

Implementació d'una intel·ligència de negoci basada en SaaS

Roger Calaf

Resum—Les tecnologies de la informació estan obrint noves vies sobre l'anàlisi de dades degut a que els magatzems de dades són cada cop més grans i per això són necessàries eines potents per al tractament i extracció d'informació. L'article següent relata el procés d'implementació d'un sistema d'intel·ligència de negoci basat en el núvol. Aquest sistema analitza dades en temps real per aconseguir mostrar estadístiques fàcils d'interpretar per a l'usuari final. També ofereix diferents nivells d'accés a les dades per a poder generar informes amb privilegis diferents. Durant aquest procés s'ha hagut de fer front a molts problemes d'integritat i administració de dades en les tasques d'unificació de fonts d'origen. Però s'ha aconseguit obtenir un sistema fàcil i intuïtiu d'utilitzar, capaç de generar vistes i indicadors clau amb gran rapidesa. Això és gràcies a aquest nou sistema de gestió de conjunts massius de dades, BigQuery. En definitiva, es detallaran els passos i les decisions preses per convertir les dades generades per un negoci en informació valuosa. Això recolzarà la presa de decisions de caràcter executiu.

Paraules clau—TIC, Big Data, Intel·ligència de negoci, SaaS, BigQuery, BIMEAnalytics, Estadístiques en temps real.

Abstract—The information technologies are opening new ways of analyzing data due to an increasing DataWarehouse and this is why we need powerful tools for treatment and information extraction. The following article describes the process of implementing a business intelligence cloud-based system. This system analyze real time data to show statistics easily interpretable by the final user. It also offers different access levels to the data so it generates reports with different privileges. During this process we had to face many problems of integrity and data management with the unification data sources tasks. However, we have managed to get an easy and intuitive system, capable of generating views and key indicators rapidly. This is thanks to this new system of managing massive datasets, BigQuery. Finally, we will detail the steps and decisions to convert the data generated in a business to valuable information. This will support the executive decision-making.

IndexTerms—TIC, Big Data, Business Intelligence, SaaS, BigQuery, BIMEAnalytics, Real time statics.



1 INTRODUCCIÓ

ELS negocis cada vegada depenen més de les tecnologies de la informació. Tota empresa, sigui petita, mitjana o gran, utilitza i té a la seva disponibilitat eines tecnològiques per a la seva gestió interna, des d'aplicacions per a generar documents fins a programes per donar suport a la logística o al control de la producció.

Un altra de les possibilitats que ofereixen les tecnologies de la informació, anomenades *TIC*, és la de transformar dades per obtenir informació. És obvi que actualment tots els processos es registren, i s'obtenen dades derivades de la seva activitat. Aquest és un dels principals motius pels que les empreses utilitzen sistemes informàtics per a la gestió. És molt important pels negocis i en molts altres àmbits disposar d'aquestes dades, que en molts casos consten d'un gran volum i baixa integritat, ja que encara que des d'un primer moment no aportin cap mena de valor, a través de varis processos de transformació i manipulació es pot extreure informació valuosa per a la presa de decisions futures.

Aquesta possibilitat ha empès a una empresa del sector de la restauració a realitzar un projecte per tal d'extreure

el màxim partit sobre les dades que generen i transformar-les en informació que ajudi a prendre futures decisions executives. El següent projecte ha estat consignat per a crear una infraestructura que es tracti amb *Big Data*¹ i ofereixi la possibilitat d'extreure informació d'aquesta font de dades a partir d'un sistema d'intel·ligència de negoci basat en *SaaS*.

L'empresa beneficiària està formada per més de 70 botigues repartides per tot l'estat, en un gran nombre localitzades aquí a Catalunya. Totes les botigues disposen d'un sistema de punt de venda que envia la informació referent a les compres realitzades pels clients vers un sistema central que emmagatzema aquestes dades en una base de dades. Es demana implantar un sistema d'intel·ligència de negoci que treballi amb les dades que arriben de les botigues per tal de poder generar estadístiques i poder conèixer la seva situació real.

La metodologia de desenvolupament ha seguit un model de tipus incremental, ja que hi han hagut fases del desenvolupament que s'han realitzat de forma seqüencial

• E-mail de contacte: roger.calaf@e-campus.uab.cat
 • Menció realitzada: *Tecnologies de la Informació*.
 • Treball tutoritzat per: Jordi Pons Aróztegui (Deic)
 • Curs 2013/14

¹ Tecnologia capaç d'emmagatzemar grans quantitats de dades, sense estructurar, en dispositius d'emmagatzematge de baix cost on les consultes de dades són molt ràpides.

² *SaaS* o Software as a Service, és el programari disponible en línia i que utilitzen els usuaris sense haver-lo de tenir instal·lat a l'ordinador. S'utilitza a través d'un navegador web.

i altres de forma iterativa. Més endavant s'entrarà en detall sobre les fases de desenvolupament, que han consistit en la realització de diverses tasques. En primer lloc s'ha fet un anàlisi de requisits per conèixer les expectatives dels interessats, un segon anàlisi per seleccionar les tecnologies necessàries, després tasques de disseny i implementació del model de dades i finalment activitats sobre els informes generats per l'aplicació d'intel·ligència de negoci. Per al correcte seguiment de la metodologia, l'equip de treball ha estat format per quatre persones però ha anat variant en funció del treball a realitzar. Una fase, degut a la seva complexitat, l'ha portat a terme una empresa externa. La seva feina ha estat desenvolupar la part d'implementació del model de dades, extreure les dades de les fonts originals i inserir-les en les noves taules a partir de les especificacions donades per nosaltres i sota la nostra supervisió. Per la meua part, esmentar la meua participació parcial en algun d'aquests processos i la meua total implicació en totes les fases restants del projecte.

Aquest article pretén ser una guia del desenvolupament del projecte realitzat per poder conèixer quins han sigut els passos efectuats i els problemes sorgits en la seva implantació real. L'article consta de tres capítols sobre el desenvolupament del projecte. El primer capítol explica la realització de les tasques principals i els problemes sorgits en aquestes. En el segon capítol es presenten els resultats obtinguts i finalment, en l'últim capítol, s'exposa una reflexió amb les conclusions extretes dels resultats del projecte.

En definitiva, l'idea d'aquest projecte és reunir el gran conjunt de dades obtingut a partir dels processos de negoci de l'empresa i transformar-lo en informació útil per a que l'empresari pugui prendre decisions de caràcter estratègic per al futur.

El projecte exposat en aquest article està desenvolupat en el marc de col·laboració entre la Universitat Autònoma de Barcelona i l'empresa Home Meal Replacement Nostrom.

