



UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR DE INGENIEROS INFORMÁTICOS

PROYECTO FIN DE CARRERA
PLAN 96

MARKETING INTELLIGENCE TRACKING TOOL

AUTOR: MARIO RODRIGÁLVAREZ GARCÍA
TUTOR: ANTONIO LATORRE DE LA FUENTE
FECHA DE PRESENTACIÓN: 9 JULIO 2018

ÍNDICE

1.	INTRODUCCIÓN	1
1.1.	CONTEXTO Y MOTIVACIONES	1
1.2.	DESCRIPCIÓN DEL PROBLEMA.....	2
1.3.	SOLUCIÓN PROPUESTA.....	3
2.	ESTADO DE LA CUESTIÓN	6
2.1.	SISTEMA ACTUAL.....	6
2.1.1.	MODELO DE NEGOCIO	6
2.1.2.	FUENTES DE INFORMACIÓN.....	8
2.1.3.	ARQUITECTURA DE SISTEMAS DE GSP.....	10
2.2.	TÉCNICAS Y METODOLOGÍAS APLICADAS EN BI.....	15
2.2.1.	DATA WAREHOUSE	17
2.2.2.	DATA MART	18
2.2.3.	PROCESOS ETL	18
2.2.4.	EL MODELADO MULTIDIMENSIONAL	19
2.2.5.	ENFOQUES PARA LA CONSTRUCCIÓN DE UN DW.....	20
2.2.6.	ESQUEMAS DE DISEÑO	22
2.2.7.	DIMENSIONES LENTAMENTE CAMBIANTES.....	23
2.3.	TECNOLOGÍAS.....	25
2.3.1.	BASES DE DATOS: MICROSOFT SQL SERVER	25
2.3.2.	PROCESOS ETL: PDI Y TOL	26
2.3.3.	CUBOS OLAP: MONDRIAN.....	30
2.3.4.	GENERACIÓN DE INFORMES: BIRT	36
2.3.5.	DESARROLLO WEB	36
2.3.6.	CONTENEDOR WEB: APACHE TOMCAT	46
3.	DESARROLLO DE LA SOLUCIÓN	47
3.1.	ANÁLISIS DE FUENTES DE INFORMACIÓN.....	48
3.1.1.	CBI	49
3.1.2.	ABI	50
3.1.3.	NET TICKETED SEGMENT.....	51

3.1.4. CRM ONE VIEW CX.....	52
3.2. MÓDULOS DE INFORMACIÓN	54
3.2.1. INFORMACIÓN COMÚN	55
3.2.2. CRM.....	60
3.2.3. BOOKINGS	61
3.2.4. TICKETING.....	68
3.3. PROCESOS DE CARGA.....	69
3.3.1. BOOKINGS	71
3.3.2. TICKETING.....	73
3.3.3. AGENCIAS DE VIAJES.....	74
3.4. APLICACIÓN WEB.....	78
3.4.1. DASHBOARD	80
3.4.2. DATA EXPLORER	95
3.4.3. REPORTS.....	98
3.4.4. MAINTENANCE	102
4. CONCLUSIONES Y LÍNEAS FUTURAS	105
4.1. CONCLUSIONES.....	105
4.2. LINEAS FUTURAS.....	106
BIBLIOGRAFÍA	108

ÍNDICE DE TABLAS

Tabla 1: Nomenclatura de los ficheros ABI y CBI	11
Tabla 2: Formato de los Office ID de Amadeus.....	12
Tabla 3: Combinación de campos empleada por GSP para identificar una oficina	13
Tabla 4: Uso de JDBC con Spring	45
Tabla 5: Información procedente de CBI.....	49
Tabla 6: Información procedente de ABI	50
Tabla 7: Información procedente de Ticketing	52
Tabla 8: Información procedente de CRM	53
Tabla 9: Descripción de una medida de hechos.....	54
Tabla 10: Descripción de una dimensión	54
Tabla 11: Dimensión Source.....	55
Tabla 12: Dimensión GDS	55
Tabla 13: Dimensión Airline	56
Tabla 14: Dimensión Vendor	56
Tabla 15: Dimensión Geography	56
Tabla 16: Dimensión Office	57
Tabla 17: Office Profile	58
Tabla 18: Dimensión Account.....	60
Tabla 19: Dimensión Segment.....	60
Tabla 20: Dimensión Match_Type.....	61
Tabla 21: Tablas de hechos agregadas de Air Bookings	63
Tabla 22: Medida Air Net Bookings	63
Tabla 23: Reservas de Hotel por tipo de proveedor	64
Tabla 24: Extracto de ABI Hotel Bookings	64
Tabla 25: Medida Hotel Net Bookings.....	65
Tabla 26: Medida Manual Hotel Bookings	65
Tabla 27: Medida Car Net Bookings	66
Tabla 28: Medida Car Manual Bookings.....	66
Tabla 29: Medida Rail Net Bookings.....	67
Tabla 30: Medida Rail SNCF Bookings	68
Tabla 31: Medida Net Ticketed	69
Tabla 32: Descripción de un proceso ETL.....	70
Tabla 33: Procesos ETL para el detalle de reservas.....	72
Tabla 34: Procesos ETL para los agregados de reservas	72
Tabla 35: Procesos ETL para el detalle de ticketing	73
Tabla 36: Procesos ETL para el agregado de ticketing	73
Tabla 37: Procesos ETL para Office	74
Tabla 38: Proceso ETL que carga el contenido procedente de CRM.....	75
Tabla 39: Lista de posibles valores del proceso "Match Type" (1).....	76
Tabla 40: Lista de posibles valores del proceso "Match Type" (2).....	77

ÍNDICE DE FIGURAS

Figura 1: Flujos de información del proyecto.....	3
Figura 2: Arquitectura de la solución propuesta.....	5
Figura 3: Modelo de negocio.....	7
Figura 4: Reservas aéreas por canal de distribución.....	7
Figura 5: Proceso de creación de CBI	9
Figura 6: Arquitectura de GSP	11
Figura 7: Interfaz de usuario construida en Microsoft Access	14
Figura 8: Ejemplo de informe Excel generado por GSP.....	15
Figura 9: Arquitectura general de una aplicación BI	16
Figura 10: Representación gráfica de un cubo de datos.....	19
Figura 11: Enfoque de Ralph Kimball	21
Figura 12: Enfoque de William H. Inmon	22
Figura 13: Esquema de estrella	22
Figura 14: Esquema de copo de nieve.....	23
Figura 15: Ejemplo de Transformación	27
Figura 16: Ejemplo de Trabajo	27
Figura 17: Interfaz gráfica de usuario de TOLBase	28
Figura 18: Acceso directo a TOL Process Manager en la barra de herramientas de TOLBase....	29
Figura 19: Interfaz gráfica de usuario de TOL Process Manager.....	29
Figura 20: Integración de Mondrian en Pentaho Business Analytics Server.....	31
Figura 21: Ejemplo de esquema XML.....	32
Figura 22: Ejecución de una consulta OLAP en Mondrian	33
Figura 23: Tabla dinámica de Pentaho Data Analyzer.....	34
Figura 24: Gráfico de barras de Pentaho Data Analyzer	34
Figura 25: Interfaz de Saiku Analytics	35
Figura 26: Interfaz de JPivot	35
Figura 27: Arquitectura de BIRT	36
Figura 28: Aplicación web con formato de escritorio	38
Figura 29: Gráfico generado con HighCharts	39
Figura 30: Diagrama de la posición de un servlet en una aplicación web	41
Figura 31: Ejemplo fichero JSP	42
Figura 32: Ejemplo MVC con Servlets y JSP.....	43
Figura 33: Drivers de JDBC	43
Figura 34: Esquema de la solución MIT.....	47
Figura 35: Arquitectura de sistemas de MIT	48
Figura 36: Segmentación de Marketing para agencias de viajes	53
Figura 39: Modelo de datos del módulo CRM	60
Figura 40: Tabla de hechos y dimensiones de Air Bookings	62
Figura 41: Tabla de hechos y dimensiones de un agregado de Air Bookings	63
Figura 42: Tabla de hechos y dimensiones de Hotel Bookings	65
Figura 43: Tabla de hechos y dimensiones de Car Bookings.....	66
Figura 44: Tabla de hechos y dimensiones de Rail Bookings	67

Figura 45: Tabla de hechos y dimensiones de Ticketing	68
Figura 46: Esquema de procesos en TOL Process Manager	70
Figura 47: Transformación creada para cargar los ficheros de CBI en tablas temporales.....	71
Figura 48: Trabajo que carga el contenido procedente de CRM	75
Figura 49: Modelo de datos de la base de datos interna de MIT	79
Figura 50: Componentes de MIT	80
Figura 51: Portada del Dashboard.....	81
Figura 53: Selectores del Dashboard.....	82
Figura 54: Botón “Reports” y mensajes informativos del Dashboard	83
Figura 55: Dashboard - Main View	83
Figura 64: Main View - Box 6.1.....	90
Figura 65: Main View - Box 6.2.....	90
Figura 66: Main View - Box 6.3.....	91
Figura 67: Main View - Box 6.4.....	91
Figura 68: Account View.....	92
Figura 69: Account View – Box 2.....	92
Figura 70: Account View – Box 6.....	93
Figura 71: Segment View.....	94
Figura 72: Segment View – Box 4.....	95
Figura 73: Segment View – Box 6.....	95
Figura 74: Visualización de un cubo en forma de tabla dinámica.....	97
Figura 75: Visualización de un cubo en forma de gráfico de barras apiladas	97
Figura 76: Generador de informes.....	98
Figura 77: Descarga de informes.....	98
Figura 78: Standard Pulse.....	99
Figura 79: Standard Pulse - Top Wins	100
Figura 80: Marketing Pulse.....	101
Figura 81: Ranking	101
Figura 82: Maintenance Interface	102
Figura 83: Filtros disponibles en Maintenance	103
Figura 84: Operaciones disponibles en Maintenance.....	104

LISTA DE ABREVIATURAS, SIGLAS Y ACRÓNIMOS

Con el fin de facilitar al lector la comprensión de este documento, se ha optado por introducir esta sección.

- ABI: Amadeus Billing Information
- ACO: Amadeus Commercial Office
- API: Application Programming Interface
- BI: Business Intelligence
- BIDT: Billing Information Data Transfer
- Bkg: Bookings
- BWS: Bayes Web Suite
- CBI: Competitive Business information
- CRM: Customer Relationship Management
- CSV: Comma-separated values
- DM: Data Mart
- DW: Data Warehouse
- ERP: Enterprise Resource Planning
- ETL: Extract, Transform y Load
- GCG: Global Customer Group
- GDS: Global Distribution System
- GSP: Global Sales Planning
- IATA: International Air Transportation Association
- JDBC: Java Database Connectivity
- JS: JavaScript
- JSON: JavaScript Object Notation
- JSP: Java Server Pages
- KPI: Key Performance Indicator
- MDX: MultiDimensional eXpressions
- MIDT: Marketing Information Data Tapes
- MIT: Marketing Intelligence Tracking
- OLAP: Online Analytical Processing
- OLTP: Online Transactional Processing
- PDI: Pentaho Data Integration
- PNR: Passenger Name Record
- SCD: Slowly Changing Dimensions
- SGBD: Sistema Gestor de Bases de Datos
- SQL: Structured Query Language
- SSIS: Microsoft SQL Server Integration Services
- TI: Tecnologías de la Información
- Tkt: Ticketing
- TOL: Time Oriented Language
- XML: Extensible Markup Language
- YTD: Year to Date

MARKETING INTELLIGENCE TRACKING TOOL

Desde los datos al conocimiento

Hoy en día las empresas generan un gran volumen de información a la cual no se suele extraer todo el potencial ya que se realiza un uso precario de las herramientas analíticas que tienen a su alcance. Estas herramientas informáticas facilitan el análisis y evaluación de la situación actual del sector, permitiendo el descubrimiento de tendencias y patrones que determinan de forma fiable la probabilidad de un suceso o resultado en el futuro, lo que permite anticiparse y posicionarse frente a nuevas oportunidades y tendencias.

Por tanto, toda empresa con fuertes necesidades de información y conocimiento sobre sí misma y el mercado en que desarrolla su negocio, no puede prescindir de todo lo que incluye el concepto de Business Intelligence (BI) o inteligencia de negocio.

BI consiste en el proceso de reunir, depurar, analizar y transformar la información en el campo del negocio para extraer un cierto conocimiento que pueda ser utilizado como ayuda a la toma de decisiones, mejorando la empresa en cuanto a estrategia, gestión, evaluación del mercado y futuras tendencias. Actividades como el marketing, ventas, bases de datos de clientes, información sobre la competencia y toda la información relevante para la empresa entran en este proceso. Una vez extraída, depurada y transformada la información en estructuras o modelos de datos optimizados para su almacenamiento, su explotación se realiza a través de herramientas analíticas que resultan de fácil uso para los usuarios, los encargados de las tomas de decisiones de la empresa.

Con este trabajo queremos ayudar a una empresa en términos de BI, por un lado mejorando el actual repositorio de sus datos comerciales y, por otro, proporcionando un fácil acceso a los mismos, en el momento oportuno, para mejorar los procesos de toma de decisiones y obtener el éxito empresarial en un mercado tan competitivo y dinámico como el que existe hoy en día.

La estructura de este documento es la siguiente:

En el capítulo 1 se introduce el problema que se pretende resolver en este proyecto.

A continuación, en el capítulo 2 analizamos el sistema existente y hacemos una revisión de la literatura, describiendo las diferentes tecnologías y metodologías de desarrollo existentes que se han utilizado a lo largo del proyecto, de manera que el lector se familiarice con términos que aparecen en los apartados posteriores.

Posteriormente, en el capítulo 3 se documentan en detalle todas las correcciones efectuadas sobre la base de datos actual, previo análisis de las fuentes de información existentes, lo que nos llevará a desarrollar una nueva versión optimizada de la misma y sus procesos de carga. La integración y acceso a los datos que nos ofrece esta nueva base de datos resulta muy importante pero no suficiente, ya que una empresa con estas necesidades precisa de una solución BI, la cual es descrita en detalle en esta sección.

Por último, en el capítulo 4 se reflejan varias ideas y conclusiones alcanzadas en el momento de finalizar el trabajo así como una serie de mejoras y sugerencias sobre las que se podría ampliar el proyecto en un futuro.

Al final del documento se encuentra la bibliografía, donde se incluyen todos los documentos y páginas web consultadas para la elaboración de la memoria del proyecto.



1. INTRODUCCIÓN

Este capítulo contiene una descripción del problema inicial, origen de todo el trabajo asociado a este proyecto de fin de carrera, así como los objetivos que se pretenden alcanzar con la solución propuesta.

En primer lugar, se contextualiza el proyecto, mencionando las necesidades y motivaciones por las que se desarrolla e indicando el resultado que se desea alcanzar al final del mismo.

A continuación, se realiza una descripción de alto nivel del problema que se pretende resolver en la que se establece una lista de objetivos, seguida de la solución propuesta para alcanzarlos y el rol del alumno en cada uno de ellos.

1.1. CONTEXTO Y MOTIVACIONES

Este trabajo aborda un proyecto de inteligencia de negocio (Business Intelligence – BI) para la ayuda a la toma de decisiones estratégicas de un departamento de marketing de la empresa Amadeus, llamado Global Sales Planning (GSP). El proyecto ha sido desarrollado por la empresa de consultoría Bayes Forecast, en la que el alumno prestaba sus servicios como Ingeniero de Datos antes de pasar a formar parte de la plantilla de Amadeus.

Amadeus es una compañía de referencia en la provisión de soluciones tecnológicas para el sector mundial de los viajes y el turismo, que se ha convertido en uno de los principales sistemas globales de distribución (Global Distribution System – GDS) del mundo. Opera en 195 mercados de todo el mundo para ofrecer el mejor contenido, productos y servicios para agencias de viajes (Travel Agencies - TA). Dentro de este ámbito de actividad, Amadeus compite con otros GDS, siendo los más importantes: Sabre, Galileo, Worldspan y Abacus.

El grupo cuenta con unos 16.000 trabajadores en todo el mundo (Amadeus IT Group, S.A., 2017), que prestan sus servicios desde:

- la sede central en Madrid,
- el centro de desarrollo de software en Sophia Antipolis, cerca de Niza (Francia),
- la base de operaciones de Erding, cerca de Munich (Alemania) y
- varias organizaciones comerciales (Amadeus Comercial Organizations – ACO) repartidas por diferentes países en todo el mundo.



El promotor de esta iniciativa, GSP, es un equipo central que dispone de una base de datos interna con información de clientes y la quiere integrar con el sistema corporativo para la gestión de clientes, CRM, donde las diferentes ACO mantienen la información de las agencias de viajes que operan en su mercado (por ejemplo, la ACO de España se encarga de las agencias de España, Portugal y Andorra).

Además, desea equipar a sus clientes internos - el Comité Ejecutivo y los equipos regionales de ventas - con una herramienta web fácil de usar, que les permita consultar la evolución de las ventas, comprender de manera detallada cuál ha sido el rendimiento de una región o mercado determinado, con la capacidad de bajar al nivel de cliente para tomar decisiones comerciales.

1.2. DESCRIPCIÓN DEL PROBLEMA

Tal y como se comenta en el punto anterior, el proyecto llevado a cabo trata de resolver un problema de BI, para el cual el cliente, el equipo GSP de Amadeus, requiere una solución que le permita identificar problemas de inconsistencia de información, proporcione un fácil acceso a los datos de ventas y permita realizar diversos análisis de los mismos con el fin de averiguar tendencias, sopesar hipótesis, iniciativas... que ayuden a la toma de decisiones.

Este cliente busca un proveedor que:

1. Evalúe la arquitectura actual de la base de datos del Data Warehouse (DW) que GSP está utilizando y construya uno nuevo que permita a Amadeus identificar y corregir, de forma optimizada, datos de registros de transacciones capturados por sus ACO. Amadeus ha requerido el uso de Microsoft SQL Server para este cometido.
2. Revise y optimice los procesos de carga de datos de ventas, de manera que se simplifique el mantenimiento de la información de clientes.
3. Desarrolle un módulo de mantenimiento de la información de clientes.
4. Desarrolle un sistema de cubos OLAP multidimensionales, para generar informes que permitan la explotación dinámica de la información de la base de datos.
5. Implante una herramienta web que integre lo anterior, de modo que permita a Amadeus tener en un mismo sitio los datos de sus clientes, reduciendo el tiempo requerido para producir los informes comerciales, según el proceso y arquitectura actuales. Los clientes internos de GSP accederán a esta herramienta para ver sus informes e indicadores comerciales relevantes con mayor flexibilidad que la actual y con capacidad de personalización.



La figura 1 muestra de forma esquemática la arquitectura y los flujos de información del proyecto:

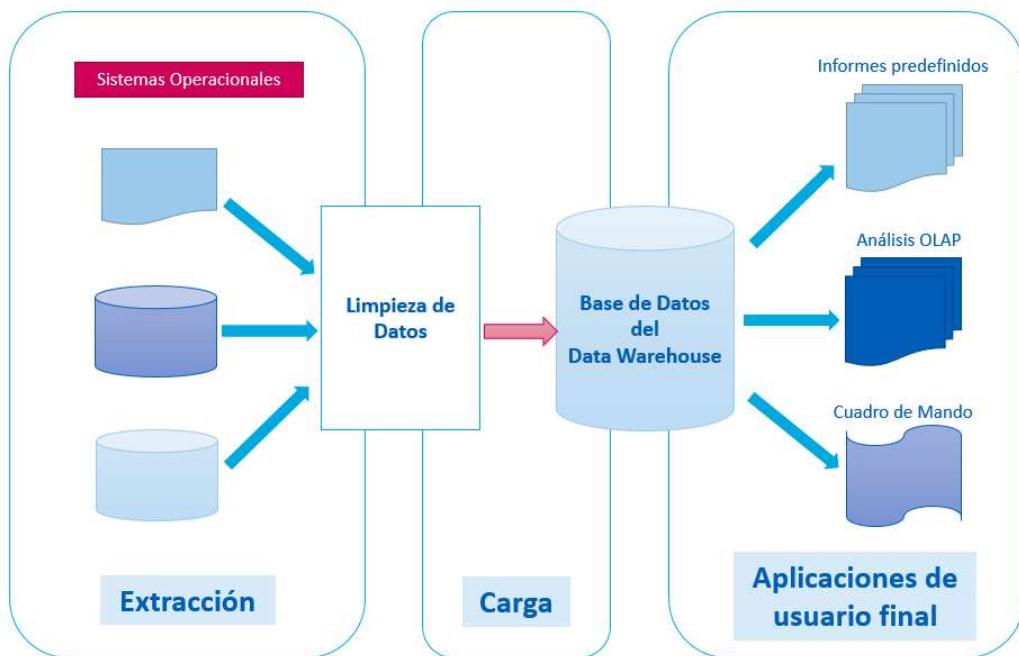


Figura 1: Flujos de información del proyecto

En resumen, es un proyecto de BI que consiste en el desarrollo de una base de datos analítica con sus correspondientes procesos de carga de datos, así como la explotación de la información a través una herramienta web.

1.3. SOLUCIÓN PROPUESTA

Bayes Forecast es una empresa de analítica de datos que desarrolla sistemas de medición, previsión y decisión con los datos de operación de sus clientes.

Entre sus soluciones ofrece la Bayes Web Suite (BWS), una herramienta web, accesible desde cualquier navegador, que combina todas las bondades de un software de BI tradicional con la funcionalidad de los modelos matemáticos desarrollados por esta empresa en un lenguaje de programación propio, llamado TOL (Time Oriented Language).

