix als ajuntaments de la seva província un catàleg de serveis que poden sol·licitar, en matèria protecció i promoció de la salut.

L'escenari de partida ha estat un conjunt de fonts de dades i aplicacions independents, posades en marxa en diferents moments evolutius de l'organisme, per donar suport als processos del mateix, però que no complien els requisits per a ser analitzades de manera eficient.

El projecte ha comportat una tasca de planificació prèvia, seguida d'una fase d'estudi acurat dels mòduls de magatzems de dades facilitats per la UOC. S'ha dut a terme l'anàlisi de les fonts de dades originals i un disseny complet del que hauria de ser la factoria d'informació corporativa, al mateix temps que s'instal·lava, s'experimentava i es valoraven les eines de *business intelligence* de *Pentaho*.

Finalment, s'ha implementat una de les tres estrelles del magatzem corporatiu dissenyat, concretament la que fa referència a les sol·licituds rebudes per l'organisme. A tal efecte, s'ha realitzat un procés acurat d'ETL amb l'eina *Kettle-Spoon* de la *suite Pentaho*. A continuació, s'han modelat els cubs ROLAP amb *Mondrian*, s'han creat i explotat les vistes d'anàlisi amb *BI Server* i finalment, s'han generat una sèrie de llistats amb *Pentaho-Report-Designer*. Les altres dues estrelles restants s'implementaran seguint el mateix procés, després de les vacances d'estiu.

Índex de continguts

1. Introducció.....	8
1.1. Justificació del TFG i context en el qual es desenvolupa: punt de partida i aportació del TFG ..	8
1.2. Objectius del TFG.....	9
1.3. Enfocament i mètode seguit.....	9
1.4. Planificació del projecte.....	10
1.5. Productes obtinguts.....	13
1.6. Breu descripció dels altres capítols de la memòria.....	13
2. Anàlisi detallat de requeriments	14
2.1. Substitució dels fulls de càlcul per un sistema operacional	14
2.2. Magatzems de dades operacionals, corporatius i departamentals.....	16
2.3. Proposta d'arquitectura d'alt nivell de la factoria d'informació corporativa.....	17
3. Disseny	19
3.1. Disseny conceptual.....	19
3.1.1. Les eines OLAP i la multidimensionalitat.....	19
3.1.2. Identificació de fets i estrelles	19
3.2. Definició de l'estrella de Sol·licituds.....	20
3.2.1. Trobar la granularitat escaient.....	20
3.2.1. Dimensions de l'estrella de Sol·licituds i atributs associats.....	20
3.2.2. Descriptors i jerarquies d'agregació.....	21
3.2.3. Definició de les mesures que interessen i de les cel·les.....	22
3.2.4. Viabilitat de l'estrella.....	22
3.2.5. Diagrama del disseny conceptual resultant	22
3.3. Definició de l'estrella d' Actuacions.....	23
3.3.1. Trobar la granularitat escaient.....	24
3.3.2. Dimensions de l'estrella d'Actuacions i atributs associats.....	24
3.3.3. Descriptors i jerarquies d'agregació.....	24
3.3.4. Definició de les mesures que interessen i de les cel·les	24
3.3.5. Viabilitat de l'estrella.....	25
3.3.6. Compartició de dimensions entre Sol·licituds i Actuacions.....	25
3.3.7. Diagrama del disseny conceptual resultant	26
3.4. Definició de l'estrella de Comunicacions.....	26
3.4.1. Trobar la granularitat escaient.....	26
3.4.2. Dimensions de l'estrella de Comunicacions i atributs associats.....	27

3.4.3.Descriptors i jerarquies d'agregació.....	27
3.4.4.Definició de les mesures que interessen i de les cel·les.....	27
3.4.5.Viabilitat de l'estrella.....	27
3.4.6.Diagrama del disseny conceptual resultant	27
3.5.Disseny lògic.....	29
3.5.1.Disseny lògic de l'estrella de Sol·licituds.....	29
3.5.2. Disseny lògic de l'estrella d'Actuacions.....	30
3.5.3. Disseny lògic de l'estrella de Comunicacions.....	30
3.6.Disseny físic	32
3.7. Disseny i descripció dels informes.....	33
3.7.1. Consultes sobre l'estrella de Sol·licituds.....	33
3.7.2. Consultes sobre l'estrella d'Actuacions	33
3.7.3. Consultes sobre l'estrella de Comunicacions	34
4. Captures de pantalla i resultats obtinguts.....	35
4.1. Procés d'extracció, transformació i càrrega ETL.....	35
4.1.1.ETL Ajuntaments.....	35
4.1.2.ETL Estats.....	38
4.1.3.ETL Proveïdors.....	39
4.1.4.ETL Equipaments.....	39
4.1.5.ETL Programes.....	39
4.1.6.ETL Sol·licituds atòmiques.....	40
4.1.7.ETL Resum sol·licituds.....	41
4.2.Modelatge de l'estrella de Sol·licituds amb Mondrian.....	42
4.3.Vistes d'anàlisi de cubs ROLAP.....	44
4.4. Llistats i informes.....	47
4.4.1.Dissenyant llistats amb BI Server.....	47
4.4.2. Pentaho Report Designer.....	48
4.4.3. Captures de pantalla dels llistats obtinguts.....	49
5.Conclusions.....	54
6.Línies de desenvolupament futur.....	55
7.Glossari	56
8. Bibliografia.....	57
9. Annex I : Entorn d'implementació i programari emprat.....	58
Maquinari.....	58
Programari.....	58
MySQL.....	58
Java.....	58

Pentaho Business Intelligence Server.....	58
Pentaho Administration Console.....	59
Pentaho Data Integration, Kettle.....	59
Pentaho Schema Workbench, Mondrian.....	59
Pentaho Report Designer.....	60

Índex d'il·lustracions

Il·lustració 1: Enfocament i mètode seguit per al TFG : Cicle de vida en cascada	10
Il·lustració 2: Diagrama de Gantt del TFG.....	12
Il·lustració 3: Diagrama d'entitat - relació de la nova aplicació que substituirà els fulls de càlcul.....	15
Il·lustració 4: Esquema dels diferents magatzems de dades proposats.....	17
Il·lustració 5: Proposta d'arquitectura d'alt nivell dels sistemes d'informació corporatius.....	18
Il·lustració 6: Proposta d'Estrelles de la Factoria d'Informació Corporativa.....	20
Il·lustració 7:: Disseny conceptual de l'Estrella de Sol·licituds.....	23
Il·lustració 8: Disseny conceptual de l'estrella d'Actuacions.....	26
Il·lustració 9: Disseny conceptual de l'estrella Comunicacions.....	28
Il·lustració10: Disseny lògic de Sol·licituds.....	30
Il·lustració 11: Disseny lògic d'Actuacions.....	30
Il·lustració 12: Disseny lògic de l'estrella de Comunicacions.....	31
Il·lustració 13: Disseny físic de la base de dades del magatzem corporatiu.....	32
Il·lust. 14: objecte tipus 'Table Input'.....	35
Il·lust. 15: Configuració d'una connexió a la BDD d'origen i obtenció de les dades.....	35
Il·lust. 17: cerca de valors dins d'una altra taula.....	36
Il·lust. 16: Icona Stream lookup.....	36
Il·lust.19:opcions de filtratge de dades.....	36
Il·lust.18: Icona per filtrar les dades.....	36
Il·lust. 20: fluxe de dades correctes i errònies.....	37
Il·lust. 21: utilitat de selecció de camps.....	37
Il·lust. 22: configuració de la taula de destinació.....	37
Il·lust. 23: Procés ETL sencer de càrrega de la taula 'ajuntament' del DW.....	38
Il·lust. 24: Procés ETL sense de càrrega d'estats al DW.....	38
Il·lust. 25: Procés ETL de proveïdors.....	39
Il·lust. 26: Procés ETL sencer dels equipaments.....	39
Il·lust. 27: Procés ETL sencer de càrrega de Programes al DW.....	39
Il·lust. 28: extracció i càrrega de la dimensió Temps.....	40
Il·lust. 29: Procés ETL sencer de càrrega de les dimensions Temps i del fet Sol·licitud Atomica.....	41
Il·lust. 30: Procés ETL sencer de càrrega del fet "Resum Sol·licituds".....	42

Il·lust. 31: Creació d'una estrella amb Schema Workbench - Mondrian.....	43
Il·lust. 33: Definició del nivell d'agregació Catàleg de la dimensió Programa.....	43
Il·lust. 32: Dimensions i agregacions.....	43
Il·lust. 34: Finestra de publicació d'una estrella al BI Server.....	44
Il·lust. 35: Menú principal del BI Server.....	44
Il·lust. 36: tria d'esquema i cub.....	44
Il·lust. 37: vista d'anàlisi.....	45
Il·lust. 38: Jerarquia d'agregació de la dimensió temps en una vista d'anàlisi.....	45
Il·lust. 39: desplegament per estat de sol·licituds.....	45
Il·lust. 40: estudi de sol·licituds per estat i programa segons setmana escollida.....	45
Il·lust. 41: OLAP Navigator.....	46
Il·lust. 42: selecció d'atributs.....	46
Il·lust. 43: exemple d'anàlisi de resultats aplicant els filtres descrits.....	46
Il·lust. 44: exemple d'anàlisi basada en l'estrella de sol·licituds atòmiques.....	46
Il·lust. 45: selecció d'un origen de dades per al llistat.....	47
Il·lust. 46: consulta que retorna camps a incloure en un report de BI Server.....	47
Il·lust. 47: nivells de detall i agrupació en els llistats.....	47
Il·lust. 48: captura de pantalla del Pentaho Report Designer.....	48
Il·lust. 49: captura de pantalla de l'entorn gràfic de disseny de llistats.....	48
Il·lustració 50: Llistat de totes les sol·licituds d'un dia, setmana, mes, trimestre, any.....	49
Il·lustració 51: Llistat de totes les sol·licituds agrupades per estat.....	49
Il·lustració 52: Llistat de sol·licituds que es troben en un estat concret passat per paràmetre.....	50
Il·lustració 53: Llistat de totes les sol·licituds agrupades per proveïdor al qual s'han assignat.....	50
Il·lustració 54: Llistat sol·licituds assignades a un proveïdor concret passat per paràmetre.....	50
Il·lustració 55: Llistat de totes les sol·licituds agrupades per ajuntament i comarca.....	51
Il·lustració 56: Llistat de sol·licituds d'ajuntament i comarca passats per paràmetre.....	51
Il·lustració 57: Llistat de totes les sol·licituds agrupades per catàleg, àrea, línia i programa.....	52
Il·lustració 58: Sol·licituds en les quals s'ha inclòs un equipament concret passat per paràmetre.....	52
Il·lustració 59: Resum de sol·licituds per setmana, mes i any, agrupades per programa i estat.....	52
Il·lustració 60: Llistat de totes les sol·licituds de la comarca de 'La Selva' fetes durant l'últim mes, agrupades per àrea, programa i línia.....	53
Il·lustració 61: Sol·licituds en estat Assignada o En Procés que s'han assignat a un proveïdor concret durant el darrer trimestre 2010.....	53

1. Introducció

1.1. Justificació del TFG i context en el qual es desenvolupa: punt de partida i aportació del TFG

L'Organisme de Salut Pública de la Diputació de Girona realitza funcions d'assistència als municipis d'aquesta província, en relació a totes les competències que en matèria de salut pública tenen atribuïdes, en el marc de la protecció i la promoció de la salut.

És un organisme públic relativament nou (2009) que ha experimentat un creixement molt gran, tant a nivell de recursos humans, com a nivell de processos i d'actuacions en matèria de les seves competències. És també un ens que aposta per la implantació de les TIC com a eina d'informació per a la bona gestió de l'organisme i assegurar la qualitat en els processos i actuacions de duu a terme.

Des del moment inicial del desplegament dels seus serveis, l'organisme va començar a organitzar la seva informació seguint una estructura lògica de fitxers (majoritàriament fulls de càlcul i documents de text) centralitzada en un servidor, dins de carpetes i subcarpetes classificades per àrees (departaments) i temes.

D'aquesta manera, cada àrea creava i manipulava les dades que necessitava, tot i que molt aviat es va detectar la duplicitat de fonts d'informació i es va fer palesa la necessitat d'unificar-ne algunes. En alguns casos, i a l'espera d'aquesta unificació, s'ha arribat a l'acord que una única persona mantindrà el full de càlcul i que la resta només l'utilitzarà per extreure'n la informació necessària i/o consultar-la.

Paral·lelament, l'organisme va posar en marxa altres aplicacions subministrades per tercers:

- un gestor d'expedients
- un CRM per guardar la informació referent al flux de comunicació amb els ajuntaments
- un registre d'entrades i sortides vinculat al gestor d'expedients
- un gestor de continguts web

Finalment, en aquest escenari, el passat 30 de novembre de 2010 neix el SIMSAP, primer Sistema d'Informació Municipal en Salut Pública, una aplicació pionera que permet als ajuntaments de la província conèixer en temps real, i de forma transparent, l'estat sanitari dels seus equipaments d'ús públic (piscina, pavelló, sorreres infantils) i saber quines actuacions s'estan duent a terme en el seu municipi per tal de prevenir i promoure la salut dels seus habitants. El SIMSAP també permet als proveïdors de serveis (empreses que realitzen analítiques i desinfeccions d'instal·lacions, o bé executen serveis de promoció de la salut) pujar documents rellevants al sistema i omplir formularis que aporten informació molt valuosa, tant per a l'organisme com per l'ajuntament.

Totes les aplicacions esmentades fins ara són sistemes d'informació operacionals dissenyats per a optimitzar la introducció de dades, però no compleixen els requisits per a ser analitzats de manera eficient. Així doncs, cal definir un o varis sistemes nous que aprofitin aquestes dades i alhora serveixin

per a interpretar-les.

En definitiva, l'aportació d'aquest projecte consisteix en la realització del disseny de la factoria d'informació corporativa adient al model de l'organisme, estudiant la conveniència o no de la creació de magatzems de dades operacionals o departamentals i, finalment, en la implementació d'un magatzem de dades corporatiu partint de les dades operacionals.

1.2. Objectius del TFG

A títol personal, l'objectiu primordial del TFG ha estat l'aprenentatge del que són els magatzems de dades, de les diferents arquitectures i implementacions possibles de la factoria d'informació corporativa, tot aplicant els coneixements adquirits directament als sistemes d'informació de Dipsalut.

A nivell de projecte, el principal objectiu ha estat, per una banda, analitzar la transformació dels fulls de càlcul que utilitzaven les diferents àrees de l'organisme, unificant-los en un sistema operacional construït sobre una base de dades relacional, i per l'altra, extreure dades clau de les bases de dades relacionals existents, transformar-les i carregar-les al magatzem de dades corporatiu per tal d'analitzar-les.

La finalitat última de tot aquest procés és facilitar la presa de decisions a l'àrea de Gerència mitjançant llistats i gràfiques que resumeixin l'estat de totes les sol·licituds rebudes i feines derivades d'aquestes, així com veure'n la evolució periòdica.

També és un objectiu estratègic el fet que els propis usuaris del magatzem puguin construir molt fàcilment els informes que desitgen amb eines de *reporting* intuïtives.

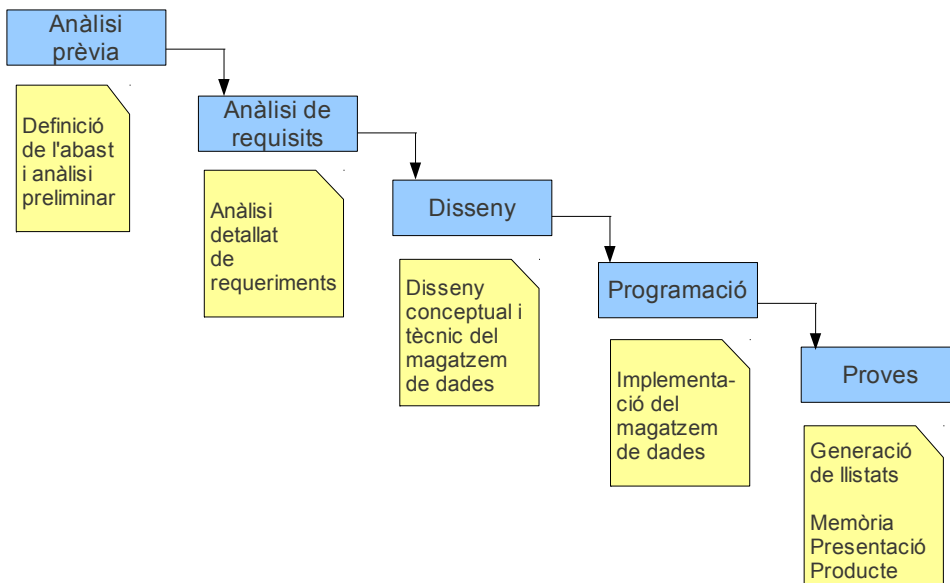
Per acabar, cal assenyalar que la consecució del projecte comporta l'aprenentatge de les eines de la *suite Pentaho BI Server*, concretament:

- Processos d'ETL amb *Kettle / Spoon*
- Creació d'esquemes en estrella definint Fets, Dimensions i Mesures amb *Mondrian*
- Creació de cubs ROLAP i vistes d'anàlisi a partir dels esquemes en estrella definits amb *Pentaho BI Server*
- Creació de llistats a partir de les dades dels esquemes amb *Pentaho Report Designer*

1.3. Enfocament i mètode seguit

L'enfocament i el mètode seguit per a la realització d'aquest projecte es basa en el model de cicle de vida clàssic, també anomenat "en cascada". A cada etapa s'obtenen uns documents que són les bases de partida de l'etapa següent i no es torna a etapes passades. En aquest cas però, no s'ha inclòs el manteniment. La fase de proves es correspon a la obtenció dels llistats esperats.

A continuació es descriuen aquestes etapes i els productes obtinguts al final de cadascuna.



Il·lustració 1: Enfocament i mètode seguit per al TFG : Cicle de vida en cascada

1.4. Planificació del projecte

Per a elaborar una planificació adequada al projecte i al temps del qual es disposava, s'han seguit els següents passos:

- Identificar els lliurables del projecte.
- Identificar les activitats de primer nivell.
- Descompondre les activitats de primer nivell en tasques i subtasques: creació de l'estructura de distribució del treball *EDT*.
- Tenir en compte les dates fixades pels lliuraments parcials (PACS) i encaixar-los en la distribució de tasques, definint-les també en el temps.
- Generar un diagrama de Gantt i una taula de planificació a partir de la informació anterior.

Els elements lliurables del projecte, que coincideixen amb les fites, són els definits en el pla docent del TFG:

- **19 de març 2011 : Pla de treball i anàlisi preliminar de requeriments**
Es tracta d'un document d'anàlisi preliminar de les fonts de dades operacionals a partir de les quals es vol construir el magatzem de dades. També inclou en detall la planificació de tot el projecte.
- **23 d'abril de 2011 : Anàlisi de requeriments i disseny conceptual i tècnic**
Correspon a un document amb l'anàlisi detallat de requeriments basat en l'anàlisi preliminar realitzat en el lliurable anterior, i complementant-lo amb la descripció i disseny del model dimensional, especificant el procés d'extracció de dades a alt nivell.
- **25 de maig de 2011: Implementació del magatzem de dades**
Es tracta d'un document que conté una explicació pas a pas de la implementació realitzada del magatzem, amb captures de pantalla. A més de construir el magatzem, carregar-hi les dades, configurar l'eina d'explotació de dades i construir els informes per a l'anàlisi de la

informació, s'ha permès l'accés *online* per a testejar el funcionament de tot el conjunt.

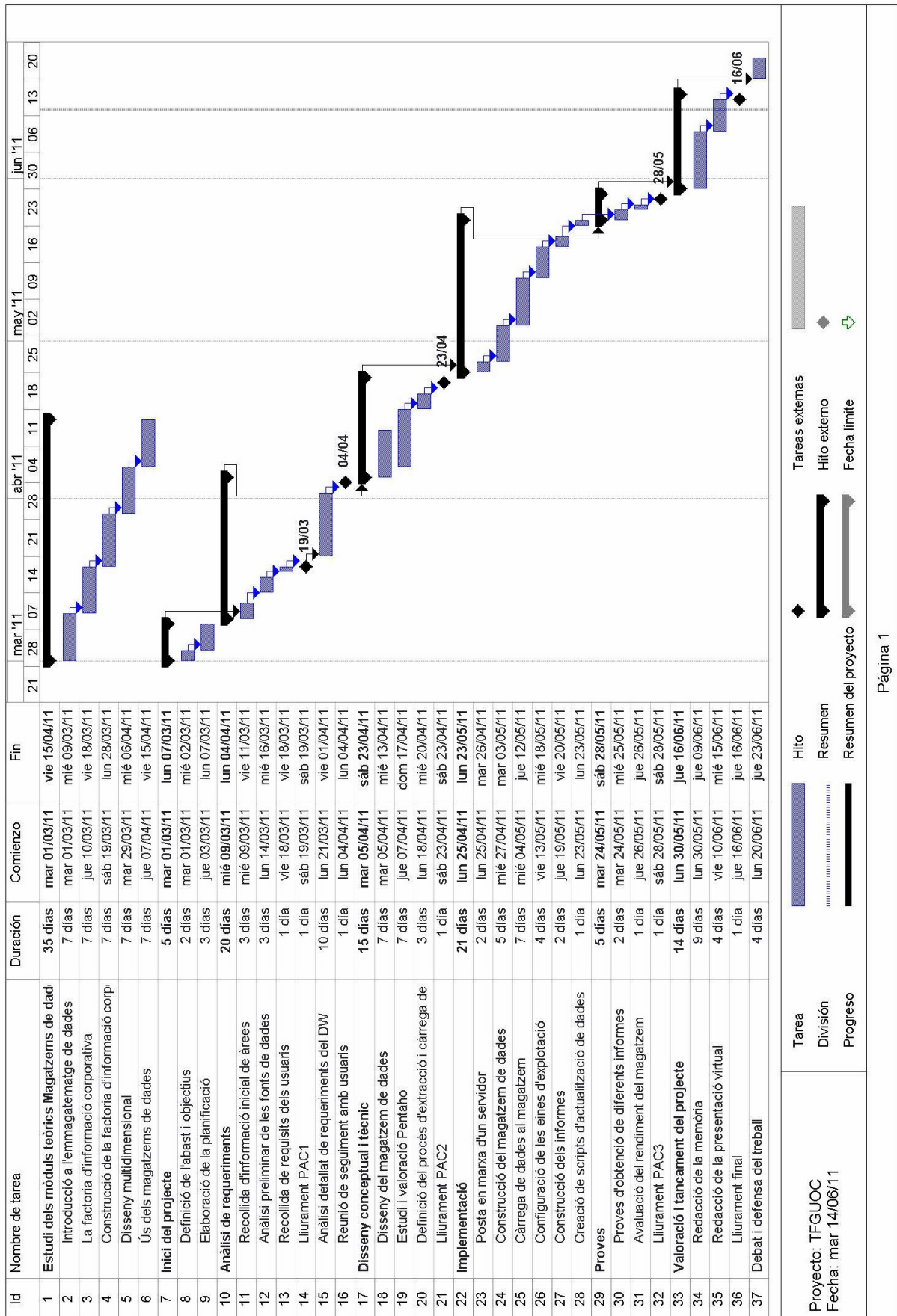
- **16 de juny : Lliurament final**

- **Memòria:** Memòria del treball seguint les indicacions del pla docent i les recomanacions de redacció de textos científicotècnics.
- **Presentació virtual:** lliurament d'un document de presentació en vídeo contenint la síntesi del treball realitzat i els resultats obtinguts.