1.1 Objectius

La implantació d'un sistema d'intel·ligència de negoci d'aquest estil comporta resoldre una gran quantitat de problemes. Les dificultats típiques d'aquests sistemes són les de tractar amb grans quantitats d'informació. Quan es treballa amb volums molt grans, el tractament de les dades és lent i costós. Una senzilla operació de trasllat de dades pot trigar hores en executar-se i ja no entrem en la consulta i extracció d'informació. Aquest sistema ha de ser capaç de comparar dades i fer consultes amb milions de línies en qüestió de segons. Amb l'objectiu de que aquest sistema estigui disponible des de qualsevol lloc i en qualsevol moment del dia, necessitem un software accessible des de qualsevol dispositiu i que estigui funcionant al núvol per poder mantenir la disponibilitat de les dades. Per altra banda, el sistema ha de ser capaç de generar estadístiques en temps real. Quan parlem de temps

real no vol dir que l'última informació utilitzada serà la del dia anterior sinó que serà la de cinc minuts abans. Aquesta dimensió temporal és acceptable ja que permet conèixer amb molt detall el funcionament de les botigues.

Una altra característica consisteix en que el sistema gestionarà la privacitat de la informació. No ha de tenir el mateix grau de coneixement l'encarregat d'una botiga que el directiu. L'encarregat de la botiga no té perquè conèixer el número final de vendes a totes les botigues, ja que només li interessarà la seva. Per tant, s'hauran de definir rols d'usuari per protegir les dades.

Per la part de visualització de la informació, el sistema ha d'incorporar estadístiques de tot tipus, des de taules per consultar estats, fins a gràfics amb indicadors claus de l'estat del negoci. Per tant, les estadístiques han de ser escollides amb detall perquè puguin ser útils per a l'empresari i el personal que treballarà amb elles. A més a més s'ha de visualitzar la informació de forma intuïtiva.

Finalment, s'ha d'aconseguir unificar totes les dades necessàries en una sola font de dades. És molt important unificar les fonts de dades originals en una sola, ja que això millorarà notablement la rapidesa de les consultes i es podrà gestionar i transformar la informació més fàcilment ja que la tindrem en un únic sistema.

Per tant, els objectius principals a assolir seran establir un sistema de presa de decisions basat en el núvol, unificar les fonts d'informació en una sola, generar les estadístiques i gràfics amb dades a temps real, gestionar una gran quantitat d'informació de manera ràpida, mantenir un control d'accés a la informació disponible i finalment, visualitzar de forma intuïtiva la informació.

1.2 Estat de l'art

Seguint l'evolució dels sistemes decisionals³ en els anys 80 neix el concepte d'intel·ligència de negoci (BI). Aquest concepte engloba un conjunt d'estratègies enfocades a recopilar, emmagatzemar, accedir i analitzar les dades per ajudar als usuaris a prendre millors decisions. Els BI neixen amb l'objectiu de satisfer varis factors de negoci com són la necessitat de recolzar els processos de presa de decisions en les organitzacions o la necessitat d'accedir a informació de major qualitat. Entrant en el concepte d'intel·ligència de negoci hi destaquen dues estratègies comunes que són els *DataWarehouse*⁴ i els processos *ETL*⁵. Els processos *ETL* es segueixen utilitzant avui dia i formen una part molt important dels BI, ja que són crucials en la integració de dades. En canvi, l'ús de sistemes de *DataWarehouse* en les intel·ligències de negoci va quedant

³ Els sistemes decisionals o DSS són sistemes d'informació orientats al suport de la presa de decisions de diferents tipus per directius de l'organització, gestors i altres professionals.

⁴ El *DataWarehouse* o magatzem de dades és una base de dades amb informació històrica d'una companyia dissenyada i estructurada per a realitzar-hi consultes eficientment.

⁵ Els processos ETL permeten a les organitzacions moure dades desde múltiples fonts, reformatar-les i netejar-les.

endarrerit a causa de l'aparició de noves fonts de dades i la necessitat d'analitzar-ho tot, des de fluxos en temps real fins a una enorme quantitat de contingut no estructurat. Aquest fet fa que moltes empreses s'adonin que estan en una era en que l'espectre de les càrregues de treball analítiques són tant amplies que no poden ser tractades utilitzant només un *DataWarehouse*. Es per això que els requeriments de negoci dicten que cal un nou entorn d'anàlisi més complex per donar suport a una gamma de càrregues analítiques de treball [1]. *Big Data* és un bon sistema per donar suport a aquestes càrregues analítiques ja que les gestiona amb una combinació de gran volum, velocitat de dades i varietat de dades.

Les grans organitzacions han adoptat l'anàlisi de grans volums de dades en quantitats apreciables. Els avantatges potencials que es poden determinar impulsant aquesta adopció són que qualsevol cosa que impliqui els clients podrà beneficiar l'anàlisi de grans dades, ja que aquests poden ajudar a desenvolupar les definicions de pèrdua de clients i altres comportaments així com una comprensió del comportament del consumidor a partir de clics. Un altre dels avantatges és que la intel·ligència de negoci podria endinsar-se en els resultats de negoci, comprendre canvis empresarials, tenir una millor planificació i la identificació de causes fonamentals dels costos. Les aplicacions analítiques també serien probables beneficiaris d'aquests anàlisis, com serien les aplicacions de detecció de frau, la quantificació de riscos o la tendència del mercat.

Tot i aquests avantatges, l'anàlisi massiu de dades també té les seves barreres. Una d'aquestes seria disposar de personal inadequat o amb poques habilitats, ja que moltes organitzacions encara són noves en l'anàlisi massiu de dades i moltes d'aquestes han desenvolupat les seves pròpies habilitats. Una altra de les barreres seria la manca de suport a les empreses, ja que pot dificultar el procés d'obtenció de dades en la seva anàlisi. Els problemes amb el programari també poden ser barreres per a l'anàlisi de dades grans quan aquestes estan localitzades als *DataWarehouse* [2].

A part, la culminació del programari com a servei (*SaaS*) ha portat la tecnologia i els canvis de negoci a proveïdors de serveis i usuaris. El model de negoci *SaaS* permet reduir costos i oblidar-se dels problemes de compatibilitat entre dispositiu i programari ja que tot està centralitzat. També s'allibera als usuaris finals d'haver d'instalar programari al seu sistema i permet treballar d'una forma còmode sense pèrdues d'informació, ja que tota la informació es manté al núvol [3].

Les intel·ligències de negoci basades en programari com a servei són una bona opció, ja que el cost és baix gràcies a que no hi ha costos de compra inicials o requeriments addicionals de personal necessaris per gestionar el sistema de BI. Tot i així, les eines d'anàlisi poden no tenir totes les característiques que el software tradicional té, encara que siguin menys complexes, poden ser menys

funcionals.