El almacenamiento de la información está centralizado, siendo el diseño y la implementación del modelo de datos de cada proyecto, una de las tareas principales del departamento de Ingeniería de Datos del que formaba parte el alumno. En este caso se desarrollará una nueva base de datos, junto con sus procesos de carga, en base a la información y requerimientos facilitados por el cliente.



La BWS fue creada por otro departamento tecnológico de Bayes Forecast, Productos, para facilitar al usuario final el acceso a los datos y modelos desarrollados por Ingeniería de Datos y el departamento de Modelación. Dispone de diferentes módulos personalizables en función de su uso, por lo que se estima que puede encajar con los objetivos de Amadeus aunque requiere el desarrollo de módulos adicionales. A continuación, se muestran los módulos originales de este producto y su papel en este proyecto:



Data Explorer: cubos OLAP que facilitan la comprensión de los datos históricos en función de los indicadores y medidas definidas por el plan de negocios del usuario.



Report Builder: permite visualizar y crear informes dinámicos que se actualizan automáticamente.



Chart Builder: crea presentaciones de los indicadores del negocio mediante un conjunto de gráficos predefinidos. Con estas visualizaciones se pretende crear un cuadro de mando que ofrezca una visualización intuitiva de la actividad comercial de Amadeus.



Simulador: ofrece al usuario la interacción directa con los modelos matemáticos creados por Bayes para aplicarlos en determinados escenarios. Es decir, permite obtener previsiones bajo diferentes escenarios construidos por los usuarios, editando los valores de los inputs en la misma interfaz. Este módulo queda fuera del alcance de este proyecto ya que Amadeus no deseaba hacer ningún tipo de previsión hasta que los datos de su negocio fueran confiables.



Forecast Manager: herramienta de trabajo que permite controlar el flujo de tareas del proceso de previsión. Está muy orientada a tareas de previsión, fuera del ámbito de este proyecto, al menos en esta fase.



TOL Dispatcher: herramienta de control de las actividades de TOL.

Adicionalmente, se requiere el desarrollo de una interfaz para el mantenimiento de los datos de cliente integrada en la BWS, como un módulo adicional.

Esta propuesta permite la generación de informes y la explotación de la información de este negocio, tal y como el equipo de GSP demanda, aunque tendrá cambiar el modo en que realiza muchas de las tareas que viene realizando.



La figura 2 resume la solución BI propuesta:

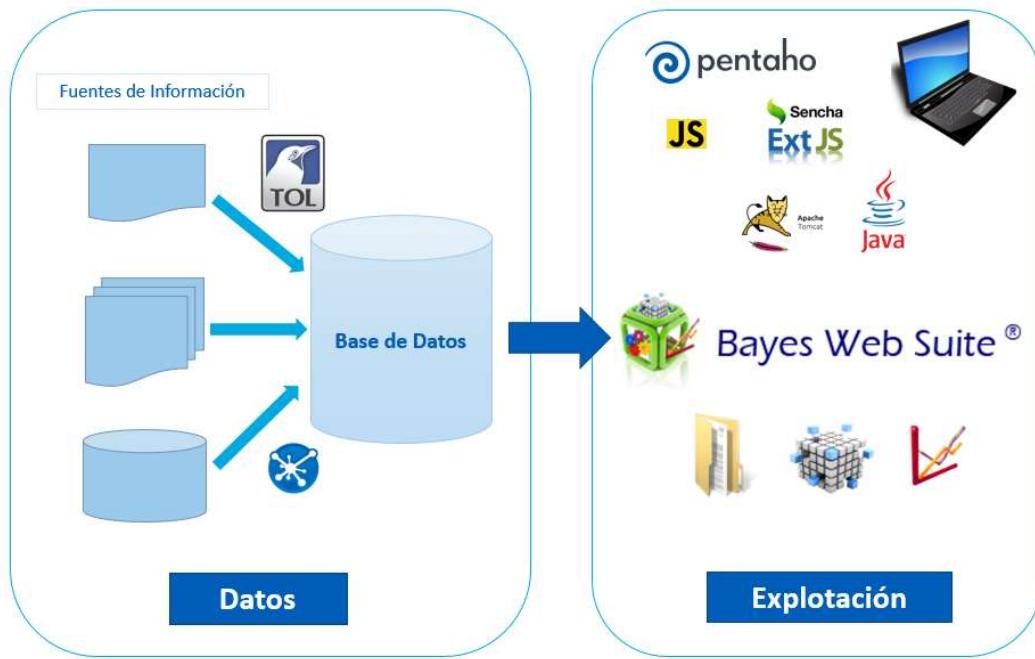


Figura 2: Arquitectura de la solución propuesta



2. ESTADO DE LA CUESTIÓN

En este capítulo se hace un repaso de la situación en el momento de emprender este trabajo, tanto a nivel del proyecto, como a nivel del estado del arte.

Comenzaremos explicando el modelo de negocio de Amadeus, las fuentes de información disponibles y la actual arquitectura de sistemas utilizada por GSP. A continuación, explicaremos las distintas técnicas y metodologías utilizadas a lo largo de la elaboración del proyecto, de acuerdo con las necesidades tecnológicas que se han requerido. Estos conceptos comprenden tanto la definición de un almacén de datos o data warehouse (DW), como lo que es un data mart (DM), y también el modelo multidimensional de datos.

Posteriormente, haremos un recorrido por la arquitectura y las tecnologías presentes en el proyecto. La primera herramienta que se explica es Microsoft SQL Server, que es el sistema gestor de bases de datos que Amadeus quería utilizar como soporte de información. A continuación, se habla de TOL, el lenguaje de programación con el que se han implementado la mayoría de los procesos de extracción, transformación y carga de datos (ETL), entrando en detalle con el framework utilizado, el TOL Process Manager. También se explica otra herramienta ETL complementaria que se ha utilizado, Pentaho Data Integration (PDI).

Finalmente, se describen las tecnologías que están presentes, de un modo u otro, en la solución propuesta. Se introducen los lenguajes, frameworks y productos presentes, tanto en el back-end (Java, Spring, BIRT, Mondrian) como en el front-end (JavaScript, Sencha Ext JS, HighCharts), dejando claro aquellos en los que el alumno no ha tenido que implementar nada. Además, se especifica el entorno en que está desplegada la BWS (Apache Tomcat instalado en Red Hat Linux).

2.1. SISTEMA ACTUAL

En este apartado se describe la situación existente en GSP al inicio del proyecto. Se explican el modelo de negocio de Amadeus, las fuentes de información disponibles y la actual arquitectura de sistemas utilizada por GSP

2.1.1. MODELO DE NEGOCIO

Cuando un usuario quiere reservar un billete de avión, puede optar por varios métodos de reserva (figura 3):



- A través de una agencia de viajes, ya sea en una de sus oficinas (ej. Viajes El Corte Inglés) o por Internet (ej. Expedia, Odigeo).
- Mediante el canal directo de una aerolínea, bien en las oficinas comerciales (Air Transport Op-ATO/Commercial Transport Op-CTO) o la página web de la aerolínea.

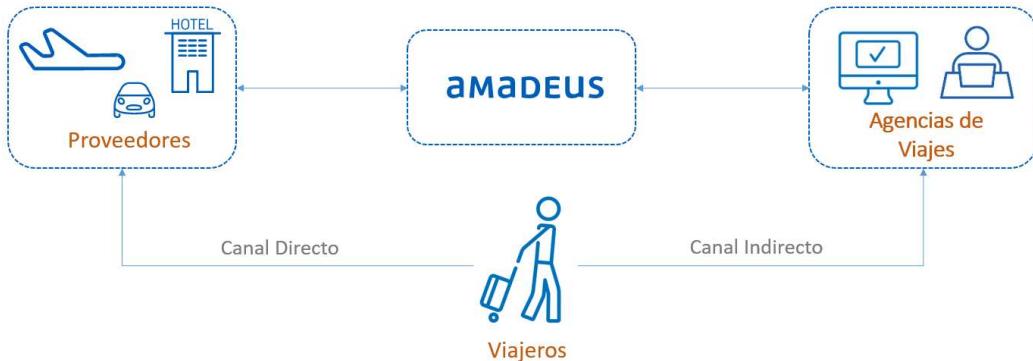


Figura 3: Modelo de negocio

La actividad comercial de Amadeus para agencias de viajes (en la figura 4, canal indirecto) se puede dividir en 2 áreas principales:

- Actividad de GDS: reservas efectuadas a través de los GDS
- Actividad de TI: productos y soluciones ofrecidas a las agencias de viajes

Cada reserva hecha por una agencia de viajes a través de un GDS de los servicios prestados por un proveedor, se define y cuenta como una reserva, que es el elemento utilizado para fines de negociación comercial con agencias de viajes y para medir el desempeño comercial de los diferentes actores de la industria.



Figura 4: Reservas aéreas por canal de distribución



Por tanto, el modelo de datos de GSP incluye información sobre proveedores (empresas según su tipo de actividad), agencias de viajes y GDS.

2.1.2. FUENTES DE INFORMACIÓN

Teniendo en cuenta el modelo de negocio descrito anteriormente, pasamos a describir el tipo de información disponible en el canal indirecto, ámbito de acción de GSP.

Los GDS disponen de dos tipos de productos analíticos de datos:

- BIDT (Billing Information Data Transfer): específico para cada proveedor (ej. una aerolínea) y sus datos de facturación.
- MIDT (Marketing Information Data Tapes): contiene información sobre reservas aéreas realizadas por agencias de viajes de todo el mundo conectadas a algún GDS, por ejemplo, el de Amadeus.

Una de las particularidades de este negocio es que los principales GDS del mundo (Amadeus, Sabre y Travelport) han acordado compartir la información de MIDT entre sí. Esto les ayuda a comprender el rendimiento del canal indirecto y desarrollar estrategias de marketing.

En MIDT está disponible casi toda la información almacenada en el registro de nombre del pasajero (Passenger Name Record - PNR), donde podemos encontrar, entre otras cosas, información del punto de venta o del itinerario, es decir, origen, conexiones, paradas, aerolíneas operativas o de marketing. Sin embargo, en nuestro caso no usamos datos personales del pasajero sino solo información de la agencia y de la aerolínea.

Dicho esto, hay que resaltar que la información de MIDT es incompleta. Solamente contiene las reservas realizadas a través de un GDS, lo que se denomina canal indirecto. Es decir, los GDS capturan transacciones de reserva generadas por agencias de viajes que utilizan su plataforma, es decir, Amadeus solo captura las reservas realizadas por agencias que utilizan el sistema de reservas de Amadeus, Sabre por agencias que utilizan el sistema Sabre, etc.

Por tanto, las reservas de pasajeros realizadas directamente (canal directo) con una aerolínea, a través del correspondiente sitio web, oficina, teléfono... no se registran en esta fuente de datos. Debido a esta circunstancia, la mayoría de aerolíneas de bajo coste (Low Cost Carrier - LCC) no aparecen en MIDT, solo aquellas que tienen acuerdos de distribución limitados con algunos GDS. Tampoco se capturan las agencias de viajes que utilizan tecnologías de reservas de "conexión directa" ya que omiten a los GDS.

Por otra parte, muchas agencias de viajes están acreditadas como agentes IATA (International Air Transport Association/Asociación de Transporte Aéreo Internacional).



Se trata de un sello de aprobación reconocido en todo el mundo que les autoriza para vender billetes nacionales o internacionales en nombre de las aerolíneas.

IATA es una asociación para la cooperación entre aerolíneas, promoviendo la seguridad, fiabilidad, confianza y economía en el transporte aéreo. Su objetivo principal es crear medios de colaboración entre las compañías aéreas y entre estas y la industria del turismo, especialmente el sector de las agencias de viajes. Tiene 54 oficinas en 53 países y representa a unas 280 aerolíneas de 120 países (International Air Transport Association, IATA by Region).

Amadeus dispone de un paquete de datos de marketing que contiene información de la comunidad mundial de agencias de viajes acreditadas como agentes IATA.

Las principales fuentes de información que GSP utiliza y que alimentan este proyecto están basadas en extractos generados a partir IATA, BIDT y MIDT por un departamento técnico de Amadeus. A continuación, encontramos una breve descripción de las mismas:

- CBI (Competitive Business Information): construido a partir de los archivos de MIDT, proporciona información precisa de la actividad (reservas aéreas) realizada en el GDS informado (Amadeus, Abacus, Travelport y Sabre) por las agencias de viajes. Esta fuente ofrece cifras mensuales de participación de mercado para un mejor análisis de la evolución de la industria, desde una perspectiva global hasta explorar en detalle la actividad de una sola oficina. La información de la oficina se complementa con los atributos obtenidos del paquete de datos de IATA. En la figura 5 se muestra un esquema del proceso de creación de esta fuente. Disponemos de datos desde enero 2005.

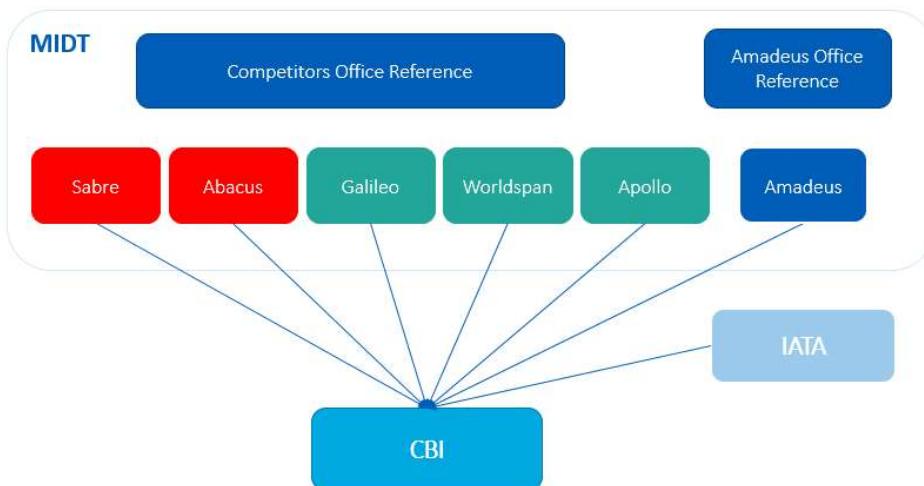


Figura 5: Proceso de creación de CBI



- ABI (Amadeus Billing Information): basado en BIDT, contiene información de facturación (billing) de Amadeus en los negocios Air, Hotel, Car y Rail. Disponemos de datos de reservas mensuales desde enero 2006.

Ambas fuentes contienen, principalmente, información de los GDS con los que se ha tramitado la reserva, datos de la oficina que ha tramitado la misma, es decir, la agencia de viajes (cliente para los GDS como Amadeus), y datos del proveedor (aerolínea, hotel, etc.) al que se le ha vendido la reserva o que al final ofrece el servicio. Al parecer, se puede haber vendido la reserva en una aerolínea (que es al que al final se le cobra) pero el vuelo puede ser operado por otra compañía aérea. En cuanto a los datos de reservas, se dispone de distintas medidas numéricas (active add, cancel, passive, modify...) pero la única que se explota es la cantidad neta (net bookings).

Las agencias de viajes pueden ser clientes de varios GDS, por lo que un mismo cliente puede estar identificado en cada GDS de diferente manera. Este es uno de los problemas que tiene GSP con estas fuentes de datos, ya que los datos asociados a la oficina, es decir, el Office Profile construido a partir de los ficheros MIDT de cada GDS y la información proporcionada por IATA, no es consistente y en muchos casos no les permite identificar fácilmente a la agencia de viajes.

2.1.3. ARQUITECTURA DE SISTEMAS DE GSP

La arquitectura interna de sistemas que usa GSP (figura 6) consiste en:

- un servidor de base de datos Microsoft SQL Server, donde existe una base de datos con la actividad comercial de Amadeus.
- archivos Microsoft Access con tablas vinculadas a la base de datos.
- archivos Microsoft Excel con macros para generar informes a partir de los archivos Access.

La información, procedente de las fuentes CBI y ABI, se recibe mensualmente en archivos CSV y se carga en tablas temporales gracias a diferentes procesos de carga básicos, implementados con la herramienta Microsoft SQL Server Integration Services (SSIS). Después, mediante la ejecución manual de varios scripts Transact-SQL, se transforma la información de las tablas temporales y pasa al modelo optimizado de base de datos de GSP.

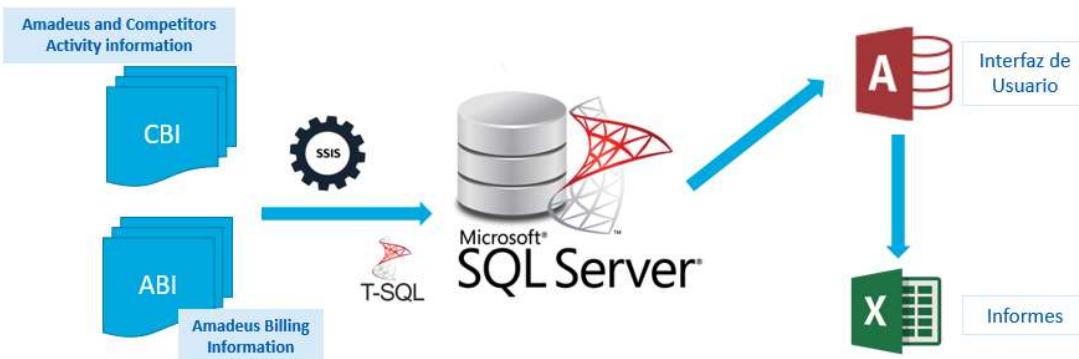


Figura 6: Arquitectura de GSP

La nomenclatura utilizada en estos ficheros es la siguiente:

YYYYMM_XXX_Region_NN_Negocio.csv

YYYY	Año al que corresponden los datos
MM	Mes al que corresponden los datos
XXX	Nombre de la fuente, en este caso CBI o ABI
Region	Región a la que pertenecen los datos (APAC, LATAM, CESE, WE, MEA, NA) o WWi en caso de que sea sin filtros de región o no se ajuste a estas
NN	Número de fichero de esa región en el mes, en caso de haber más de uno
Negocio	Air, Car, Hotel, Rail. Dado que estos ficheros se almacenan en directorios diferentes según el negocio, entendemos que no es obligatoria esta nomenclatura

Tabla 1: Nomenclatura de los ficheros ABI y CBI

En los procesos de carga de los datos de reservas de viajes, entre otras acciones, se agrupan las oficinas de las agencias presentes en CBI y ABI, a partir de los campos del Office Profile, con el propósito de asignar correctamente a cada agencia las reservas de viaje que ha realizado.

Respecto a la agrupación de agencias de viajes realizada por GSP, a continuación ofrecemos una visión más detallada de la situación actual. Resulta necesario para comprender los procesos del sistema.

En el Office Profile de la información de negocio disponemos de un campo llamado GDS Office Code / Office ID, cuyo nombre nos haría pensar que es un identificador único pero existen inconsistencias que impiden identificar de manera única a una oficina con este ID.



1. Cada GDS maneja sus propios Office ID por lo un mismo ID puede existir en distintos países, o incluso en el mismo, identificando a diferentes agencias de viajes. Entre los competidores de Amadeus tenemos a Travelport, fruto de la unión de Galileo y Worldspan. Para este GDS, tenemos casos en que el mismo ID aparece en diferentes países para agencias de viajes independientes.
2. Los Office ID se reciclan (en Amadeus han de pasar 6 meses desde el último uso) por lo que el mismo ID puede haber sido utilizado por más de un cliente a lo largo del tiempo. Esto provocará que tengamos un mismo ID con varios Office Name y/o IATA y, por tanto, perteneciente a diferentes agencias.
3. Una agencia de viajes puede ser un agente acreditado por IATA, en cuyo caso tendrá asociado un código IATA pero esta circunstancia puede ser temporal o, como sucede en el caso de agencias pequeñas que suscriben acuerdos con otras más grandes (ej. Thomas Cook, Carlson Wagonlit, Hogg Robinson, Aerticket...) de manera puntual, puede ir cambiando en función del código IATA de la agencia global.

Por otra parte, el formato de los Office ID de Amadeus y la competencia es diferente. Los de Amadeus (tabla 2) tienen una longitud fija de 9 caracteres alfanuméricos, cuya nomenclatura es la siguiente: XXXYYCNNN

CÓDIGO	DESCRIPCIÓN
XXX	Código IATA que identifica la ciudad o aeropuerto en que está ubicado el terminal
YY	Código corporativo para aerolínea, agencia de viajes, hotel, automóvil, etc.
C	Código corporativo para identificar el tipo de oficina
0	System user airlines
1	Participating airlines
2,3	Travel agencies
4	Hotel Provider
5	Car provider
6	Tour Provider
7	Surface provider
8	Other Provider
9	Wholesaler
NNN	Número de oficina

Tabla 2: Formato de los Office ID de Amadeus

Debido a estas razones, se requiere la utilización de otros campos para identificar a una oficina de manera única y asignarla a la agencia de viajes que corresponda.

Además, aunque los Office ID de Amadeus aparecen en CBI y ABI, no coincide al 100% la información de los campos del Office Profile de ambas fuentes. En ocasiones, tenemos el mismo Office ID con diferentes valores:



- Country Code. CBI contiene el país en el que está ubicada la oficina mientras que ABI se rige por cuestiones de facturación, habiendo casos en que una agencia de Andorra (AD en CBI) se factura en España (ES en ABI).
- City. En CBI aparecen valores que no se corresponden con el código del Office ID, totalmente alineado en ABI. Suelen deberse a errores en el procesamiento de MIDT para generar CBI.