- **20 a 23 de juny : Defensa del treball i participació en els debats**

Consisteix en respondre preguntes sobre el treball realitzat i participar en debats.

Tal com es pot observar en el diagrama de Gantt adjunt, les fites coincideixen amb les dates esmentades i l'estructura de distribució del treball es basa en les fases del model de cicle de vida en cascada, afegint-hi subtasques que no es solapen en el temps, excepte l'estudi dels mòduls de magatzems de dades, que s'han hagut de fer en paral·lel amb les altres tasques.



Il·lustració 2: Diagrama de Gantt del TFG

1.5. Productes obtinguts

Els productes obtinguts durant la implementació del projecte són:

- Una base de dades relacional anomenada “*magatzem*”, creada amb el sistema gestor de bases de dades *MySQL*, que conté les taules que modelen l'estrella de Sol·licituds del magatzem corporatiu dissenyat en el punt 3 d'aquesta memòria.
- Els fitxers de definició dels processos ETL que es poden executar cada vegada que es vulgui per tal d'extreure, transformar i carregar el magatzem de dades a partir dels sistemes operacionals de Dipsalut.
- Una plataforma de proves però completament funcional, de codi obert, basada en el conegut servidor d'aplicacions Jboss, que és el *Pentaho BI Server*. A més de portar implementat un mòdul d'autenticació i seguretat, aquesta plataforma ofereix un *front-office* molt complet per a la creació de tot tipus de vistes d'anàlisi i la realització de consultes i informes basades en les estrelles implementades.
- Els llistats i les vistes d'anàlisi corresponents a l'estrella de Sol·licituds, que es poden obrir, executar i provar en la plataforma de proves fins el 23 de juny, a l'adreça següent <http://80.32.31.86:8080/pentaho> (usuari: consultor, password: uoc).

1.6. Breu descripció dels altres capítols de la memòria

El capítol segon analitza la duplictat de fonts de dades i planteja la substitució dels nombrosos fulls de càlcul que utilitzen les diferents àrees de l'organisme per un sistema operacional que servirà per alimentar el magatzem de dades corporatiu. També analitza quina hauria de ser l'estructura de la factoria d'informació corporativa, argumentant la conveniència o no de construir magatzems operacionals i departamentals.

El capítol tercer correspon al disseny sencer del magatzem de dades corporatiu. Primerament, s'ha completat el disseny conceptual, definint els tres fets / estrelles que l'ens vol analitzar: Sol·licituds, Accions i Comunicacions. Per a cadascuna d'elles, s'ha definit una granularitat adient, les dimensions, els descriptors, les jerarquies d'agregació i les mesures, tot estudiant la viabilitat de cada estrella. A continuació, s'ha traduït el disseny conceptual multidimensional a un disseny lògic relacional i finalment, s'ha construït el diagrama del model físic, a partir del qual s'ha generat el script de creació de la base de dades.

El quart capítol conté la explicació detallada dels processos d'ETL de l'estrella de Sol·licituds, la descripció del modelatge de cubs, i finalment, les vistes d'anàlisi ROLAP i els llistats definits al disseny, tot complementat amb captures de pantalla.

Per acabar, es presenten les conclusions finals, es comenten les possibles línies de desenvolupament futur, un glossari de termes utilitzats, i es relaciona tota la bibliografia consultada. En últim lloc, es descriu a l'Annex I el programari utilitzat, així com la seva instal·lació i configuració.

2. Anàlisi detallat de requeriments

2.1. Substitució dels fulls de càlcul per un sistema operacional

Tot i que aquest punt s'aparta lleugerament del tema de magatzems de dades, és necessari incorporar-lo en el context de l'anàlisi per tal de que s'entengui quina serà l'estructura global de la factoria d'informació corporativa.

En el procés d'anàlisi de requeriments de totes les àrees de l'organisme, s'ha arribat a les conclusions següents:

- La majoria de dades que es volen analitzar es troben ja actualment en el Simsap i en el CRM.
- Entre tots els fulls de càlcul detectats, no n'hi ha cap que s'hagi de mantenir imperativament.
- A més, hi ha una gran quantitat de taules que són comunes a les diferents àrees de treball.

La principal conclusió que se'n extreu és que cal desenvolupar un projecte a mida, durant el qual es passarà la informació dels fulls de càlcul a una base de dades relacional, i que permetrà l'accés dels usuaris de les diferents àrees a les dades que necessiten, mitjançant permisos i perfils.

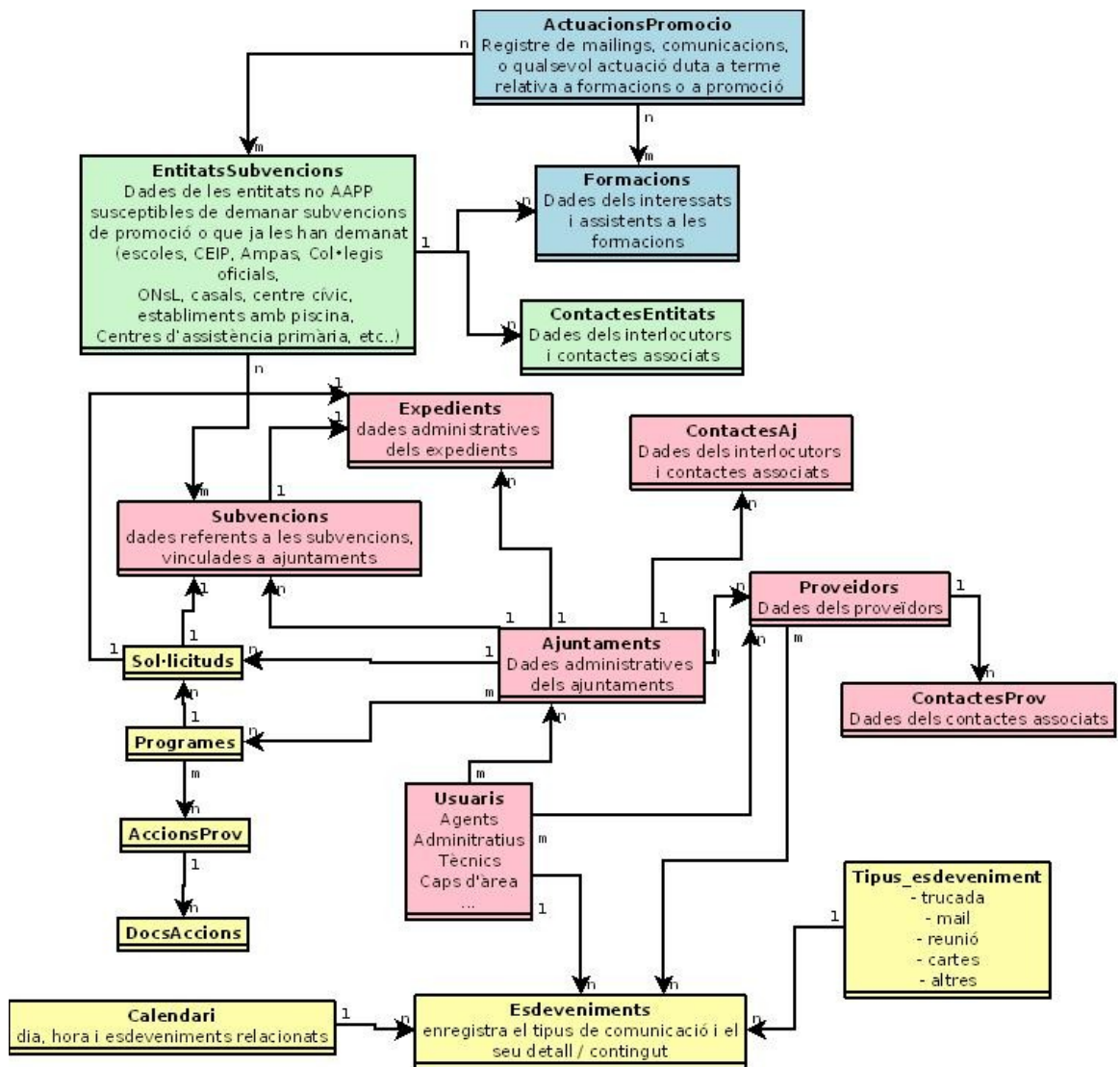
Aquesta aplicació nova permetrà el manteniment de registres -altes, baixes i modificacions- a certs usuaris, mentre que d'altres només les podran consultar o llistar. Hauria d'estar disponible 7 dies x 24 hores i s'hauria de permetre l'accés tant des de dins com des de fora de l'organisme. Es tracta doncs d'una aplicació en línia que permetrà l'accés mitjançant navegadors. Caldrà preveure un servidor per allotjar-la i planificar el projecte, tasca que queda fora de l'abast d'aquest treball.

A més, hi ha la dificultat afegida de decidir, d'entre les diferents fonts d'informació heterogènies existents, quines són les més adequades per a cadascuna de les dades que es vol tenir al magatzem corporatiu. Es tracta per tant de decidir quines es descarten i quines s'agafen com a més actualitzades, més ben estructurades i precises.

En aquest procés de depuració de dades caldrà corregir valors inconsistents, afegir valors als camps no definits, i detectar i corregir la informació duplicada.

Es tracta doncs d'un altre projecte independent dins del marc de la construcció del magatzem de dades corporatiu, ja que moltes de les dades d'aquest sistema operacional s'hi podran incorporar posteriorment.

A continuació es detalla un diagrama entitat-relació d'alt nivell de la nova aplicació que s'hauria de desenvolupar.



Il·lustració 3: Diagrama d'entitat - relació de la nova aplicació que substituirà els fulls de càlcul

Significat dels colors utilitzats

- Promoció de la salut
- Protecció de la salut
- Promoció + Administració + Difusió
- Promoció + Administració + Protecció + Difusió

2.2. Magatzems de dades operacionals, corporatius i departamentals

Tot seguit s'ha analitzat la conveniència o no de la creació de magatzems de dades operacionals i departamentals previs a la implementació del magatzem de dades corporatiu.

Els magatzems de dades departamentals guarden únicament les dades històriques que cada departament considera interessants. Estan dissenyats per a que els usuaris de cada departament obtinguin un temps de resposta molt bo en l'execució d'informes, però alhora representen una dificultat afegida per a mantenir els processos independents de transformació i càrrega de dades.

Segons el mòdul de la UOC "*La factoria d'Informació corporativa*", el magatzem de dades corporatiu no és apropiat per als usuaris finals, perquè està dissenyat per a gestionar i integrar grans quantitats de dades que, juntament amb l'excés d'usuaris, degraden el temps de resposta.

El més habitual doncs, és construir un magatzem de dades corporatiu genèric i únic, que guardi tota la informació susceptible de ser analitzada, i posteriorment, a partir d'aquest, alimentar altres magatzems de dades departamentals que serviran de base per a les consultes dels analistes, tot evitant els problemes de rendiment descrits.

De moment, a Dipsalut no es fa palesa la necessitat d'implementar magatzems de dades departamentals. El nombre d'usuaris totals és de 30 persones i la concurrència màxima dels mateixos és aproximadament de l'ordre de 6. D'altra banda, la gran majoria de departaments comparteixen pràcticament la mateixa informació, tot i que distribuïda de forma redundat, i el nombre de taules de dades específiques -que només ha de consultar un departament concret- no és significatiu. Tot això porta a pensar que segurament, en un futur no gaire llunyà, sí que caldrà construir alguns magatzems de dades departamentals, sobretot pel que fa al departament d'Administració.

D'altra banda, es decideix per raons de temps, centrar el projecte en la implementació del magatzem de dades corporatiu, ja que aquest servirà de base per a futurs magatzems específics.

Atès que el magatzem de dades corporatiu no està permanentment actualitzat, sinó que les dades es carreguen habitualment durant les nits o els caps de setmana per tal de no afectar el rendiment de les aplicacions de treball operacionals, cal tenir present que s'està prioritant la consulta de dades històriques sobre la consulta de dades completament actualitzades.

Sobre aquest punt, s'ha de destacar que és possible que un grup d'usuaris de l'organisme vulgui consultar dades integrades en un moment precís, raó per la qual, cal avaluar les opcions.

Tal com s'ha dit abans, efectuar consultes que sumin o facin la mitjana de certes dades directament sobre les bases de dades dels sistemes operacionals, representa una degradació important del rendiment de les aplicacions. Una opció -sobretot pel que respecta a les dades del SIMSAP- consisteix en construir un magatzem de dades operacional, orientat al tema i integrat com qualsevol magatzem de dades, però renunciant a tenir les dades historiades.

Aquest magatzem operacional s'ha d'alimentar de les bases de dades operacionals, replicant el contingut de les taules com a mínim un cop al dia, fent ús d'un programari específic que mantingui diverses rèpliques de les taules.

Un cop es té el magatzem operacional funcionant, el pas d'aquestes dades cap al magatzem corporatiu, tot afegint la temporalitat, pot consistir simplement en un bolcatge de les dades cada nit. D'aquesta manera, el magatzem operacional i el corporatiu es complementarien.

Amb el primer seria possible estudiar les dades actualitzades i amb el segon estudiar els històrics de les mateixes dades.

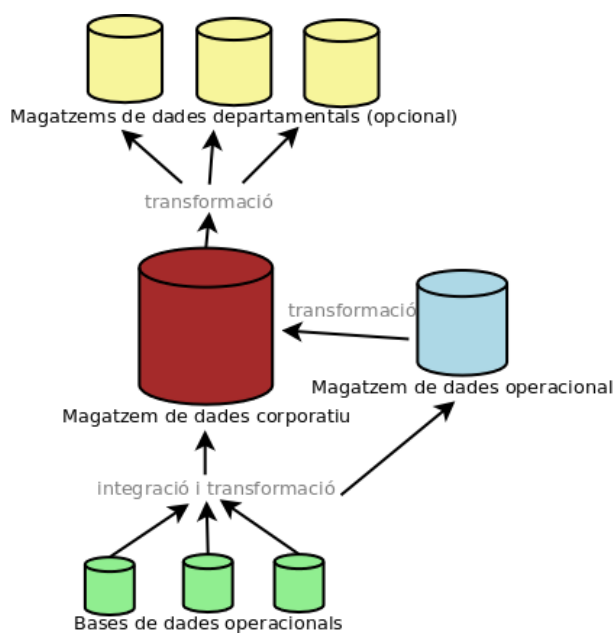
Una altra opció consisteix en programar les actualitzacions del magatzem de dades corporatiu de forma estratègica, és a dir, a unes determinades hores durant les quals se sap que no hi ha pràcticament ningú treballant amb les bases de dades operacionals. Per exemple, entre les 14 i les 15 h, els accessos al SIMSAP es redueixen a una mitjana de cinc, repartides principalment entre ajuntaments que consulten informació i proveïdors de serveis que actualitzen informes, i entre les 15 i les 16 h pràcticament no hi accedeix ningú. Per tant, es podria programar les actualitzacions del magatzem corporatiu en aquesta franja horària, sempre i quan el procés no durés més de mitja hora. Un cop acabat, el magatzem corporatiu ja contindria les darreres dades actualitzades i a més historiadades. Qualsevol informe, consulta o anàlisi que es desitja obtenir amb dades actualitzades, s'hauria de fer posteriorment al darrer procés de sincronització.

A priori, sembla més interessant la primera opció, ja que la segona suposa una interrupció del servei durant un temps, tot i que en una franja horària en la qual no afecta massa els usuaris. Tot i així, aquesta mateixa franja és també la que s'utilitza per actualitzar els canvis a nivell de programació (passar modificacions de l'entorn de desenvolupament a l'entorn de producció) i no hauria de coincidir amb el procés de sincronització del magatzem de dades corporatiu. A més de complementar-se, el magatzem operacional i el corporatiu estan dissenyats per a dues finalitats diferents i si els combinem en una estructura comuna, es degradarà el temps de resposta per a tots dos.

Sembla doncs més net construir un magatzem operacional cercant una solució de programari per al manteniment de rèpliques de taules en temps real, sense que afecti el rendiment de l'aplicació principal.

2.3. Proposta d'arquitectura d'alt nivell de la factoria d'informació corporativa

A més de tenir com a objectiu principal la construcció d'un magatzem de dades corporatiu, aquest projecte té també la finalitat de definir una nova estructura de sistemes d'informació per tal de treballar i analitzar unes dades unificades, precises i fiables.



Il·lustració 4: Esquema dels diferents magatzems de dades proposats

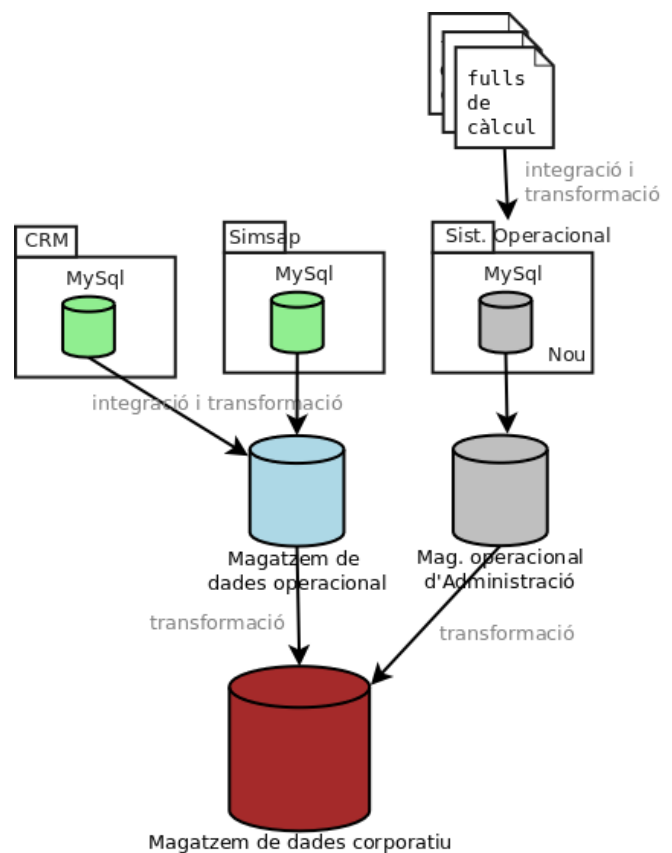
Per una banda, cal construir una aplicació operacional per a substituir els diversos fulls de càlcul existents. Per l'altra, s'ha dissenyat un magatzem de dades operacional basat en la informació del SIMSAP. Aquest magatzem es centra en l'obtenció i l'emmagatzematge de totes les dades fins al màxim nivell de detall possible.

El magatzem de dades corporatiu s'alimenta del magatzem de dades operacional afegint-hi la temporalitat. I en últim lloc, és possible -tot i que no necessari en aquests moments- que calgui posar en marxa magatzems de dades departamentals.

La complexitat del projecte fa que s'hagi de valorar com a altament recomanable l'opció de dividir-lo en projectes més petits i autònoms de tal manera que no necessitin d'altres projectes per a entrar en producció i alhora proporcionin per si mateixos un valor afegit a l'organisme. També cal tenir en compte el factor de l'escalabilitat de cadascun d'aquests projectes, per tal de permetre el creixement progressiu tant de la infraestructura física com de la operacional i l'augment del nombre d'usuaris.

De fet, aquests petits projectes, la suma dels quals conformen la *factoria d'informació* de l'organisme, ja han estat identificats en els punts anteriors:

- construcció d'una aplicació basada en la informació recollida en els diversos fulls de càlcul
- construcció d'un magatzem de dades operacional basat en el SIMSAP
- construcció d'un segon magatzem de dades operacional a partir de l'aplicació que substituirà els fulls de càlcul i a partir de la informació introduïda en el CRM.
- construcció del magatzem de dades corporatiu
- construcció dels magatzems de dades departamentals (opcional)



Il·lustració 5: Proposta d'arquitectura d'alt nivell dels sistemes d'informació corporatius

3. Disseny

3.1. Disseny conceptual

3.1.1. Les eines OLAP i la multidimensionalitat

Actualment, les eines OLAP es presenten com a millor opció per a resoldre les necessitats analítiques de les organitzacions. El seu nom prové de les sigles *Online Analytical Processing*, és a dir, de la possibilitat de processar consultes en línia, responnent a la majoria de consultes de forma eficient i ràpida, oferint eines d'anàlisi estadística i generació automatitzada d'informes, tot afegint-hi seguretat en l'accés a les dades.