2 METODOLOGIA

En aquest apartat exposaré, a grans trets, les diferents fases del projecte, explicant tan els problemes que han fet endarrerir-lo, com les valoracions sobre les tasques realitzades i condicions a tenir en compte en futures aplicacions.

Com he comentat en l'anterior apartat, s'ha seguit una metodologia de tipus incremental, ja que les fases de disseny del model de dades i dels informes han estat de forma iterativa. Això, és degut a que s'han d'anar actualitzant. En el model de dades per exemple, cada cop s'afegia una font de dades o una taula nova i es feia la selecció dels camps importants. Les altres fases, en canvi, han seguit una línia seqüencial d'inici a final. Començaré explicant cada una de les principals fases en ordre d'execució, començant primer per la fase d'anàlisi de requisits i acabant per la fase de generació d'informes amb l'aplicació d'intel·ligència de negoci.

2.1 Anàlisi de requisits

En aquesta fase inicial és quan es pren contacte amb el client, realitzant una reunió on les dues parts interessades es posen d'acord en els requisits i les funcionalitats que ha de complir el projecte. També s'hi negocia el cost de producció i la duració d'aquest, per poder complir els terminis previstos.

Els requisits pactats més destacats en aquest projecte van ser, primer de tot, una duració màxima de tres mesos per a la seva implantació, ja que el client tenia pressa en disposar de nova informació per conèixer l'estat de la seva empresa. També es va pactar escollir una tecnologia que funcionés al núvol (*SaaS*), per seguir amb la mateixa idea de negoci, ja que s'implantarà un nou sistema de TPV basat en *SaaS* a les botigues. El client necessitava que l'aplicació tingués diferents polítiques d'accés a les dades, perquè no tothom tinguí accés a la mateixa informació, per tant es van pactar nivells de privilegi tan a les dades com als *dashboards*⁶.

Un dels requisits principals va ser amb quines dades es volia muntar el sistema i es va estipular muntar un historial de dades amb totes les dades que hi havia en aquell moment a les bases de dades fins a l'actualitat. Quan parlo del terme actualitat, em refereixo a les dades que es generen el mateix dia. Partint d'aquesta base, també es va acordar la franja de temps per aconseguir l'obtenció de dades a temps real. La franja de temps que es va decidir va ser d'uns cinc minuts de diferència, un temps raonable per aconseguir un efecte de temps real.

A part de la negociació de la franja de temps real, va ser un requisit indispensable carregar les dades històri-

⁶ És el tauler de control d'un BI, en ell hi apareixen tots els tipus d'estadístiques i indicadors que es volen monitoritzar.

ques en hores on la font de dades origen estigués en baixa càrrega per evitar col·lapsar-la. Un altre dels requisits per dur a terme aquestes càrregues va ser planificar aquestes connexions al núvol de l'empresa externa que s'encarregaria de muntar l'infraestructura de temps real.

Finalment, es van concertar dos requisits més. Un va consistir en sessions de formació, a càrrec de l'empresa creadora de la intel·ligència de negoci, amb la finalitat de conèixer i aprendre a utilitzar l'aplicació. L'altre requisit va consistir en poder traspasar la tecnologia al client un cop la infraestructura estigués acabada. D'aquesta manera, el client podria retocar-la en cas de necessitar alguna dada addicional.

Amb els requisits acordats i amb les dues parts enteses poden continuar realitzant-se els processos i activitats estipulats en el projecte.

2.2 Selecció de les eines

Per tal de fer efectius els requisits demanats pel client, és molt important una reflexió prèvia sobre l'elecció i l'ús de les tecnologies que s'utilitzaran. Com s'ha comentat abans en l'apartat de requisits, un dels requisits principals que es demana és que aquesta eina ha de funcionar en el núvol. Això significa que des de qualsevol lloc i amb qualsevol dispositiu l'usuari podrà consultar la informació que necessita. Per aquest motiu, s'ha realitzat una cerca per escollir la tecnologia que millor s'adapta a les necessitats del producte i s'han analitzat les solucions següents:

- *Business Objects, Business Intelligence Workhouse (SAP⁷)*. És una eina molt robusta, però està basada en SAP, que requereix molta feina de monitorització; el desenvolupament és llarg i el cost molt elevat. No és òptima per treballar amb *Big Data*.
- *RoamBI*. És una de les eines més potents que existeixen alhora de generar vistes i crear estadístiques, ja que és molt flexible en aquest sentit. Però no permet gaires funcionalitats en els processos de recollida i normalització de dades.
- *QlickView*. És una eina ràpida, econòmica i fàcil d'integrar. Si tenim present que està al núvol, s'adapta molt bé al model de producte que s'ha dissenyant.
- *Cognos*. És una solució de BI desenvolupada per IBM. Ha quedat obsoleta, és necessari molt temps de desenvolupament per treballar-hi i és molt costosa.
- *Bime*. És una tecnologia molt flexible ja que s'adapta a gran varietat de fonts de dades; és senzill d'utilitzar i fàcil d'implementar. Igual que *QlickView* també treballa al núvol. A més a més,

aquesta té una manera molt millor de representar la vista de dades.

Després d'analitzar les diferents solucions, es va optar per utilitzar la tecnologia *Bime*, ja que les seves característiques són les que s'equiparen més als requisits que s'exigeixen. Com comentàvem abans, és una tecnologia molt flexible que suporta gran quantitat de tipus diferents de fonts de dades i funciona com a *SaaS*. Però les grans diferències que destaquen d'aquest software són la seva capacitat d'homogeneïtat a l'hora de tractar diferents fonts de dades i els múltiples models de visualització d'estadístiques. També es va valorar molt positivament la manera que té de controlar l'accés a les dades. Això és interessant, ja que permet establir nivells de privilegis segons convingui. Concretament, un usuari pot tenir accés a la consulta d'un tipus de dades que un altre usuari no pot tenir accés.

2.2.1 Intel·ligència de negoci

Per entendre els processos que s'han emprats és bàsic conèixer el funcionament dels sistemes d'intel·ligència de negoci. Aquests, principalment, es componen de tres nivells o capes:

- *Data Source*. Aquesta capa s'encarrega de fer una còpia en brut de totes les fonts de dades, ja que les dades poden arribar des de múltiples orígens diferents.
- *Extract Data*. Aquesta capa és on es normalitzen les dades per a poder-les tractar (*processos ETL*). Específicament, les dades es classifiquen en atributs, mesures i dates. En aquest nivell és on també es defineixen els noms dels camps de dades.
- *View Layer*. Aquesta capa és on es visualitzen les dades a través de taules o gràfics diversos. Poden generar estadístiques i visualitzar-les de múltiples formes.