A pesar de existir inconsistencias de este tipo se ha intentado tener una dimensión común de Office. GSP emplea 12 campos para este cometido. Se trata de una tabla que contiene el listado de combinaciones de los siguientes campos:

Country Code	GDS Code	GDS Office Code / Office ID		GDS Office Name / Office Name		IATA Number
Account Name	Trade Name	Address Line 1	Address Line 2	Address Line 3	Postal Code	City of Agency

Tabla 3: Combinación de campos empleada por GSP para identificar una oficina

La granularidad de esta dimensión es enorme y el rendimiento de las consultas a nivel oficina no es óptimo.

Respecto al agrupamiento de estas oficinas en agencias de viajes, durante el proceso de carga se clasifican automáticamente las oficinas nuevas en base a los valores de los campos que contienen el nombre de agencia (Office Name, Account Name y Trade Name). Básicamente, se comparan estas oficinas con las antiguas para encontrar similitudes. De no ser así, se crea una nueva agencia de viajes con el nombre que viene en Account Name.

Dado que la actual base de datos de GSP no utiliza ningún tipo de información procedente de CRM, parte del equipo se dedica a revisar cada mes las oficinas asignadas automáticamente en el sistema, para garantizar la calidad de los datos. Aplican sus propias reglas internas, comprobando la información de los campos de la dimensión office. Por ejemplo, si se dispone de IATA Number, se comprueba en la página web de IATA el nombre de la agencia y se corrige, si es necesario; o se busca en Internet si existe alguna agencia de viajes en alguno de los campos de dirección de la oficina. Del mismo modo, validan manualmente la información disponible en CRM, o directamente preguntan a la ACO y a los gerentes de cuentas.

El equipo de GSP es el responsable de la gestión de esta agrupación. Para este propósito, los usuarios se conectan a la base de datos de Microsoft SQL Server a través de tablas vinculadas de un archivo de Microsoft Access (figura 7) que además dispone de varios formularios. Contiene formularios simples para modificar atributos o campos directamente en las tablas vinculadas y formularios más complejos, codificados en



Visual Basic for Applications (VBA), para realizar tareas de mantenimiento en la agrupación de oficinas que no están automatizadas.

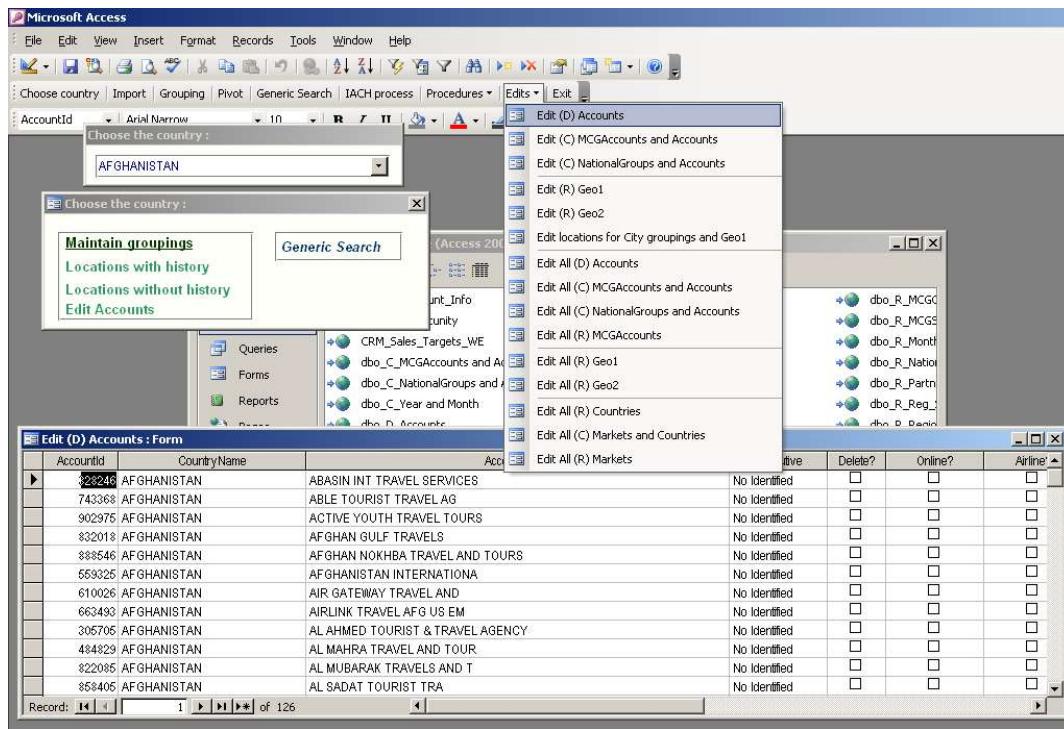


Figura 7: Interfaz de usuario construida en Microsoft Access

GSP entrega mensualmente una serie de informes estándar a las unidades comerciales basados en Microsoft Excel (figura 8), en los que suele comparar 2 períodos de tiempo (ejemplo: datos de 2017 vs 2016, mes en curso contra el mes anterior o contra el mismo mes en el año anterior). El proceso para producir estos informes requiere:

1. Ejecución de un script Transact-SQL para generar en tablas agregadas los datos mensuales de reservas por agencia de viajes.
2. Generación en Microsoft Access, mediante la ejecución de macros VBA, de tablas temporales con los datos que se utilizarán en el informe, en función de los períodos de tiempo que se desean comparar.
3. Creación de los informes, en formato Excel, a través de la ejecución de diversas macros VBA que se conectan a las tablas temporales del archivo Access del paso anterior y generan diversos ficheros con los datos correspondientes.



WESTERN EUROPE (SOUTH) Pulse the 'Sharemeter'							
April 2012 4 months comparation (Jan12 - Apr12)							
Prepared by: Global Sales Planning							
GDS Booking Volumes ('000)							
(figures in thousands)							
4 months at end April 2011							
Grand Total	Amadeus	Galileo	Sabre	W/Span	Abacus	Total	
Online	39,723	15,462	6,612	3,928	-	65,724	
MCG	10,726	1,791	868	1,395	-	14,781	
Shared	9,304	1,386	868	1,078	-	12,636	
Above 100	430	-	-	279	-	709	
10 to 100K	411	-	-	277	-	689	
Below 10K	13	-	-	-	-	13	
Negative	6	-	-	1	-	7	
Local	992	405	1	38	-	1,436	
Above 100	926	373	-	36	-	1,336	
10 to 100K	62	29	1	(0)	-	91	
Below 10K	5	3	-	2	-	10	
Negative	-	-	-	-	-	(169)	
Offline	28,996	13,670	5,743	2,533	-	50,943	
MCG	7,973	3,886	1,447	54	-	13,361	
Shared	5,823	678	360	82	-	6,943	
Above 100	5,211	437	294	82	-	6,025	
10 to 100K	568	227	60	-	-	855	
Below 10K	44	13	6	-	-	63	
Negative	-	-	-	-	-	-	
Local	15,200	9,106	3,926	2,397	-	30,639	
Above 100	5,478	3,921	758	1,315	-	11,472	
10 to 100K	5,624	2,575	1,631	643	-	10,473	
Below 10K	4,035	2,605	1,498	416	-	8,554	
Negative	63	5	49	22	-	139	
Growth in volume							
Grand Total	Amadeus	Galileo	Sabre	W/Span	Abacus	Total	
Online	418	(561)	(85)	(728)	-	(957)	
Growth in %							
Grand Total	Amadeus	Galileo	Sabre	W/span	Abacus	Total	
Online	7.9%	-10.3%	1.2%	-20.5%	0.0%	2.6%	

Figura 8: Ejemplo de informe Excel generado por GSP

Por tanto, estos informes Excel no son dinámicos y su creación implica un procesamiento manual significativo. Una vez creados, se almacenan manualmente en una carpeta compartida con permiso restringido o se envían a los usuarios finales por correo electrónico.

2.2. TÉCNICAS Y METODOLOGÍAS APLICADAS EN BI

Hoy en día, el término BI, Business Intelligence o inteligencia de negocio, es uno de los más empleados a nivel empresarial cuando hablamos de gestión de datos. Las empresas generan con su actividad una gran cantidad de información que se ha convertido en un activo muy importante. Esto les permite comprender su pasado, su presente y decidir sobre su futuro, proporcionándoles así una clara ventaja competitiva frente a empresas que carecen de este tipo de soluciones.

BI comprende el conjunto de herramientas orientadas a extraer y transformar los datos almacenados en los sistemas de la empresa, propiciando su análisis y conversión en información y conocimiento, lo que se traduce en una visión más profunda del comportamiento de ciertos aspectos de su negocio. De esta manera, se otorga a las compañías el apoyo que necesitan en la toma decisiones para resolver las preguntas de negocio rápidamente.



Centrándonos en una visión más técnica y atendiendo a la arquitectura general de una aplicación BI tradicional, la figura 9 muestra el esquema básico en un proyecto de este tipo (Dertiano, ¿Qué es Business Intelligence?, 2014):

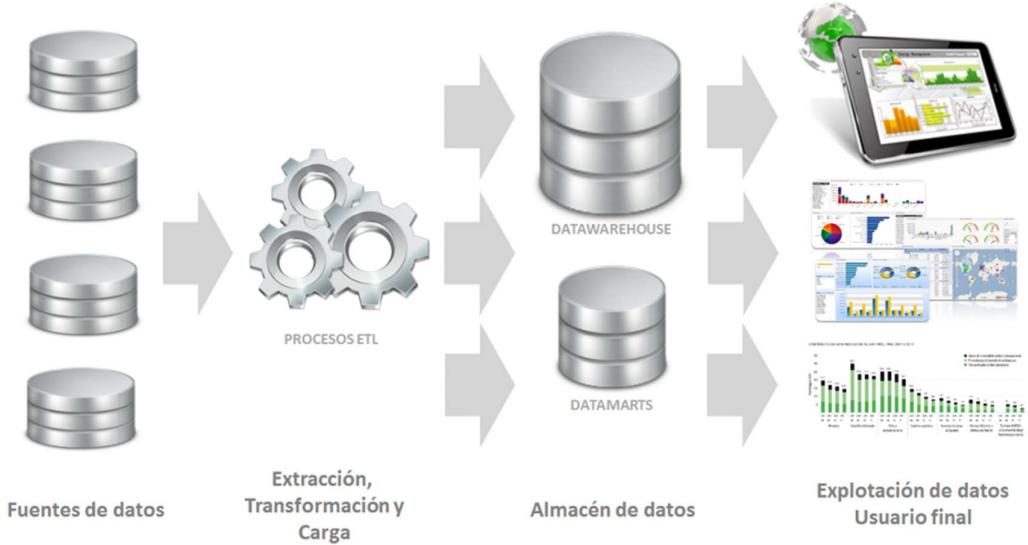


Figura 9: Arquitectura general de una aplicación BI

Un proyecto BI consiste en extraer los datos de las distintas fuentes que se encuentren en la compañía, unificándolos y transformándolos, en base a las futuras preguntas a las que se quiera dar respuesta, para cargarlos finalmente en un sistema integral centralizado. A continuación, entran en juego las herramientas de explotación, especializadas en la visualización de la información, propiciando su análisis a través de informes, cuadros de mando, etc. para dar respuesta a las preguntas de negocio.

Normalmente, los datos fuente provienen de entornos transaccionales (sistemas OLTP: Online Transactional Processing), enfocados a que cada operación (transacción) trabaje con pequeñas cantidades de filas, y a que ofrezcan una respuesta rápida. Como ejemplo de este tipo de sistemas, podemos citar cualquier aplicación con la que el usuario interactúa para introducir datos operacionales al sistema, es decir, un ERP (Enterprise Resource Planning), un CRM (Customer Relationship Management), aplicaciones financieras, departamentales, de ventas, etc.

Los procesos de extracción, transformación y carga de datos se conocen por las siglas ETL (Extract, Transform y Load) y el sistema integral centralizado mencionado anteriormente, recibe el nombre de data warehouse (DW) o almacén de datos, aunque también podría tratarse de un data mart (DM) que, compartiendo los aspectos técnicos y las características generales de un data warehouse, está orientado a un proceso o un área de negocio específica.



2.2.1. DATA WAREHOUSE

Un data warehouse (DW) o almacén de datos es un repositorio de datos corporativos estratégicos, tácticos y operativos en el que se integra información depurada de las diversas fuentes que hay en la organización y que los diseñadores del DW consideran que puede ser de interés (Ramos, 2016).

Esta información debe ser homogénea y fiable, se almacena de forma que permita su análisis desde muy diversas perspectivas y, a su vez, facilite la consulta y la extracción de información de grandes volúmenes de datos con unos tiempos de respuesta óptimos.

Según Bill Inmon, un DW se caracteriza por ser (Wikipedia, Data warehouse):

- **Orientado a temas:** los datos están organizados por temas para facilitar el entendimiento por parte de los usuarios, de forma que todos los datos relativos a un mismo elemento de la vida real queden unidos entre sí. Por ejemplo, todos los datos de un cliente pueden estar consolidados en una misma tabla, todos los datos de los productos en otra, y así sucesivamente.
- **Integrado:** los datos se deben integrar en una estructura consistente, debiendo eliminarse las inconsistencias existentes entre los diversos sistemas operacionales. La integración es uno de los factores esenciales para la calidad de los datos en un DW y, por lo tanto, esencial para los resultados obtenidos con el fin de soportar el proceso de la toma de decisiones. Algunas de las inconsistencias más comunes que nos solemos encontrar son: en nomenclatura, en unidades de medida, en formatos de fechas, múltiples tablas con información similar (por ejemplo, varias aplicaciones con tablas de clientes). Por otra parte, la información se debe estructurar en diversos niveles de detalle para adecuarse a las necesidades de consulta de los usuarios.
- **Histórico (variante en el tiempo):** los datos, que pueden ir variando a lo largo del tiempo, deben quedar reflejados de forma que al ser consultados reflejen estos cambios y no se altere la realidad que había en el momento en que se almacenaron, evitando así la problemática que ocurre en los sistemas operacionales, que reflejan solamente el estado de la actividad de negocio presente. Un DW debe almacenar los diferentes valores que toma una variable a lo largo del tiempo, registrando además la fecha en la que ese valor estuvo o está vigente.
- **No volátil:** la información de un DW debe mantenerse para futuras consultas, es decir, ha de ser permanente, de sólo lectura, a no ser que se requiera la corrección de errores. Por tanto, solamente existen dos tipos de operaciones: la carga de los datos y el acceso a los mismos.



2.2.2. DATA MART

Un DW puede contener información muy diversa de los distintos departamentos de una organización. Muchas veces, uno de estos departamentos, o un área específica del negocio, requiere que la porción de datos que a ellos les interesa reciba un tratamiento diferente, ya sea para elaborar una estructura óptima de los datos desde la que analizar la información contenida, o por motivos de seguridad, de mejora de prestaciones, etc... Dichas estructuras orientadas a una solución analítica se llaman data marts (DM) (Ramos, 2016).

Esto no significa que un DM sea más fácil de construir. Un DM puede ser un subconjunto lógico, a través de vistas sobre los datos del DW, o físico, migrando y transformando los datos del DW a una nueva base de datos. Por tanto, la principal diferencia entre estos dos tipos de bases de datos, DW y DM, será su alcance.

2.2.3. PROCESOS ETL

Un DW/DM se carga periódicamente y en él se unifica información procedente de múltiples fuentes, creando una base de datos que cumple una serie de características descritas anteriormente. Esto implica la existencia de una serie de procesos que leen los datos de las diferentes fuentes, los transforman y adaptan al modelo que hayamos definido, los depuran y limpian, y los introducen en esta base de datos de destino.

Estos procesos que organizan el flujo de los datos que alimentan el DW/DM se conocen por las siglas ETL (Extract, Transform y Load), ya que se encargan de la extracción de fuentes de datos heterogéneas, transformación en un modelo único de representación del negocio y carga en el DW (Wikipedia, Extract, transform, load).

A continuación se detallan brevemente las partes de un proceso ETL:

Extracción	Esta fase consiste en la obtención de datos de las fuentes de origen. Estos orígenes o fuentes de datos pueden ser: bases de datos relacionales, bases de datos no relacionales, ficheros, etc. Otro aspecto a considerar es si optamos por una actualización completa, que es mucho más simple, o si optamos por una actualización incremental, que es lo más conveniente.
Transformación	En esta fase se realizan los cálculos, agregaciones, transformaciones de campos..., que se necesiten para preparar la información a almacenar en el DW. Los datos deben ser depurados para eliminar inconsistencias, discrepancias y duplicidades. Estas transformaciones suelen conllevar cambios con respecto a la estructura de origen para adaptarla al destino, cambios en el contenido de los valores de origen y creación de nuevos valores en las filas de destino.
Carga	En esta parte del proceso se recogen los datos procedentes de la fase de transformación y se incorporan finalmente a la base de datos de destino.



2.2.4. EL MODELADO MULTIDIMENSIONAL

El modelado multidimensional es el nombre de una técnica de diseño lógico que a menudo se utiliza para realizar el diseño de los DW. La idea subyacente es que toda organización se puede representar como un cubo cuyas celdas contienen valores medidos y cuyas aristas representan las dimensiones naturales de los datos (figura 10). Es obvio que se permiten más de tres dimensiones, con lo que, en puro rigor, no es un cubo, sino un hipercubo, aunque sean los términos de cubo y cubo de datos los más utilizados en la literatura (Menasalvas Ruiz, 2007):



Figura 10: Representación gráfica de un cubo de datos

Un modelo dimensional se compone de tablas de:

- Hechos: tabla que almacena medidas para medir el negocio, como las ventas, los ingresos, el número de reclamaciones o el coste de las mercancías.
- Dimensiones: tablas que contienen varias columnas y atributos utilizados para describir los distintos aspectos de un proceso de negocio. En otras palabras, una tabla de dimensión almacena detalles acerca de los hechos. Por ejemplo, si desea determinar los objetivos de ventas, puede almacenar los atributos de dichos objetivos en una tabla de dimensiones. De esta manera, se pueden agrupar objetivos de ventas anuales por país, producto o cliente, y dichas agrupaciones se almacenarán en tablas de dimensiones.

Las tablas de hechos contienen claves foráneas a las tablas de dimensiones, de manera que cada fila de datos de la tabla de hechos se relaciona con sus correspondientes dimensiones y niveles. Mientras que una dimensión consiste en una o más tablas que contienen claves y atributos que describen los valores de los datos.

En ocasiones, pueden existir dimensiones que se relacionen entre sí jerárquicamente, siendo una de ellas un conjunto más agregado de otra (por ejemplo, una dimensión provincia y una dimensión comunidad autónoma donde los posibles valores de una se pueden agrupar y conformar los posibles valores de la otra). A nivel de almacenamiento, estas jerarquías pueden estar normalizadas, formando una tabla diferente por cada



nivel jerárquico, o desnormalizadas, acumulando en una misma tabla una columna para cada nivel.

Una práctica habitual, a la par que recomendable en el diseño de un DW, es la coexistencia de tablas de hechos principales y tablas de hechos agregadas en función del modo en que se utilice la información en la organización. Las tablas de hechos agregadas son tablas con un grano de información más grueso que las tablas principales, y que permiten acceder a los resultados de ciertas operaciones de un modo inmediato.

Dependiendo de las necesidades de información de la organización, algunas tablas agregadas pueden dejar de ser necesarias y puede requerirse la definición de otras nuevas. No obstante, la utilización de agregados debe ser transparente al usuario, de modo que los usuarios no sabrán si la información obtenida proviene de la tabla de hechos detalle (principal) o de una tabla agregada, aunque experimentarán unos tiempos de respuesta mejores.

2.2.5. ENFOQUES PARA LA CONSTRUCCIÓN DE UN DW

Para la construcción de un DW predominan dos enfoques, basados en las filosofías de Ralph Kimball y Bill Inmon. (Rangarajan, 2016) (Abramson, 2002)

Kimball defiende una metodología de trabajo “Bottom-up”, cuyo procedimiento consiste en empezar por pequeños componentes, DM departamentales, que se van construyendo a medida que son requeridos para desarrollar soluciones de acceso de información. De esta forma, el modelo irá evolucionando a estructuras superiores que conformen el DW. Es decir, para Kimball, un DW no es más que la unión de los diferentes DM de una organización.

La principal ventaja de este enfoque es que, al estar formado por pequeños DM orientados a la extracción rápida mediante la consulta y generación de informes, el DW al completo puede ser explotado directamente por las herramientas de reporting y análisis de datos sin la necesidad de estructuras intermedias.

El argumento que avala esta filosofía es que las preguntas procedentes de las unidades de negocio son impredecibles, de manera que el DW tiene que ser flexible para dar respuesta a todas ellas rápidamente, garantizando la exploración de los datos y la navegación a través de jerarquías desde datos agregados hasta información desagregada. A este tipo de arquitectura, representada en la figura 11, Kimball lo denomina “arquitectura en bus” y los cuatro pasos fundamentales que se han de seguir para construir este tipo de base de datos son (Dertiano, Arquitectura BI (Parte III): El enfoque de Ralph Kimball, 2015):



- Identificación del proceso de negocio que se pretende estudiar
- Definición de la granularidad de los datos
- Selección de las dimensiones y atributos
- Identificación de los hechos o métricas

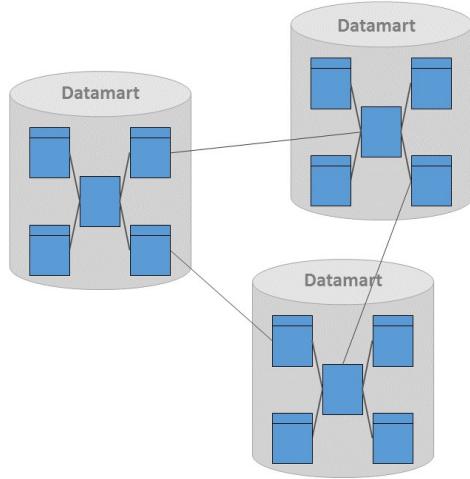


Figura 11: Enfoque de Ralph Kimball

Por otra parte, Bill Inmon propone un enfoque “Top-Down” (Dertiano, Arquitectura BI (Parte II): El enfoque de William H. Inmon, 2015). La figura 12 muestra este enfoque, que consiste en diseñar primero una solución global para la compañía y, una vez que tenemos el DW generado de esta manera, crear los DM para las áreas de negocio que lo necesiten. De esta manera, se asegura la integración de la información ofrecida. Normalmente, esto requiere más tiempo de desarrollo, ya que implica que se tengan que realizar posteriormente los DM asociados a las soluciones analíticas y además puede que se terminen incluyendo en el DW datos que nunca serán consultados por carecer de valor para la toma de decisiones, ya que los objetivos finales de extracción de información no son prioritarios al inicio.