Però sense dubte, la principal característica de tota eina OLAP és la multidimensionalitat, mitjançant la qual s'aconsegueix presentar la informació als analistes en termes de **fets i dimensions**, en lloc de taules, atributs i claus foranes.

Les **dimensions** estan íntimament relacionades amb les dades descriptives i defineixen un espai multidimensional que s'anomena **cub**. Dins el cub, es col·loquen els fets o dades mètriques que es volen analitzar. Cada posició en el cub es coneix com a **cel·la**, que no s'ha de confondre amb les cel·les d'un full de càlcul, ja que una cel·la d'un cub s'identifica pels valors de les dimensions d'anàlisi que conté. Les dimensions s'utilitzen principalment per seleccionar i agregar dades a diferents nivells de detall. Es podrien equiparar al diferents punts de vista des dels quals es poden analitzar les dades.

Els **fets**, per la seva banda, representen un tema objecte d'anàlisi. Cada fet conté un conjunt de **cel·les** relacionades per agregacions i tenen atributs anomenats **mesures**. Els fets i el seu conjunt de dimensions formen el que s'anomena, en termes de disseny conceptual de magatzems de dades, una **estrella**.

A continuació, es descriu el procés seguit per a realitzar el disseny conceptual multidimensional del magatzem de dades corporatiu, tenint en compte les dades del Sistema d'Informació Municipal en Salut Pública, que aglutina la majoria d'informació que es necessita analitzar. De fet, el projecte sencer de la factoria d'informació corporativa s'ha dividit en diferents projectes més petits, i el present treball se centra en la construcció del magatzem corporatiu a partir de les dades del SIMSAP.

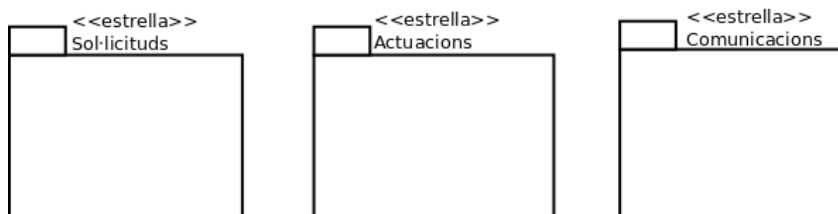
3.1.2. Identificació de fets i estrelles

Per poder construir una estrella per a cada fet, cal identificar tots els fets susceptibles de ser estudiats.

D'entrada, l'organisme vol analitzar un tema concret que parteix d'un procés de negoci molt conegut per totes les àrees: les **Sol·licituds**. Els ajuntaments poden sol·licitar serveis i subvencions a Dipsalut. Aquest és doncs el punt de partida de tot anàlisi que permetrà estudiar les sol·licituds i les tendències dels ajuntaments.

D'altra banda, les sol·licituds deriven en actuacions i resultats que duen a terme els proveïdors de serveis, i que es podrien catalogar com a un mateix fet. Per tant, apareix un segon candidat a fet / estrella que permetrà estudiar els proveïdors i la feina que duen a terme: les **Actuacions**.

Finalment, cal estudiar tot el tema de **Comunicacions** entre ajuntaments i Dipsalut. Es tracta d'un altre fet diferenciat que permetrà analitzar els canals de comunicació amb els ajuntaments i les consultes més freqüents.



Il·lustració 6: Proposta d'Estrelles de la Factoria d'Informació Corporativa

3.2. Definició de l'estrella de Sol·licituds

3.2.1. Trobar la granularitat escaient

El **grànul** és la cel·la més petita que es vol tenir disponible. Habitualment, es defineixen com a grànuls les instantànies diàries d'un estat o qualsevol esdeveniment molt freqüent (per exemple, les vendes diàries d'una botiga). Atès que un grànul molt petit fa que la base de dades sigui molt gran i en canvi, un grànul molt gran implica perdre informació, és molt important triar adequadament el grànul d'una estrella.

Una possible granularitat consistiria en una imatge diària de les sol·licituds, tot i que cal avaluar la seva viabilitat, i en cas de ser necessari, es canviaria a una imatge setmanal.

Existeix però una solució mixta, ja que per a cada programa del catàleg de serveis que es pot sol·licitar, s'obre un període de sol·licitud limitat en el temps -amb data d'inici i data final- i la majoria de programes -més del 50%- són sol·licitables durant el primer quadrimestre de cada any. Per tant, una possibilitat seria utilitzar una granularitat diària durant aquest període, i una granularitat setmanal durant la resta de l'any.

El magatzem de dades s'ha posat en marxa un cop tancat el primer quadrimestre. Al llarg del segon quadrimestre, es compararà el volum en bytes generat amb el que s'ha previst en l'estudi de viabilitat, i si és acceptable, es mantindrà la granularitat diària per la resta de l'any. Si fos massa gran, es passaria a definir una granularitat setmanal a partir del cinquè mes i fins a acabar l'any.

D'aquesta manera, en lloc de guardar 365 dies, s'estaria guardant com a màxim 250, ja que no té sentit enregistrar moviments en caps de setmana. I en cas de confirmar la granularitat diària pel primer quadrimestre i setmanal per a la resta, estaríem parlant de $22 \text{ dies} \times 4 + 4 \times 8 = 120$ imatges de les dades.

En qualsevol cas, l'estudi de viabilitat de l'apartat 2.3.5 confirmarà una granularitat o una altra.

3.2.1. Dimensions de l'estrella de Sol·licituds i atributs associats

Les dimensions de l'estrella de Sol·licituds són **Temps, Programa, Estat, Ajuntament, Proveïdor, i Equipaments**. D'altra banda, per a cada dimensió, cal definir qualsevol atribut que pugui ser útil per a seleccionar, agrupar o simplement posar com a capçalera d'un informe.

A continuació es descriuen les dimensions i atributs de cadascuna d'elles.

La dimensió de **Temps**, que és molt habitual en qualsevol estrellada, té diferents nivells d'agregació: *dia, setmana, mes, trimestre i any*. Per tant, aquests són els atributs de la dita dimensió.

La dimensió de **Programa** està relacionada amb l'àrea a la qual pertany una sol·licitud -Protecció o Promoció- i a més, cal saber la *tipologia*, és a dir, saber si són de suport econòmic o de serveis. Els programes també es poden agrupar en catàlegs, els quals tenen un any associat. Per exemple, el catàleg 2011 agrupa tots els programes que es poden sol·licitar a partir del 30 de novembre 2010 fins el 31 de desembre 2011. Els atributs són per tant *àrea, tipologia i catàleg*.

La dimensió **Estat** amb un únic atribut defineix els diferents *estats* en que es pot trobar una sol·licitud (en esborrany, pendent de validar, pendent de registrar, registrada).

La dimensió **Ajuntament** tindrà varis atributs que permetran seleccionar dades a partir d'ells mateixos. Els ajuntaments estan associats a un únic municipi, i els municipis es poden agrupar en comarques, però no cal anar més enllà, atès que Dipsalut opera a nivell de la Província de Girona. Els atributs són doncs *nom ajuntament, municipi, i comarca*.

A més, en el cas dels Ajuntaments, aquests s'haurien de poder agrupar segons els rangs del nombre d'habitants (menys de 1500 habitants, entre 1500 i 5000, entre 5000 i 10000, entre 10000 i 50000, i més de 50000). Cal doncs preveure un atribut més per al *nombre d'habitants*.

Pel que fa als **Proveïdors**, l'atribut que ens permetria seleccionar-los seria senzillament el *nom de l'empresa*. No calen dades geogràfiques ni es poden agrupar proveïdors sota cap concepte.

Finalment, la dimensió **Equipaments** està relacionada amb les instal·lacions o equipaments de cada municipi que s'han inclòs en les diferents sol·licituds per tal de dur-hi a terme actuacions (per exemple, en una analítica d'aigua d'una piscina, l'equipament seria la piscina). Els atributs que defineixen els equipaments són doncs el *nom de l'equipament, el municipi al qual pertanyen, i el tipus d'equipament* del que es tracta.

3.2.2. Descriptors i jerarquies d'agregació

Les jerarquies d'agregació són els atributs que permeten agrupar registres, mentre que els descriptors permetran seleccionar-los.

La dimensió Temps ha de permetre seleccionar o agrupar per a *dia, setmana, mes, trimestre i any*.

Els descriptors que permeten seleccionar registres de la dimensió Programa són *l'àrea, la tipologia, i el catàleg*, i alhora aquests mateixos atributs són també d'agregació, ja que tots els programes es poden agrupar per àrea, tipologia i catàleg.

La dimensió Estat només té un descriptor i cap jerarquia d'agregació.

Els registres de la dimensió Ajuntament es poden seleccionar per *nom, municipi, comarca i número d'habitants* -és a dir pels descriptors-. L'atribut *comarca* estableix les agregacions possibles.

Pel que fa a la possibilitat d'agrupar o seleccionar ajuntaments per rangs d'habitants (menys de 1500, de 1500 a 5000, de 5000 a 10000, de 10000 a 50000 i més de 50000), es decideix que, a l'hora d'obtenir llistats on aparegui aquesta dada, serà l'usuari el que introduirà en unes caselles un valor mínim i màxim per al número d'habitants. D'aquesta manera, s'evita haver d'incloure aquests rangs com a columnes de la taula d'Ajuntament i s'estalvia espai.

La dimensió **Proveïdor**, té només un descriptor, el *nom de l'empresa*.

La dimensió **Equipaments** té descriptors definits per *nom de l'equipament, ajuntament al que*

pertanyen, i tipus d'equipament. Aquest últim definirà també una jerarquia d'agregació.

3.2.3. Definició de les mesures que interessin i de les cel·les

Cal obtenir el *nombre total de sol·licituds* segons els descriptors i jerarquies d'agregació seleccionats. Es tracta doncs d'una mesura derivada.

Una altra mesura que interessa és el *nombre d'equipaments* inclosos en les sol·licituds. Es pot obtenir a nivell de sol·licitud atòmica i també com a mesura derivada.

Les cel·les que interessin són la sol·licitud atòmica, un resum de sol·licituds on es faci el sumatori de totes les sol·licituds que compleixen els requisits marcats pels descriptors, i un resum d'equipaments, on també es fa el sumatori a nivell d'equipaments municipals.

A més, es guarda l'identificador de cada sol·licitud -idSollicitud- que prové del sistema operacional, ja que aquest identificador permet anar a buscar el detall de la sol·licitud al SIMSAP si es considera interessant.

3.2.4. Viabilitat de l'estrella

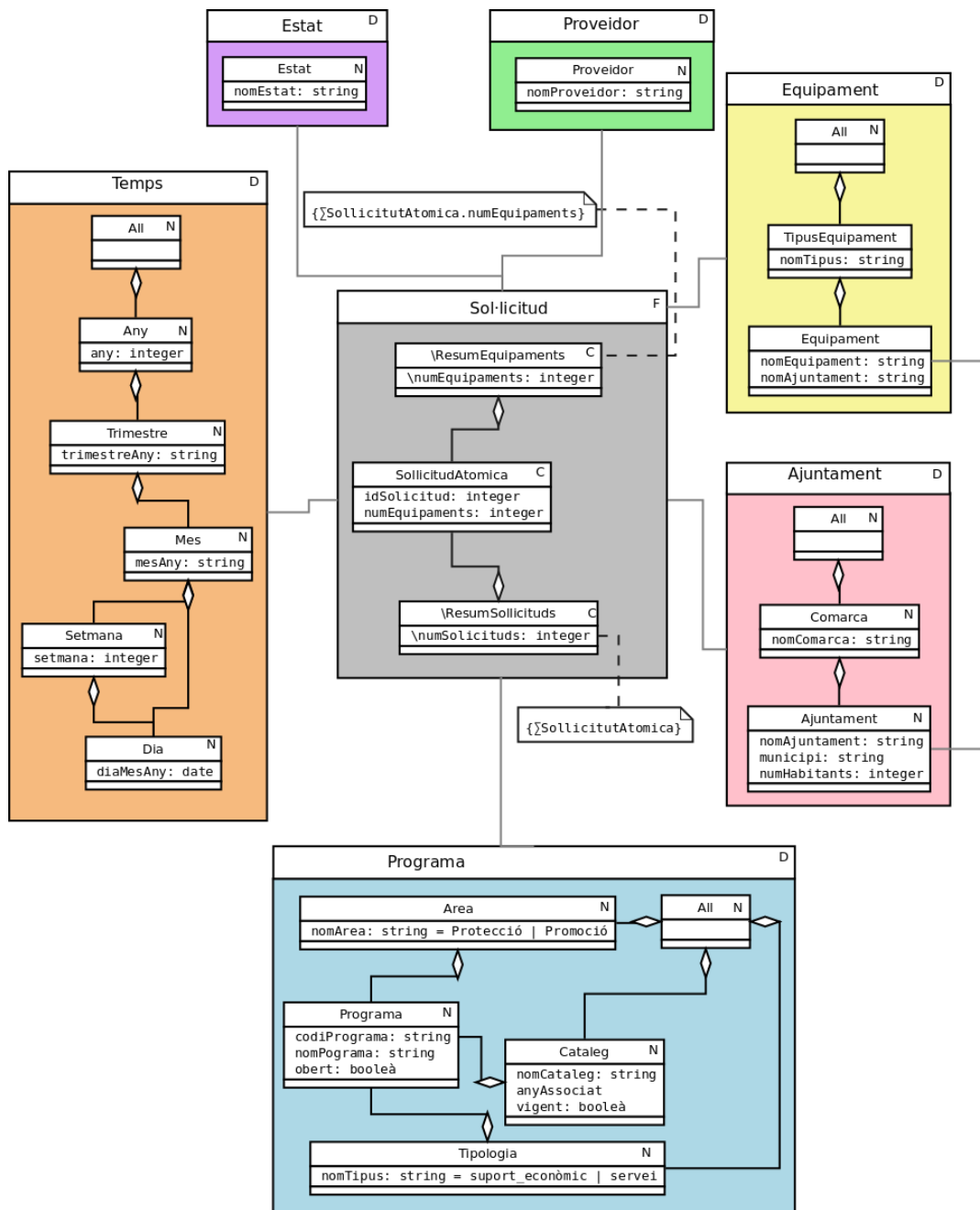
Tal com s'ha explicat en el punt 3.2.1, es vol guardar una imatge diària de les dades, al menys, durant el primer semestre. A partir d'aquí, s'estudiarà la possibilitat de fer-ho setmanalment.

Anem a veure com influiria una imatge diària durant tot l'any. D'entrada, els caps de setmana no compten, reduint-se el nombre de dies a 250. D'altra banda, el catàleg de serveis del 2011 ofereix 22 programes en total. Els estats en què es pot trobar una sol·licitud són 8 (en esborrany, pendent de validar, pendent de registrar, registrada, avaluant, bloquejada, en procés, programa finalitzat). Per a cada programa, normalment hi ha un únic proveïdor que duu a terme aquell servei, tot i que hi han excepcions. I finalment, hi han 221 municipis que poden fer sol·licituds.

Es tracta doncs de $250 \times 22 \times 8 \times 1 \times 221 = 9724000$ cel·les. Tenint en compte que cada quantitat del Fet ocupa 2 bytes i que es necessiten 6 identificadors (un per a cada dimensió) de 4 bytes cada un, la cel·la ocuparia $2 \times 4 + 6 \times 4 = 32$ bytes.

Finalment, s'obté un espai de $9724000 \times 32 = 311.168.000$ bytes / 1024 = 303.875 Kbytes / 1024 = **296 Mb**, resultat que, malgrat ser el màxim possible, és totalment acceptable. Sabem a més que no tots els ajuntaments sol·liciten totes les sol·licituds, per tant, aquesta fita és la màxima anual i ben segur el volum total sempre estarà per sota.

3.2.5. Diagrama del disseny conceptual resultant



Il·lustració 7.: Disseny conceptual de l'Estrella de Sol·licituds

3.3. Definició de l'estrella d' Actuacions

A continuació es defineix l'estructura de l'estrella d'Actuacions, que aporta informació valuosa sobre les feines derivades de les sol·licituds i que duen a terme els proveïdors. Cada programa del catàleg de serveis porta associades unes determinades accions que sempre s'han de fer i d'altres que són optatives. Per exemple, en un programa de detecció i prevenció de legionel·la, cal dur a terme les següents accions obligatòries, que han estat definides per l'equip tècnic de protecció de la salut de Dipsalut :

- Descripció i implantació dels plans de gestió: autocontrol, manteniment, neteja i desinfecció
- Analítica de recompte legionel·la
- Neteja i desinfecció de la instal·lació
- Visita d'assessorament i avaluació de la implantació dels plans de gestió (1 cada trimestre)

Cal remarcar que algunes d'aquestes accions tenen periodicitat: algunes només es duen a terme un cop l'any i d'altres més d'una, com és el cas de la darrera acció.

A més, hi han altres accions que són optatives i que pot ser no es duran a terme mai, com la neteja i desinfecció de la instal·lació per presència de legionel·la (si no es detecta legionel·la no cal fer-la).

En aquest disseny, totes les accions possibles vinculades a un programa s'anomenen **Accions**. Un cop el proveïdor inicia una acció, aquesta passa a tenir la denominació d'**Actuació**, raó per la qual s'estudia el fet de les Actuacions.

3.3.1. Trobar la granularitat escaient

La granularitat triada en l'estrella de Sol·licituds ha estat a nivell diari. Pel que fa a les Actuacions, es mantindrà la mateixa granularitat per poder estudiar les variacions en les accions a nivell diari, sempre i quan es demostrï la seva viabilitat en l'apartat corresponent.

3.3.2. Dimensions de l'estrella d'Actuacions i atributs associats

Les dimensions de l'estrella d'Actuacions són pràcticament les mateixes que les de l'estrella Sol·licituds. És a dir, tenim **Temps**, **Estat**, **Programa**, **Ajuntament**, **Proveïdor**, i **Equipaments** que no varien i mantenen els mateixos atributs.

No obstant, hi ha una diferència entre els estats d'una sol·licitud i els estats d'una actuació, ja que no són els mateixos i tenen denominadors diferents. Una opció per resoldre aquesta diferència, és definir una dimensió nova anomenada **EstatActuacio**, separant-la de la dimensió Estat de les sol·licituds. Aquesta dimensió EstatActuacio, igual l'altre, tindrà un únic atribut anomenat *nomEstat*.

D'altra banda, apareix una nova dimensió que fa referència a les **Accions**. Aquestes es podrien considerar com a nivell d'agregació dins la dimensió **Programa**, ja que unes determinades accions es poden agrupar per programa, però posteriorment, en implementar el disseny lògic i físic, la taula Programa tindria unes noves columnes que són totalment innecessàries pel que fa a l'estudi de les sol·licituds, i a més, el volum de la taula creixeria, doncs per a cada programa, hi hauran tantes files com accions associades a aquell programa. De l'altra manera, tenim els programes en una taula, i les accions en una altra, amb un camp que referencia al programa al que pertanyen.

3.3.3. Descriptors i jerarquies d'agregació

Els descriptors i jerarquies d'agregació són els mateixos que per l'estrella de Sol·licituds. Només s'explicitarà en aquest apartat els descriptors de la nova dimensió Accio. Una acció tindrà uns descriptors que permetran seleccionar instàncies i que seran el *nom de l'acció*, el *programa* al qual pertanyen, si és *obligatòria* o no, i la seva *periodicitat*.

3.3.4. Definició de les mesures que interessin i de les cel·les

Cal obtenir el *nombre total d'actuacions* segons els descriptors i jerarquies d'agregació seleccionats. Aquesta serà per tant una mesura derivada.

Les cel·les que interessin són les que defineixen una actuació atòmica i un resum d'actuacions on es faci el sumatori de totes les actuacions que compleixen els requisits marcats pels descriptors.

A més, es guardaran dos identificadors provinents del sistema operacional que permetran anar a buscar una actuació concreta o la sol·licitud de la qual prové. Aquests identificadors seran : *idActuacio* i *idSollicitud* i ni serviran per agrupar ni per seleccionar.

3.3.5. Viabilitat de l'estrella

Sabem que es vol guardar l'històric de 250 dies, i que hi han 22 programes, però hi han algunes dades incertes que poden fer variar molt els càlculs. Aquestes dades són la quantitat d'equipaments per als quals els ajuntaments sol·liciten actuacions i per cadascun dels quals s'hauran de dur a terme totes les accions obligatòries associades al programa.

Per exemple, un ajuntament de la costa tindrà probablement moltes piscines d'ús públic, mentre que un poble petit d'alta muntanya només en tindrà una o dues, i a més, no se sap si demanarà les actuacions per tots els equipaments o només per alguns.

Per a estimar un volum aproximat, caldrà partir de la quantitat total d'equipaments sol·licitats l'any anterior i aplicar-hi un increment del 30% ja que enguany hi han ajuntaments nous que s'han adherit al catàleg de serveis de Dipsalut. Aquests ajuntaments nous estan sol·licitant programes que l'any passat no havien sol·licitat. L'any passat doncs, hi van haver un total de 2181 equipaments i aplicant-hi l'increment estimat, serien uns 2835 equipaments.

D'altra banda, hi han programes que tenen més de 10 accions associades i d'altres que només en tenen una. Per a poder efectuar el càlcul de viabilitat, cal fer una mitjana d'accions per programa. Aquesta resulta en 5 accions per programa.

Finalment, les actuacions poden estar en 4 estats diferents: en execució, finalitzada, validada per Dipsalut, facturada.

Es tracta doncs de $250 \times 22 \times 2835 \times 5 \times 4 = 311850000$ cel·les. Tenint en compte que cada quantitat del fet ocupa 2 bytes i que es necessiten 7 identificadors (un per a cada dimensió) de 4 bytes cada un, la cel·la ocuparia $2 \times 2 + 7 \times 4 = 32$ bytes.