En la següent figura, *Figura 1*, es representen aquestes tres capes lògiques. Com es pot observar, una aplicació d'intel·ligència de negoci és en essència una eina que s'encarrega de visualitzar les dades d'una manera agradable davant la vista de l'usuari, ja que la gran potència per fer consultes i treure estadístiques l'aporta el motor de base de dades. En conjunts de dades massius són molt importants els temps de resposta en les consultes a les bases de dades.

⁷ Sistema d'informació que permet a les companyies executar i optimitzar diferents processos de negoci.

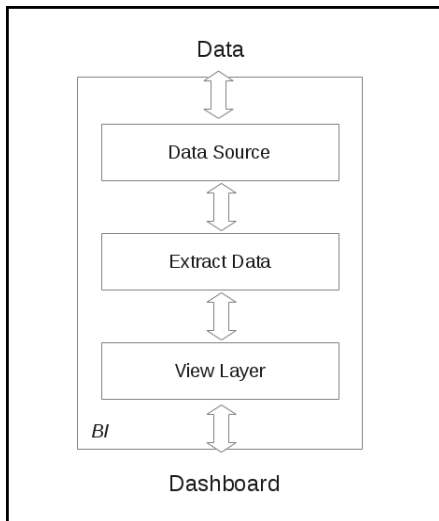


Figura 1. Esquema d'un BI

Bime gestiona la capa de Data Source de tal manera que es creen connexions per a poder realitzar còpies en brut (*Full Load*⁸ i *Delta Load*⁹) de totes les fonts de dades originals. Aquestes còpies s'emmagatzemen en fonts de dades preparades per oferir rapidesa en les consultes. *Bime* ofereix diversos sistemes d'emmagatzematge per a la consulta ràpida segons el volum de dades:

- *Sistema original*. No es realitza cap còpia en brut. Es consulta directament de la font original. No permet tenir les dades ràpidament, tot i així és útil per fonts de dades petites.
- *Déjà vu*. Les dades sempre estan en línia. És més ràpid que la font de dades original. Suporta fins a un màxim de 200.000 línies.
- *BimeDB*. Permet tenir sempre les dades en línia. És molt més ràpid que la font de dades original i permet escalar grans conjunts de dades. Suporta un màxim de 50.000.000 línies.
- *Google BigQuery*. Les dades sempre estan disponibles. És la tecnologia que suporta més volum de dades. No hi ha límit d'emmagatzematge, permeten taules de bilions de línies amb consultes molt eficients.

En efecte, l'opció escollida va estar totalment condicionada pel volum de les fonts de dades. En aquest cas es treballa amb taules de bases de dades de més de 60 milions de línies, per tant la solució que presenta Google és la més adient.

Per tant, les dades per generar estadístiques, sorgeixen d'una connexió amb la font de dades *BigQuery* que és l'encarregada d'aportar rapidesa en les consultes.

⁸ Un full load és una càrrega total d'una taula de la base de dades.

⁹ Un delta load és una càrrega parcial d'una taula de la base de dades. Les dades es carreguen en funció d'un *offset* de temps.

2.2.2 Google BigQuery

En essència, *BigQuery* és una solució *cloud* que permet emmagatzemar i realitzar consultes sobre grans volums de dades [4], aquestes característiques fan que pugui classificar-se com un sistema de *Big Data*.

Google gestiona grans conjunts de dades constantment, això significa que disposa d'una infraestructura força desenvolupada per tractar aquests conjunts massius. Amb *BigQuery*, posen aquesta tecnologia a l'abast de tothom, ja que *BigQuery* és la implementació externa del nucli tecnològic de la companyia, anomenat Dremel.

Dremel és un servei que permet realitzar consultes SQL amb conjunts de dades massius i obtenir resultats en segons. Aquest utilitza un servei de consulta paral·lela massiva al núvol, que paral·lelitzava la consulta i l'executa en desenes de milers de servidors simultàniament.

Pel que fa a la seva arquitectura, hi ha dues tecnologies principals que li aporten aquest rendiment:

- *Emmagatzematge en columnes*. Les dades es separen en valors de columna i aquests es guarden en volums d'emmagatzematge diferents. Aquest sistema fa possible aconseguir una alta compressió i minimitzar l'escaneig [5].
- *Arquitectura en arbre*. S'utilitza per a l'enviament de consultes i per a la unió dels resultats produïts pels milers de màquines en uns segons.

L'únic desavantatge destacable està en la tècnica d'emmagatzematge en columnes, ja que no és eficient si s'han d'actualitzar registres, per això Dremel no suporta aquests tipus de funcionalitats.

En conseqüència, *BigQuery* i Dremel comparteixen la mateixa arquitectura i característiques de rendiment. Per aquest motiu les funcionalitats d'actualitzar registres no estan implementades i això fa que s'hagin de realitzar modificacions a través de subconsultes.

2.3 Definició del model de dades

El model de dades s'ha construït a partir de diverses fonts de dades, unificant les més importants en una única font, per així poder aconseguir rapidesa en la generació d'estadístiques. Com s'ha comentat anteriorment, la base de dades que es va escollir per guardar les dades, tractar-les i unificar-les és la de *BigQuery*. En la *Figura 2* es representa l'esquema descrit:

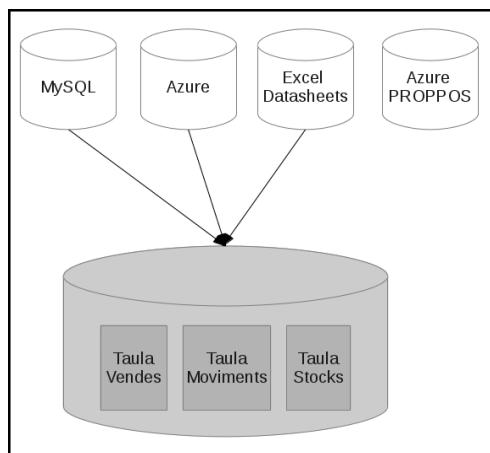


Figura 2. Esquema de dades

Totes aquestes fonts contenen dades importants per a l'extracció d'informació. No es van extreure totes les taules amb tots els camps sinó que es va fer una selecció dels camps i de les taules necessàries. Per altra banda, els noms de les fonts de dades mencionades són purament simbòlics i només serveixen per poder anomenar-les i diferenciar-les. Aquest esquema també recull una font de dades anomenada *azurePROPPOS* que s'integrarà en un futur. Més endavant explicarem millor la seva utilitat.