En cuanto a la estructura interna del DW, para Inmon la prioridad es que el modelo de datos esté normalizado (3FN: tercera forma normal). El proceso de normalización consiste en aplicar una serie de reglas o normas a la hora de establecer las relaciones entre los diferentes objetos dentro de la base de datos. Con este proceso de normalización, se consiguen muchos beneficios, como evitar la redundancia de los datos, mantener su integridad referencial, facilitar el mantenimiento de las tablas y disminuir el tamaño de la base de datos. Sin embargo, a diferencia de los DW desnormalizados, la explotación de los datos exige el empleo de consultas SQL más complejas, lo que puede dificultar el análisis directo de la información y el uso de las herramientas de informes.

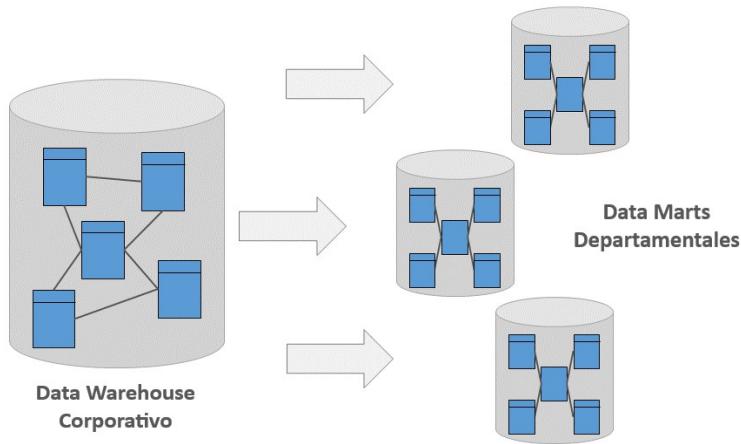


Figura 12: Enfoque de William H. Inmon

2.2.6. ESQUEMAS DE DISEÑO

Un esquema de diseño es una forma de representar gráficamente un modelo de datos a través de un diagrama. En el diseño de un DW, podemos optar por dos tipos de esquemas:

- Esquema de estrella: consta de una sola tabla de hechos central rodeada por un conjunto de tablas de dimensiones, con la particularidad de que cada dimensión es una tabla de una sola dimensión. Estos esquemas son útiles para consultas simples por lo que es una opción utilizada generalmente en las tablas agregadas para acelerar las consultas.

En la figura 13 se muestra un esquema de estrella con una sola tabla de hechos y cuatro tablas de dimensiones (IBM Knowledge Center, Esquemas de estrella):



Figura 13: Esquema de estrella



- Esquema de copo de nieve: en este modelo de datos las tablas de hechos están conectadas a tablas de dimensiones de niveles múltiples, es decir, cada dimensión puede tener más de una tabla que representa exactamente un nivel en una jerarquía. Estos esquemas son útiles para consultas complejas. Las tablas detalladas son ejemplos comunes de este paradigma.

En la figura 14 se muestra un esquema de copo de nieve con dos dimensiones, cada una con tres niveles (IBM Knowledge Center, Esquemas de copo de nieve):

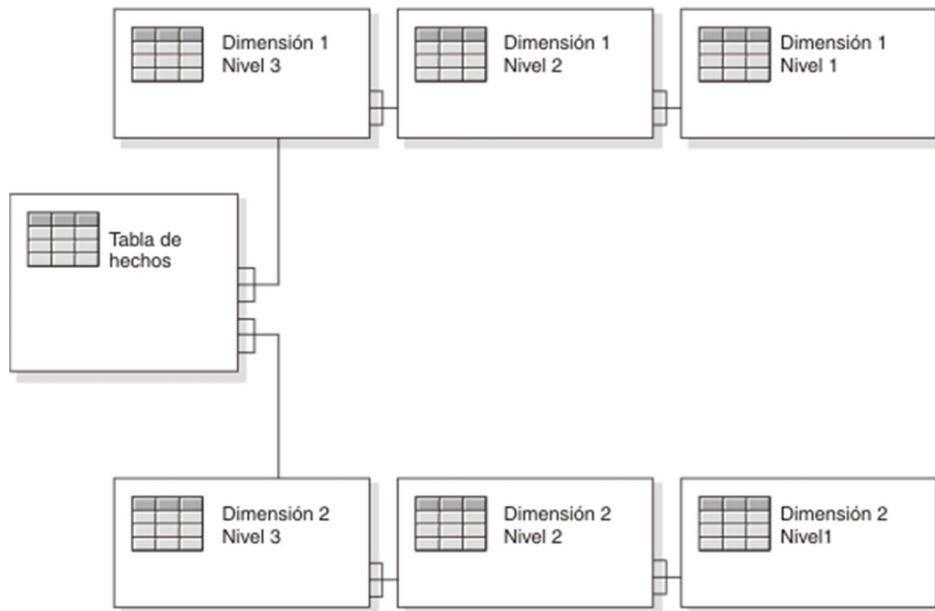


Figura 14: Esquema de copo de nieve

2.2.7. DIMENSIONES LENTAMENTE CAMBIANTES

Otro aspecto a tener en cuenta en el diseño de un DW es la gestión del tiempo. En el mundo real, dimensiones como cliente o empleados pueden cambiar a lo largo del tiempo sus datos de contacto (teléfono, correo electrónico, residencia...), se casan, se divorcian, tienen más hijos, etc. Ante estas circunstancias, es preciso tomar una decisión sobre cómo manejar estos cambios ya que una de las principales responsabilidades del DW es representar la historia correctamente. (Kimball, 2008)

La respuesta a esta problemática, definida como dimensiones lentamente cambiantes (SCD - Slowly Changing Dimensions), no consiste en incorporar todo en la tabla de hechos o hacer todas las dimensiones dependientes del tiempo, puesto que esto complicaría el modelo de datos con consecuencias desastrosas de pérdida de inteligibilidad y rendimiento. Desde un punto de vista práctico, cuando ocurren estos cambios, se puede optar por seguir alguna de estas dos grandes opciones:



- Registrar el historial de cambios
- Reemplazar los valores que sean necesarios

Inicialmente, Ralph Kimball expuso tres estrategias para tratar las SCD: tipos 1, 2 y 3; pero a través de los años esta lista se ha ampliado con más tipos SCD, por ejemplo: tipos 4 y 6. (Ross M. , 2013)

A continuación se describen algunas estrategias SCD:

- Tipo 0: No hacer nada. Según este enfoque, el valor del atributo de dimensión nunca cambia, por lo que los hechos siempre se agrupan por este valor original.
- Tipo 1: Sobrescribir. Este tipo consiste en sobrescribir el dato antiguo con el nuevo. Es sencillo de implementar ya que, si bien no guarda los cambios históricos, tampoco requiere ningún modelado especial y no necesita que se añadan nuevos registros a la tabla. Se considera apropiado en casos en donde la información histórica no sea importante de mantener, tal como sucede cuando se debe modificar el valor de un registro porque tiene errores de ortografía.
- Tipo 2: Añadir fila. Registra datos históricos mediante la creación de múltiples registros con claves diferentes. Con el tipo 2 se preserva toda la historia mediante la inserción de un nuevo registro cada vez que se hace un cambio. Esta estrategia requiere que se agreguen columnas adicionales a la tabla de dimensión, para que almacenen el historial de cambios (fecha de inicio y fin de la vigencia).
- Tipo 3: Añadir columna. Registra los cambios utilizando columnas separadas. Esta estrategia requiere incorporar a la tabla de dimensión una columna adicional por cada una de las columnas de las que se desea mantener un historial de cambios, por lo que el número de cambios está limitado al número de columnas dedicadas para mantenimiento de datos históricos.
- Tipo 4: Tabla de histórico separada. La función básica de esta técnica es almacenar en una tabla adicional los detalles de cambios históricos realizados en una tabla de dimensión. En este caso, una tabla almacena los datos actuales y se utiliza una tabla adicional para almacenar los cambios de valores. Esta tabla adicional indicará, por ejemplo, qué tipo de operación se ha realizado (Insert, Update, Delete), los campos afectados y en qué fecha.
- Tipo 6: Híbrido. Este método combina los estrategias originales: 1, 2 y 3 (1 + 2 + 3 = 6), de ahí su nombre.

En resumen, lo único que hemos de tener en cuenta a la hora de diseñar la estrategia de carga de una tabla de dimensión lentamente cambiante es si queremos guardar la historia de los cambios o no. (Carrera, 2017) (Kimball & Ross, 2013)



2.3. TECNOLOGÍAS

En los siguientes apartados se describen las tecnologías presentes en el proyecto.

2.3.1. BASES DE DATOS: MICROSOFT SQL SERVER

Microsoft SQL Server es un sistema gestor de bases de datos (SGBD) relacional desarrollado por la empresa Microsoft. Puede ser configurado para utilizar varias instancias en un mismo equipo. Asimismo, cada instancia administra varias bases de datos, del sistema o de datos de usuario.

Además, a partir de la versión 2005, permite dividir el contenido de cada una de las bases de datos en esquemas. Trabajar con esquemas permite disponer de más flexibilidad en la administración de los permisos de objeto de base de datos. Un esquema es un contenedor con nombre para objetos de base de datos, que permite agrupar objetos en espacios de nombres independientes. La sintaxis de asignación de nombres para hacer referencia a los objetos es:

Server.Database.DatabaseSchema.DatabaseObject

El lenguaje de desarrollo utilizado por este SGBD relacional es Transact-SQL (T-SQL), una implementación del estándar ANSI del lenguaje SQL (Structured Query Language). T-SQL permite realizar las operaciones clave en SQL Server, incluyendo la creación y modificación de esquemas de base de datos, creación de tablas y definición de relaciones entre ellas, inserción y modificación de datos en la base de datos, así como la administración del servidor como tal. Además, dispone de control de transacciones (TRANSACTION, COMMIT, ROLLBACK,...), declaración de variables, procesamiento condicional del código (IF, ELSE, CASE...), manejo de errores, cursos para el procesamiento de cada fila, etc.

Para acelerar la recuperación de datos de una tabla o vista, Microsoft SQL Server dispone de índices. Un índice es una estructura en disco o en memoria, asociada con una tabla o vista. No obstante, si se utiliza un gran número de índices en una tabla, el rendimiento de las instrucciones INSERT, UPDATE, DELETE y MERGE se verá afectado, ya que todos los índices deben ajustarse adecuadamente a medida que cambian los datos de la tabla.

Microsoft SQL Server también tiene la opción de definir y ejecutar procedimientos almacenados, escritos en scripts T-SQL. A estos procedimientos se les da un nombre y si se requiere, se le pueden definir parámetros de entrada, de manera que posteriormente se invocan como si fuera una función.



2.3.2. PROCESOS ETL: PDI Y TOL

Actualmente existen muchas herramientas orientadas al diseño y desarrollo de procesos ETL sin la necesidad de programarlas exclusivamente en código. Algunas son muy potentes, como IBM Datastage, Oracle Data Integrator o SQL Server Integration Services, pero también tienen un alto coste de licencias. Como alternativa, tenemos algunas herramientas Open Source como Talend o Pentaho Data Integration (PDI).

En la empresa Bayes Forecast, se apuesta por el desarrollo de procesos ETL combinando la herramienta gráfica de Pentaho, PDI, con la programación en TOL, un lenguaje creado en la compañía, muy orientado a su área de negocio: análisis estadístico y modelación de procesos.

Una de las razones por las que se utiliza PDI es su compatibilidad con la mayoría de SGBD existentes en el mercado (PostgreSQL, Oracle, MySQL, Microsoft SQL Server...). Para este proyecto en particular, dado que Amadeus utiliza Microsoft SQL Server, se planteó la posibilidad de usar SQL Server Integration Services (SSIS), la herramienta ETL disponible en esta suite de Microsoft. Sin embargo, puesto que inicialmente el proyecto se desarrolló en los servidores de Bayes Forecast, esta opción fue descartada debido a que SSIS solo está disponible en las ediciones de pago desde que fue lanzado con Microsoft SQL Server 2005, para reemplazar a Data Transformation Services (DTS), la herramienta ETL previa de Microsoft SQL Server (MikeRayMSFT, 2017).

PENTAHO DATA INTEGRATION

Pentaho Data Integration (PDI), anteriormente conocido como Kettle (Bernabeu, 2010), es una herramienta gráfica para crear procesos ETL que importen datos desde diferentes orígenes de datos (Excel, CSV, XML, SGBD, etc.) a una gran variedad de sistemas de almacenamiento de información. Estas tareas son típicas en procesos de migración, integración de terceros, explotación de Big Data, limpieza de datos, análisis y perfilado de datos, etc.

Proporciona una interfaz visual muy intuitiva, donde podemos diseñar Trabajos (Jobs, ficheros XML con extensión KJB) y Transformaciones (Transformations, ficheros XML con extensión KTR), que finalmente constituyen los procesos ETL. A continuación, se describen las principales características de estos elementos (Fernández, 2017):

- Transformación: consiste en una colección de pasos que representan una tarea ETL (ver figura 15). Cada paso es una operación particular sobre datos y están conectados entre sí a través de saltos, visualmente son flechas, que representan el flujo de información de datos, es decir, el flujo de salida de un paso y el flujo de entrada de otro. Los pasos se ejecutan de manera simultánea y asíncrona.

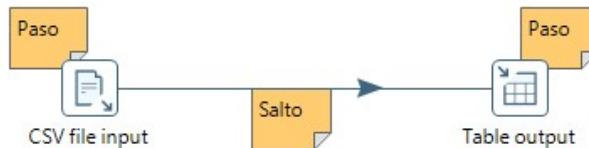
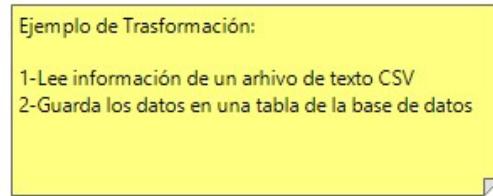


Figura 15: Ejemplo de Transformación

- Trabajo: consta de una o más transformaciones que serán ejecutadas secuencialmente (ver figura 16). La ejecución de cada entrada de trabajo presenta una salida de status, que puede ser evaluada para continuar con la ejecución del proceso o no. Puede contener otros trabajos que efectúen otras transformaciones.



Figura 16: Ejemplo de Trabajo

TOL

TOL (Time Oriented Language) es un lenguaje de programación interpretado, creado por la empresa Bayes Forecast, que se enfoca en los procesos de análisis de series temporales y en el desarrollo de procesos estocásticos. (TOL Project, Welcome to TOL Project)

Este lenguaje admite el paradigma de programación estructurada, en el que en el código se pueden definir funciones que serán alcanzables en todo el ámbito de la instancia de ejecución una vez declaradas, así como variables globales siempre y cuando éstas se definan fuera del ámbito de las funciones (si se definen dentro de ellas serán variables



de ámbito local). De la misma manera, si se definen variables dentro de estructuras condicionales o de bucle, serán solamente válidas mientras la instancia se encuentre dentro de dicha condición o bucle.

TOL dispone de un tipo de datos llamado Nameblock que está pensado para ofrecer un espacio de definición o ámbito local y permanente donde crear variables y funciones. Dichas estructuras admiten herencia múltiple, pudiendo recibir del Nameblock padre la misma estructura de variables y funciones, adquiriendo así polimorfismo. Con estas estructuras se puede construir también código que siga el paradigma de programación orientado a objetos.

TOLBase es una interfaz gráfica de usuario que permite evaluar el código TOL, editar archivos TOL y también administrar proyectos completos escritos en TOL. Algunas de las funcionalidades disponibles en este entorno de programación son: resaltado de sintaxis, analizador sintáctico, librería de consulta de funciones y compilación parcial de una parte del código seleccionada. Además, dispone de un gestor de conexiones a bases de datos enlazado con el gestor ODBC de Windows.

Este programa también es fácilmente escalable en cuanto a funcionalidad, pudiendo añadir fácilmente al entorno del proyecto extensiones en TOL e interfaces gráficas creadas con Tcl/Tk.

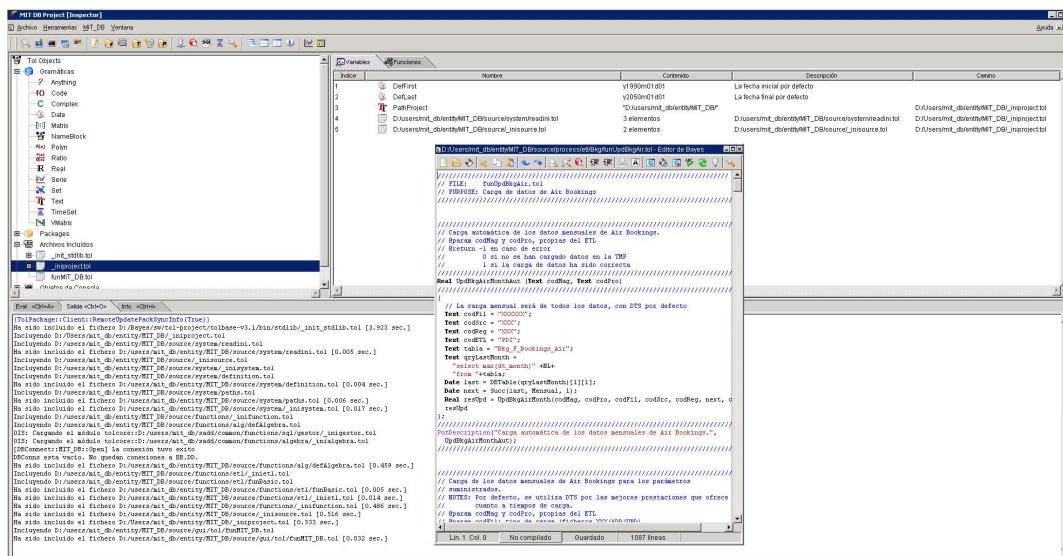


Figura 17: Interfaz gráfica de usuario de TOLBase

TOL PROCESS MANAGER

TOL Process Manager es un framework codificado en TOL e integrado en los proyectos de la empresa Bayes Forecast, que consiste en un interfaz que facilita la ejecución de los procesos ETL desarrollados en TOL. Permite que cualquier usuario sin conocimientos en



programación de procesos ETL pueda ejecutarlos sin dificultad. Para ello, proporciona tres formas de ejecutar un proceso:

- Ejecución manual desde la interfaz gráfica de usuario. Se dispone de un acceso directo en la barra de herramientas de TOLBase (ver figura 18) para abrir una ventana en la que se pueden elegir tanto los procesos que se quieran lanzar como los parámetros de entrada de los mismos.



Figura 18: Acceso directo a TOL Process Manager en la barra de herramientas de TOLBase

La figura 19 muestra las diferentes ventanas que ofrece este interfaz. En la primera ventana se selecciona el módulo de información donde se ha agrupado el proceso, en la segunda se selecciona el nombre del proceso que se quiere actualizar y finalmente aparecerán las ventanas para la selección de los parámetros de entrada.

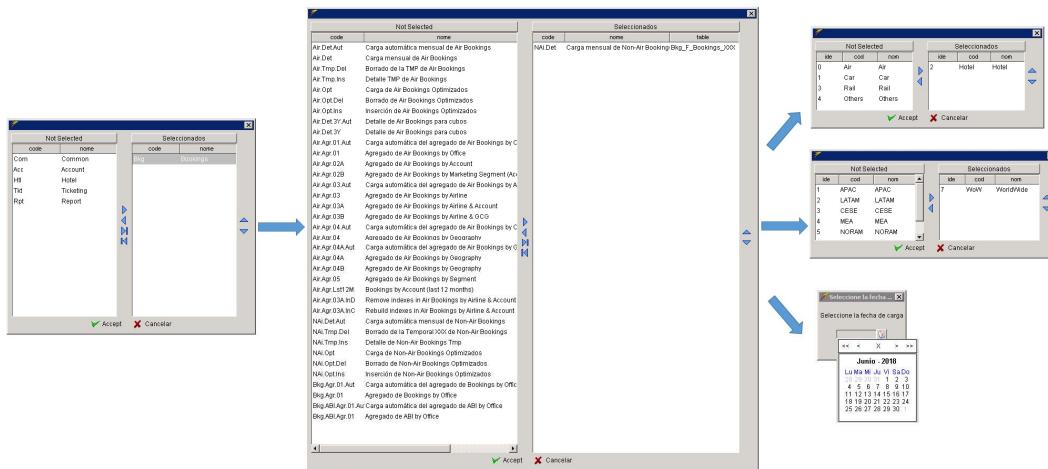


Figura 19: Interfaz gráfica de usuario de TOL Process Manager

- Ejecución manual desde un script en TOL. En caso de querer ejecutar un gran número de procesos, o un proceso de manera repetitiva, el uso de la interfaz gráfica resulta ineficiente. Para estos casos, se puede utilizar un script en TOL en el que se reúnen las llamadas a las funciones.
- Ejecución automática, programando su ejecución a través del programador de tareas del sistema operativo. Este tipo de ejecución sólo es válida para aquellos procesos que no requieren parámetros de entrada, o que estos se calculan previamente. Por ejemplo, si se tiene un proceso que introduce datos diarios, se puede tener un proceso que los actualice introduciendo la fecha o el rango de



fechas que se quieren cargar y además se puede tener su proceso análogo automático que en base a la última fecha cargada en el sistema y la información disponible en la fuente de datos, cargue los datos que falten.