Finalment, s'obté un espai anual de $311850000 \times 28 = 9979200000$ bytes / 1024 = 9745312 Kbytes / 1024 = **9517 Mb = 9,29 GB**. Si es vol mantenir aquesta granularitat, caldrà un servidor dedicat per al magatzem de dades corporatiu. No es descarta reduir la granularitat a nivell de setmana, tot i que inicialment, es vol poder estudiar el fet sense perdre informació.

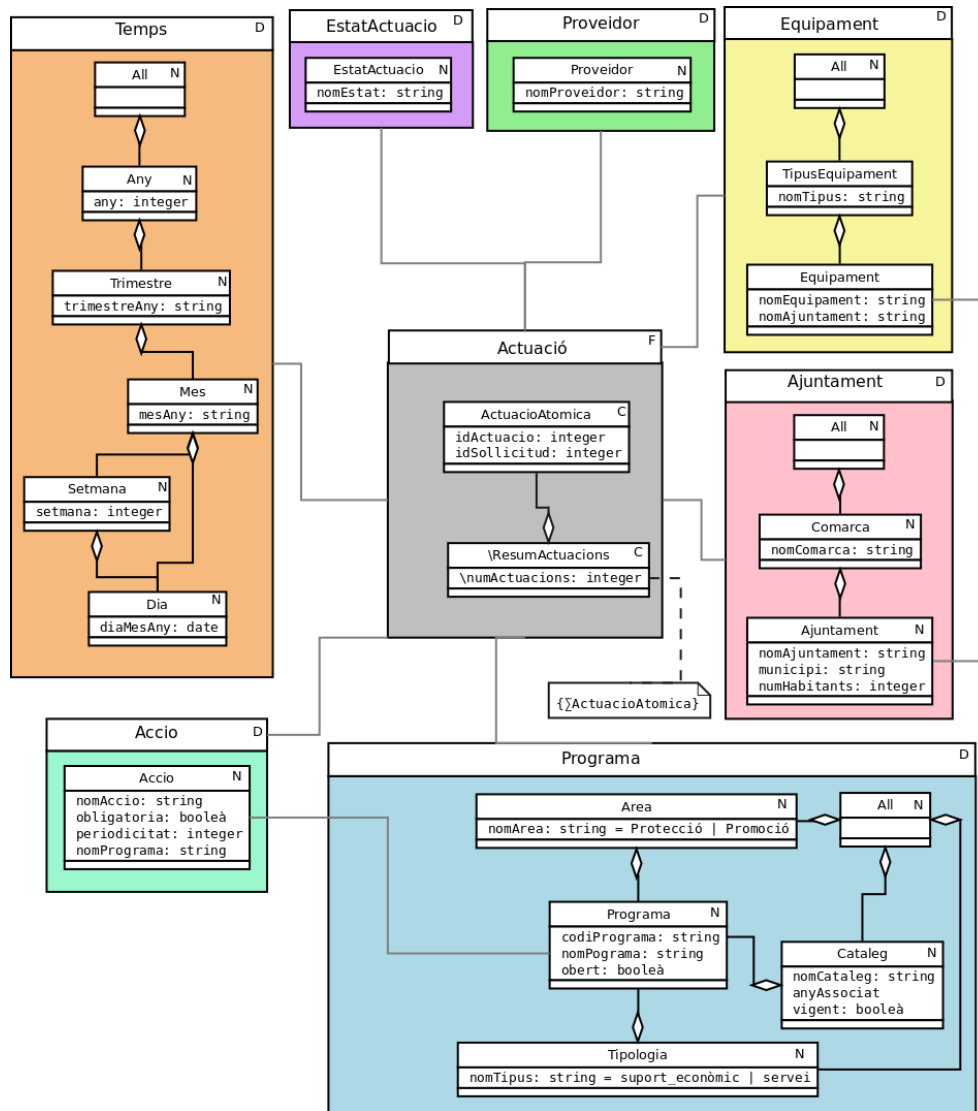
3.3.6. Compartició de dimensions entre Sol·licituds i Actuacions

La gran majoria de dimensions de les estrelles de Sol·licituds i Actuacions són les mateixes. Repetir-les en un mateix magatzem de dades no tindria sentit i malbarataria l'espai de la base de dades física. Es poden utilitzar les mateixes taules de dimensió en estrelles diferents sempre i quan tinguin la mateixa forma. Es diu que cal "*conformar*" les dimensions per poder fer possible el *Drill-across*¹ i per

¹ *Drill-across*: Operació multidimensional que, donat un espai, canvia les dades que s'hi mostren. Canvia d'un tema d'anàlisi a un altre.

estalviar feina de disseny i administració. D'aquesta manera, les taules dels fets (SollicitudAtomica i ActuacioAtomica) es podrien substituir l'una per l'altre sense modificar la resta de les consultes.

3.3.7. Diagrama del disseny conceptual resultant



Il·lustració 8: Disseny conceptual de l'estrella d'Actuacions

3.4. Definició de l'estrella de Comunicacions

Es vol estudiar el fet de les comunicacions per saber quins ajuntaments realitzen més consultes als Agents de Salut Pública, quin tipus de consulta són (segons el programa, l'àrea o la tipologia), i quin mitjà utilitzen per a fer consultes (correu electrònic, reunions presencials o telèfon).

3.4.1. Trobar la granularitat escaient

Novament, i per tal de no perdre nivell de detall, es proposa una granularitat diària. S'estudiarà a l'apartat corresponent la viabilitat d'aquesta granularitat.

3.4.2. Dimensions de l'estrella de Comunicacions i atributs associats

Algunes de les dimensions de l'estrella de Comunicacions són les mateixes que les de l'estrella Sol·licituds i Actuacions. Concretament, comparteix les dimensions de **Temps, Ajuntament, i Programa**.

Sorgeix una dimensió nova anomenada **Mitjà**, que tindrà un únic atribut de tipus *string* i que recollirà el mitjà utilitzat en la comunicació.

3.4.3. Descriptors i jerarquies d'agregació

Pel que fa a Temps, Ajuntament i Programa, els descriptors i jerarquies d'agregació són els mateixos que per l'estrella d'Actuacions. La dimensió Mitjà només tindrà un descriptor que permetrà seleccionar instàncies del fet.

3.4.4. Definició de les mesures que interessin i de les cel·les

Cal obtenir el *nombre total de comunicacions* segons els descriptors i jerarquies d'agregació seleccionats. Aquesta serà per tant una mesura derivada.

Les cel·les que interessin són les que defineixen una comunicació atòmica, i un resum de comunicacions on es faci el sumatori de totes les comunicacions que compleixen els requisits marcats pels descriptors.

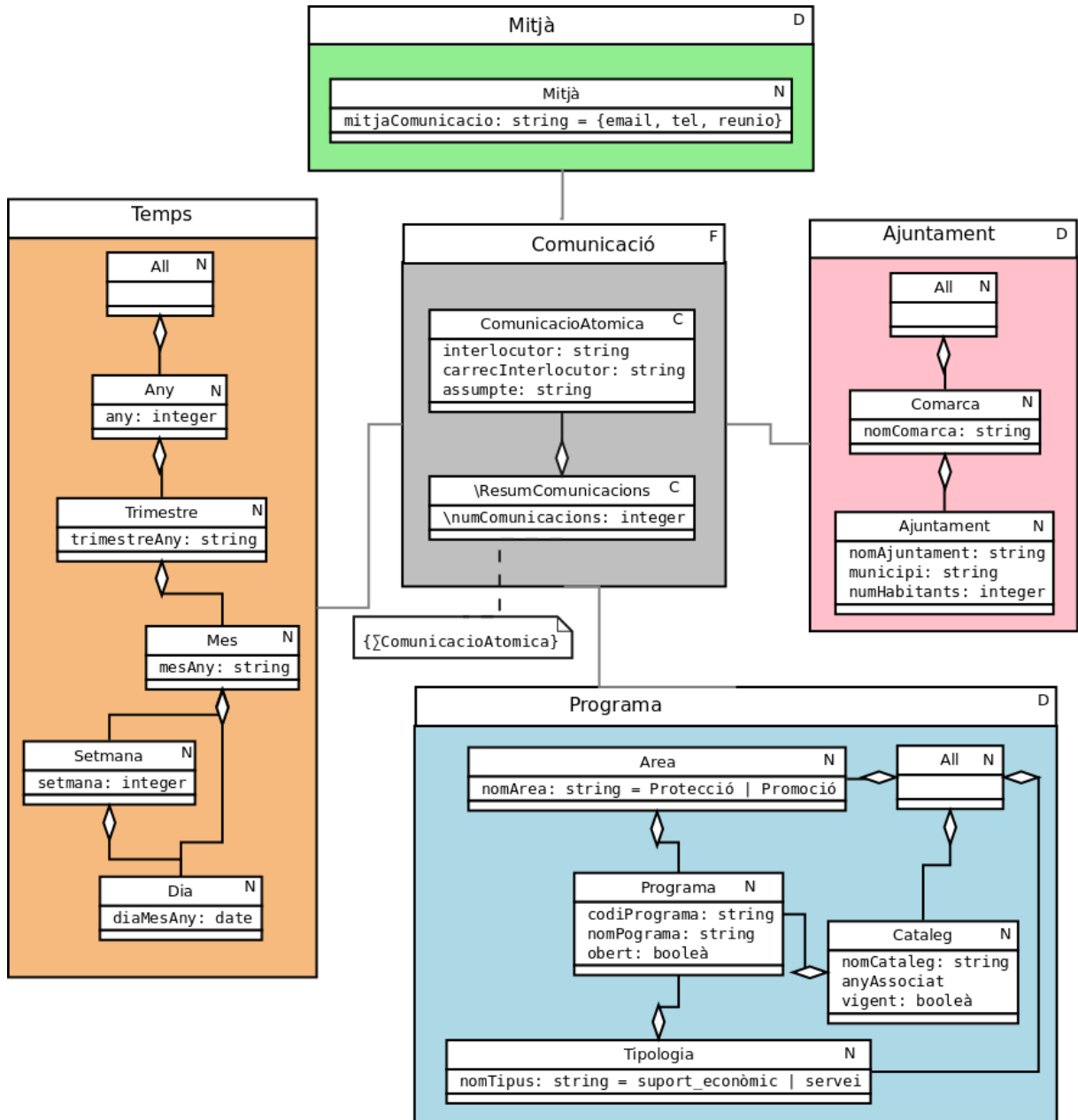
3.4.5. Viabilitat de l'estrella

Tenim 221 ajuntaments que realitzen una mitjana de 4 consultes mensuals, la majoria de vegades sobre algun dels 22 programes, que poden fer-se per 3 mitjans diferents (email, telèfon o reunió) i volem guardar informació de 250 dies.

Es tracta doncs de $250 \times 221 \times 22 \times 4 \times 3 = 14586000$ cel·les. Tenint en compte que cada quantitat o atribut del fet ocupa 2 bytes i que es necessiten 4 identificadors (un per a cada dimensió) de 4 bytes cada un, la cel·la ocuparia $2 \times 4 + 4 \times 4 = 24$ bytes.

Finalment, s'obté un espai anual màxim de $14586000 \times 24 = 350064000$ bytes / 1024 = 341859 Kbytes / 1024 = **334 Mb anuals** . Efectivament, la granularitat diària és viable.

3.4.6. Diagrama del disseny conceptual resultant



Il·lustració 9: Disseny conceptual de l'estrella Comunicacions

3.5. Disseny lògic

En aquest punt cal traduir el disseny conceptual multidimensional OLAP a un disseny lògic relacional ROLAP².

Per a fer-ho, s'agafa cada estrella per separat, creant una taula per a cada dimensió i una altra per al fet, en la qual cada fila representa una cel·la de l'espai multidimensional. Pel que fa a les jerarquies d'agregació, queden recollides en els valors dels atributs de les taules de dimensió.

Finalment, les taules resultants es relacionen entre elles mitjançant claus foranes. Per reduir la grandària de les taules, es defineixen substituïts de les claus primàries, que en lloc de ser cadenes de text, consisteixen en identificadors únics enters (de 4 bytes).

Les taules de dimensions es creen a partir d'unes dades que no varien pràcticament mai. En canvi, a les taules de fets s'aniran inserint files regularment.

3.5.1. Disseny lògic de l'estrella de Sol·licituds

Temps (idTemps, diaMesAny, setmana, mesAny, trimestreAny, any)

Estat (idEstat, nomestat)

Proveïdor (idProveïdor, nomProveïdor)

Programa (idPrograma, codiPrograma, nomPrograma, obert, nomTipus, nomArea, nomCatalog, anyAssociat, vigent)

Ajuntament (idAjuntament, nomAjuntament, municipi, numHabitants, nomComarca)

Equipament (idEquipament, nomEquipament, nomAjuntament, nomTipus)

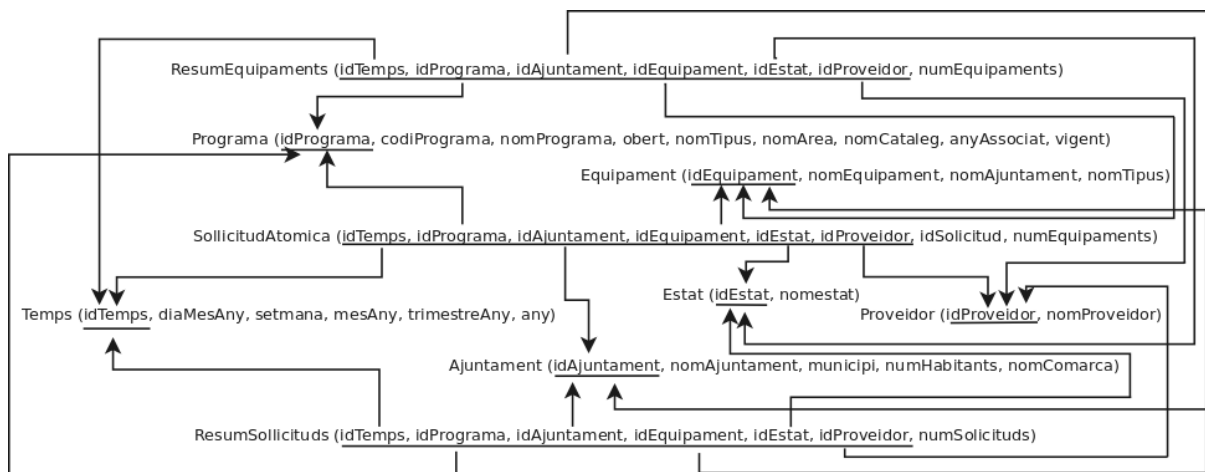
SollicitudAtomica (idTemps, idPrograma, idAjuntament, idEquipament, idEstat, idProveïdor, idSollicitud, numEquipaments)

ResumEquipaments (idTemps, idPrograma, idAjuntament, idEquipament, idEstat, idProveïdor, numEquipaments)

ResumSollicituds (idTemps, idPrograma, idAjuntament, idEquipament, idEstat, idProveïdor, numSollicituds)

Cal remarcar que tenim més d'una cel·la per al mateix fet, el que origina una estructura coneguda com a **Floc de neu**, enlloc d'una estrella.

² ROLAP: Relational Online Analytical Processing



Il·lustració 10: Disseny lògic de Sol·licituds

3.5.2. Disseny lògic de l'estrella d'Actuacions

No cal repetir les taules Temps, Proveïdor, Ajuntament i Equipament, ja que no varien respecte de l'estrella de sol·licituds.

Les noves taules que cal definir són:

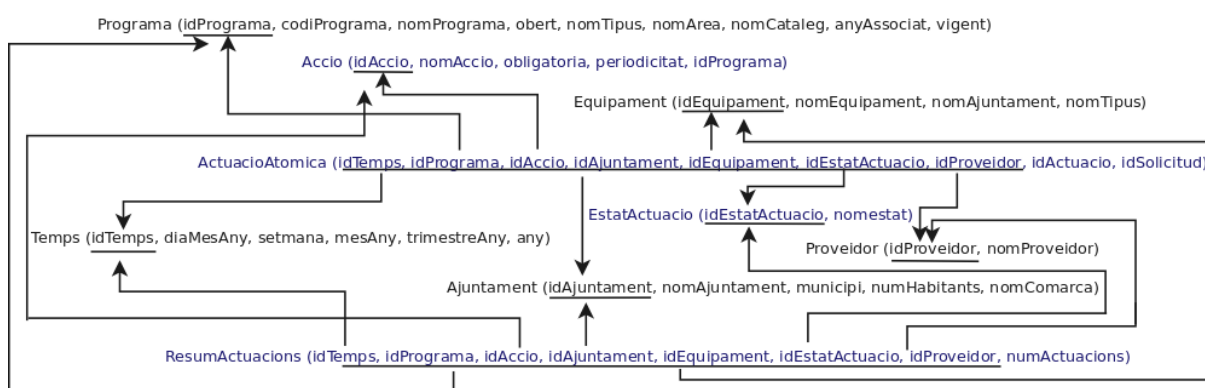
Accio (IdAccio, nomAccio, obligatoria, periodicitat, idPrograma)

EstatActuacio (IdEstatActuacio, nomestat)

ActuacioAtomica (IdTemps, IdPrograma, IdAccio, IdAjuntament, IdEquipament, IdEstatActuacio, IdProveïdor, IdActuacio, IdSol·licitud)

ResumActuacio (IdTemps, IdPrograma, IdAccio, IdAjuntament, IdEquipament, IdEstatActuacio, IdProveïdor, numActuacions)

En la figura de sota, les noves taules apareixen en color blau marí, i les que s'aprofiten de l'estrella de Sol·licituds en color negre.



Il·lustració 11: Disseny lògic d'Actuacions

3.5.3. Disseny lògic de l'estrella de Comunicacions

Novament, es mantenen invariables les taules Temps, Ajuntament i Programa, respecte al floc de neu

d'Actuacions i Sol·licituds.

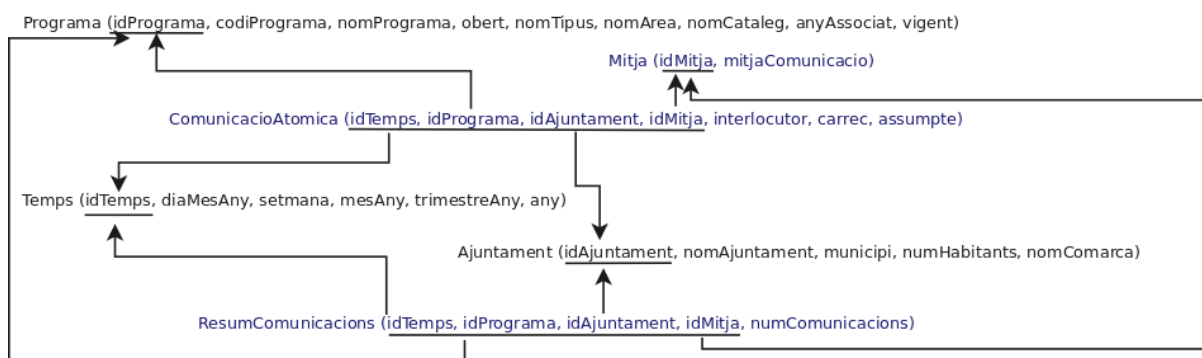
Les úniques taules noves que s'han de definir són la que representa la dimensió Mitjà i les taules dels fets:

Mitja (idMitja, mitjaComunicacio)

ComunicacioAtomica (idTemps, idPrograma, idAjuntament, idMitja, interlocutor, carrec, assumpte)

ResumComunicacions (idTemps, idPrograma, idAjuntament, idMitja, numComunicacions)

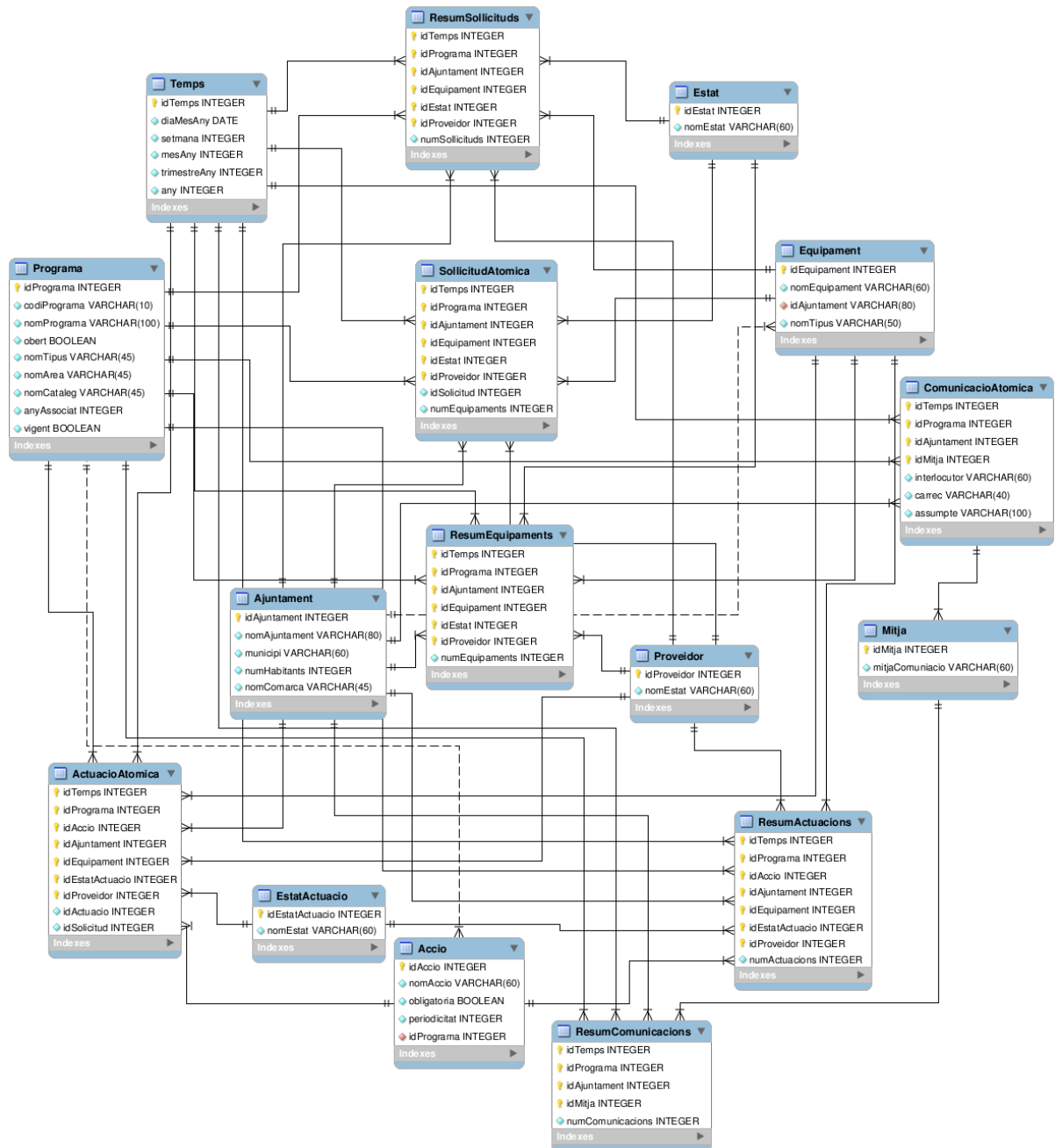
En la figura de sota que representa el floc de neu de Comunicacions, les noves taules apareixen en color blau marí, i les que s'aprofiten del model lògic de Sol·licituds i Actuacions es mantenen en color negre.