A *BigQuery*, només hi resideixen les dades valuoses per nosaltres. Ens referim a les dades que ens permeten generar informació i crear estadístiques. Per això, es va tractar minuciosament el procés de dissenyar el model de dades, ja que s'havia de seleccionar la informació correctament perquè sinó després, en fases posteriors, podríem tenir moltes complicacions a l'hora de crear *dashboards*¹⁰ i obtenir estadístiques interessants, ja que tindríem dades innecessàries.

Un cop definit el model de dades que s'extraurà de les fonts originals, es van definir les taules destí on s'emmagatzemen les dades consultades per l'aplicació d'intel·ligència de negoci. Aquestes taules consisteixen en taules planes¹¹ amb desenes de columnes que relacionen les dades directament, permetent que les cerques de dades siguin molt ràpides ja que no s'ha d'accedir a altres taules per a obtenir-les. Les tres taules s'anomenen **Vendes**, **Moviments** i **Stocks** i els seus camps no s'han escollit aleatòriament, sinó que hi ha hagut un procés minuciós d'anàlisi per conèixer quines dades són necessàries per realitzar certes estadístiques i sobretot quines relacions es volen obtenir entre elles.

La taula Vendes és la més gran de totes. En ella hi ha contingudes les dades relacionades amb les vendes de productes de cada botiga, per tant, aquí s'emmagatzemen les dades relacionades amb els tiquets de compra, les botigues, els articles i els clients. Per fer-nos una idea de

les dimensions d'aquesta taula, només la taula tiquets conté al voltant de 50 milions de files.

La taula de Moviments conté tota la informació respectiva als moviments de productes a les botigues. En ella s'hi guarden les dades relacionades amb els productes, els moviments de comandes i les botigues.

A la tercera taula, la taula de Stocks, s'hi emmagatzemen les dades referents als estocs de productes de les botigues. Conté les dades sobre els productes, estocs i botigues.

Hi ha dades que es repeteixen entre les diferents taules, això és necessari perquè aquestes taules funcionen de forma independent. Això vol dir que no es crearà mai una estadística que utilitzi dues taules a la vegada. En el cas que aquestes dades s'haguessin de modificar, aquesta duplictat seria un problema feixuc ja que això implicaria modificar tots els valors redundants, però no és un problema, ja que el nostre objectiu és únicament consultar les dades i aquesta estructura ens permet fer-ho d'una manera directa.

Per a l'elecció de les dades a emmagatzemar va ser molt important la reunió amb el client. Ens van comentar que en un futur deixarien d'utilitzar la font de dades *MySQL*, perquè implantarien un nou sistema de punt de venda, *PROPPOS*¹², i que aquest emmagatzemaria les dades al núvol (*azurePROPPOS*). Aquest fet va condicionar el desenvolupament del projecte ja que va implicar tenir valors amb el mateix significat en fonts de dades diferents (*MySQL* i *azurePROPPOS*). A causa d'això, el procés de definició d'aquestes taules no va ser tant trivial com sembla, ja que és va tenir en compte de no adaptar les taules per a una font de dades en concret. S'ha mantingut un nivell d'abstracció prou elevat com per poder integrar les dades sense cap complicació addicional.

Arribats en aquest punt, on s'havia definit un model més o menys definitiu, és quan es deriva el procés d'implementació del model a l'empresa externa, *BIME Analytics* perquè l'integri dins el *BigQuery*.

2.4 Implementació del model de dades

La fase d'implementació, ha estat un procés llarg i meticulós, és el procés més crític de tot el projecte. Per tant, ha estat necessari un control i una verificació constant de les dades que s'han integrat en les noves taules de *BigQuery*. Per la integració de les dades, es va tenir present el volum de dades i sobretot la tecnologia que s'estava utilitzant, ja que com veurem més endavant a vegades s'arriba gairebé al límit de la tecnologia.

Primerament, si es volen importar dades a *BigQuery*, s'ha de crear un contenidor (*Dataset*) que contindrà totes les taules. Un cop creat el contenidor, hi ha diverses mane-

¹⁰ És una eina de visualització de dades que mostra l'estat actual de les mètriques i indicadors clau de rendiment per a una empresa.

¹¹ És el resultat d'unir múltiples taules relacionades per aconseguir una única taula amb tota la informació.

¹² És un terminal punt de venda al núvol que s'ajusta a cada comerç i ofereix totes les eines necessàries per a l'administració i gestió del negoci

res d'importar les dades: en fulls de càlcul, programàticament, Google Cloud. Cada cop que s'importen dades, aquestes queden guardades en un fitxer comprimit dins el compte del núvol de Google (*GoogleCloud*). El mètode d'importació utilitzat aquí va ser emmagatzemant les dades a Google Cloud ja que es va fer a través d'una connexió que pots crear a *Bime*. Un cop hi ha totes les dades importades dintre el *BigQuery*, s'han de tractar totes des d'allà dins. Una característica comentada anteriorment és que en aquest sistema les taules són només de lectura (read only), això significa que per canviar algun valor s'ha de tornar a construir tota la taula. El gran avantatge és que crear una taula nova és relativament ràpid. En qüestió de segons *BigQuery* és capaç de duplicar milions de files. D'aquesta forma és com es va realitzar el procés de materialització de les taules planes.

El procés de materialitzar la taula de Vendes amb totes les dades referents a les fonts de dades originals ha estat bastant llarg i ha anat acompanyat de múltiples processos de comprovació d'integritat. Paral·lelament, es va treballar per a què totes les taules importades realitzessin un *delta load* actualitzant les dades.

En aquesta part del projecte es van haver de superar diversos imprevistos causats pel poc manteniment dels sistemes que s'utilitzen. Un cop l'empresa externa va acabar i va executar el *pipe*¹³ per a la càrrega de dades en temps real, ens van comunicar que hi havia algunes inconsistències en el sistema. Fent una sèrie de tests a la base de dades *MySQL* es va detectar que entre càrregues hi havia informació que es repetia.

Per poder comprendre millor la importància d'aquestes duplicitats, s'ha de conèixer l'origen de la informació. Entre d'altres dades, a la base de dades de *MySQL* s'hi emmagatzema un atribut existent a totes les taules importants, parlem del camp *modificado*, que és el que s'utilitza per escollir les files de les taules que es carregaran al *BigQuery* amb la càrrega de dades a temps real.

El camp *modificado* és de tipus *timestamp*¹⁴ i en ell s'hi guarda la data i l'hora en la que s'ha inserit la fila dins la base de dades. En un principi, la hipòtesi més clara de perquè es repetien les files era perquè aquest camp estava en mode *ON UPDATE CURRENT_TIMESTAMP* [6]. Això significa que quan la fila en concret s'actualitza aquest camp també canvia amb la data i l'hora d'aquell instant. Aquest fet provoca que quan s'executa el *delta load*, com que el camp *modificado* canvia amb una altra data diferent, en la següent execució del *delta* torna a agafar la fila i la puja al *BigQuery*.