Por otra parte, si el proceso lo requiere, este framework se puede enlazar con otras herramientas de procesos ETL, como PDI, llamándolas a través del intérprete de comandos del sistema operativo y pasando como parámetros los parámetros de entrada del proceso TOL.

Las razones para usar este framework son:

- Organizar todos los procesos en módulos de información, clasificándolos por el tipo de información que manejan, de acuerdo con el modelo de datos.
- Tener un punto de entrada común, TOL Process Manager, de manera que la ejecución de cada uno de los procesos pase por el mismo punto, dando la posibilidad de generar un log, medir tiempo, etc.
- División conceptual de los procesos principales en funciones. Cada proceso tiene un nombre o un código único que lo identifica frente al resto de procesos del proyecto. Todos los procesos ETL deben ser implementados mediante funciones a las que se llamará mediante el punto de entrada común. Estas funciones deben tener definidos como primer y segundo parámetro de entrada dos variables de tipo texto, que en el momento del lanzamiento se llenarán con el nombre del módulo y con el nombre del proceso, respectivamente.
- División lógica de los procesos en pasos. Cada paso tiene una descripción, un cuerpo, que es lo que se ejecuta, y un resultado.
- Mantener un registro de la ejecución de los procesos.
- Posibilidad de ofrecer al usuario los distintos procesos existentes para que elija cual/es quiere lanzar, en cuyo caso son solicitados los valores de los argumentos de esos procesos.
- Posibilidad de programar el lanzamiento de los procesos

2.3.3. CUBOS OLAP: MONDRIAN

Mondrian es un servidor OLAP (Online Analytical Processing) escrito en Java, que permite analizar grandes cantidades de datos utilizando las expresiones de cálculo del lenguaje MDX (MultiDimensional eXpressions), es decir, haciendo una exploración dimensional de los datos.

Esta plataforma pertenece a la empresa Pentaho. A su vez, Pentaho fue adquirido por Hitachi Data Systems en 2015 y en septiembre de 2017 se convirtió en parte de Hitachi



Vantara, una nueva compañía que unifica las operaciones de Pentaho, Hitachi Data Systems y Hitachi Insight Group (Wikipedia, Pentaho).

Debido a estos cambios, la oferta que Pentaho ofrece en su portfolio ha ido cambiando considerablemente en los últimos años. Dispone de una serie de productos que proporcionan integración de datos, procesos ETL, servicios OLAP, generación de informes, cuadros de mando o minería de datos. Pentaho los ofrece como componentes individuales o a través de una solución completa de BI, actualmente denominada Pentaho Business Analytics Server (Hitachi Vantara, Pentaho Business Analytics). Existen versiones gratuitas (Community Edition) y de pago (Enterprise Edition) de la plataforma BI y de cada producto, entre los que se encuentra integrado Mondrian. La figura 20 muestra esta integración.

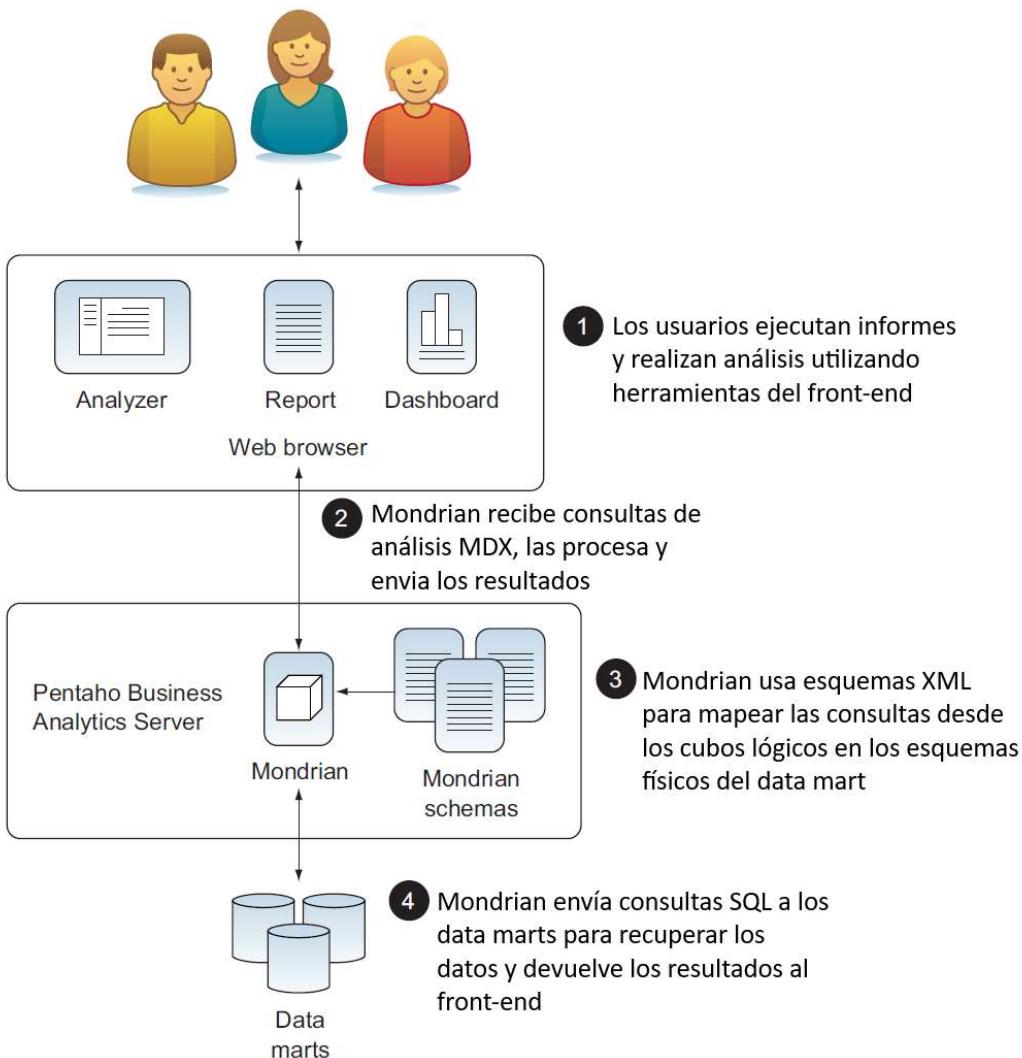


Figura 20: Integración de Mondrian en Pentaho Business Analytics Server



En Mondrian, la definición de los cubos OLAP se realiza mediante ficheros XML, esquemas, donde se declaran tres tipos de elementos: dimensiones, cubos y cubos virtuales. En la figura 21 se muestra un ejemplo de estos esquemas.

Un esquema está asociado a una fuente de datos específica, por lo que se puede enlazar el motor OLAP con varias bases de datos relacionales, construyendo un esquema para cada una de ellas.

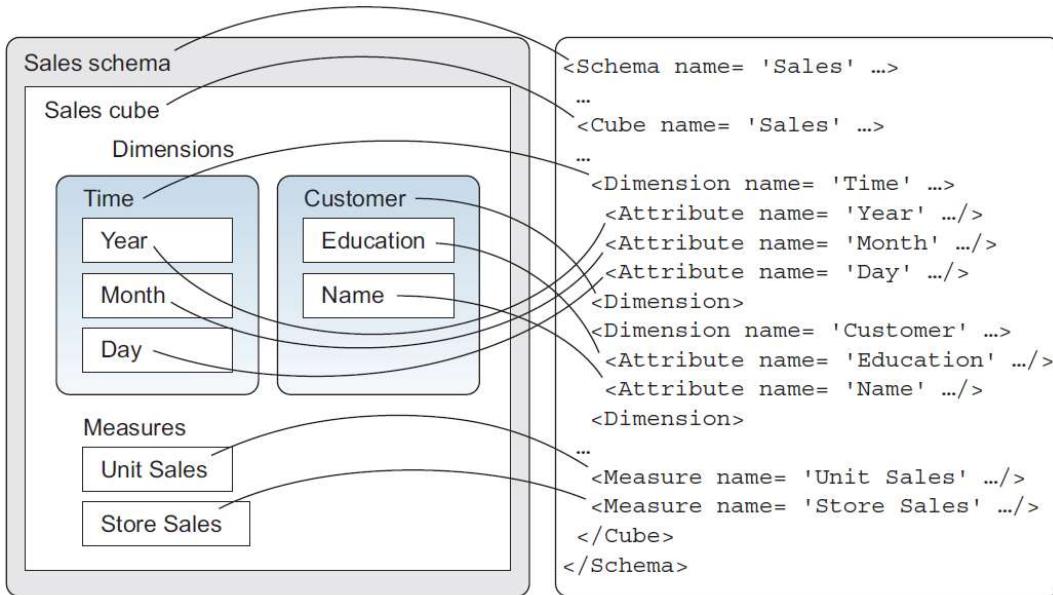


Figura 21: Ejemplo de esquema XML

Por otra parte, Mondrian proporciona una API a través de la cual una aplicación cliente puede ejecutar consultas MDX para crear visualizaciones de datos e informes que después pueden ser integrados en aplicaciones web.

Existen varios front-end para Mondrian que hacen uso de esta API y ofrecen una solución analítica fácil de usar que permite a los usuarios consultar los cubos OLAP definidos sin necesidad de tener conocimientos de MDX ni de bases de datos. Esto se debe a que Mondrian es un motor ROLAP (Relational OLAP) que solamente se encarga de traducir la consulta MDX generada con la selección de la información de un cubo en una o varias consultas SQL desde las que extrae el dato.

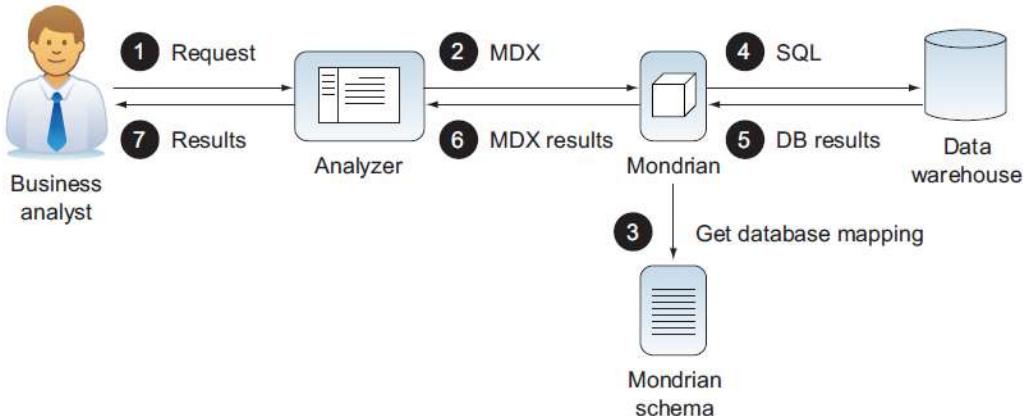


Figura 22: Ejecución de una consulta OLAP en Mondrian

Las consultas SQL se realizan en el momento que se solicita información, con lo que si el SGBD se encuentra en otra máquina o la información consultada no está optimizada, por ejemplo mediante el uso de tablas agregadas, puede que el tiempo de respuesta no sea óptimo. Para paliar posibles problemas de rendimiento, Mondrian gestiona las últimas consultas efectuadas por los usuarios en una caché intermedia, con el fin de agilizar el análisis repetitivo de la misma información.

El interfaz ofrecido por estos front-end permite visualizar y analizar la información en forma de tabla dinámica o mediante gráficos. Se muestra un listado de las dimensiones y medidas disponibles en el cubo que, a su vez, se pueden arrastrar y soltar como columnas, filas, filtros o dentro del propio cubo de acuerdo con la visualización deseada por el usuario. También hay opciones para dar un formato condicional a los datos mostrados, según el valor que tenga y visualizar subtotales o realizar otros cálculos agregados que no se hayan incluido en la definición del cubo.

Los front-end de Mondrian más populares son:

- Pentaho Data Analyzer (figuras 23 y 24): disponible en Pentaho Business Analytics Server, la versión Enterprise Edition (EE) de Pentaho.
- Saiku Analytics (figura 25): permite la visualización de los datos en formato de tabla dinámica y gráficos, tanto en la versión CE como en la EE. Se distribuye, bien con Mondrian embebido como una aplicación independiente en formato WAR (Web Application Archive) que se puede desplegar en un contenedor web, por ejemplo Apache Tomcat, o como un plugin de la plataforma Pentaho Business Analytics que se puede descargar desde el Marketplace de Pentaho.
- JPivot (figura 26): disponible en la versión Community Edition (CE) de Pentaho, ofrece unas prestaciones muy básicas.



The screenshot shows the Pentaho Data Analyzer interface with a dynamic table. The table has columns for YEAR ID (2003, 2004, 2005) and TERRITORY (APAC, EMEA, Japan, NA). The table is titled "SALES". A context menu is open over the "SALES" column, showing options like "Color Scale: Green-Yellow-Red", "Color Scale: Red-Yellow-Green", etc. The left sidebar lists various dimensions and measures, and the bottom right has "Report Options...".

YEAR ID	2003	2004	2005
TERRITORY	SALES	SALES	SALES
APAC	253134	345308	147680
EMEA	1660315	2389129	929829
Japan	243773	168479	42921
NA	1359757	1821247	671057

Figura 23: Tabla dinámica de Pentaho Data Analyzer

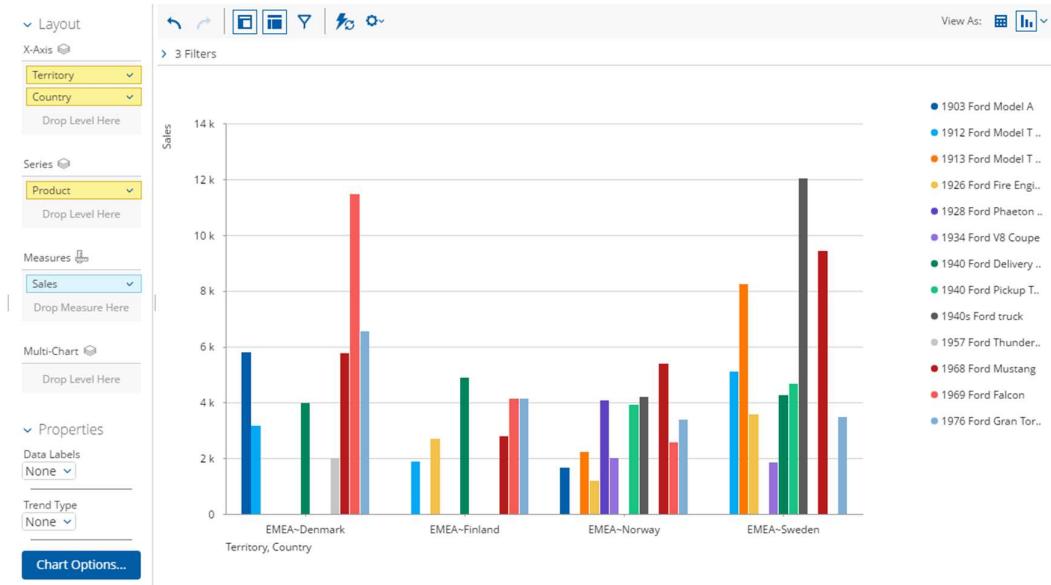


Figura 24: Gráfico de barras de Pentaho Data Analyzer

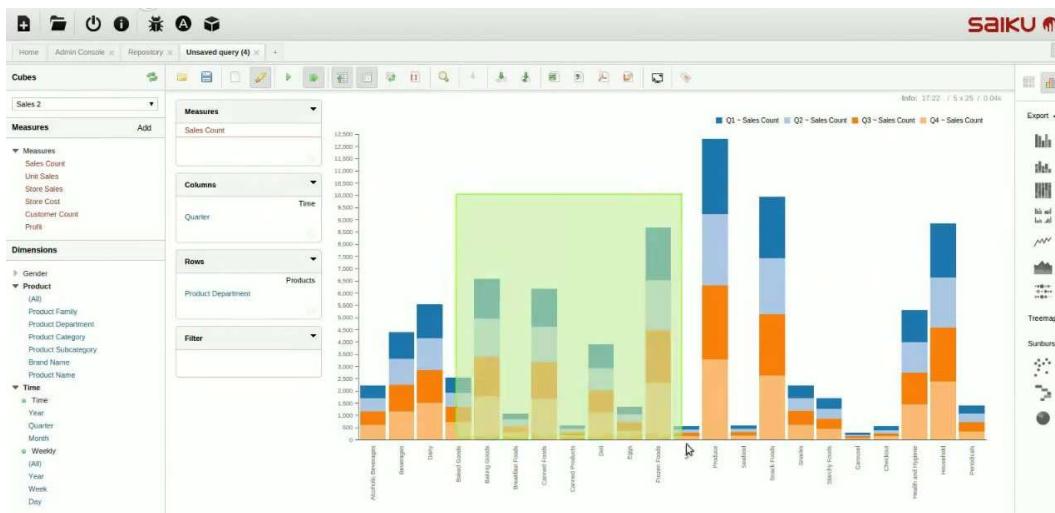


Figura 25: Interfaz de Saiku Analytics

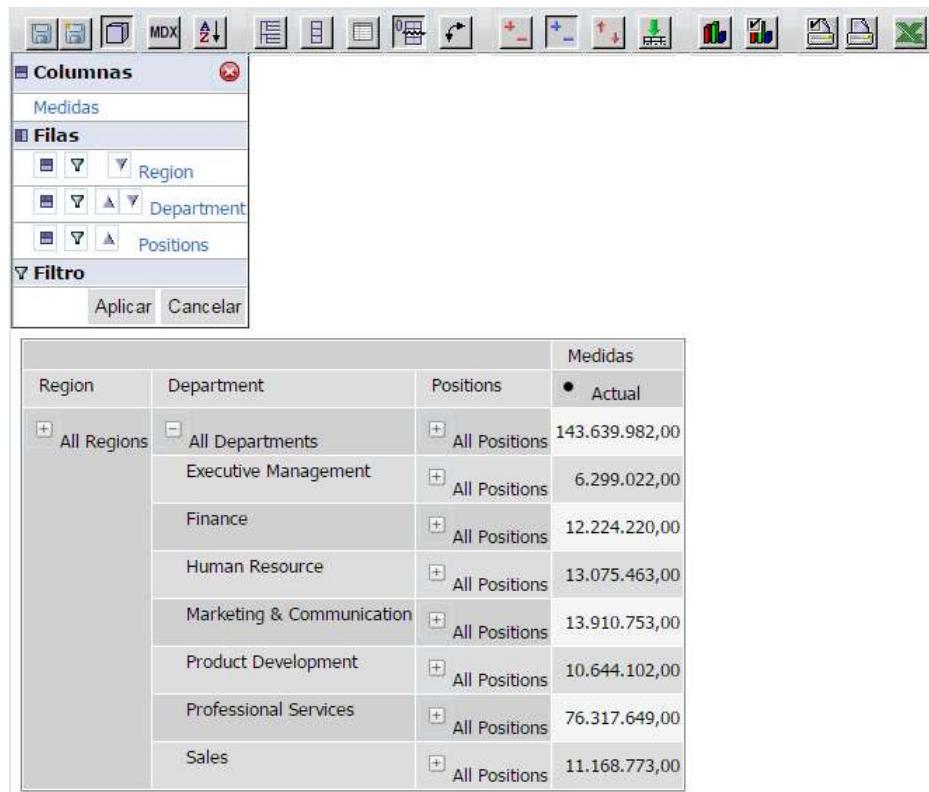


Figura 26: Interfaz de Jpivot



2.3.4. GENERACIÓN DE INFORMES: BIRT

BIRT es un proyecto Open Source que proporciona una potente plataforma para crear visualizaciones de datos e informes que después pueden ser integrados en aplicaciones Java autónomas o aplicaciones web. El proyecto se desarrolla oficialmente dentro de la Fundación Eclipse, un consorcio independiente sin ánimo de lucro y formado por los proveedores de la industria de software y una comunidad de código abierto (Wikipedia, Business Intelligence and Reporting Tools).

BIRT, tal y como muestra la figura 27, tiene dos componentes principales: un diseñador de informes visuales (Report Designer) que está basado en el entorno de desarrollo integrado Eclipse, y un componente de rutina (Report Engine) para generar informes que pueden ser puestos en uso en cualquier entorno Java.

Los diseños de informes BIRT se hacen en ficheros XML y pueden acceder a diferentes fuentes de datos.

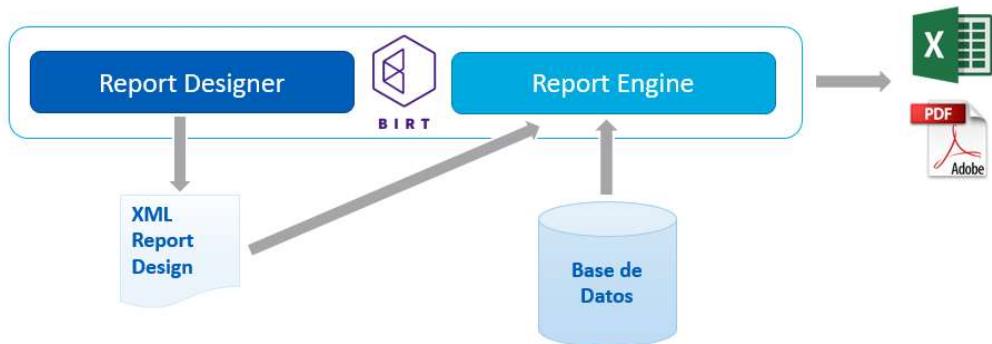


Figura 27: Arquitectura de BIRT

2.3.5. DESARROLLO WEB

Existen varios tipos de tecnologías que forman parte del proceso de desarrollo y diseño web:

- Front-End: las tecnologías del lado del cliente son las que se interpretan en el navegador web. Estos incluyen HTML (HyperText Markup Language), CSS (Cascading Style Sheets) y JS (JavaScript).
- Back-End: las tecnologías de back-end son aquellas que son interpretadas por el servidor que almacena el código de la página web y lo envía al navegador web de un usuario cuando lo solicita.