Il·lustració 12: Disseny lògic de l'estrella de Comunicacions

3.6. Disseny físic

El present disseny físic inclou tota la base de dades del magatzem, integrant els dissenys lògics dels tres flocs de neu : Sol·licituds, Actuacions i Comunicacions.



Il·lustració 13: Disseny físic de la base de dades del magatzem corporatiu

En el lliurament del producte -fitxer `adesset_producte.zip`- s'inclou el script de creació del magatzem de dades, concretament de les taules involucrades en l'estrella de Sol·licituds.

3.7. Disseny i descripció dels informes

3.7.1. Consultes sobre l'estrella de Sol·licituds

Amb el floc de neu resultant es poden efectuar una gran varietat de consultes:

A nivell de sol·licituds atòmiques podem obtenir:

- Llistat de totes les sol·licituds d'un dia, setmana, mes, trimestre, any
- Llistat de totes les sol·licituds agrupades per estat, o de les sol·licituds que es troben en un estat concret seleccionat per l'usuari.
- Llistat de totes les sol·licituds agrupades per proveïdor al qual s'han assignat, o de les que han estat assignades a un proveïdor concret seleccionat per l'usuari.
- Llistat de totes les sol·licituds agrupades per ajuntament i comarca o bé, seleccionant un ajuntament o comarca, obtenir les sol·licituds de l'ens o comarca seleccionats.
- Llistat de totes les sol·licituds agrupades per catàleg, àrea, programa i/o línia, i a més, a partir d'un programa, àrea, catàleg o tipologia seleccionats per l'usuari, obtenir les sol·licituds que corresponen a les condicions triades.
- Llistat de totes les sol·licituds agrupades per tipologia d'equipaments, o bé, obtenir les sol·licituds en les quals s'ha inclòs un equipament concret.

A més, totes les condicions anteriors de selecció i agregació es poden combinar entre elles, de tal manera que es podria obtenir per exemple:

- Llistat de totes les sol·licituds de la comarca de 'La Selva' fetes durant l'últim mes, agrupades per àrea, programa i línia.
- Llistat de totes les sol·licituds en estat de 'assignada' o 'en procés' que s'han assignat a un proveïdor concret durant el darrer trimestre 2010.

A nivell de la cel·la ResumSollicituds i ResumEquipaments, es podrien obtenir les mateixes combinacions de consultes, és a dir:

- Nombre total de sol·licituds d'un dia, setmana, mes, trimestre o any.
- Quantes sol·licituds es troben en un estat concret, o un resum de nombre de sol·licituds per estat i programa.

3.7.2. Consultes sobre l'estrella d'Actuacions

Amb el floc de neu resultant i compartint dimensions amb el de Sol·licituds, es podrà fer un Drill-across, intercanviant els dos fets d'estudi.

A nivell d'actuacions atòmiques podem obtenir:

- Llistat de totes les actuacions d'un dia, setmana, mes, trimestre, any
- Llistat de totes les actuacions agrupades per estat, o de les actuacions que es troben en un estat concret seleccionat per l'usuari.
- Llistat de totes les actuacions agrupades per proveïdor, o del proveïdor que han seleccionat

els usuaris.

- Llistat de totes les actuacions agrupades per ajuntament i comarca o bé, seleccionant un ajuntament o comarca, obtenir les actuacions de l'ens o comarca seleccionats.
- Llistat de totes les actuacions agrupades per catàleg, àrea, programa i/o tipologia, i a més, a partir d'un programa, àrea, catàleg o tipologia seleccionats per l'usuari, obtenir les actuacions que corresponen a les condicions triades.
- Llistat de totes les actuacions lligades a un tipus d'acció concreta.
- Llistat de totes les actuacions agrupades per tipologia d'equipaments, o bé, obtenir les actuacions en les quals s'ha inclòs un equipament concret.

Igual que passava amb el model de les Sol·licituds, totes les condicions anteriors de selecció i agregació es poden combinar entre elles, de tal manera que es podria obtenir per exemple:

- Llistat de totes les actuacions fetes durant l'últim mes a la comarca del Gironès, agrupades per àrea, programa, tipologia i nom de l'acció.
- Llistat de totes les actuacions en estat de 'validada per Dipsalut' o "facturada" d'un proveïdor concret durant el darrer trimestre, i agrupades per tipologia.

A nivell de la cel·la ResumActuacions, es podrien obtenir les mateixes combinacions de consultes, és a dir:

- Nombre total d'actuacions d'un dia, setmana, mes, trimestre o any.
- Quantes actuacions es troben en un estat concret, o un resum de nombre d'actuacions per estat i programa.

3.7.3. Consultes sobre l'estrella de Comunicacions

A nivell de comunicacions atòmiques podem obtenir:

- Llistat de totes les comunicacions d'un dia, setmana, mes, trimestre, any
- Llistat de totes les comunicacions agrupades per mitjà, o de les comunicacions efectuades mitjançant un mitjà seleccionat per l'usuari.
- Llistat de totes les comunicacions agrupades per ajuntament i comarca o bé, seleccionant un ajuntament o comarca, obtenir les actuacions de l'ens o comarca seleccionats.
- Llistat de totes les comunicacions agrupades per catàleg, àrea, programa i/o tipologia, i a més, a partir d'un programa, àrea, catàleg o tipologia seleccionats per l'usuari, obtenir les comunicacions relacionades.

A nivell de resum, s'obtindrà la quantitat total de comunicacions, d'acord amb els descriptors triats per l'usuari que executi l'informe.

4. Captures de pantalla i resultats obtinguts

4.1. Procés d'extracció, transformació i càrrega *ETL*.

L'eina *Spoon* de *Kettle-Pentaho* (veure Annex I per a més referències) facilita enormement el modelatge de processos d'extracció, transformació i càrrega de dades, tot i que és l'eina més complicada d'utilitzar de la suite de *Pentaho*. A continuació es descriu el procés seguit per a carregar l'estrella de Sol·licituds.

4.1.1. ETL Ajuntaments

En primer lloc, s'han arrossegat a la pantalla les icones '*Table Input*'. Aquests objectes modelen una taula d'origen i permeten aprofitar connexions existents a una base de dades del MySQL. En aquest cas, hi han dues bases de dades configurades:



Ajuntaments SIMSAP

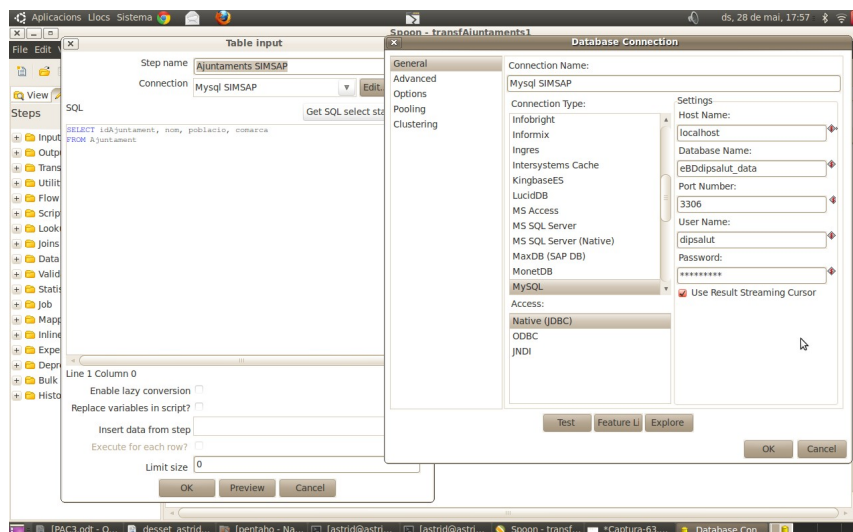
Il·lustr. 14: objecte tipus '*Table Input*'

eBDDipsalut_data : base de dades replicada del SIMSAP

magatzem : que conté les taules definides en el disseny de l'estrella Sol·licitud i Accions

Per tant, cal arrossegar la icona de '*Table input*', reanomenar-la i fer-hi doble clic per obrir el quadre de diàleg que es pot observar a la il·lustració 16. Si ja existeix una connexió configurada, es pot recuperar amb el desplegable '*Connection*', en cas contrari, s'ha de configurar com a nova, tal com es mostra en la mateixa il·lustració, en la finestra flotant de la dreta.

A continuació es pot observar com es configuren les connexions i s'obtenen els camps desitjats. Cal haver creat un usuari al MySQL que tingui accés a les bases de dades involucrades en el magatzem. En aquest cas, s'ha utilitzat l'usuari **dipsalut**.



Il·lustr. 15: Configuració d'una connexió a la BDD d'origen i obtenció de les dades

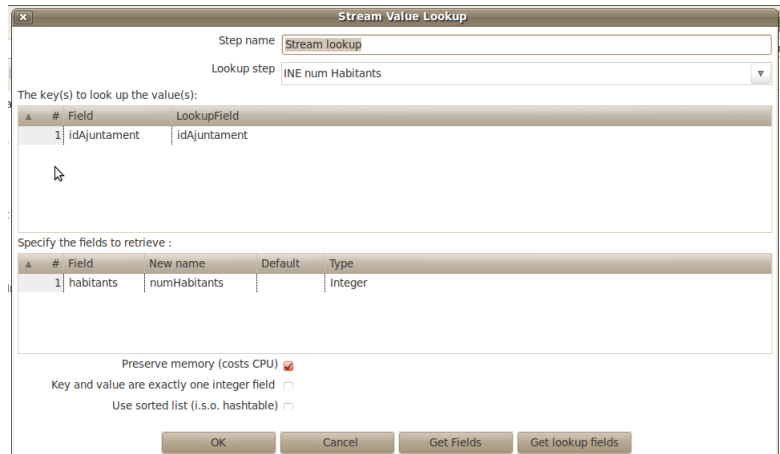
D'altra banda, en el SIMSAP, el nombre d'habitants de cada municipi està emmagatzemat en la taula INE, ja que cada any, l'Institut Nacional d'Estadística publica el cens actualitzat de la població, per tant, tot i que la taula Ajuntament conté un camp amb el nombre d'habitants, a la realitat, s'ha d'anar a buscar aquesta dada per al darrer any censat. En conseqüència, cal afegir aquesta taula i programar que s'agafi el camp *numHabitants* d'aquesta última i no de la primera.

A tal efecte s'ha utilitzat la icona 'Stream lookup' que es lliga a les dues taules d'origen mitjançant salts o 'Hops'. Per a unir les dues taules amb el 'Stream lookup'



Il·lustr. 16: Icona Stream lookup

només cal clicar la icona d'origen i dibuixar amb el ratolí una fletxa cap al destí, mentre es manté premuda la tecla *Shift* en tot el procés.



Il·lustr. 17: cerca de valors dins d'una altra taula

Un cop la icona 'Stream lookup' està unida a les dues taules d'origen, es fa doble clic a sobre i es configuren els camps clau que relacionen ambdues taules, en aquest cas, *idAjuntament*. També cal especificar els camps que ens interessa obtenir (*habitants*) i a més, interessa renombrar-lo a *numHabitants* per a que coincideixi amb el camp de la taula de destí del magatzem.

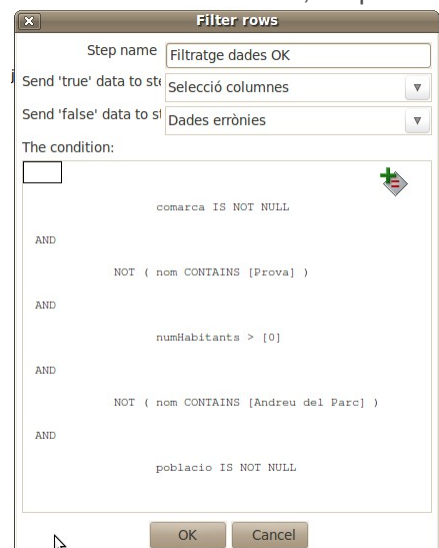
En qualsevol icona es pot fer clic amb el botó dret del ratolí i visualitzar quins camps entren i quins camps surten de la transformació en qüestió. També es poden previsualitzar els resultats, lo qual és molt útil per a detectar possibles errors en les dades i tractar-los.

En el cas dels ajuntaments, gràcies a la vista prèvia de resultats, s'ha detectat que es tenien en compte un parell d'ajuntaments ficticis que s'havien fet servir durant les proves del SIMSAP, i que aquests no tenien correctament definida la comarca o la població. Per a solucionar aquest cas s'ha utilitzat la icona 'Filter rows'.



Il·lustr. 18: Icona per filtrar les dades

A més per a excloure un ajuntament també fictici que sí contenia municipi i comarca, tot i que inventats, ha calgut afegir la condició *NOT nom (CONTAINS [Andreu del Parc])*.

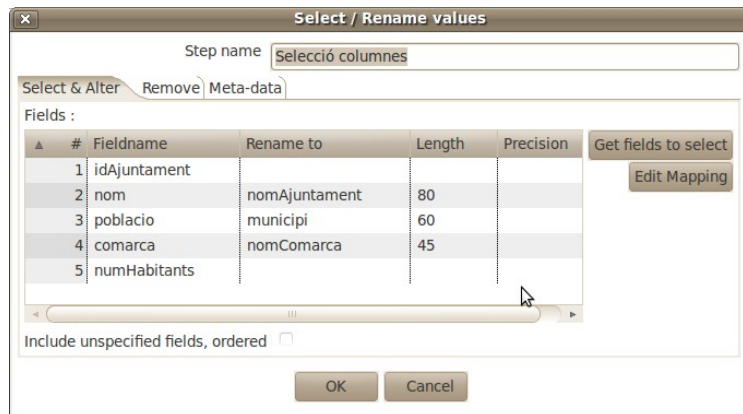


Il·lustr. 19:opcions de filtratge de dades



Il·lust. 20: fluxe de dades correctes i errònies

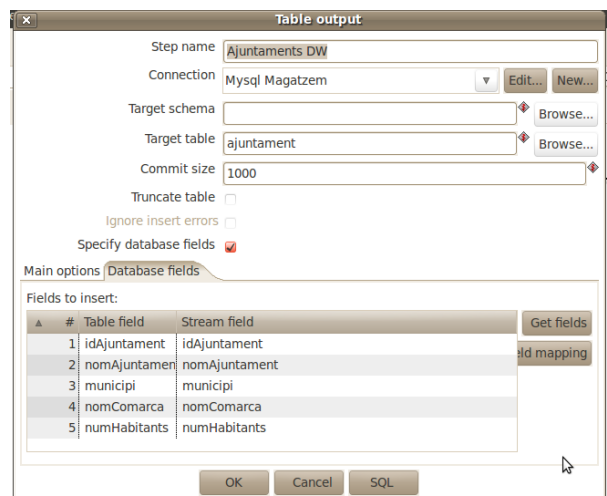
En la il·lustració de la dreta, el camp *idAjuntament* roman sense canvis, el camp nom es dirà *nomAjuntament* i ocuparà 80 caràcters per adaptar-lo a la taula de destí, *poblacio* es dirà *municipi* i *comarca* es dirà *nomComarca*, alhora que *numHabitants* no canvia. Hi ha una pestanya per afegir metadades que no s'ha utilitzat.



Il·lust. 21: utilitat de selecció de camps

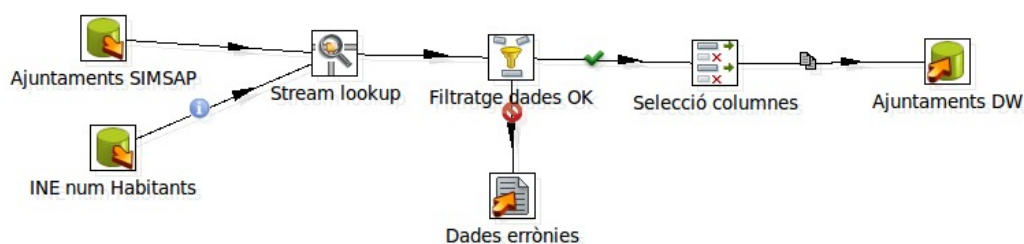
Finalment, cal afegir la taula de destinació. A tal efecte, *Spoon* proveeix una icona nomenada '*Table output*' que permet mapejar els camps entrants i sortints. Cal fer-hi doble clic i configurar la connexió a la base de dades.

En la il·lustració 23, s'ha configurat una connexió a *localhost*, base de dades *magatzem*. La taula de destinació és *ajuntament* i s'han mapejat els camps que es diuen igual. És convenient prémer el botó SQL per comprovar que els camps entrants i sortints són compatibles. En cas negatiu, apareixerà una sentència SQL que proposa una alteració de la taula de destí, per tal d'ajustar-la als camps entrants. Com que el tamany dels camps i el nom dels mateixos s'han ajustat en el moment de seleccionar-los -*Select rows*-, només resta prémer OK.



Il·lust. 22: configuració de la taula de destinació

La tasca d'extracció, transformació i càrrega queda doncs definida de la següent manera:



Il·lustr. 23: Procés ETL sencer de càrrega de la taula 'ajuntament' del DW

Spoon permet desar aquestes transformacions i recuperar-les posteriorment. Aquesta transformació s'adjunta en el ZIP del producte lliurat, dins la carpeta *Spoon* i amb el nom *transfAjuntaments1.ktr*. Cal tenir en compte que si es vol provar el procés en local, s'han d'obrir les taules d'origen i destí i configurar les connexions a Mysql emprant un usuari i passwords existents en el sistema. Els fitxers *ktr* només es poden obrir amb *Spoon*.

4.1.2. ETL Estats

La càrrega d'estats de les sol·licituds presenta diverses particularitats que fan que no sigui trivial. En primer lloc, en la base de dades d'origen -SIMSAP- hi ha dues taules diferents que identifiquen els estats de les sol·licituds. D'una banda, hi ha els estats que pot veure un ajuntament, i en segon lloc, els estats que pot veure Dipsalut o un proveïdor de serveis.

Els ajuntaments només poden veure 4 estats: (en esborrany, pendent de validar, pendent de registrar i registrada). A partir del moment en que la sol·licitud està registrada, pot evolucionar a diferents estats (en procés, avaluant, denegada, etc).

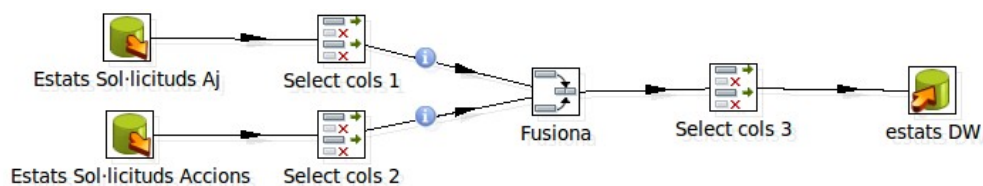
Així doncs, és necessari obtenir els 4 primers estats de la taula *StatusAj* del SIMSAP, i la resta d'estats de *StatusGral*. En la finestra de configuració de connexions i selecció de taules, la sentència SQL que permet obtenir els 4 primers és:

```
SELECT idStatusSol, statusAj FROM StatusSol WHERE idStatusSol<5
```

La que permet obtenir la resta:

```
SELECT idStatusGral, statusDip FROM StatusGral WHERE idStatusGral>=5 ORDER BY idStatusGral
```

Tal com es pot observar, s'utilitza la icona de *Select cols* per a seleccionar les mateixes columnes i s'uneix el resultat amb *Fusiona* (*Merge rows*). Finalment, es carrega la taula d'estats sense més complicacions.



Il·lustr. 24: Procés ETL sense de càrrega d'estats al DW

Aquesta transformació s'adjunta en el ZIP del producte, a la carpeta *Spoon*, fitxer *transfEstats1.ktr*

4.1.3. ETL Proveïdors

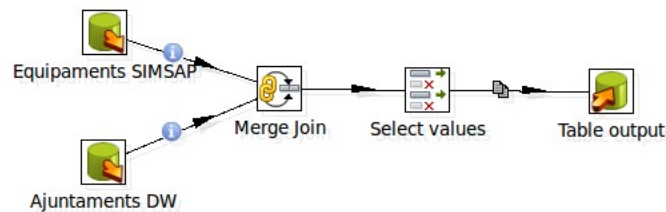
El procés ETL de Proveïdors és extremadament senzill i les eines emprades ja han estat descrites. Es poden consultar els passos seguits en el fitxer *transfProveïdors1.ktr*.