La primera solució va consistir en eliminar aquest mode però no va sorgir efecte ja que el camp *modificado* seguia actualitzant-se, així que es va haver de buscar una

alternativa. Aquesta alternativa va consistir en crear un procediment (*trigger*¹⁵) que omplís un camp amb la data i l'hora actual en el moment de creació de la fila. Aquest procediment s'executa cada cop que s'insereix una fila a la taula. La solució es va aplicar correctament, però de la mateixa manera que la primera no va arreglar el problema, aquesta tampoc ho va aconseguir. Arribats a aquest punt es van continuar fent tests i proves en la base de dades i es va determinar que eren molt pocs els tiquets que es repetien, la qual cosa era bastant estranya. Es van analitzar els arxius de *log*¹⁶ del servidor de *BBDD* i es va descobrir que els *TPVs* sincronitzaven les files utilitzant la instrucció *REPLACE*.

Aquest fet provoca que els tipus de dades *TIMESTAMP* canviïn. Per tant, al no poder fixar el camp *modificado* s'arribà a la conclusió que s'estava tocant sostre amb la tecnologia ja que no hi havia cap forma de destriar els duplicats automàticament abans de realitzar el *delta load*. La decisió final que es va prendre va ser la de realitzar la mateixa càrrega que totes les altres taules, o sigui, realitzar un *delta load* quan el sistema estigui baixa càrrega.

Aquesta situació va provocar un fracàs en l'assoliment de l'objectiu de carregar dades en temps real. Tot i així, no s'ha fracassat totalment, ja que s'arribarà a assolir amb la integració del nou punt de venda. En un període a curt termini s'estima actualitzar els *TPVs* de les botigues amb el nou sistema de punt de venda basat en el núvol. Això significa que les dades generades pels *TPVs* s'enviaran a un nou sistema de base de dades. Per tant, això fa que s'hagi de construir un *pipe* per a la càrrega de dades en temps real per aquest sistema de gestió que, finalment, farà que es puguin consultar els informes i les estadístiques en temps real.

Paral·lelament a aquest problema de sincronització dels *TPVs*, va sorgir un altre de molt semblant. Quan es materialitzava la taula *Vendes* unint la taula de clients per poder disposar de la informació respectiva als clients, es produïa un error en l'execució d'aquest procés, que era degut a que la taula de clients contenia també duplicats. Hi havia múltiples instàncies de clients, clients falsos i, fins i tot, clients diferents amb el mateix número de targeta. Això va comportar un procés de correcció de dades que es va realitzar aplicant diverses consultes a la font de dades original.

El procés va consistir en una sèrie de consultes, dependent del nombre de duplicats. En general, si es vol eliminar la duplicitat en una taula, s'han d'agrupar les files a través de les seves claus primàries i contar el nombre de registres que comparteixen la mateixa clau. A partir d'aquí, quan tens el conjunt de registres duplicats agrupats per la seva clau primària, és on es fa l'eliminació de

¹³ El *pipe* o canal és una infraestructura lògica però on hi circulen dades que és volen transmetre cap a un destí.

¹⁴ El *timestamp* és un tipus de dades comú a les bases de dades i serveix per emmagatzemar la data i l'hora.

¹⁵ Un *trigger* (disparador) és un objecte amb nom dintre la base de dades el qual s'associa amb una taula i s'activa quan apareix en aquesta taula un event en particular.

¹⁶ Els arxius de *log* són un registre oficial d'events durant un rang de temps en particular.

duplicats. Quan els duplicats són idèntics s'utilitza una funció de comparació per escollir-ne un d'ells per a no ser eliminat, per exemple, podríem utilitzar la funció de *Max()*, ja que escull el valor màxim o en tot cas una de les dues instàncies si el valor és igual, tot depenent del sistema de gestió de bases de dades que s'utilitzi. Aquest seria, en el millor dels casos, ja que, com ha passat aquí, hi havia duplicats amb informació diferent en cada camp. Si succeeix aquesta situació, s'han de reassignar els registres repetits amb un altra clau primària manualment sinó és vol perdre informació. En aquest cas es va seguir el procediment d'eliminar i exportar-los en un còpia de seguretat per a reassignar-los posteriorment amb una clau primària nova.

Amb la taula *Vendes* implementada, en el procés de verificació de la integritat de les dades es va descobrir un problema referent als codis de targeta, en les taules de tiquets de *MySQL*. El problema sorgia quan s'unificava una taula de *Vendes* parcial amb la taula de clients carregada al *BigQuery*. Aquestes dues taules s'uneixen a través d'un índex que representa el codi de targeta de client. Un cop realitzat el procés d'unificació, s'observava que molts registres no incorporaven les dades referents al client del número de targeta. Aquest és un tipus de problema referent a la consistència de les dades, ja que els números de targeta aparentment coincidien fins que es va descobrir que l'atribut que representa el codi de targeta client de la taula parcial de *Vendes*, contenia espais en blanc. Visualment el codi de targeta semblava correcte, però internament, no tenia res a veure, ja que el mateix número en una taula, en l'altra contenia múltiples espais en blanc. Encara que el problema fos difícil de detectar, la solució no es va allunyar de la trivialitat, ja que aplicant una funció nativa de *MySQL* (*TRIM [6]*) es van poder corregir les dades esborrant els caràcters corruptes.

Tots aquests problemes comentats aquí, sorgeixen del mateix principi. És una qüestió fonamental prendre's el temps necessari a l'hora de definir l'esquema d'un model de dades, s'ha de tenir en consideració possibles migracions de dades en el futur. Per això, és molt important realitzar una reflexió prèvia abans de definir un model per així aconseguir una estructura prou abstracte com per evitar problemes futurs en la integritat o migració de les dades emmagatzemades.

Un altre aspecte força interessant a tenir en compte i valorat gràcies als problemes sorgits, són els testos i verificacions d'integritat de les dades. És realment pràctic realitzar bateries de proves sobre les dades i buscar incoherències en els resultats d'una consulta, ja que estem parlant de molts valors i seria difícil i costosa una comprovació rutinària d'aquests.