La tecnología de back-end generalmente crea código HTML y envía ese código al navegador. Entre estas tecnologías de back-end, se incluye Java Enterprise Edition (Java EE).

Cada lenguaje o tecnología de front-end y back-end cumple un rol específico dentro del proceso de desarrollo de aplicaciones web. Estas tecnologías, a menudo, interactúan para producir la página web o la aplicación completa, pero tienen un papel distinto.

HTML

HTML está diseñado para formatear un documento. No se usa para producir el aspecto o el estilo de un documento. El marcado HTML indica el propósito de cada elemento, no su posición o aspecto en la página. HTML está diseñado para que un documento se pueda reutilizar en todos los medios sin cambiar el código. Es posible que escuche el término HTML semántico para indicar HTML que se produce solo con el propósito del contenido en mente.

A partir de octubre de 2014, HTML5 es la quinta revisión final y completa del estándar HTML del World Wide Web Consortium (W3C). La versión anterior, HTML 4, se estandarizó en 1997.

Los principales objetivos de HTML5 han sido mejorar el lenguaje con soporte para la última tecnología multimedia, a la vez que mantener su fácil legibilidad para los humanos y se entiende de manera consistente por computadoras y dispositivos (navegadores web, analizadores sintácticos, etc.). HTML5 pretende incluir no solo HTML 4, sino también XHTML 1 y DOM Level 2 HTML. (Wikipedia, HTML5)

CSS

CSS es la capa de diseño. Interpretado por el navegador, determina la apariencia visual de los elementos. Esto incluye tanto la apariencia del elemento en sí como su posición relativa a otros elementos en la página.

Cascading Style Sheets (CSS) es un lenguaje de hojas de estilo utilizado para describir el aspecto y el formato de un documento escrito en un lenguaje de marcado. Aunque se usa con mayor frecuencia para cambiar el estilo de las páginas web y las interfaces de usuario escritas en HTML y XHTML, el lenguaje se puede aplicar a cualquier tipo de documento XML, incluidos XML simple, SVG y XUL. Junto con HTML y

JavaScript, CSS es una tecnología fundamental utilizada por la mayoría de los sitios web para crear páginas web visualmente atractivas, interfaces de usuario para aplicaciones web e interfaces de usuario para muchas aplicaciones móviles.

CSS está diseñado principalmente para permitir la separación del contenido del documento de la presentación del documento, incluidos elementos como el diseño, los colores y las fuentes. Esta separación puede mejorar la accesibilidad del contenido,



proporcionar más flexibilidad y control en la especificación de las características de presentación, permitir que varias páginas HTML compartan el formato especificando el CSS relevante en un archivo .css separado y reducir la complejidad y la repetición en el contenido estructural.

CSS permite separar las instrucciones de presentación del contenido HTML en una sección de archivo o estilo independiente del archivo HTML. (Wikipedia, Cascading Style Sheets)

JAVASCRIPT

JavaScript es la tecnología de front-end que se ocupa de la interactividad del usuario. El código JavaScript se puede usar para validar la entrada del usuario, crear efectos desplegables, cambiar dinámicamente estilos CSS e incluso realizar cálculos complejos.

Existen muchos frameworks y librerías escritos en JavaScript. Un framework de JavaScript difiere de una librería en su flujo de control: una biblioteca ofrece funciones para ser llamadas desde nuestro código, mientras que un framework define el diseño de la aplicación completa.

Algunos frameworks de JavaScript siguen el paradigma de modelo-vista-controlador (MVC) diseñado, entre otros propósitos, para mejorar la calidad del código y la capacidad de mantenimiento.

En este proyecto, hemos empleado el framework Ext JS, propiedad de Sencha, y la librería de gráficos HighCharts, creada por la empresa noruega Highsoft. (Wikipedia, JavaScript)

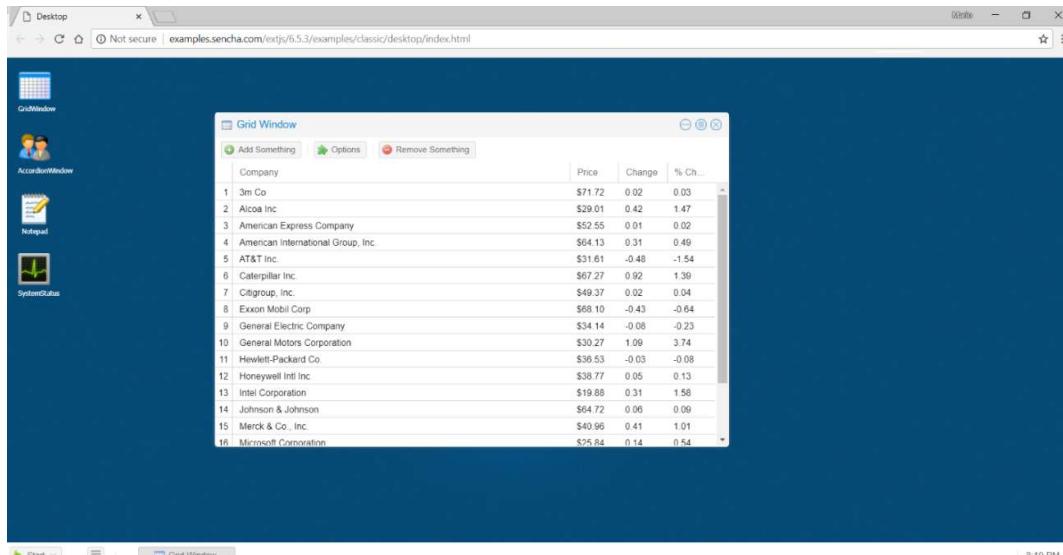


Figura 28: Aplicación web con formato de escritorio



EXT JS

Ext JS provee un mecanismo de creación y extensión de clases muy sencillo de utilizar. Con este framework se pueden definir métodos, atributos, alias y propiedades de una forma mucho más sencilla que con JavaScript estándar. Además, provee una forma sencilla de realizar herencia múltiple y simple entre los diferentes elementos de una aplicación.

Entre las características más potentes y populares de Sencha Ext JS, encontramos la simulación de una aplicación de escritorio totalmente personalizable, como el ejemplo de la figura 28.

HIGHCHARTS

Highcharts es una librería de gráficos (ver figura 29) que se ha desarrollado activamente desde 2009. Es una de las librerías más potentes del mercado, que facilita la incorporación de gráficos interactivos optimizados para todo tipo de navegadores y dispositivos móviles.

Además, cuenta con una sólida documentación y diferentes plugins para simplificar su uso en los frameworks JavaScript más populares de la industria, como es el caso de Sencha Ext JS.



Figura 29: Gráfico generado con HighCharts



JAVA

Java es un lenguaje de programación de propósito general, concurrente, orientado a objetos, que fue diseñado específicamente para tener tan pocas dependencias de implementación como fuera posible. Su intención es permitir que los desarrolladores de aplicaciones escriban el programa una vez y lo ejecuten en cualquier dispositivo, lo que quiere decir que el código que es ejecutado en una plataforma no tiene que ser recompilado para correr en otra.

Actualmente es uno de los lenguajes de programación más populares en uso, particularmente para aplicaciones web, en el lado del servidor/back-end.

En los programas de Java, las clases son los componentes básicos. Al definir una clase, se describen todas las partes y características de uno de esos bloques de construcción. Para usar la mayoría de las clases, han de crearse objetos. Un objeto es una instancia de tiempo de ejecución de una clase en la memoria. Todos los valores de los diversos objetos de todas las clases representan el estado de tu programa.

Java EE es la edición usada para producir aplicaciones web de nivel empresarial. Fue definida por los creadores de Java (Sun Microsystems, adquirida por Oracle en 2010) y representa el núcleo de algunas de las aplicaciones web más grandes del mundo.

A continuación, se realiza una breve descripción de dos tecnologías ampliamente utilizadas, en nuestra aplicación también, para la implementación de servicios web en Java: Java Servlets y Java Server Pages (JSP). Además, se explicará la API (Application Programming Interface) de JDBC (Java Database Connectivity) empleada para conectarse a una base de datos.

JAVA SERVLETS

Los servlets de Java son programas que se ejecutan en un servidor web o de aplicaciones y actúan como una capa intermedia entre una solicitud proveniente de un navegador web u otro cliente HTTP y bases de datos o aplicaciones en el servidor HTTP (ver figura 30).

Esta tecnología proporciona una forma de generar páginas web dinámicamente, utilizando las ventajas de la programación en Java. Permite recopilar información de los usuarios a través de formularios de páginas web, presentar registros obtenidos de alguna base de datos u otra fuente, manejar peticiones concurrentes, etc.

Podríamos decir que sirven para el mismo propósito que los programas implementados usando la interfaz de puerta de enlace común (CGI - Common Gateway Interface), pero los servlets ofrecen varias ventajas en comparación con el CGI:



- Se ejecutan dentro del espacio de direcciones de un servidor web. No es necesario crear un proceso separado para manejar cada solicitud del cliente.
- Son independientes de la plataforma porque están escritos en Java.
- La funcionalidad completa de las librerías de Java está disponible para un servlet. Por ejemplo, puede comunicarse con applets, bases de datos u otro software a través de una API como JDBC.
- Utilización de cookies, pequeños datos en texto plano que pueden ser guardados en el cliente. La API de servlets permite un manejo fácil y limpio de ellas.

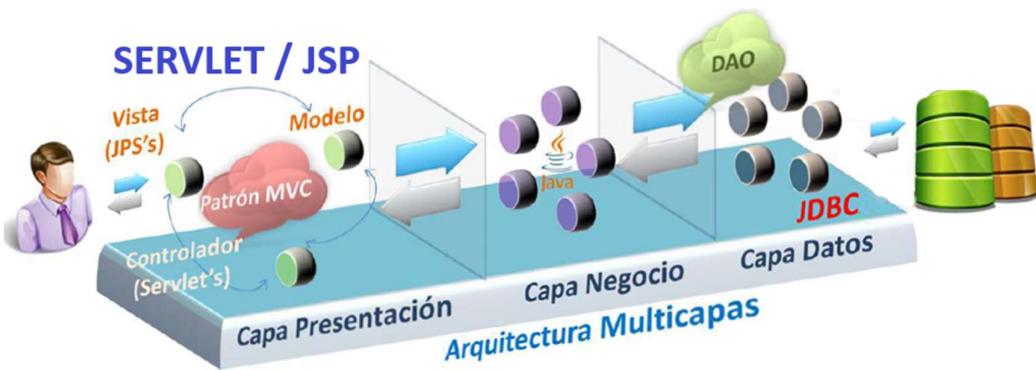


Figura 30: Diagrama de la posición de un servlet en una aplicación web

JAVA SERVER PAGES

JSP es una tecnología de script en el servidor, similar a los servlets, que ayuda a los desarrolladores de software a crear páginas web dinámicas basadas en HTML y XML, entre otros tipos de documentos. Por ejemplo, permite agregar contenido dinámico a un archivo HTML mediante código escrito en Java dentro del archivo JSP, haciendo uso de etiquetas especiales, *tags*, que son procesados por el servidor web antes de enviarlos al cliente.

Estas páginas tienen un aspecto similar al de una página HTML, pero se transforman en clases Java (como el ejemplo de la figura 31). Concretamente, el código JSP se compila la primera vez que se le invoca o cuando sufre modificaciones, generando las clases Java servlets correspondientes y en adelante es el código del nuevo servlet el que se ejecuta produciendo como salida el código HTML que compone la página web de respuesta. Por tanto, un compilador de JSP es un programa que analiza este tipo de ficheros y los transforma en Java Servlets ejecutables.



```
<%@ page
language="java"
contentType="text/html; charset=UTF-8"
pageEncoding=UTF-8
import="java.util.List"
import="com.bayesforecast.bwsuite.usermodel.model.*"
import="org.springframework.web.context.WebApplicationContext"
import="org.springframework.web.context.support.WebApplicationContextUtils"
%><%@ taglib prefix="fmt" uri="http://java.sun.com/jsp/jstl/fmt"
%><%@ taglib prefix="c" uri="http://java.sun.com/jsp/jstl/core"
%><%>
String baseUrl = request.getContextPath() + "/index.jsp";

String modules = "";
UserModel user = (UserModel) request.getAttribute("LoggedUserInfo");
for (ModuleModel module : user.getModules()) {
    modules += "" + module.getCode() + ":" + 1;
}
modules += "\HelpWindow":1;

String env = System.getenv("BWSENV");
if (env == null) {
    env = "Test Environment";
} else if ("Production".equals(env)) {
    env = " "; // espaço obrigatório
} else {
    env += " Environment";
}
%><!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en">
<head>
<title><fmt:message key="project.code" /> - <fmt:message key="project.title" /></title>
<meta http-equiv="content-type" content="text/html; charset=UTF-8" />
<meta http-equiv="X-UA-Compatible" content="IE=EmulateIE8">
<!--[if IE]><script type="text/javascript" src="resources/js/dygraph/excanvas.js"></script><![endif]-->
<!--<meta http-equiv="X-UA-Compatible" content="IE=8" /> -->
<base href="<%=baseUrl%>" />
<link rel="shortcut icon" href="resources/favicon/favicon.ico" type="image/x-icon" />
<link rel="stylesheet" href="resources/css/styles.css" type="text/css" />
<link rel="stylesheet" href="resources/css/customer/<fmt:message key="project" />.css" type="text/css" />
<script type="text/javascript" src="resources/js/dygraph/dygraph-dev.js"></script>

<script type="text/javascript">
//Global Session for the Diagnosis
var workspace = {};
</script>
<script type="text/javascript">
bwsuite=<%modules%>;
</script>
```

Figura 31: Ejemplo fichero JSP

Algunos contenedores JSP admiten la configuración de la frecuencia con la que el contenedor comprueba las marcas de tiempo del archivo JSP para ver si la página ha cambiado. Normalmente, esta marca de tiempo se establecerá en un intervalo corto (quizás segundos) durante el desarrollo del software, y un intervalo más largo (quizás minutos, o incluso nunca) para una aplicación web desplegada.

Cuando se construyen aplicaciones con JSP hay que procurar desacoplar la lógica de presentación de la lógica de negocio. Para esto, se suele utilizar el denominado esquema MVC (modelo-vista-controlador) de la figura 32. En otras palabras, lo que realizamos es una consulta a un componente, normalmente un servlet, que realiza las labores complejas y delega sobre un JSP la labor de presentación (Wikipedia, JavaServer Pages) (Global Mentoring, Curso de Java con JDBC).

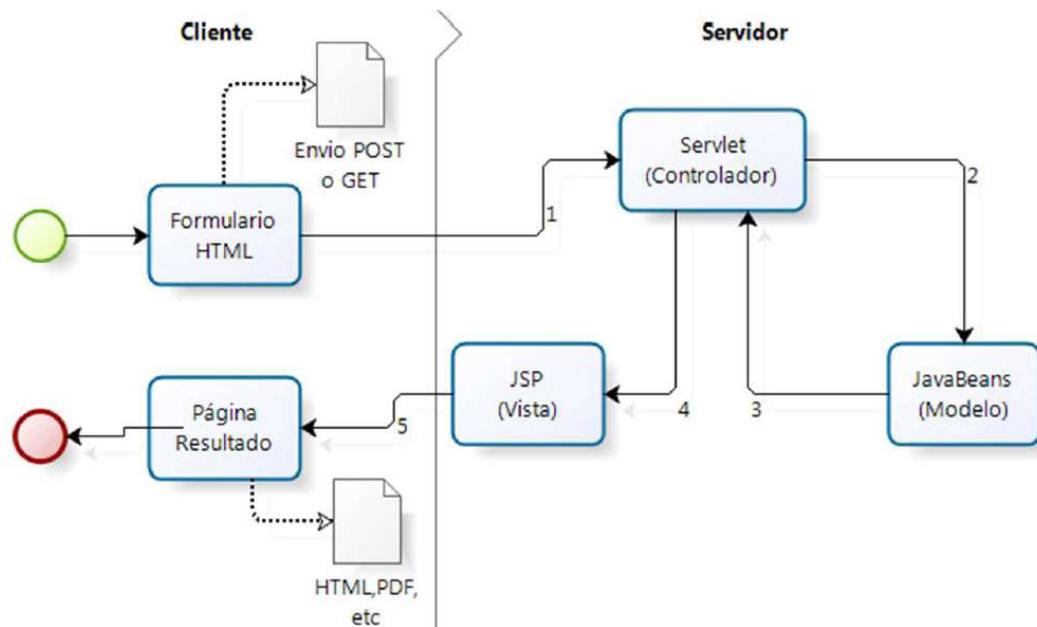


Figura 32: Ejemplo MVC con Servlets y JSP

JDBC

La API de JDBC (Java Database Connectivity) es el estándar de la industria para la conectividad entre el lenguaje de programación Java y una amplia gama de bases de datos basadas en SQL y otras fuentes de datos tabulares, como hojas de cálculo o archivos planos.

Para conectarse a la base de datos se necesita un controlador o driver. Un driver es la implementación de la especificación del API de JDBC definido por Java para un SGBD concreto: Microsoft SQL Server, Oracle, MySQL, etc. La figura 33 muestra un esquema con varios ejemplos de drivers JDBC:

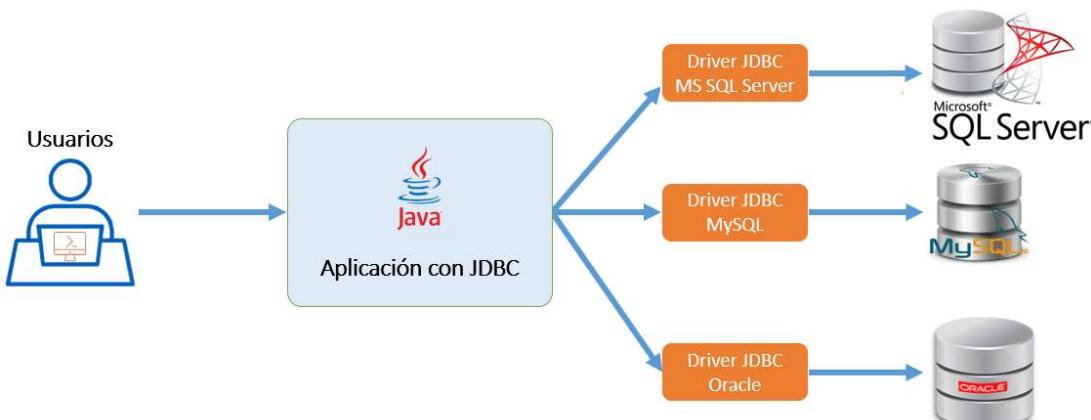


Figura 33: Drivers de JDBC



Mediante esta abstracción se obtienen las siguientes ventajas:

- Permite la conexión y acceso simultáneo a datos de distintas fuentes heterogéneas utilizando una interfaz común.
- El desarrollo de las aplicaciones queda independizado de los detalles de implementación propios de las operaciones de conexión a base de datos remotas.
- Permite realizar un cambio de la base de datos subyacente a una aplicación sin provocar cambios de adaptación en la aplicación, mientras que exista un driver JDBC para la nueva base de datos.

En resumen, JDBC y Java presentan como ventaja fundamental sobre el resto de entornos de programación de bases de datos, la posibilidad de desarrollar aplicaciones independientes de la plataforma de ejecución e independientes del SGBD utilizado, proporcionando al desarrollador transparencia frente al gestor (Oracle, JDBC Overview) (Grupo Atrium, Master Certificado Experto Java EE) (Global Mentoring, Curso de Java con JDBC).

SPRING

Hoy en día, existen muchos frameworks que ayudan en el proceso de creación de páginas web en Java haciendo uso del patrón de diseño MVC. Bayes Forecast se ha decantado por el uso de uno de estos frameworks: Spring.

Spring es un framework para el desarrollo de aplicaciones para la plataforma Java. Se trata de un proyecto de código abierto, por lo que existe una gran comunidad en torno a la actualización y mejora del framework. Está muy extendido en un uso empresarial ya que permite gestionar aplicaciones de gran tamaño y, sobre todo, facilita la comunicación entre la propia aplicación y los diferentes servicios externos sobre los que se conecta.

Las características principales del framework pueden ser utilizadas por cualquier aplicación Java, pero hay extensiones para construir aplicaciones web en la parte superior de la plataforma Java EE. Aunque el marco no impone ningún modelo de programación específico, se ha vuelto popular en la comunidad Java como alternativa, reemplazo o incluso agregado al modelo Enterprise JavaBeans (EJB).

Dentro del framework Spring disponemos de diversos módulos, los cuales ofrecen un amplio rango de servicios. No hace falta contar con todos sino que podemos contar únicamente con aquellos que realicen las funciones que necesitemos para nuestros proyectos. Spring está diseñado para no ser intrusivo, lo que significa que su lógica de dominio el código generalmente no tiene dependencias en el marco mismo. Algunos de estos módulos son:



- Spring Core: se trata del módulo principal. Es el único obligatorio para el trabajo con el framework. Contiene la inyección de dependencias y la configuración y uso de objetos Java.
- AOP: para trabajar con la Programación Orientada a Aspectos.
- DAO: nos permite la interacción con los sistemas de bases de datos.
- MVC: para gestionar toda la parte web.
- JMS: lleva toda la parte de la mensajería entre las diferentes partes.
- Security: para el control de la seguridad.