Il·lustr. 25: Procés ETL de proveïdors

4.1.4. ETL Equipaments

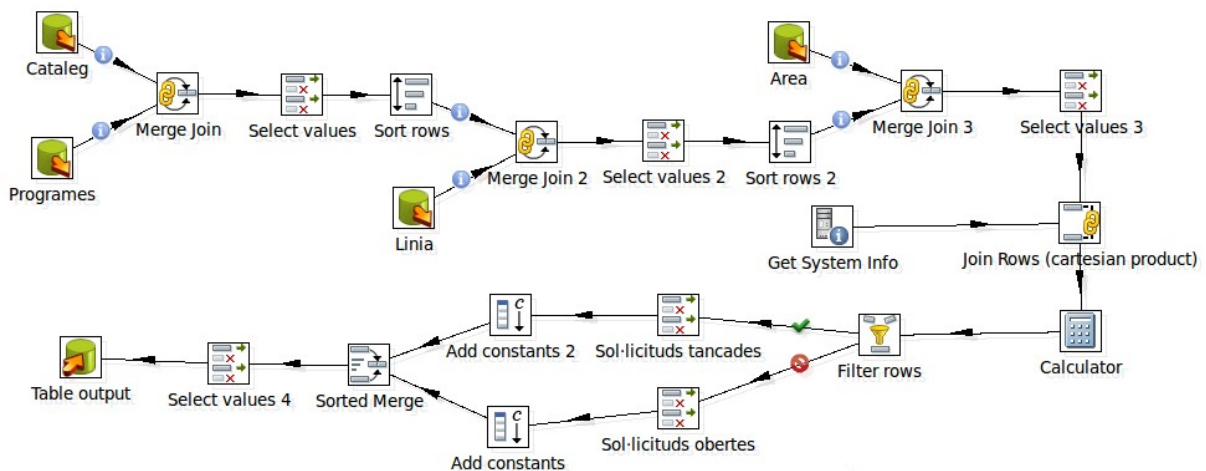
Per a carregar els equipaments, ha fet falta relacionar-los amb l'ajuntament al qual pertanyen. La manera més adient és mitjançant una sentència *Join* que es modela amb la icona *Merge Join*. Posteriorment, es seleccionen les columnes i es carreguen a la taula equipament. Es poden consultar els passos seguits en el fitxer *transfEquipaments1.ktr* de la carpeta *Spoon* del ZIP del producte.



Il·lustr. 26: Procés ETL senzill dels equipaments

4.1.5. ETL Programes

Per a dur a terme aquest procés d'ETL, ha estat necessari relacionar quatre taules d'origen: Cataleg, Programa, Àrea i Línia. No es tornaran a explicar les transformacions ja descrites en processos ETL anteriors, com són el *Select values* o *Merge Join*. En canvi, cal explicar les transformacions noves que apareixen a continuació.



Il·lustr. 27: Procés ETL senzill de càrrega de Programes al DW

Sort rows serveix per aplicar criteris d'ordenació a les dades sortints. És necessari ordenar-les per a poder concatenar diferents operacions de *Merge join*, tal com alerta la pròpia aplicació si no es fa.

Cal parar atenció a la icona *Get System Info* i *Calculator*, ja que la primera serveix per obtenir la data del sistema i la segona per calcular la diferència entre la data de finalització d'un programa que consta a la taula d'origen i la data del sistema obtinguda en el pas anterior. Si la diferència és negativa, és que el període de sol·licitud està tancat, en cas contrari, encara està obert i per tant, es poden rebre noves sol·licituds. Aquest càlcul és necessari per poder carregar un valor al camp *obert* de la taula *programa*. Pel que fa a aquest valor, s'utilitza la icona *Add constants* dins la qual es defineix una constant textual que val 'S' per al cas en que el període de sol·licitud estigui obert i 'N' per al cas contrari.

Es poden consultar els passos següents en el fitxer *transfProgrames1.ktr* de la carpeta *Spoon* del ZIP entregat.

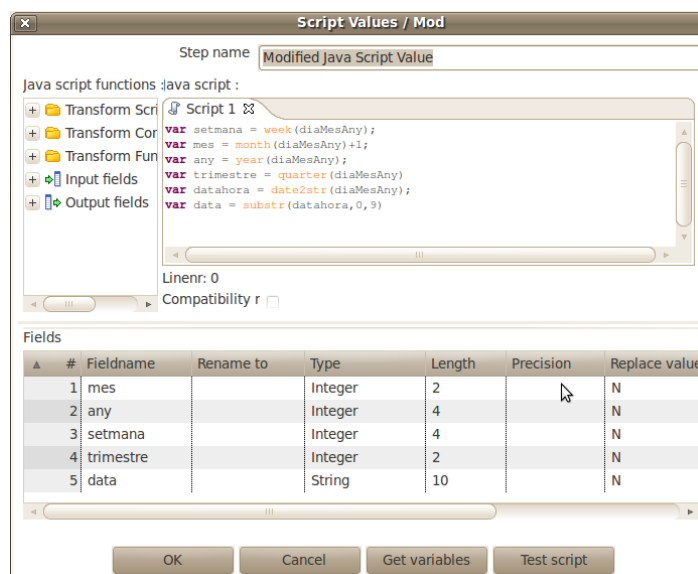
4.1.6. ETL Sol·licituds atòmiques

Novament, es procedeix a explicar noves transformacions que no han aparegut abans. Pel que fa a les sol·licituds atòmiques, presenten la particularitat d'haver de carregar prèviament la taula *temps* del magatzem abans de poder inserir registres a la taula del fet, ja que es necessari un camp *idTemps*.

Per tal d'obtenir aquest camp *idTemps*, s'utilitza una icona de tipus *Table input* que realitza la consulta SQL següent:

```
SELECT distinct(dataSolicitud) as diaMesAny FROM SollicitutServei ORDER BY dataSolicitud
```

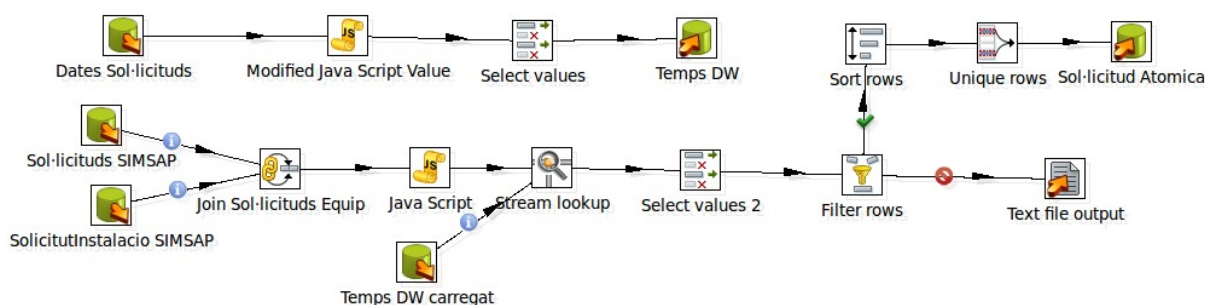
Un cop obtinguda la relació de dates de sol·licituds, es pot extreure el dia, mes, any, setmana i trimestre mitjançant javascript, tal com es pot observar a la figura de sota:



Il·lust. 28: extracció i càrrega de la dimensió Temps

Un cop carregada la dimensió de temps, es pot executar una segona transformació que consisteix en relacionar mitjançant *Join* les sol·licituds amb els equipaments que s'hi han inclòs. Amb javascript es transforma la data de la sol·licitud en un *String* i mitjançant *Stream lookup* es compara aquesta data amb les dates carregades a la dimensió de temps anterior. D'aquesta manera, es pot relacionar la data de cada sol·licitud amb el seu *idTemps* corresponent. Es seleccionen els camps, es filtren els errors, i finalment, es carreguen a la taula Sol·licitud Atomica.

Es poden consultar els passos següents en el fitxer *transfSollicituds1.ktr* de la carpeta *Spoon* del ZIP entregat.



Il·lustr. 29: Procés ETL sencer de càrrega de les dimensions Temps i del fet Sol·licitud Atomica

4.1.7. ETL Resum sol·licituds

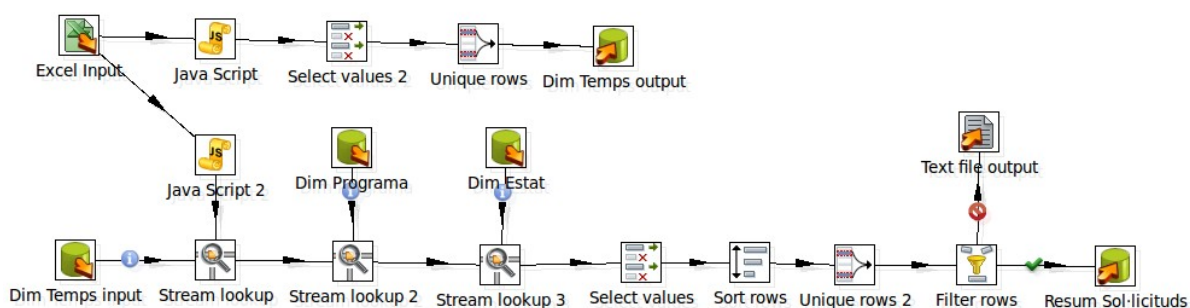
Aquest cas és lleugerament diferent dels altres, ja que es vol obtenir una “foto” setmanal resumida de les sol·licituds del darrer catàleg (any 2011), agrupades per programa i estat. Com que el magatzem de dades no estava implementat, cada divendres des del mes de gener 2011 s'ha anat obtenint els resultats mitjançant la consulta SQL següent que es bolcava en un full de càlcul:

```
select pr.codiPrograma, st.statusAj, count(*) as num from SollicitutServei ss
inner join StatusSol st on ss.idStatusSol=st.idStatusSol
inner join Programa pr on pr.idPrograma=ss.idPrograma and ss.idAjuntament<244
group by ss.idPrograma, st.statusAj
order by pr.codiPrograma, pr.idCataleg, st.idStatusSol desc
```

Per tant, en aquesta transformació es parteix d'un full d'Excel.

Primer es carrega la dimensió *Temps* del magatzem extraient les dates del full de càlcul, desant trimestre, mes i any, i finalment, eliminant les dates duplicades.

Un cop carregats els *idTemps* a la taula corresponent, es cerca aquest valor comparant-lo amb la data del full de càlcul i es passa el resultat al pas següent. A continuació es cerca *l'idPrograma* dins la dimensió *Programa*, on el camp *codiPrograma* coincideixi amb el que hi ha al full de càlcul. Tot seguit es fa el mateix amb la dimensió *Estat*. Un cop tenim els tres “id” necessaris dins el flux de dades (*idTemps*, *idPrograma* i *idEstat*), s'ordena el resultat, s'eliminen duplicats si n'hi han i es filtren els registres on *idPrograma* es nul. Per acabar, es carrega la taula *resumsollicituds* de la base de dades *magatzem*.



Il·lustr. 30: Procés ETL sencera de càrrega del fet "Resum Sol·licituds"

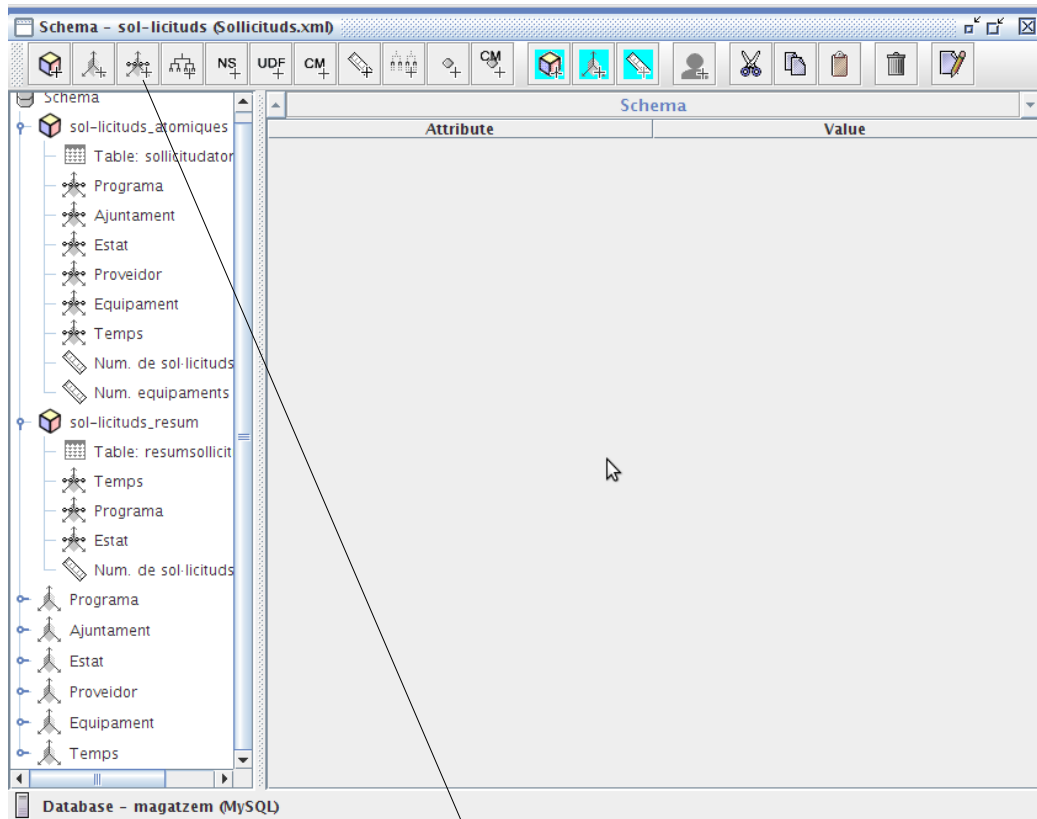
S'ha decidit eliminar els camps *idAjuntament* i *idEquipament* del Fet que ens ocupa, ja que no aportaven res de significatiu pel que fa al resum de sol·licituds.

Es poden consultar els passos seguits en el fitxer *transfResumSollicituds1.ktr* de la carpeta *Spoon* del ZIP entregat.

4.2. Modelatge de l'estrella de Sol·licituds amb *Mondrian*

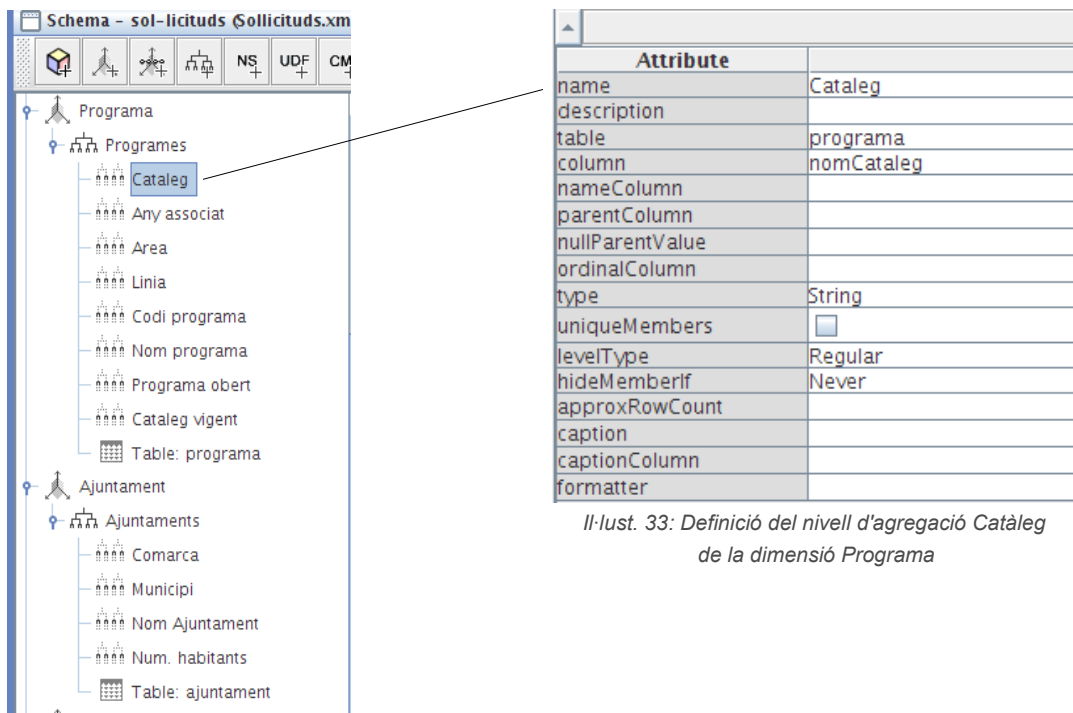
L'eina *Schema Workbench*, també coneguda com a *Mondrian*, permet definir de manera gràfica les estrelles d'un magatzem de dades per a estirar-les posteriorment des del BI Server i analitzar cubs.

Primer s'arrossega la icona del cub i s'omplen els valors per a cada atribut. Tot seguit, es poden crear les dimensions i mesures arrossegant les icones corresponents i definint els valors.



Il·lustr. 31: Creació d'una estrella amb Schema Workbench - Mondrian

Cada dimensió pot ser compartida amb altres cubs, per això es creen penjant de l'esquema principal. Un cop creades, s'arrossega una icona de "Dimension usage" per a cada dimensió del cub, alhora que es configuren els nivells d'agregació i la taula que li correspon, tal com es pot veure a les figures 33 i 34.



Il·lustr. 33: Definició del nivell d'agregació Catàleg de la dimensió Programa

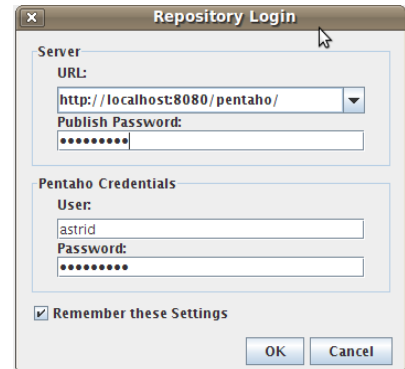
Il·lustr. 32: Dimensions i agregacions

En definir les mesures, simplement es tria d'un desplegable si la operació a efectuar es *sum*, *count*, *avg*, etc.

Finalment, cal publicar el model al BI Server anant a *File->Publish*.

L'usuari que publiqui esquemes ha d'estar creat prèviament al *P.Admin Console* i tenir els privilegis adients. Després de publicar-lo, cal anar a la interfície BI Server i refrescar el *Repository cache* i les *metadades* (Tools -> Refresh -> Repository...)

És convenient desar el model en un fitxer XML. En aquest cas, s'adjunta el model resultant al fitxer *Sollicituds.xml* de la carpeta *Mondrian* del ZIP entregat.



Il·lustr. 34: Finestra de publicació d'una estrella al BI Server

4.3.Vistes d'anàlisi de cubs ROLAP

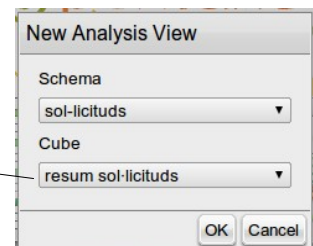
Per a analitzar les dades de l'estrella de sol·licituds, cal autenticar-se al Pentaho BI Server en el localhost, és a dir, a <http://localhost:8080/pentaho> o bé en l'entorn de proves temporal <http://80.32.31.86:8080/pentaho>. Es recomana utilitzar preferiblement el Google Chrome com a navegador, ja que el Firefox dona algun problema en entorns d'Ubuntu.

Apareix un menú amb quatre grans apartats:



Il·lustr. 35: Menú principal del BI Server

Cal escollir *New Analysis View* i triar *sol·licituds* en el desplegable *Schema*. En el segon desplegable on s'especifica el cub, podem triar qualsevol dels dos cubs creats: *resum-sol·licituds* o *sol·licituds* (atòmiques).



Il·lustr. 36: tria d'esquema i cub

Tot seguit, apareix la vista d'anàlisi següent:

Temps	Programes	Estats	Measures
All Temps	All Programa.Programess	All Estat.Estatss	Num. de sol·licituds 15.130

Slicer:

Il·lustr. 37: vista d'anàlisi

Per exemple, si es fa clic sobre el botó + al costat de *All temps*, només apareixerà l'any 2011 ja que el resum de sol·licituds setmanal només es disposa per a l'any en curs.

Si es descendeix un esglaó més en la jerarquia d'agregació, es passa a nivell de mesos, i consecutivament a nivell setmanal.

Es pot constatar que quan s'arriba a nivell setmanal, el nombre de sol·licituds es correspon amb la realitat.

Suposant que ara es volgués saber en quin estat estaven les sol·licituds la segona setmana de gener, només caldria clicar el botó + al costat de *All Estats* en la línia corresponent.

D'aquesta manera, es sabria ràpidament que al finalitzar la segona setmana de gener hi havien 109 sol·licituds iniciades, 64 de les quals estaven en esborrany, 21 pendents de registrar, 9 pendents de validar i 15 registrades.

Si a més es desplegués la jerarquia de programes d'aquesta setmana, es podria saber que, en la línia de Sanitat ambiental, s'havien iniciat 45 sol·licituds, 26 de les quals corresponien al programa Pt01 i 19 al Pt02.

També es podria saber per a cada programa, en quin estat estaven les sol·licituds.

Però aquí no acaben els avantatges d'aquestes vistes, ja que es poden aplicar filtres i triar els atributs estictament necessaris.

El número de sol·licituds és molt elevat perquè es sumen els totals de sol·licituds de cada data capturada.

Les tres primeres columnes corresponen a les dimensions relacionades. El botó + al costat de cada dimensió, permet navegar pels diferents nivells d'agregació i estudiar els resultats.