Hi ha tres punts principals on començar a buscar problemes d'inconsistència. El primer punt i el més lògic consisteix en agrupar els registres en claus primàries per a la cerca de registres duplicats. El segon punt va enfocat en la comprovació dels formats de dades per a que siguin els

correctes. Ja que un valor de caràcter numèric contingut en un camp amb format cadena de caràcters pot contenir caràcters inadequats que corrompin la integritat de la dada. L'altre punt principal consisteix en la cerca de coherència en les dades, això implica una cerca en busca de patrons a les estadístiques generades a través de la intel·ligència de negoci. Un exemple concret consistiria a contar el nombre de tiquets generats cada dia durant sis mesos. Si observem que en un dia el nombre de tiquets es desvia de la mitjana estàndard, hi ha una alta probabilitat de que la informació sigui errònia.

2.5 Disseny d'informes i indicadors clau

Per extreure informació útil sobre les dades emmagatzemades s'ha de fer un estudi previ per conèixer quines són les necessitats informatives i què es pretén obtenir amb les dades materialitzades a la taula per a la realització dels *dashboards*. Han estat necessàries tres tipus diferents de vistes o estadístiques: Estats, Semàfors i KPIs.

Els anàlisis d'estats són un simple punt de vista de les vendes realitzades en un instant. Hi ha determinats moments en que és necessari conèixer l'estat d'una botiga o saber els productes més demanats, etc. Aquí exposo els següents estats que són útils conèixer:

- *Hores de venda.* S'actualitzen al moment i mostren la botiga, l'import total dels tiquets des d'aquell moment, el número de tiquets i l'hora de l'últim tiquet. Hi ha diferents vistes sobre cada moment (avui, ahir, l'últim mes) seleccionat i sobre els diferents tipus de botigues, com les botigues iguals¹⁷. Aquest anàlisi tracta de conèixer les situacions de vendes per botiga en l'instant determinat.
- *Productes.* Mostren l'anàlisi de productes venuts classificats per famílies i subfamílies, mostrant l'import i el nombre de vendes acumulades.
- *Clients.* És important per l'empresa l'estudi dels usuaris. Aquesta vista mostra els clients nous per botiga i els clients perduts per botiga.

Els semàfors són un sistema per conèixer l'evolució d'una botiga en el temps. Mostren una comparació entre dates diferents. Els tiquets realitzats avui comparats amb els mateix dia un any anterior, una setmana abans i la mitjana en sis mesos. Si la comparació amb alguna d'aquestes dates és negatiu, es mostra un quadre vermell indicant una baixada de les vendes, pel contrari es mostra un quadrat verd. En la següent figura ens mostra un exemple de la situació concreta:

STORE	TODAY			ONE YEAR AGO		ONE WEEK AGO		AVERAGE (six months)	
	Tickets	Import	Last ticket done	Result	Tickets/Import	Result	Tickets/Import	Tickets/Result	Result
RAMBLA	394	1708	16:29:01	419	1749	454	1932	1970	87 %
BACERODÀ	215	972	16:01:07	172	749	186	819	815	119 %

Figura 3. Exemple d'un semàfor

¹⁷ Són les botigues que es poden comparar en el mateix període de temps que es mostren.

Les analítiques KPI o indicadors clau d'exercici mostren un conjunt d'estadístiques en un format concret referents a varis tipus d'informació. Existeixen entre d'altres:

- *Vendes sense targeta client.* Aquest tipus de vendes són totes les vendes de productes que realitza l'empresa sense estar pagats amb la targeta client.
- *Mermes.* Són els productes de rebuig que han estat registrats a causa d'haver sofert algun tipus de moviments que els ha malmès o bé perquè han passat la data de caducitat.
- *Estoc de sortida.* Indica l'increment de manca de productes a les botigues.
- *Diferències d'inventari.* Aquest indicador ens mostra la importància de les diferències quan els articles es compten. Ens mostra els resultats de les dades que volem en contra d'un altre període de temps que podem seleccionar.

Gràcies a tots aquests tipus de vistes, es poden generar *dashboards* que ofereixen informació de vital importància per a gestionar diferents parts de l'empresa, ja sigui en els departaments de màrqueting o en la part de logística per conèixer quin productes diferents s'han d'enviar a cada botiga.

3 RESULTATS

Amb la implementació de la taula de Vendes a dins del *BigQuery* ja s'extreu informació a través del programa de *BI*, per tant això suposa un grau de satisfacció ja que aquest era l'objectiu principal del projecte. A més a més els nivells d'assoliment dels objectius han estat quasi tots al 100%. Anem a analitzar l'assoliment dels diferents objectius en detall.

L'objectiu d'establir un sistema de presa de decisions basat en *SaaS* s'ha aconseguit ja que amb la taula *Vendes* materialitzada s'estan realitzant diferents estadístiques i indicadors per a la presa de decisions sobre el negoci.

L'objectiu d'unificar les fonts d'informació també s'ha assolit correctament ja que totes les fonts de dades originals han acabat integrant les dades necessàries a *BigQuery*. Gràcies aquest fet hem pogut materialitzar la taula *Vendes* i podrem materialitzar la taula de *Stocks* i la de *Moviments*.

Per contra, l'objectiu de treballar amb dades en temps real no s'ha pogut assolir. En un principi es van aconseguir carregar dades en temps real, però més tard es va parar el sistema a causa d'inconsistències en els processos de *delta load*. La càrrega a temps real es va suspendre, però amb l'arribada del nou *TPV*, els *delta loads* a temps real tornaran a ser possibles, ja que la forma en que entren els tiquets a la base de dades original és totalment diferent a l'anterior.

L'objectiu de gestionar gran quantitat d'informació de manera ràpida també està assolit ja que es una funcionalitat que ens dona *BigQuery*. La diferència entre un temps de resposta a *MySQL* i el temps de resposta a *BigQuery* és abismal ja que utilitza l'emmagatzematge per columnes i paral·lelitzat la consulta amb milers de servidors.

També s'ha aconseguit mantenir un control d'accés a la informació disponible, ja que el programa de *BI* ofereix diferents nivells d'accessibilitat, des de protecció de visibilitat davant algun camp de dades o grups de dades fins a la protecció d'un *dashboard* sencer. Una de les grans característiques és que permet vincular la compte de Google amb la compte d'accés al *dashboard*. Estalviant així haver de memoritzar paraules de pas addicionals.

Finalment, s'ha aconseguit visualitzar de forma intuïtiva la informació gràcies a les diferents possibilitats que ofereix *Bime* de mostrar la informació d'una manera entenedora. Tot i així, és el responsable de crear el *dashboard* qui ha d'aconseguir visualitzar la informació amb més o menys claredat, per a què tothom pugui entendre els resultats que mostren aquesta informació.

Vist l'assoliment dels objectius en detall, podem assegurar uns resultats força positius. Amb les dades materialitzades a la taula *Vendes* s'han generat els diferents estats, semàfors i *KPIs*. Entre les taules d'estats que s'han pogut generar hi ha totes les vistes referents a les vendes realitzades en botigues sobre els diferents períodes de temps a seleccionar. També s'han creat vistes sobre les vendes dels productes i tota la informació sobre els clients nous i els clients perduts per les botigues.