El núcleo de Spring está basado en un patrón de diseño llamado Inversión de Control (IoC). En un flujo normal es el usuario el que invoca los procedimientos de las diferentes librerías o clases del proyecto dependiendo de la funcionalidad. Con el uso del patrón IoC, es el framework el que se encarga de invocar el código creado por el usuario. Bajo este patrón, el código de alto nivel se encuentra dentro del propio framework y el usuario se encarga de generar el código de más bajo nivel. Podemos ver un claro ejemplo de este paradigma en la siguiente tabla, donde indicamos las diferentes acciones y responsables respecto al uso de JDBC con Spring (Imagina Formación, Curso de Desarrollo de Aplicaciones Java con Spring):

ACCIÓN	SPRING	DESARROLLADOR
Definir los parámetros de conexión		x
Abrir la conexión	x	
Especificar la sentencia SQL		x
Declarar parámetros y proporcionar valores de los parámetros		x
Elaborar y ejecutar la sentencia	x	
Configurar el bucle para iterar a través de los resultados (si los hay)	x	
Hacer el trabajo para cada iteración		x
Procesar cualquier excepción	x	
Manejar las transacciones	x	
Cerrar connection, statement y resultset	x	

Tabla 4: Uso de JDBC con Spring

En resumen, si utilizamos Spring correctamente, estaremos preparando nuestra aplicación con unas capas bien delimitadas y siguiendo unas buenas prácticas de programación (Wikipedia, Spring Framework).



2.3.6. CONTENEDOR WEB: APACHE TOMCAT

Apache Tomcat, también conocido como Jakarta Tomcat o simplemente Tomcat, es un contenedor web de código abierto desarrollado y mantenido por una comunidad abierta de desarrolladores de la Apache Software Foundation (ASF) y voluntarios independientes. Está publicado bajo la licencia Apache License 2.0 e implementa Java Servlet y JavaServer Pages (JSP), entre otras especificaciones Java EE (Enterprise Edition de Oracle Corporation, aunque creado originalmente por Sun Microsystems) (Wikipedia, Apache Tomcat).

En este proyecto, se ha utilizado la versión 6.0.29 de Tomcat que requiere Java versión 1.6 y consta de los siguientes componentes:

- Catalina: es un contenedor de servlet.
- Coyote: es un conector HTTP para Tomcat. Esto permite que Catalina también actúe como un servidor web simple que sirve archivos locales como documentos HTTP.
- Jasper: es un motor que compila JSP convirtiéndolas en servlets que pueden ser manejados por Catalina. En tiempo de ejecución, Jasper detecta cambios en los archivos JSP y los recompila.

Tomcat no es un servidor de aplicaciones aunque, a menudo, se presenta en combinación con el servidor web Apache. No obstante, en este proyecto no ha sido necesario combinar ambas tecnologías ya que Tomcat puede funcionar como servidor web por sí mismo. En sus inicios, existió la percepción de que el uso de Tomcat de forma autónoma era sólo recomendable para entornos de desarrollo y entornos con requisitos mínimos de velocidad y gestión de transacciones. Hoy en día, ya no existe esa percepción y es usado como servidor web autónomo en entornos con alto nivel de tráfico y alta disponibilidad.

Dado que Tomcat fue escrito en Java, funciona en cualquier sistema operativo que disponga de la máquina virtual Java.



3. DESARROLLO DE LA SOLUCIÓN

En este apartado de la memoria se describen todas las partes del proyecto desarrolladas, así como todas las acciones emprendidas para la elaboración de la solución que da respuesta al problema planteado.

En primer lugar se ha realizado un análisis de las fuentes de información disponibles. Cuando recibimos los datos del cliente nuestra primera tarea es entender y analizar detenidamente las fuentes de información y los datos disponibles para formar y organizar los diferentes módulos que compondrán el proyecto. Un módulo de información es una agrupación conceptual de la información que organiza los datos bajo una misma idea de negocio. Básicamente, se muestra el diseño de la base de datos del sistema, con las tablas de hechos y dimensiones asociadas.

Posteriormente, se explican los procesos ETL que se han desarrollado, haciendo énfasis en el proceso de clasificación de oficinas y en los que construyen las tablas de hechos, tanto detalle como agregadas, con las que mejoramos el rendimiento de las consultas SQL utilizadas para explotar la información.

Finalmente, se explica la aplicación web MIT, basada en la BWS Suite de la empresa Bayes Forecast, que se alimenta de las tablas de hechos y dimensiones generadas en los procesos ETL. Se describen todos los módulos que forman parte de la herramienta y la participación del alumno en el desarrollo y mantenimiento de cada una de ellos.

En la figura 34 se muestra un esquema de la solución MIT, donde podemos ver las fuentes de información que se cargan en el DM del proyecto mediante procesos ETL, así como la aplicación web con la que interactúan los usuarios finales para extraer la información disponible.

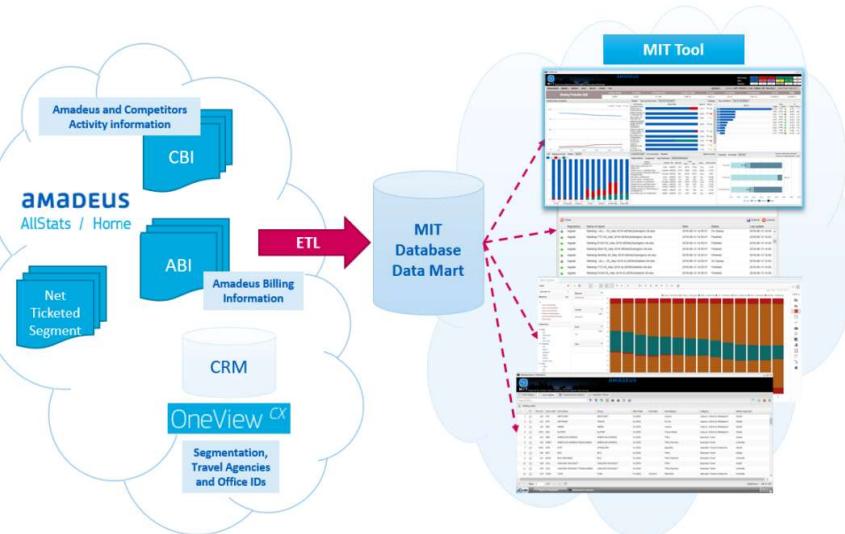


Figura 34: Esquema de la solución MIT



Por otra parte, antes de entrar en detalles, es recomendable conocer la arquitectura de los sistemas que forman parte de MIT (figura 35):

- Un servidor con sistema operativo Windows Server 2008, en el que tenemos un servidor de base de datos Microsoft SQL Server 2012, donde reside el DM descrito en los primeros apartados de este capítulo.
- Un servidor de “cálculo” con sistema operativo Windows Server 2008, en el que se ejecutan los procesos ETL. Es decir, en este servidor tenemos instalado TOL y PDI.
- Dos servidores con sistema operativo Red Hat Linux, en el que tenemos la versión 6 del contenedor de servlets Apache Tomcat sobre el que se ejecuta la aplicación web.
- Un almacenamiento conectado en red, (NAS - Network Attached Storage), de 200 GB, donde se almacenan los informes generados desde la aplicación web, de manera que puedan ser compartidos por los usuarios.

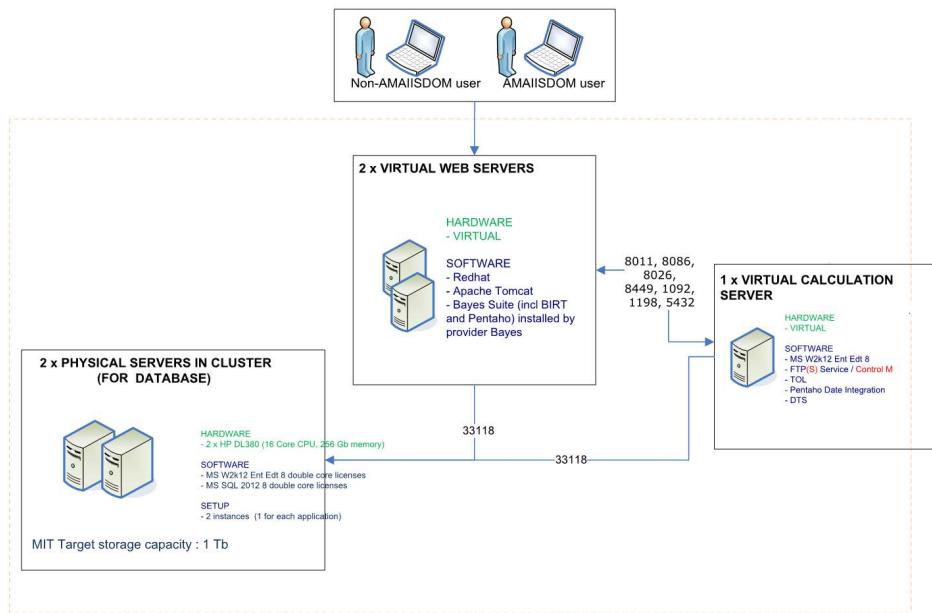


Figura 35: Arquitectura de sistemas de MIT

3.1. ANÁLISIS DE FUENTES DE INFORMACIÓN

En este apartado se analiza la información disponible del negocio, describiendo todas las fuentes de información utilizadas en el proyecto. Básicamente, son ficheros de texto plano, en formato CSV, que alimentan las distintas dimensiones y medidas presentes en nuestro DM a través de diferentes procesos ETL que se explicarán en futuros apartados.



3.1.1. CBI

Tal y como se explicó en el apartado 2.1.2. Fuentes de Información, CBI (Competitive Business Intelligence) proporciona información mensual de la actividad aérea realizada por las agencias de viajes a través de los principales GDS existentes en el mercado: Amadeus, Abacus, Travelport y Sabre. El contenido de los ficheros CBI es el siguiente:

CAMPO	DESCRIPCIÓN
Year	Año al que corresponden las reservas aéreas
Month	Mes al que corresponden las reservas aéreas
Country Code	Código ISO del país en que se encuentra ubicada la oficina de la agencia de viajes
GDS Code	Código del GDS donde la agencia ha realizado la reserva
GDS Name	Nombre del GDS donde la agencia ha realizado la reserva
GDS Office Code	Código de oficina para la agencia que realiza la reserva. Específico del GDS por lo que, en caso de coincidencia entre dos GDS, puede que no representen lo mismo
GDS Office Name	Nombre de oficina para la agencia que realiza la reserva. Específico del GDS por lo que, en caso de coincidencia entre dos GDS, puede que no representen lo mismo
IATA Number	Identificador de agencia en IATA, en caso de ser una agencia acreditada por esta entidad
Account Name	El nombre oficial y legal de la agencia registrado
Trade Name	The Business Name of the Agency "Doing Business As" Name
Address Line 1	First Address Line of the agency
Address Line 2	Second Address Line of the agency
Address Line 3	Third Address Line of the agency
City of Agency	City of agency – free form text from CRS, should be different from an CRS to another
Postal Code	The postal code of the agency
Provider Code	Código utilizado en el GDS para definir al proveedor de transporte de pasajeros, es decir, la aerolínea
Provider Name	Nombre del proveedor de transporte de pasajeros, es decir, el nombre comercial de la aerolínea
EOS Activity Booking Indicator	Este indicador marca las reservas de EOS, aquellas que fueron hechas en un ATO / CTO pero emitidas en una agencia de viajes. El GDS Office Code asignado a la reserva de EOS es el de la agencia de viajes
Active Add Bookings	Reservas hechas por un GDS en su propio sistema.
Active Cancel Bookings	Reservas canceladas hechas por un GDS en su propio sistema.
Passive Add Bookings	Reservas creadas originalmente en otro sistema GDS.
Passive Cancel Bookings	Reservas canceladas originalmente creadas en otro sistema GDS.
Late Cancel Billable	Suma de reservas canceladas (activas y pasivas) y nulas con fecha de transacción mayor o igual al día de salida
Non-Billable Active Bookings	La reserva activa modificada no es facturable. Solo aplicable para Amadeus
Non-Billable Passive Bookings	La reserva activa pasiva modificada no es facturable. Solo aplicable para Amadeus
Net Bookings	Net bookings = Active Add Bookings + Passive Add Bookings + Late Cancel Bookings - Active Cancel Bookings - Passive Cancel Bookings - Excluded Office Net Bookings

Tabla 5: Información procedente de CBI



3.1.2. ABI

ABI (Amadeus Billing Information) contiene información de facturación de las transacciones efectuadas por las agencias de viajes a través de los sistemas de Amadeus en los negocios Air, Hotel, Car y Rail.

En el apartado 2.1.2. Fuentes de Información, se mencionó que está basado en BIDT, una fuente de datos suministrada por los GDS a sus proveedores de servicios de viajes para su procesamiento y análisis automatizado, con el fin de optimizar el negocio.

El formato de todos los ficheros ABI es muy similar, aunque con diferencias en las medidas de cada negocio:

CAMPO	DESCRIPCIÓN
Comunes	Year
	Month
	Billing Country
	Office ID
	Office Name
	Office Type
	Vendor Code
	IATA Number
Air	Provider Code
	Air added bookings
	Air canceled bookings
	Air modified bookings
Car	Air net bookings
	Car net bookings
	Car vouchers
Hotel	Car manual bookings
	Hotel net bookings
	Hotel nights
Rail	Manual hotel bookings
	Rail add bookings
	Rail cancel bookings
	Rail modify bookings
	Rail net bookings

Tabla 6: Información procedente de ABI



3.1.3. NET TICKETED SEGMENT

La fuente de datos denominada “Net Ticketed Segment” proporciona datos de actividad con respecto a la emisión, reemisión, cancelación y restablecimiento de los tickets que se realizan mensualmente en el sistema central de Amadeus por parte de las oficinas de las agencias de viajes registradas como clientes de Amadeus.

La granularidad de esta información es la misma que en el caso de las reservas de viajes, es decir, a nivel de oficina y proveedor, para así poder comparar el dato de reservas con este y ver si realmente se ha emitido el ticket.

La nomenclatura utilizada en estos ficheros es la siguiente: YYYYMM_TKT_WoW.csv

- YYYY: año al que corresponden los datos
- MM: mes al que corresponden los datos

Mientras que el contenido es el siguiente:

CAMPO	DESCRIPCIÓN
Year	Año al que corresponde la información
Month	Mes al que corresponde la información
Billing Country	Código ISO del país en que se encuentra ubicada la oficina de la agencia de viajes
Office ID	Código de oficina para la agencia que realiza la reserva. Específico del GDS por lo que, en caso de coincidencia entre dos GDS, puede que no representen lo mismo
Office Name	Nombre de oficina para la agencia que realiza la reserva. Específico del GDS por lo que, en caso de coincidencia entre dos GDS, puede que no representen lo mismo
Office Type	Código utilizado por Amadeus para diferenciar Agencias de Viajes (02) y ATO/CTO de proveedores (01)
Vendor Code	Código del socio (normalmente una ACO de Amadeus) que recibe la tarifa de distribución de Amadeus, según los acuerdos definidos entre Amadeus y el proveedor
IATA Number	Identificador de agencia en IATA, en caso de ser una agencia acreditada por esta entidad
Provider Code	Código del proveedor, en este caso una aerolinea, que proporciona el servicio de transporte aéreo
Net ticketed segments - Coupons	Número total de Net Coupons (Add coupon- Cancel Coupons – Refund Coupons)
Non Electronic Active ticketed segments	El número total de ATB, TAT y Manual (registro manual) permite realizar reservas activadas (no pasivas) con el localizador de registros PNR, excluyendo STP, Electronic, STP Electronic y TYs con ticketing de segmentos
Non Electronic Passive ticketed segments	Número total de reservas pasivas emitidas por ATB, TAT y Manual (registro manual) con localizador de registros PNR, excluyendo STP, Electronic, STP Electrónico y segmentos con TY emitidos manualmente
Electronic ticketed Segments	Número total de reservas con ticket electrónico, excluyendo reservas con ticket electrónico STP
STP Ticketed segments	Número total de reservas con etiquetas ATB, TAT y Manual (registro manual) STP (Satellite Ticket Printer) con localizador de registros PNR, excluyendo las reservas con ticket electrónico STP
STP Electronic ticketed segments	Número total de reservas emitidas por STP electrónicos



CAMPO	DESCRIPCIÓN
TY mode ticketed segments	Número total de reservas emitidas en modo TY sin localizador de registros PNR
Cancelled "non electronic" Active ticketed segment	Número total de reservas activas canceladas con ticket
Cancelled "non electronic" Passive ticketed segment	Número total de reservas pasivas con ticket cancelado
Cancelled Electronic ticketed segment	Número total de reservas emitidas electrónicas canceladas
Refunded "non electronic" Active ticketed segments	Número total de reservas ATB, TAT y Manual (registro manual) reembolsadas, activadas (no pasivas) con el localizador de registros PNR, excluyendo las reservas STP reembolsadas, Electronic, STP Electronic y manual TY
Refunded "non electronic" Passive ticketed segments	Número total de reservas pasivas emitidas por ATB, TAT y Manual (registro manual) con localizador de registros PNR, excluyendo las reservas STP reembolsadas, electrónicas, STP electrónicas y TY manuales.
Refunded Electronic ticketed segments	Número total de reservas con ticket electrónicas reembolsadas, excluyendo reservas con boleto electrónico STP reembolsadas
Refunded STP ticketed Segments	Número total de reservas devueltas STB (Impresora de tickets satelitales) ATB, TAT y Manual (registro manual) con localizador de registros PNR, excluyendo los segmentos con ticket electrónico STP reembolsado
Refunded STP Electronic ticketed Segments	Número total de reservas con tickets electrónicos de STP reembolsados
Manual Ticketed Segment	Número total de reservas emitidas manualmente por Amadeus
Cancelled STP "non electronic" ticketed segment	Número total de reservas emitidas canceladas STP no electrónicas
Cancelled STP "electronic" ticketed segment	Número total de reservas emitidas canceladas STP electrónicas
Non-categorized tickets	Número total de reservas emitidas que no pertenecen a una de las categorías anteriores

Tabla 7: Información procedente de Ticketing

3.1.4. CRM ONE VIEW CX

Esta fuente es una extracción procedente del software CRM (Customer Relationship Management) empleado por Amadeus para la administración de la relación con las agencias de viajes, incluso si no son clientes, ya que se dispone de información de otros GDS.

El uso de esta aplicación está dirigido a entender, anticipar y responder a las necesidades de los clientes actuales y potenciales de la empresa para poder hacer crecer el valor de la relación. Este software guía a los comerciales de la empresa para que sepan agrupar a las agencias de viajes con el propósito de mejorar las relaciones comerciales, centrándose específicamente en la retención de estos clientes y, en última instancia, impulsando el crecimiento de las ventas.

Los trabajadores de Amadeus recopilan en CRM diferentes datos, tales como:



- información de contacto de las agencias de viajes.
- listado de puntos de venta de cada cliente, identificados mediante los mencionados Office ID de cada GDS.
- registro mediante oportunidades de las etapas del proceso de venta así como los recursos humanos asignados a cada agencia, etc.

Además, a fin de planificar y coordinar las ventas, existe una segmentación de agencias de viajes que se aplica a nivel mundial. Dicha segmentación contiene los segmentos y sub-segmentos de la figura 36.

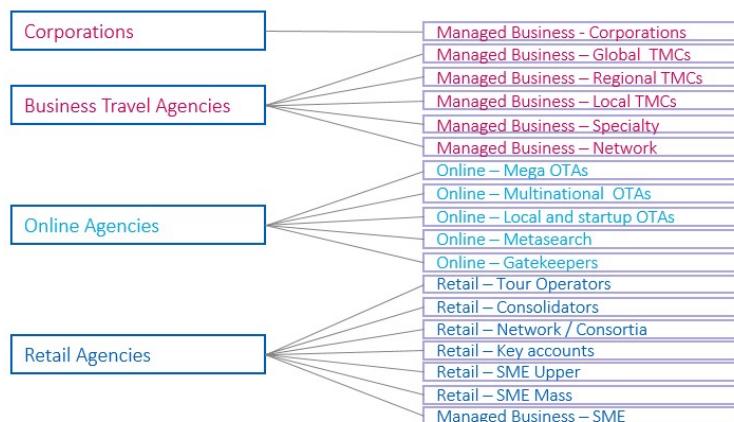


Figura 36: Segmentación de Marketing para agencias de viajes

En el ámbito de este proyecto, la información que nos resulta relevante es la jerarquía de cuentas, es decir, el listado de oficinas asociadas a cada master account / agencia de viajes, así como la segmentación central de marketing de cada agencia. De ahí que los ficheros procedentes de esta fuente contengan la información recogida en la tabla 8:

CAMPO	DESCRIPCIÓN
Master Account Number	Central Account Number del master account
Master Account Name	Nombre de la agencia de viajes
Office ID	Código de oficina utilizado por la agencia para realizar reservas en el GDS. Específico de cada GDS
Office Name	Nombre de la oficina que utiliza el Office ID. Específico de cada GDS
GDS Code	Código del GDS al que pertenece el Office ID
GDS Name	Nombre del GDS al que pertenece el Office ID
IATA Number	Identificador de agencia en IATA, en caso de ser una agencia acreditada por esta entidad
Country Code	Código ISO del país en que se encuentra ubicada la oficina de la agencia de viajes
Marketing Segmentation	Sub-segmento de la agencia de viajes, según la Segmentación de Marketing
Start Date	Fecha de inicio de la asociación Office ID-Master Account
End Date	Fecha de finalización de la asociación Office ID-Master Account

Tabla 8: Información procedente de CRM



3.2. MÓDULOS DE INFORMACIÓN

Se ha decidido dividir la información del proyecto en 4 módulos: bookings (reservas), ticketing y CRM. Además, existe otro módulo con información común.