Per exemple, si es fa clic sobre el

Temps	Programes	Estats	Measures
All Temps	All Programa.Programess	All Estat.Estatss	Num. de sol·licituds 15.130
2011	All Programa.Programess	All Estat.Estatss	15.130
1	All Programa.Programess	All Estat.Estatss	528
2	All Programa.Programess	All Estat.Estatss	109
3	All Programa.Programess	All Estat.Estatss	174
4	All Programa.Programess	All Estat.Estatss	245
2	All Programa.Programess	All Estat.Estatss	2.289
3	All Programa.Programess	All Estat.Estatss	3.725
4	All Programa.Programess	All Estat.Estatss	4.372
5	All Programa.Programess	All Estat.Estatss	4.216

Il·lustr. 38: Jerarquia d'agregació de la dimensió temps en una vista d'anàlisi

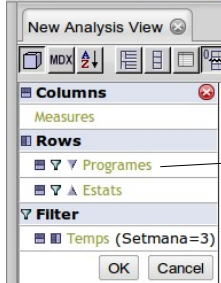
Temps	Programes	Estats	Measures
1	All Programa.Programess	All Estat.Estatss	528
2	All Programa.Programess	All Estat.Estatss	109
		Esborrany	64
		Pendent registrar	21
		Pendent validar - supervisor	9
		Registrada	15

Il·lustr. 39: desplegament per estat de sol·licituds

Temps	Programes	Estats	Measures
2	All Programa.Programess	All Estat.Estatss	109
	Sanitat ambiental	All Estat.Estatss	45
		Esborrany	25
		Pendent registrar	8
		Pendent validar - supervisor	3
		Registrada	9
	Pt01	All Estat.Estatss	26
		Esborrany	14
		Pendent registrar	5
		Pendent validar - supervisor	1
		Registrada	6
	Pt02	All Estat.Estatss	19
		Esborrany	11
		Pendent registrar	3
		Pendent validar - supervisor	2
		Registrada	3
	Seguretat alimentària	All Estat.Estatss	2
		Esborrany	1
		Pendent registrar	1
		Pendent validar - supervisor	0
		Registrada	0
	Pt13	All Estat.Estatss	2
		Esborrany	1
		Pendent registrar	1
		Pendent validar - supervisor	0
		Registrada	0

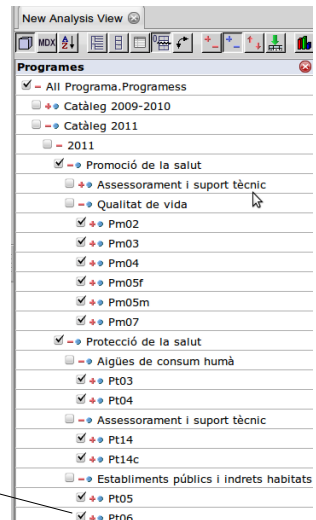
Il·lustr. 40: estudi de sol·licituds per estat i programa segons setmana escollida

Si cliquem la icona *Open OLAP Navigator*, es poden definir filtres com el de la figura de l'esquerra, en el qual definim quines col·lumnes volem que apareguin (en aquest cas només Programes i Estats). A més es filtra per Temps, concretament per la setmana 3, tot i que aquesta columna no apareixerà a la vista d'anàlisi per que està a sota de *Filter*.



Il·lustr. 41: OLAP Navigator

Pel que fa a Programes, no ens interessa tota la jerarquia d'agregació sinó només l'àrea i el codi del programa. Per a seleccionar només aquests atributs, es clica sobre Programes i es marquen les caselles d'acord amb el que es vol.



Il·lustr. 42: selecció d'atributs

Programes	Estats	Measures
• All Programa.Programess	• Num. de sol·licituds	
+	Esborrany	82
	Pendent registrar	35
	Pendent validar - supervisor	19
	Registrada	38
+	Esborrany	10
	Pendent registrar	1
	Pendent validar - supervisor	2
	Registrada	4
+	Esborrany	10
	Pendent registrar	1
	Pendent validar - supervisor	2
	Registrada	4
+	Esborrany	72
	Pendent registrar	34
	Pendent validar - supervisor	17
	Registrada	34
+	Esborrany	10
	Pendent registrar	6
	Pendent validar - supervisor	1
	Registrada	3
+	Esborrany	23
	Pendent registrar	8
	Pendent validar - supervisor	4
	Registrada	7

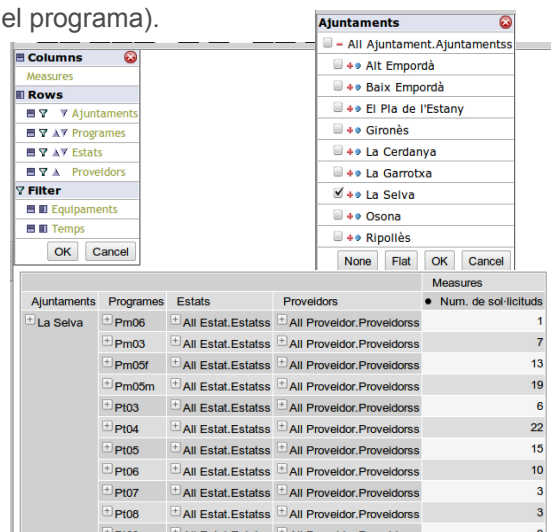
Il·lustr. 43: exemple d'anàlisi de resultats aplicant els filtres descrits

D'aquesta manera, es poden obtenir totes les combinacions possibles descrites en el document de disseny del magatzem de dades.

Suposant que es vol saber quins programes han sol·licitat els ajuntaments de la comarca de la Selva durant el 2011, caldria primer de tot escollir una nova vista d'anàlisi basada en el cub *sol·licituds*.

Tot seguit, anant a l'*OLAP navigator*, caldria posar a sota de *Filter* els atributs que no ens interessa estudiar (en aquest cas, Equipaments i Temps). Clicant a *Rows-> Ajuntaments* es pot filtrar per la comarca desitjada. Finalment, clicant a *Rows-> Programes*, es marquen els atributs que volem que es mostrin (només el codi del programa).

Si en aquest punt es volgués veure els ajuntaments, només caldria clicar el botó + al costat del nom de la comarca, i si a més es volgués saber en quin estat estan aquestes sol·licituds, caldria clicar el + al costat de *All Estats*.



Il·lustr. 44: exemple d'anàlisi basada en l'estrella de sol·licituds atòmiques

4.4. Llistats i informes

4.4.1. Dissenyant llistats amb BI Server

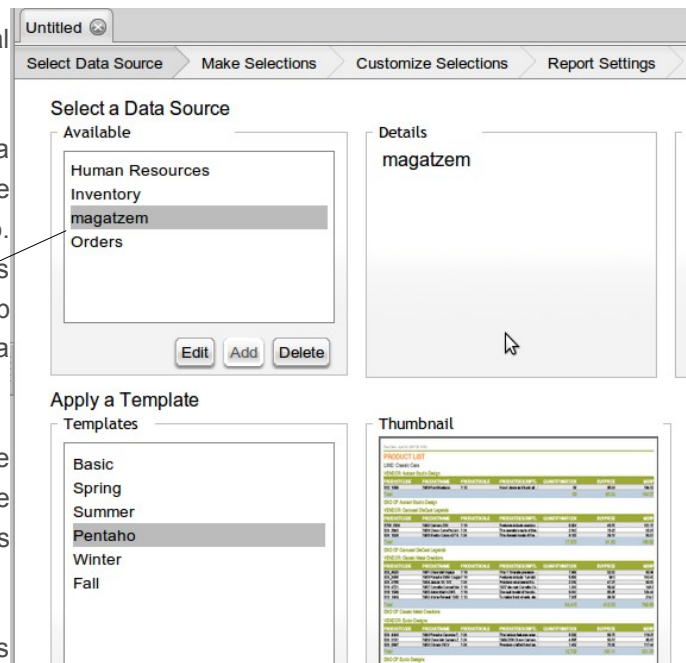
El Pentaho BI Server porta integrat un dissenyador senzill de llistats tot i que es pot utilitzar també el Pentaho Report Designer descrit en el punt següent, el qual permet construir llistats de manera molt gràfica.

Per a crear un report amb el BI Server, cal escollir *New Report* del menú principal.

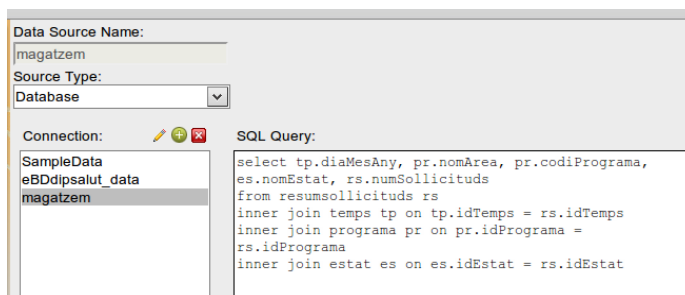
Tot seguit apareix la vista de la figura de la dreta. Apareixen per defecte les bases de dades d'exemple que venen amb l'aplicació. Cal definir una connexió a la base de dades *magatzem* clicant el botó *Add*. Un cop configurada, ja apareixerà a la llista i la podrem seleccionar.

Quan es configura la connexió a la base de dades, cal indicar una consulta SQL que retornarà els camps que s'inclouran en els llistats.

Per al cas del llistat resum de sol·licituds agrupades per data, programa i estat, la consulta és la següent:



Il·lustr. 45: selecció d'un origen de dades per al llistat

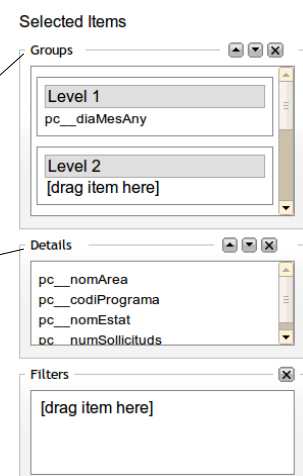


Il·lustr. 46: consulta que retorna camps a incloure en un report de BI Server

En el pas següent, cal especificar quins camps agruparan seccions i quins seran camps de detall de llistat. Això es fa simplement arrossegant els camps de la figura de sota dins el requadre *Groups* i *details* de la figura 48.

Només cal anar avançant i triant les diferents opcions que ens presenta l'assistent d'informes fins a la obtenció del llistat corresponent. Posteriorment, es pot desar l'informe o imprimir-lo en PDF.

Es pot previsualitzar el resultat de la consulta SQL i en cas de que sigui correcte, continuar amb la definició del llistat.

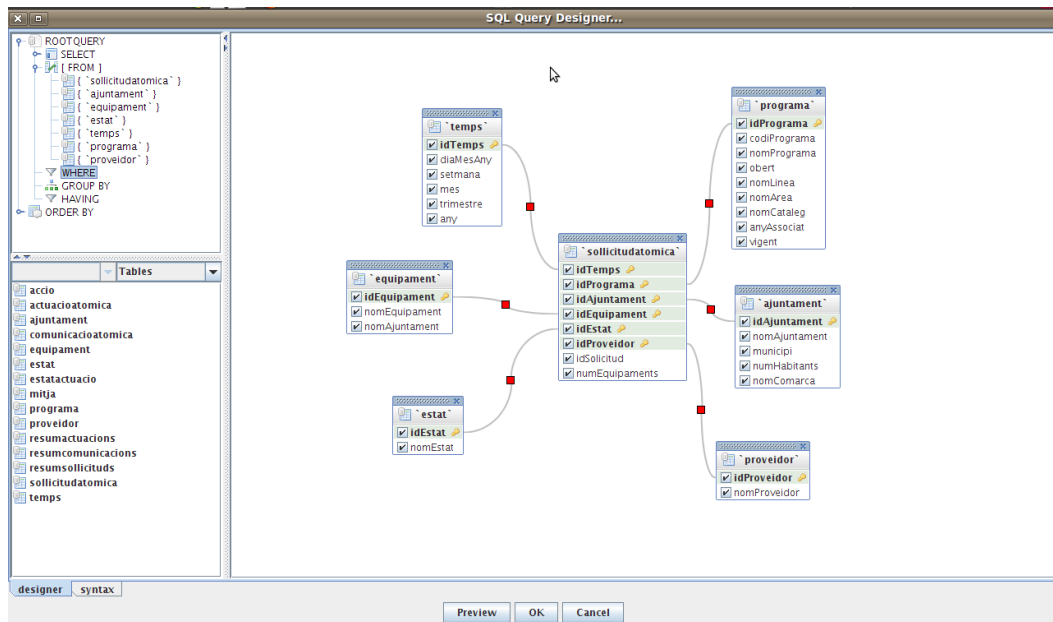


Il·lustr. 47: nivells de detall i agrupació en els llistats

Els llistats de proves realitzats amb aquesta eina es troben a la carpeta *Reports* del ZIP del producte.

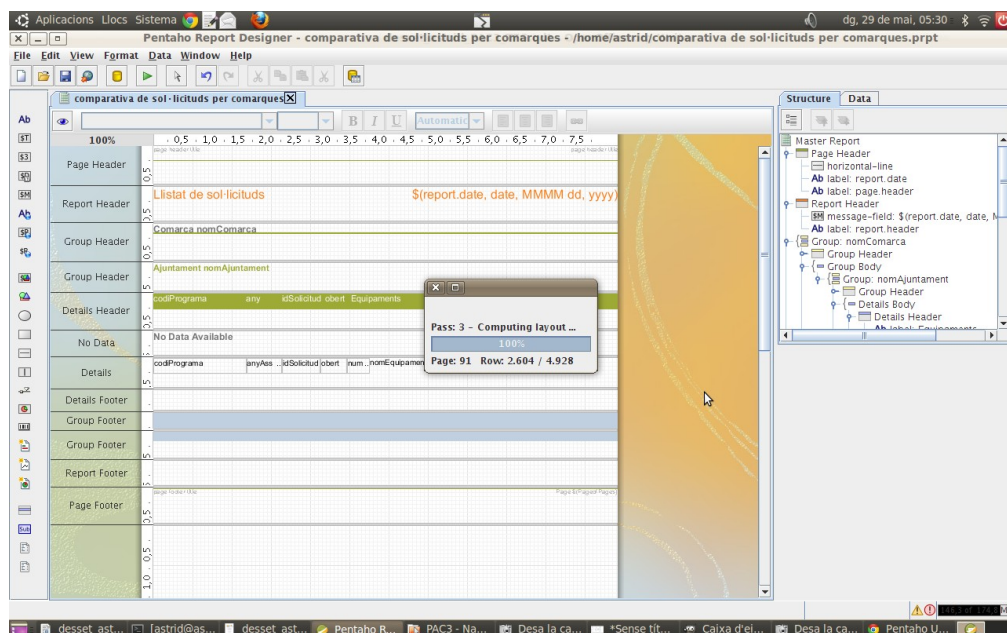
4.4.2. Pentaho Report Designer

Aquesta eina complementa i amplia la utilitat de reports del BI Server. Permet dissenyar llistats relacionant les taules de forma gràfica, especificant els camps a seleccionar i com s'han d'agrupar i/o ordenar.



Il·lustr. 48: captura de pantalla del Pentaho Report Designer

També permet dissenyar de forma gràfica les capçaleres, columnes, grups de detall, etc.



Il·lustr. 49: captura de pantalla de l'entorn gràfic de disseny de llistats

4.4.3. Captures de pantalla dels llistats obtinguts

Tots els llistats presentats a continuació es corresponen amb els que es van definir a l'apartat 3.7.1 i es poden trobar en format PDF dins del fitxer ZIP del producte lliurat, a la carpeta *Reports*.

page header title

page header title

Llistat de totes les sol·licituds d'un dia, setmana, mes, trimestre, any

Any 2.010

Trimestre 1

Mes 2

Setmana 7

Dia 18/02/10

Codi Programa	Municipi	id.Sol·licitudEstat
Pt01	La Bisbal d'Empordà	203 Finalitzat

Trimestre 2

Mes 6

Setmana 26

Dia 28/06/10

Codi Programa	Municipi	id.Sol·licitudEstat
Pt02	La Bisbal d'Empordà	204 Finalitzat

Dia 30/06/10

Codi Programa	Municipi	id.Sol·licitudEstat
Pt01	Llançà	211 Finalitzat
Pt01	Llagostera	221 Finalitzat
Pt01	Camprodon	213 Finalitzat
Pt09	Torroella de Montgrí	222 Finalitzat

Il·lustració 50: Llistat de totes les sol·licituds d'un dia, setmana, mes, trimestre, any

Llistat de totes les sol·licituds agrupades per estat

100%

page header title

page header title

Llistat de totes les sol·licituds agrupades per estat

Estat de sol·licituds : Esborrany

Municipi	codi Programa	id Sol.Data sol.	nº equipaments
Vilamacolum	Pm03	1.628 15/12/10	0
Sant Pere Pescador	Pt02	1.709 11/01/11	0
La Bisbal d'Empordà	Pt13	1.731 12/01/11	0
La Bisbal d'Empordà	Pt14	1.732 12/01/11	0
Avinyonet de Puigventós	Pm02	1.903 26/01/11	0
Sarrià de Ter	Pm06	1.920 27/01/11	0
Pontós	Pt04	1.932 28/01/11	2
Sant Pere Pescador	Pm03	1.940 31/01/11	0
Ullastret	Pt03	1.976 02/02/11	0
Ullastret	Pt14	1.977 02/02/11	0
Esponellà	Pt13	1.980 02/02/11	0
Breda	Pt09	2.001 03/02/11	5
Breda	Pm06	2.002 03/02/11	0
Llançà	Pt13	2.123 09/02/11	0
Guis de Cerdanya	Pt04	2.178 10/02/11	2
Castelló d'Empúries	Pm02	2.202 10/02/11	0
La Pera	Pm06	2.275 14/02/11	0
Garrigàs	Pm06	2.296 15/02/11	0
Alp	Pt03	2.301 15/02/11	5
Alp	Pt14	2.430 18/02/11	0
Garrigòles	Pt04	2.443 18/02/11	0
Fontcoberta	Pm03	2.486 23/02/11	0
Les Lloses	Pt03	2.507 25/02/11	1

Il·lustració 51: Llistat de totes les sol·licituds agrupades per estat

Selecció d'estat: **Denegat**

Auto-Update on selection

Llistat sol·licituds que es troben en un estat concret passat per paràmetre

Estat de sol·licituds : Denegat

Municipi	codi Programa	id Sol>Data sol.	nº equipaments
Celrà	Pm07	1.52608/11/10	0
Urús	Pm07	1.55711/11/10	0
Cantallops	Pm07	1.55811/11/10	0
Vilablareix	P109	1.94531/01/11	0
Banyoles	P108	2.88411/04/11	0

Il·lustració 52: Llistat de sol·licituds que es troben en un estat concret passat per paràmetre

Llistat de totes les sol·licituds agrupades per proveïdor al qual s'han assignat

Proveïdor CECAM

codi Programa P101

Ajuntament	id Sol·licitud	nº equip.
Ajuntament d'Agullana	2.437	1
Ajuntament d'Aiguaviva	393	1
Ajuntament d'Aiguaviva	2.148	0
Ajuntament d'Alp	2.286	1
Ajuntament d'Amer	2.036	1
Ajuntament d'Anglès	280	4
Ajuntament d'Anglès	1.639	1
Ajuntament d'Arbúcies	300	4
Ajuntament d'Arbúcies	1.617	0
Ajuntament d'Hostalric	438	1
Ajuntament d'Hostalric	1.860	0
Ajuntament d'Ogassa	343	1
Ajuntament d'Ogassa	1.778	0
Ajuntament d'Ullastret	1.874	0
Ajuntament de Banyoles	306	6
Ajuntament de Banyoles	2.251	3
Ajuntament de Bàscara	2.304	0
Ajuntament de Begur	291	1
Ajuntament de Begur	2.329	0
Ajuntament de Bescanó	2.403	1
Ajuntament de Blanes	285	9

Il·lustració 53: Llistat de totes les sol·licituds agrupades per proveïdor al qual s'han assignat

Selecció de proveïdor: **UTE Prodaisa - Aigües de Girona**

Auto-Update on selection

Llistat sol·licituds assignades a un proveïdor concret passat per paràmetre

Proveïdor UTE Prodaisa - Aigües de Girona

codi Programa P104

Ajuntament	id Sol·licitud	nº equip.
Ajuntament d'Agullana	866	2
Ajuntament d'Aiguaviva	759	2
Ajuntament d'Alp	789	2
Ajuntament d'Anglès	862	2

Il·lustració 54: Llistat sol·licituds assignades a un proveïdor concret passat per paràmetre

page header (1) (1)

Llistat de totes les sol·licituds agrupades per ajuntament i comarca

Comarca Alt Empordà

Ajuntament d'Agullana

codi Programa	id Sol·licitud	Data sol·licitud	Estat
Pm05f	2.174	10/02/11	En procés
Pm06	2.940	28/04/11	Esborrany
PI01	2.437	18/02/11	En procés
PI02	2.821	28/03/11	En procés
PI03	361	15/07/10	Finalitzat
PI03	2.168	10/02/11	En procés
PI04	866	27/09/10	Finalitzat
PI04	2.169	10/02/11	Assignable
PI05	2.864	06/04/11	Assignada
PI06	650	09/09/10	Finalitzat
PI06	2.929	14/04/11	Assignada
PI09	326	08/07/10	Finalitzat
PI09	2.171	10/02/11	En procés

Ajuntament d'Albanyà

codi Programa	id Sol·licitud	Data sol·licitud	Estat
Pm05f	2.471	22/02/11	Assignada
PI03	372	19/07/10	Finalitzat
PI03	1.884	25/01/11	En procés

Il·lustració 55: Llistat de totes les sol·licituds agrupades per ajuntament i comarca

Selecció una Comarca **La Selva**

Auto-Update on selection

page header (1) (1)

Llistat de sol·licituds per ajuntament i comarca passats per parametre

Comarca La Selva

Ajuntament d'Amer

codi Programa	id Sol·licitud	Data sol·licitud	Estat
Pm05m	2.698	11/03/11	Assignada
Pm07	1.298	18/10/10	En procés
PI01	2.036	07/02/11	En procés
PI04	2.052	07/02/11	Assignable
PI10	838	27/09/10	En procés