Un altre tipus d'estadística completada han estat els semàfors, ja que havent construït l'historial de vendes des de principis del 2006 fins a l'actualitat s'han generat els semàfors per cada instant de temps seleccionat.

En canvi, referent als indicadors clau, només s'han pogut generar els indicadors que poden sortir a través de la taula *Vendes*. Els indicadors referents a l'administració de productes encara no s'han pogut implementar ja que són necessàries les taules de *Moviments* i *Stocks*, que encara no estan materialitzades, però no trigarà gaire a obtenir-se una versió definitiva.

4 CONCLUSIÓ

En aquest apartat farem una valoració de la tecnologia utilitzada, repassarem l'assoliment d'objectius i exposarem quines són les línies de continuació del projecte.

Avui en dia els sistemes d'intel·ligència de negoci són una eina indispensable per a qualsevol empresa. Gràcies a aquesta tecnologia les empreses, a partir de totes les dades obtingudes dels seus sistemes d'informació, disposen d'indicadors per avançar-se als esdeveniments i prendre decisions de caràcter executiu per a obtenir guanys i evitar males inversions.

Una part important d'aquest mèrit també se l'emporten els sistemes de *Big Data*, ja que són els grans posseïdors de les dades que es generen i els responsables de la rapidesa en l'execució de les consultes que realitzen aquestes eines d'intel·ligència de negoci.

En aquest projecte ha estat un gran encert l'elecció de *BigQuery* com a sistema final d'unificació de totes les dades, ja que ens ha permès assolir els objectius plantejats. És un sistema molt robust i ràpid, es poden moure i modificar les dades molt fàcilment i la seva interfície és molt intuïtiva. A més a més ofereix una API per a l'execució de comandaments de manera programàtica en varis llenguatges. En definitiva es tracta d'una tecnologia molt potent. Tot i així, durant la construcció de les taules planes, va generar diversos errors interns que van haver de corregir els seus enginyers. És una eina potent però que encara està en fase de desenvolupament.

Un altre gran encert ha estat l'elecció de *Bime* com a programari d'intel·ligència de negoci. Cal destacar la seva flexibilitat i potència, ja que en el moment de crear una connexió, ell mateix ja fa la discriminació del que són mesures, dates i atributs. També presenta un tauler de creació d'estadístiques molt fàcil d'utilitzar, que fa que la corba d'aprenentatge sigui baixa. A més a més hi ha força documentació i el suport tècnic és bo, amb un temps de resposta a les incidències molt curt. Tot i això, també hem trobat alguns inconvenients. Disposa d'una versió d'escriptori una mica inconsistent ja que produeix bastants errors que en la versió en línia no es donen. Un altre inconvenient el trobem a l'hora de crear mesures i fer comparacions entre dates anteriors, ja que no ho permet amb totes les mesures que es volen. De totes formes, l'inconvenient més molest és que no totes les modificacions queden guardats, com les modificacions en els *dashboards* i que després no tenen efecte. De la mateixa forma que el *BigQuery*, aquesta és una eina puntera, que encara necessita acabar-se de desenvolupar.

Sobre les dades obtingudes de les fonts d'origen és molt important destacar que si es pretén utilitzar tecnologies de *Big Data* i programes d'intel·ligència de negoci, és molt important disposar d'un sistema origen molt ben definit. Amb això ens referim a que com més fàcil sigui d'abstreure el model de dades, més fàcil serà adaptar-lo en un altre sistema. És per això que els sistemes que es construeixen avui en dia han de tenir en compte aquests aspectes, ja que sinó es fa així, i es vol implantar un sistema de *BI*, la feina d'adaptació podria complicar-se innecessàriament. També és molt important planificar el manteniment i control de la integritat de les dades, per evitar que les dades es tamin intractables per culpa de duplicacions o dates incorrectes. El desenvolupament del nostre projecte ha patit retards importants degut precisament a aquests problemes.

través d'una empresa també han sorgit problemes pel que fa als retards en la implementació. Les comunicacions entre nosaltres i l'empresa externa han resultat molt lentes. El període de temps entre notificar un error i corregir-lo en molts casos ha estat de dies. Un altre dels problemes ha estat el de treballar amb un calendari molt ajustat, això fa que es treballi sota una pressió innecessària, que en moltes ocasions pot derivar amb problemes en el projecte.

Podem dir que hem assolit els objectius plantejats al inici del projecte. Hem aconseguit implementar un sistema de *BI* basat en *SaaS*, intuïtiu i fàcil d'utilitzar, amb dades a temps real i amb perfils d'usuaris amb diferents privilegis d'accés a les dades i *dashboards*.

Tot i que les parts planificades s'han realitzat correctament, queden obertes diferents línies de millora. Pel que fa al futur més proper, el projecte seguirà amb la tasca de materialització de la taula de *Moviments i Stocks*, per aconseguir nous indicadors, i paral·lelament amb la inserció de les dades del nou *TPV*, el *PROPPPOS*, a les taules planes.

A més llarg termini es millorarà la integritat de les dades, eliminant duplicats, i s'integraran dades més antigues a les taules planes (del 1999 fins al 2006). Per últim, el procés més ambiciós consisteix en la integració del *BI* de *Bime* a dins de *PROPPPOS*. Realitzant aquest últim procés es crearia una eina molt potent que combinaria el sistema de punt de venda basat en *SaaS* i el sistema de suport per a la presa de decisions.

AGRAÏMENTS

Vull donar les gràcies a tota la gent que m'ha ajudat a escriure i redactar aquest article sobretot al Jordi, el meu tutor del treball, per la paciència que ha tingut. També vull agrair a l'empresa on he realitzat el projecte, per la gran oportunitat que m'han donat.

BIBLIOGRAFIA

- [1] M. Ferguson, "Architecting a Big Data platform for analytics", Intelligent Business Strategies, Wilmslow Cheshire, UK, 2012
- [2] P. Russom. "Big Data Analytics", TDWI Research, Renton, WA, 2011
- [3] R.Sharma, M.Sood, "Cloud SaaS and Model Driven Architecture", in Proc. of the International Conference on Advanced Computing and Communication Technologies. 2011
- [4] K. Sato, "An Inside Look at Google BigQuery", Google Inc., Mountain View, CA, 2012
- [5] S. Harizopoulos, D. Abadi, P. Boncz. "Column-Oriented Database Systems", VLDB Tutorial, 2009
- [6] MySQL 5.0 Reference Manual. URL: <http://dev.mysql.com/doc/refman/5.0/es/triggers.html>