Ajuntament d'Anglès

codi Programa	id Sol·licitud	Data sol·licitud	Estat
Pm03	1.861	21/01/11	Finalitzat
Pm05m	2.569	02/03/11	Assignada
PI01	280	05/07/10	Finalitzat
PI01	1.639	16/12/10	En procés
PI02	283	05/07/10	Finalitzat
PI02	1.640	16/12/10	En procés
PI04	862	27/09/10	Finalitzat

Il·lustració 56: Llistat de sol·licituds d'ajuntament i comarca passats per paràmetre

page header 111a page header 111a

Listat de totes les sol·licituds agrupades per catàleg, àrea, programa i línia

Catàleg Catàleg 2009-2010

Area Promoció de la salut

Línea Alimentació i activitat física

Programa Parcs urbans de salut

Ajuntament	id.Sol. Estat
Ajuntament d'Aiguaviva	1.079 En procés
Ajuntament d'Alp	1.246 En procés
Ajuntament d'Arbúcies	1.085 En procés
Ajuntament d'Argelaguer	1.272 En procés
Ajuntament d'El Far d'Empordà	1.027 En procés
Ajuntament d'El Port de la Selva	1.026 En procés
Ajuntament d'Espolla	1.263 En procés
Ajuntament d'Espenyllà	1.023 En procés
Ajuntament d'Hostalric	1.020 En procés
Ajuntament d'Ogassa	1.142 En procés
Ajuntament d'Olot	1.135 En procés
Ajuntament d'Ullà	1.018 En procés
Ajuntament de Banyoles	1.267 En procés
Ajuntament de Besalú	1.050 En procés
Ajuntament de Bescanó	1.028 En procés
Ajuntament de Biure	1.035 En procés
Ajuntament de Blanes	1.080 En procés

Il·lustració 57: Llistat de totes les sol·licituds agrupades per catàleg, àrea, línia i programa

Selecció un equipament **Ajuntament d'Anglès - Piscina municipal**

Auto-Update on selection

page header 111a page header 111a

Sol·licituds en les quals s'ha inclòs un equipament concret

Municipi Anglès

Nom Equipament Piscina municipal

codi Programa	id Sol·licitudData	Estat
PI01	1.639/16/12/10	En procés
PI05	892/28/09/10	Finalitzat
PI01	280/05/07/10	Finalitzat

Il·lustració 58: Sol·licituds en les quals s'ha inclòs un equipament concret passat per paràmetre

page header 111a page header 111a

Nombre de sol·licituds per setmana, mes i any, agrupades per programa i estat

any 2,011
trimestre1
mes1
setmana2

codi Programa	nomEstat	numSol·licituds
Pm03	Esborany	7
Pm03	Pendent registrar	1
Pm03	Pendent validar - supervisor	2
Pm03	Registrada	2
PI01	Esborany	14
PI01	Pendent registrar	5
PI01	Pendent validar - supervisor	1
PI01	Registrada	6
PI02	Esborany	11
PI02	Pendent registrar	3
PI02	Pendent validar - supervisor	2
PI02	Registrada	3
PI03	Esborany	10
PI03	Pendent registrar	1
PI03	Pendent validar - supervisor	1
PI03	Registrada	1
PI04	Esborany	16
PI04	Pendent registrar	6
PI04	Pendent validar - supervisor	1
PI04	Registrada	0
PI09	Esborany	4

Il·lustració 59: Resum de sol·licituds per setmana, mes i any, agrupades per programa i estat

page header 0/02 page header 0/02

Listat de totes les sol·licituds de la comarca de 'La Selva' fetes durant l'últim mes, agrupades per àrea, programa i línia.

Comarca de La Selva

Ajuntament d'Hostalric

nomÀrea	nomLínia	codiPrograma	idSol·licitud
Protecció de la salut	Establiments públics i indrets habitats	Pi06	2.899

Ajuntament de Blanes

nomÀrea	nomLínia	codiPrograma	idSol·licitud
Protecció de la salut	Establiments públics i indrets habitats	Pi07	2.941

Ajuntament de Caldes de Malavella

nomÀrea	nomLínia	codiPrograma	idSol·licitud
Protecció de la salut	Establiments públics i indrets habitats	Pi05	2.914
Protecció de la salut	Establiments públics i indrets habitats	Pi06	2.910

Ajuntament de La Cellera de Ter

nomÀrea	nomLínia	codiPrograma	idSol·licitud
Protecció de la salut	Establiments públics i indrets habitats	Pi05	2.902

Il·lustració 60: Llistat de totes les sol·licituds de la comarca de 'La Selva' fetes durant l'últim mes, agrupades per àrea, programa i línia

Sol·licituds en estat Assignada o En Procés del darrer trimestre 2010

100%

Proveïdor Itinerant

Estat Assignada

Municipi	idSol·licitud	codi Programa
Alp	2.316	Pm03
Camprodon	1.665	Pm03
L'Armentera	2.232	Pm03
Llívia	2.948	Pm03
Peralada	1.725	Pm03
Sant Hilari Sacalm	930	Pm03
Vilamalla	2.946	Pm03
Vilobi d'Onyar	2.334	Pm03

Estat En procés

Municipi	idSol·licitud	codi Programa
Viladamat	1.644	Pm03
Vilafant	1.505	Pm03

Proveïdor Labaqua

Estat Assignada

Municipi	idSol·licitud	codi Programa
Cervià	2.442	Pm03

Il·lustració 61: Sol·licituds en estat Assignada o En Procés que s'han assignat a un proveïdor concret durant el darrer trimestre 2010

5. Conclusions

El TFG que s'ha exposat en els capítols previs és el resultat de l'estudi de diverses matèries de l'Enginyeria en Informàtica i d'aplicar les competències adquirides durant el transcurs de la carrera. En conseqüència, aquest treball es basa en dur a la pràctica els continguts de diferents assignatures, aplicant una reflexió a cada pas i després de cada avenç, per tal de saber què es fa, com i perquè.

També és un treball d'investigació, d'autoaprenentatge i d'experimentació amb un programari de codi obert que no coneixia amb anterioritat, i que ha resultat ser molt potent i flexible a l'hora de configurar i d'utilitzar. La *suite* de *Pentaho* és un projecte madur, amb una extensa comunitat de gent que hi participa i contribueix al seu desenvolupament. També permet definir connexions amb una gran varietat de sistemes gestors de bases de dades estàndards (Oracle, Mysql, SQL Server, Access, Postgre, etc.), o bé llegir i importar les dades de diversos formats de fitxers. A més, la seva arquitectura és extensible, doncs ofereix un connector específic per a SAP i és força fàcil desenvolupar connectors a mida.

La part que he trobat més interessant, per la seva complexitat i alhora interfície sumament intuïtiva, ha estat la implementació dels processos ETL amb *Kettle-Spoon*. Hi veig un potencial enorme en aquesta eina, no només per als magatzems de dades, sinó per qualsevol tipus d'extracció o transformació de dades, i fins i tot per a integracions entre aplicacions heterogènies.

A més, el cost d'adquirir una eina ETL comercial és elevat al principi, però representa un estalvi a llarg termini, doncs simplifica el manteniment i les modificacions posteriors a la posada en marxa. En contrapartida, el cost de desenvolupar una eina ETL a mida va creixent a mesura que els requeriments canvien i l'organisme creix, doncs l'esforç augmenta en relació a aquest creixement. Una eina de codi obert com *Kettle*, combina el millor de les dues opcions, doncs no té costos elevats d'adquisició, i redueix sensiblement els costos de manteniment de millores o canvis.

També m'ha agradat experimentar amb les vistes d'anàlisi i els cubs publicats al *Pentaho BI Server*, desplegant atributs, aplicant filtres i utilitzant les jerarquies d'agregació.

La motivació personal ha estat doble ja que, per una banda es tractava d'implementar un magatzem corporatiu real i directament aplicable a l'organisme per al qual treballa, i per l'altra, d'iniciar-me en el fascinant i desconegut món dels magatzems de dades i del *business intelligence*.

Per concloure, cal remarcar que s'ha pogut implementar un exemple complet d'una estrella, concretament la de Sol·licituds, incloent el procés detallat d'ETL, el modelatge de cubs ROLAP, i finalment, els llistats esperats i vistes d'anàlisi. No s'han implementat les estrelles d'Accions ni de Comunicacions per raons de temps i perquè realment, són molt similars, sinó quasi bé idèntics als esquemes i processos descrits en el capítol 4.

6. Línies de desenvolupament futur

Sense cap mena de dubte, queda molt per fer encara per a completar la factoria d'informació corporativa de Dipsalut. Caldrà desenvolupar l'aplicació que substituirà els fulls de càlcul i també analitzar quins llistats o vistes d'anàlisi es volen obtenir a partir d'aquest nou sistema d'informació, però ja es tracta d'un altre projecte dins del marc de la càrrega de dades al magatzem corporatiu.

S'ha parlat de la conveniència d'implementar un magatzem de dades operacional i de possibles magatzems de dades departamentals. Totes aquestes línies de desenvolupament futur s'hauran d'iniciar en breu, després de la implementació de les dues estrelles restants d'Accions i Comunicacions.

Les actualitzacions diàries del magatzem corporatiu mitjançant la creació de tasques programades anomenades *Jobs* emprant l'eina *Kettle-Pan*, han quedat pendents de programar. També cal cercar solucions de programari per al manteniment de rèpliques de taules en temps real, per poder mantenir actualitzat el magatzem operacional basat en el SIMSAP.

D'altra banda, en el transcurs de la investigació i l'autoaprenentatge de les eines de *Pentaho*, he vist per sobre una eina que m'ha semblat molt útil però en la qual no he pogut profunditzar.

Es tracta dels *dashboards* o indicadors, també coneguts com a **Quadres de comandament**, que permeten saber de manera gràfica i intuïtiva els esdeveniments clau que afecten a l'organisme i l'estat de les diferents mesures que li interessen, per tal de recolzar la presa de decisions. Dit d'una altra manera, els indicadors permeten traduir l'estratègia de l'organisme en un conjunt ampli de mesures d'actuació, proporcionant l'estructura necessària per a un sistema de mesura i gestió d'aquesta estratègia.

Per exemple, pel que fa a les sol·licituds, amb un indicador similar a un velocímetre es podria saber si s'està esgotant el termini per sol·licitar un servei i hi han menys sol·licituds en curs que l'any anterior. Un altre exemple d'indicador, en aquest cas relatiu a les accions dels proveïdors, podria consistir en definir un semàfor que es posi de color verd quan el proveïdor seleccionat està pujant regulament i de forma actualitzada la informació referent a les netejes i desinfeccions d'instal·lacions, o bé es posi de color vermell en cas contrari.

En definitiva, cal continuar explorant les possibilitats que ofereixen aquestes eines de codi obert en matèria de *business intelligence* i magatzems de dades, ja que tot el que s'ha exposat com a resultat del present TFG, malgrat comportar moltíssimes hores de feina i d'estudi, sembla només la punta de l'iceberg.

7. Glossari

business intelligence: Habilitat de transformar les dades en informació, i la informació en coneixement, i així optimitzar el procés de presa de decisions dels directius.

descriptor: Atribut d'una dimensió utilitzat per a seleccionar instàncies dels fets.

dimensió: Punt de vista utilitzat en l'anàlisi d'un cert fet. En una base de dades relacional, cada camp en un registre representa una dimensió.

estrella: Conjunt de dimensions relacionades amb un fet.

ETL: Extraction, Transformation and Loading. Procés d'extracció de les dades de les fonts originals, transformació de les mateixes i càrrega dins el magatzem de dades. Perquè el seu nom fos totalment descriptiu s'hauria d'anomenar component d'obtenció, transformació, depuració, integració, transport i càrrega .

factoria d'informació corporativa: Conjunt d'elements de programari i maquinari que ajuden en l'anàlisi de dades per a prendre decisions. sigla: FIC

fet: Objecte d'anàlisi.

granularitat: Grandària d'un objecte respecte a un altre.

magatzem de dades corporatiu: Conjunt de dades que guarda integrades totes les dades històriques de l'empresa.

magatzem de dades departamental: Conjunt de dades que resol les necessitats d'anàlisi d'un cert departament o conjunt d'usuaris.

magatzem de dades operacional: Conjunt de dades integrat i orientat al tema, però sense dades històriques. S'acostuma a utilitzar com a pas intermedi en la construcció del magatzem de dades corporatiu.

mesura: Dada numèrica associada a un esdeveniment que volem analitzar.

OLAP: On-Line Analytical Processing . Categoria de tecnologia de programari que permet als analistes, gestors i executius de millorar el seu coneixement de les dades mitjançant l'accés ràpid, consistent i interactiu a una àmplia varietat de possibles vistes d'informació que ha estat transformada des de les dades operacionals per a reflectir la dimensionalitat real de l'empresa com l'entén l'usuari.

ROLAP: Eina OLAP implementada sobre un sistema gestor de bases de dades relacional.

SIMSAP: Sistema d'Informació Municipal en Salut Pública

8. Bibliografia

"Tabla de dimensión". *Wikipedia* [en línia]. [Data de consulta: 19 d'abril de 2011].

<http://es.wikipedia.org/wiki/Tabla_de_dimensi%C3%B3n>

Carina Roldán, M (2010, abril). "Installing Pentaho Data Integration with MySQL". Instal·lant Pentaho Data Integration amb MySQL. *Packt Publishing* [en línia]. [Data de consulta 10 d'abril de 2011].

<<http://www.packtpub.com/article/installing-pentaho-data-integration-mysql>>

Carina Roldán, M (2009, febrer). "Pentaho Data Integration (Kettle) Tutorial". Tutorial de Kettle. *Pentaho Community* [en línia]. [Data de consulta 16 d'abril de 2011].

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

Moran, D (2010, 26 de maig). "Getting Started". Per començar. *Pentaho Community* [en línia]. [Data de consulta 9 d'abril de 2011].

<<http://wiki.pentaho.com/display/EAI/Getting+Started>>

Batchelor, M.; Bleuel, J. (2011, 9 de febrer). "Latest Pentaho Data Integration (aka Kettle) Documentation". Última documentació disponible sobre Pentaho Data Integration. *Pentaho Community* [en línia]. [Data de consulta 20 d'abril de 2011].

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

Cornelius, Jake (2010). Pentaho Data Integration Overview [flash video en línia] *Pentaho Community* (durada 17'27").

<http://www.pentaho.com/products/demos/PDI_overview/PDI_overview.html>

Rius Gavidia, A.; Serra Vizern, M. (2011). *Introducció a l'emmagatzematge de dades*. Barcelona: UOC ("Magatzem de dades", P02/11048/00008).

Abelló Gamazo, A.; Samos Jiménez, J. (2011). *La factoria d'informació corporativa*. Barcelona: UOC ("Magatzem de dades", P02/11048/00009).

Samos Jiménez, J. (2011). *Construcció de la factoria d'informació corporativa*. Barcelona: UOC ("Magatzem de dades", P02/11048/0000910).

Abelló Gamazo, A. (2011). *Disseny multidimensional*. Barcelona: UOC ("Magatzem de dades", P02/11048/00009).

9. Annex I : Entorn d'implementació i programari emprat

Maquinari

Per a implementar el projecte i efectuar les proves s'ha utilitzat un portàtil HP amb processador Intel Core2 Duo T5470 a 1,6GHz i 2 GB de RAM.

El sistema operatiu emprat és un Ubuntu 10.04 amb nucli 2.6.32.31-generic.

L'entorn descrit és un entorn de desenvolupament. S'està preparant a Dipsalut un entorn de producció amb un servidor basat en tecnologia de virtualització, on una de les màquines virtuals allotjarà el magatzem de dades de forma dedicada.

Programari

Tot el programari utilitzat que es descriu a continuació és de codi obert. Si bé *Pentaho* ofereix una versió comercial amb suport i diferents serveis de valor afegit, existeix una versió *community* que es pot descarregar de la URL <http://sourceforge.net/>, tenint cura d'escollir els descarregables adients al sistema operatiu de la màquina on s'allotjaran les aplicacions.

MySQL

Cal tenir degudament instal·lat el sistema gestor de bases de dades MySQL. En aquest cas s'ha utilitzat la versió 5.1.41 for debian-linux-gnu. També s'han emprat com a suport les eines de gestió de bases de dades *MySQL Workbench 5.2.25* i *MySQL Query Browser 1.2.12*.

Java

Cal tenir instal·lat el *Java Development Kit*. Per al cas que ens ocupa, s'ha emprat la versió 1.6.0_20 del JDK.

Pentaho Business Intelligence Server

El Pentaho BI Server és un servidor d'aplicacions de *Business Intelligence* basat en el conegut *Jboss Application Server*. Ha estat especialment programat com a *front-office* per dotar els usuaris dels magatzem de dades d'una interfície d'accés a consultes, vistes d'anàlisi, cubs OLAP, informes i indicadors.

El BI Server utilitzat en aquest projecte és el 3.7.0 (fitxer *biserver-ce-3.7.0-stable.tar.gz* descarregat de *Sourceforge*). La seva instal·lació és molt senzilla doncs es tracta de desempaquetar el fitxer tar.gz a */opt/pentaho/*. Aquesta acció crea dues carpetes: *biserver-ce* i *administration-console*.

Dins la carpeta *biserver-ce*, hi ha un script *-set-pentaho-env.sh-* que permet configurar les variables d'entorn per al BI Server. Finalment cal executar el script *start-pentaho.sh* que arrenca el servidor. En qualsevol cas, el procés d'instal·lació d'aquesta aplicació està molt documentat a Internet, a la mateixa web de Pentaho, i també al *Youtube*.

Un cop instal·lat i configurat, el BI Server és accessible des de qualsevol navegador al port 8080 del *localhost*, és a dir, que obrint la URL <http://localhost:8080/pentaho> es mostra la pantalla de login, on es poden escollir uns usuaris predefinits d'exemple.

Per tal de facilitar als consultors l'accés i proves de l'entorn d'implementació sense haver-se d'instal·lar les aplicacions, la màquina emprada per al desenvolupament del projecte serà accessible a <http://80.32.31.86:8080/pentaho> fins al 23 de juny 2011 (usuari: **consultor**, password: **uoc**).

En qualsevol cas, els consultors podran també provar de reproduir l'entorn i les dades, mitjançant la instal·lació de tota la **suïte BI de Pentaho** en un ordinador amb SO Windows o Linux i carregant la base de dades amb els scripts facilitats.

Pentaho Administration Console

L'aplicació Pentaho Administration Console permet gestionar els usuaris que accediran al BI Server i dotar-los de permisos de visualització i consulta dels diferents magatzems. A més, permet definir connexions a bases de dades, en aquest cas a MySQL, mitjançant drivers JDBC (cal disposar d'alguna versió de *mysql-connector-java*).

No cal fer res addicional per a instal·lar aquesta consola, doncs forma part del BI Server. Per a executar-la, cal engegar el servei mitjançant el script *start-pac.sh* que es troba a */opt/pentaho/administration-console*. Un cop fet això, es pot accedir a la consola d'administració mitjançant la URL <http://localhost:8099> usuari **admin** i password **password**. Evidentment, cal canviar aquestes credencials i crear nous usuaris, a part dels que ja venen amb les dades d'exemple. És important assenyalar que per a que la consola funcioni adequadament, cal engegar primer el BI Server tal com s'explica en el punt anterior.

Per al projecte que ens ocupa, ja s'ha creat a l'entorn de desenvolupament, mitjançant la dita consola, l'usuari específic **consultor** amb tots els privilegis, que podrà accedir al BI Server de proves.

Pentaho Data Integration, *Kettle*

Kettle, també coneguda com a PDI -Pentaho Data Integration-, és l'aplicació que permet dur a terme les operacions d'ETL -Extract, Transform and Load- de manera gràfica i intuïtiva, així com programar tasques repetitives de càrrega. La versió emprada en aquest projecte és la 4.1.0 que es pot descarregar de *Sourceforge* (fitxer *pdi-ce-4.1.0-stable.tar.gz*). Un cop desempaquetat el fitxer anterior, es crea una carpeta anomenada *data-integration*. Per a executar l'aplicació específica que permet modelar transformacions, cal utilitzar el script *spoon.sh*.

Pentaho Schema Workbench, *Mondrian*

Per tal de modelar una estrella dins del BI Server i fer-la accessible al mateix per tal d'utilitzar-les en vistes de cubs o llistats, cal una eina addicional anomenada *Mondrian* - Pentaho Schema Workbench.

Aquesta eina permet crear de manera gràfica una estrella o floc de neu, afegir-hi dimensions, fets i nivells d'agregació. El model resultant es desa en un fitxer XML que la mateixa aplicació permet publicar al BI Server. Un cop publicada l'estrella en qüestió, un usuari autenticat dins el BI Server, pot estirar les dades de l'estrella per fer-hi tota mena de consultes.

En aquest projecte s'ha utilitzat la versió 3.2.2 de Mondrian (fitxer *psw-ce-3.2.2.14148.tar.gz* descarregat de *Sourceforge*). Un cop desempaquetat el fitxer anterior dins de */opt/pentaho*, es crea una carpeta anomenada *schema-workbench*. S'ha de copiar el driver de mysql dins de *schema-workbench/drivers* – en aquest cas s'ha utilitzat el fitxer *mysql-connector-java-5.1.6.jar* – i finalment, executar el script *workbench.sh* per tal d'iniciar l'aplicació que permet modelar el cub ROLAP així com publicar-lo.

Pentaho Report Designer

Una altra de les eines emprades és el Pentaho Report Designer que permet modelar llistats de tota mena a partir de les taules definides en el magatzem de dades. Per a aquest projecte s'ha descarregat i instal·lat el fitxer *prd-ce-mac-3.8.0-stable.tar.gz* -versió 3.8.0 de Pentaho Report Designer. Un cop desempaquetat, cal obrir la carpeta resultant i executar el script *report-designer.sh*.