Los módulos de bookings y ticketing tienen diferentes tablas de hechos que hacen referencia a una serie de tablas de dimensiones comunes. Dependiendo de la granularidad de la tabla de hechos, encontramos esquemas de copo de nieve o estrella.

- Copo de nieve: en este esquema podemos encontrar más de una tabla para cada dimensión. Se ha utilizado para las tablas de hechos principales o detalle, ya que contienen datos que no cambiarán con el tiempo y no requieren recargas pero sí consultas complejas. Por ejemplo, el número de reservas producidas por un office ID en un proveedor en un mes concreto no cambiará, mientras que esa producción por agencia de viajes sí es cambiante ya que un vendedor puede cambiar la jerarquía Office ID-Master Account en CRM, afectando al histórico de datos.
- Estrella: se ha empleado en las tablas de hechos agregadas, generalmente para acelerar las consultas, ya que contiene solo una tabla para cada dimensión.

Tanto para las tablas de información común como para cada uno de los módulos, se adjunta un esquema gráfico de las mismas, seguido de una descripción de su contenido, con el formato de las tablas 9 y 10:

NOMBRE DE TABLA	Nombre de la tabla de hechos
NOMBRE	Nombre de las medidas de negocio
DESCRIPCIÓN	Descripción de las medidas de negocio
TIPO	Tipo de datos
AGREGADO	Operación de agregado de la medida

Tabla 9: Descripción de una medida de hechos

NOMBRE DE TABLA	Nombre de la tabla que actúa como dimensión
NOMBRE	Nombre de la dimensión
DESCRIPCIÓN	Descripción de la dimensión
TIPO	Tipo de datos
CARDINALIDAD	Cardinalidad de la dimensión

Tabla 10: Descripción de una dimensión



3.2.1. INFORMACIÓN COMÚN

Las dimensiones comunes utilizadas en el DM son las que se describen a continuación:

SOURCE

NOMBRE DE TABLA	Com_D_Source
NOMBRE	Source
DESCRIPCIÓN	Fuente de la que procede la reserva de viajes (booking)
TIPO	Texto
CARDINALIDAD	Baja

Tabla 11: Dimensión Source

Esta dimensión se ha creado para identificar el fichero fuente del que procede la información de bookings.

El volumen de datos de CBI es bastante grande por lo que recibimos varios ficheros por cada región geográfica (ej. APAC, NORAM, LATAM...). En ocasiones, se han recibido datos del mismo país en diferentes ficheros, es decir, por duplicado, por lo que se decidió identificar la fuente de cada registro, de manera que es posible anticiparse a ese tipo de errores en el proceso de carga de datos.

Por otra parte, en las tablas de hechos de bookings recogemos datos de CBI (actividad de Amadeus y sus competidores) y ABI (facturación, solo de Amadeus) que comparten medidas (ej. Net Bookings) por lo que resulta necesario diferenciarlas. No obstante, en algunos agregados, se ha optado por eliminar la dimensión y duplicar las medidas para CBI y ABI.

Se trata de una dimensión estática, aunque puede ocurrir que se reciba un fichero con un nombre de región nuevo y haya que actualizarla, aunque el volumen de cambios es muy reducido.

GDS

NOMBRE DE TABLA	Com_D_GDS
NOMBRE	GDS - Global Distribution System
DESCRIPCIÓN	GDS en el que se ha realizado la reserva de viajes (booking)
TIPO	Texto
CARDINALIDAD	Baja

Tabla 12: Dimensión GDS

Esta dimensión es estática, ya que la lista de GDS no ha sufrido cambios desde el inicio del proyecto, aunque los procesos de carga comprueban si existe algún nuevo GDS en los ficheros fuente.



AIRLINE

NOMBRE DE TABLA	Com_D_Airline
NOMBRE	Airline
DESCRIPCIÓN	Aerolínea que proporciona el servicio de transporte aéreo
TIPO	Texto
CARDINALIDAD	Baja

Tabla 13: Dimensión Airline

Esta dimensión contiene el listado de aerolíneas que aparecen tanto en los datos de bookings como de ticketing.

VENDOR

NOMBRE DE TABLA	Com_D_Vendor
NOMBRE	Vendor
DESCRIPCIÓN	ACO de Amadeus que recibe la tarifa de distribución de Amadeus
TIPO	Texto
CARDINALIDAD	Baja

Tabla 14: Dimensión Vendor

Esta dimensión contiene el listado de socios de Amadeus, normalmente ACO, que reciben la tarifa de distribución de Amadeus, según los acuerdos definidos entre Amadeus y el proveedor. Es aplicable a los datos de ABI y ticketing.

GEOGRAPHY

NOMBRE DE TABLA	Com_D_Geography_ML/Hie/Level
NOMBRE	Geography
DESCRIPCIÓN	Niveles geográficos: región, subregión, mercado, country...
TIPO	Texto
CARDINALIDAD	Baja

Tabla 15: Dimensión Geography

El listado de países está basado en la norma ISO 3166-1 alpha-2, que proporciona códigos de 2 letras para los nombres de países. Por tanto, podríamos pensar que esta dimensión es estática, con una carga inicial no incremental del listado de países, pero no es así debido a los cambios organizacionales de la empresa, que requieren la agrupación de países en diferentes niveles geográficos (región, subregión, mercado...).



La idea inicial fue desnormalizar esta dimensión, incluyendo para cada país los valores de los niveles superiores pero finalmente se ha optado por crear una jerarquía geográfica con un nivel básico, país, y N niveles superiores, ya que es más cambiante de lo esperado. No obstante, se dispone de una vista materializada con los niveles más usados.

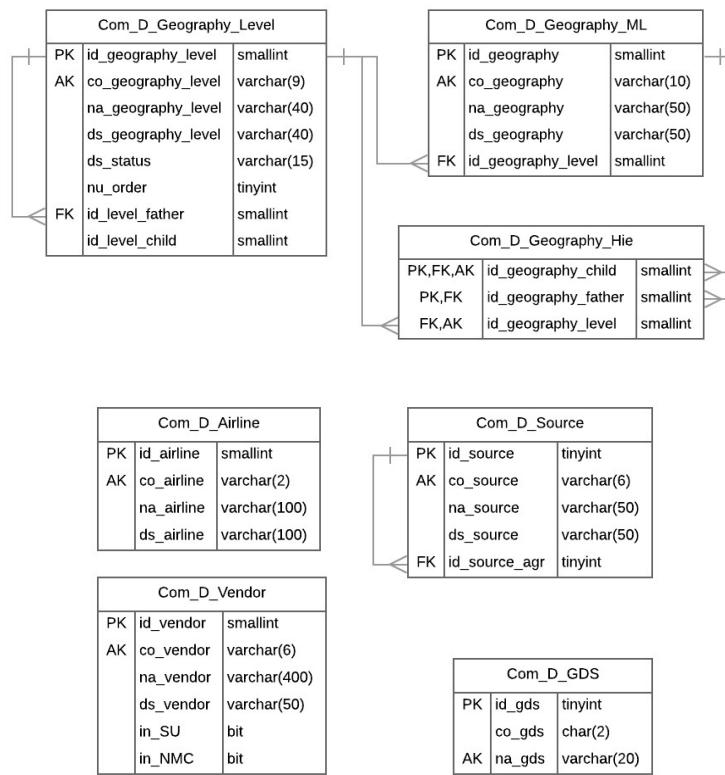


Figura 37: Modelo de datos de las dimensiones comunes

OFFICE

NOMBRE DE TABLA	Com_D_Office
NOMBRE	Office
DESCRIPCIÓN	Oficina que produce la reserva en el GDS de Amadeus
TIPO	Texto
CARDINALIDAD	Alta

Tabla 16: Dimensión Office



La información de oficinas procede de los campos que forman parte del denominado Office Profile:

CBI	ABI y TICKETING
Country Code	Billing Country
GDS Code	Amadeus
GDS Office Code	Office ID
GDS Office Name	Office Name
No Disponible	Office Type
IATA Number	IATA Number
Account Name	
Trade Name	
Address Line 1	
Address Line 2	
Address Line 3	
City of Agency	
Postal Code	

Tabla 17: Office Profile

Tal y como se comentó en el capítulo 2, disponemos de un campo Office ID cuyo nombre nos haría pensar que es un identificador único pero resulta insuficiente para poder identificar de manera consistente a cada oficina.

Por otra parte, la combinación de los 12 campos del Office Profile, método empleado por GSP para construir esta dimensión, resulta redundante e ineficiente. Con esa lógica encontramos muchas oficinas prácticamente iguales, con diferencias mínimas (en los campos Address, City, Postal Code...), que realmente pertenecen a la misma agencia de viajes.

Teniendo en cuenta esto, decidimos usar también una serie de campos para identificar a una oficina de manera inequívoca, pero solo aquellos que observamos que eran determinantes en la asociación oficina-agencia de viajes. Estos campos son: Office ID, Office Name, IATA Number y Country Code.

El resto de campos, a pesar de que van cambiando poco a poco a lo largo del tiempo, encontramos que no son determinantes para crear un nuevo registro en nuestra dimensión.

Por tanto, se trata de una dimensión lentamente cambiante (SCD) dado que parte de la información almacenada sobre una oficina es estática y el resto de campos son más dinámicos. Concretamente, se ha considerado que es:



- SCD de tipo 1, para los cambios en Office Type procedentes de ABI ya que en CBI este campo no existe. Es decir, en caso de existir un cambio en este campo, se sobrescriben los valores existentes en el maestro, Com_D_Office, con los nuevos valores.
- SCD de tipo 4, para el resto de campos (Account Name, Trade Name, Addresses, City and Postal Code). En este caso, se guarda el historial de cambios utilizando fechas de vigencia en una tabla auxiliar, Com_D_Office_Vig, con una clave foránea al maestro. En esta tabla auxiliar, cada registro contiene los valores de estos campos, las fechas en que están vigentes y el motivo por el que se ha insertado el registro: nueva oficina, baja de oficina (deja de aparecer en los datos mensuales), cambio en uno o varios de estos campos tratados como SCD.

Las figuras 37 y 38 muestran el modelo de datos de las tablas que conforman las dimensiones comunes.

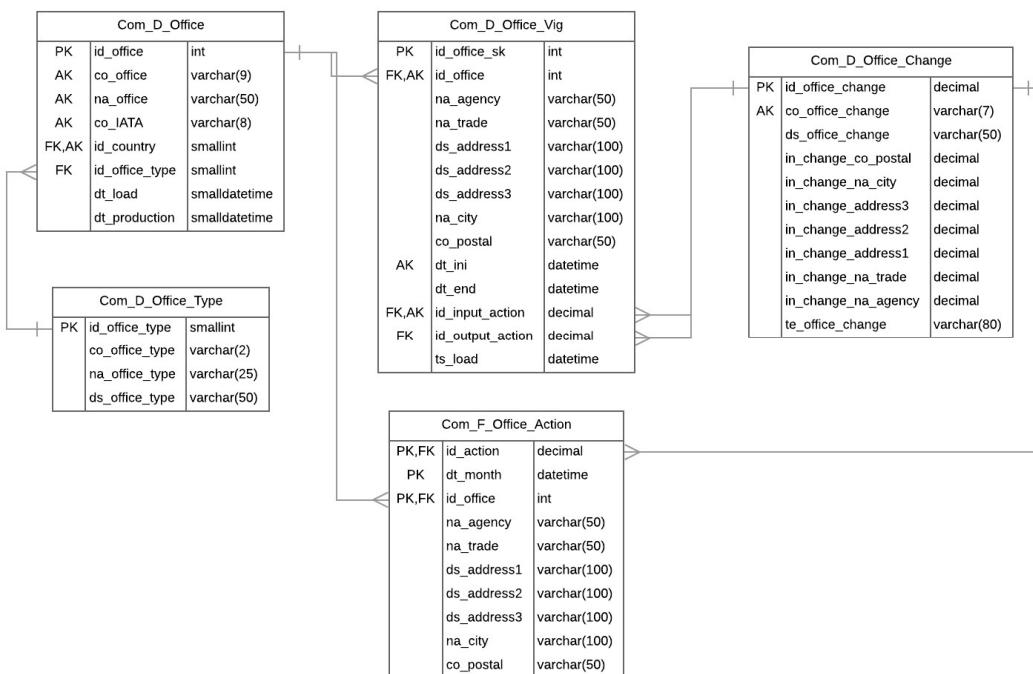


Figura 38: Modelo de datos de la dimensión común Office



3.2.2. CRM

Este módulo de información contiene los datos procedentes de la herramienta CRM de la empresa.

Básicamente, tenemos un listado de agencias de viajes (Master Accounts en CRM) clasificadas según la segmentación de marketing y la lista de oficinas asociadas a cada agencia.

NOMBRE DE TABLA	CRM_D_Account
NOMBRE	Account
DESCRIPCIÓN	Master Account, procedente de CRM, que identifica a la agencia de viajes que realiza reservas en algún sistema GDS
TIPO	Texto
CARDINALIDAD	Alta

Tabla 18: Dimensión Account

NOMBRE DE TABLA	CRM_D_Segment_ML/Hie/Level
NOMBRE	Segment
DESCRIPCIÓN	Segmentación de Marketing. Dispone de varios niveles jerárquicos
TIPO	Texto
CARDINALIDAD	Baja

Tabla 19: Dimensión Segment

En los ficheros de CRM recibimos el par *Master Account-Marketing Segmentation* aunque en realidad se trata del sub-segmento, el nivel básico de segmentación. Después, la agrupación de sub-segmentos no está consensuada globalmente en la empresa, por lo que se ha optado por crear una jerarquía de niveles en lugar de desnormalizar esta dimensión. De esta manera, la agrupación de sub-segmentos resulta más flexible, dependiendo de la audiencia de nuestros datos.

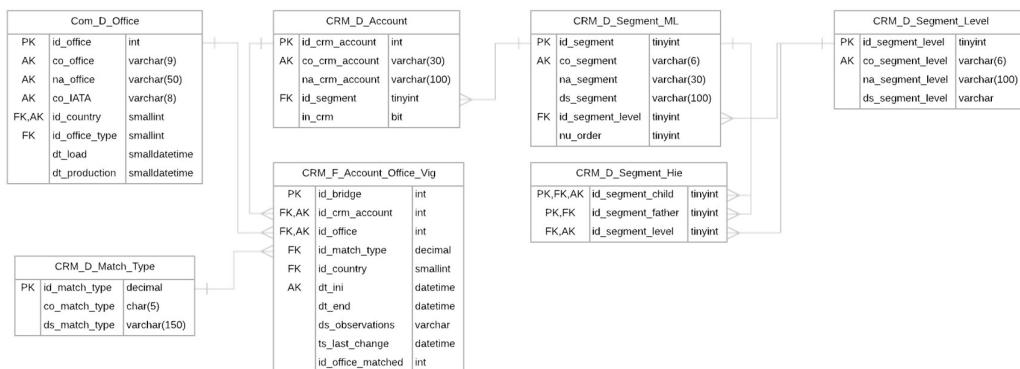


Figura 39: Modelo de datos del módulo CRM



En ocasiones necesitamos describir eventos en nuestro modelo de datos a través de una tabla de hechos sin ningún hecho, ninguna medición, simplemente con claves foráneas a las dimensiones, de modo que podamos obtener información acerca de lo que ha ocurrido.

En el caso de CRM, el histórico de oficinas asociadas a agencias de viajes es un buen ejemplo para aplicar esta técnica: se ha optado por crear una tabla de hechos (CRM_F_Account_Office_Vig) con fechas de vigencia de la relación entre CRM Account (agencia de viajes) y Office (oficinas).

Por otra parte, la identificación y gestión de clientes llevada a cabo por los vendedores en el sistema CRM no es totalmente confiable. No está disponible en todos los países del mundo e incluso, en ocasiones, no cubre a todas las oficinas presentes en CBI. Algunas veces, porque son oficinas de la competencia que el vendedor no sabe o no quiere asignar, mientras que otras veces, aparecen nuevas oficinas que tardarán un tiempo en asignarse. Debido a estas circunstancias, se necesita:

1. Incorporar la agrupación o clasificación alternativa que GSP mantiene en su actual sistema para aquellos países que no disponen de CRM, de ahí la columna in_CRM de la dimensión CRM_D_Account para indicar si es una agencia procedente de CRM.
2. Asignar las oficinas que no tienen CRM Master Account. Para ello, tal y como veremos en el siguiente apartado, dedicado a [procesos ETL](#), se ha implementado un proceso que busca oficinas similares en el histórico de oficinas, en base a unos criterios predefinidos en la dimensión Match_Type, para aplicar el mismo enlace oficina-agencia.

NOMBRE DE TABLA	CRM_D_Match_Type
NOMBRE	Match_Type
DESCRIPCIÓN	Determina el grado de similitud entre 2 oficinas
TIPO	Texto
CARDINALIDAD	Baja

Tabla 20: Dimensión Match_Type

3.2.3. BOOKINGS

El módulo de Bookings engloba toda la información de reservas procedente de CBI (Air) y ABI (Air, Car, Hotel y Rail). Teniendo en cuenta la estructura y el volumen de datos de cada negocio, se ha optado por crear tablas de hechos para cada uno de ellos y diferentes agregados, según los casos de uso identificados.



AIR

Los datos de Air Bookings constan de varias tablas de hechos con distinta granularidad. Por un lado, tenemos una tabla de hechos principal con el máximo detalle (Bkg_F_Bookings_Air) y, por otro, las tablas de agregados con diferentes agrupaciones.

La tabla de hechos principal (figura 40) contiene referencias a las dimensiones comunes: GDS, Source, Geography, Airline, Vendor y Office. En el caso de querer tener datos a nivel Segment o Account, es preciso obtener esa información a partir de Office y sus relaciones con las tablas de CRM, por lo que nos encontramos ante un esquema de copo de nieve.

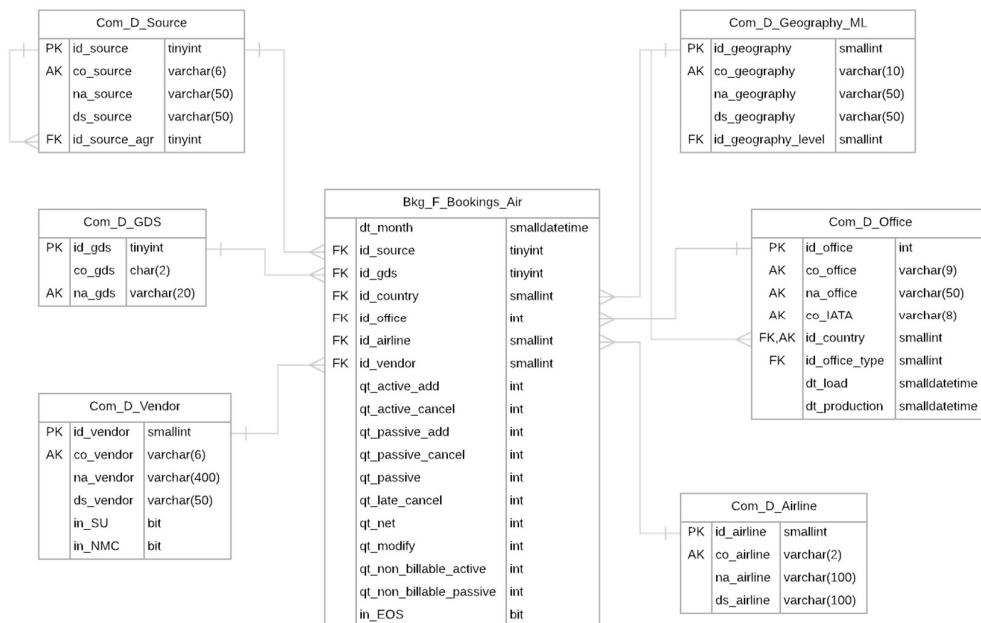
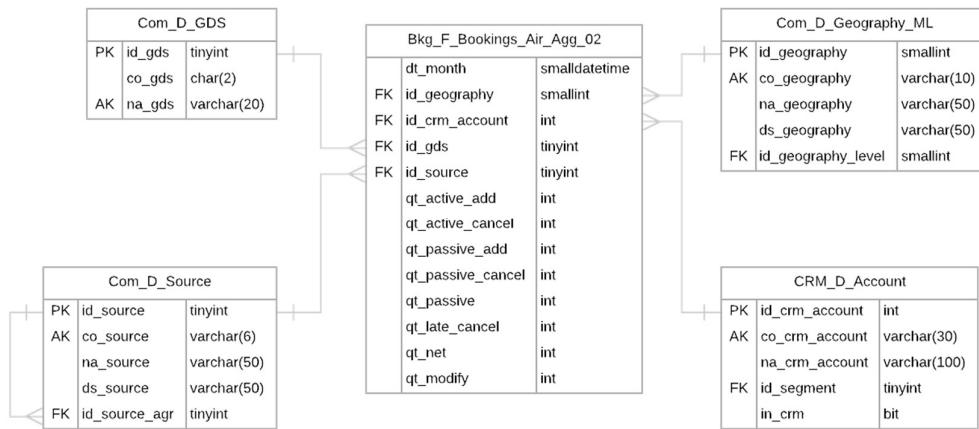


Figura 40: Tabla de hechos y dimensiones de Air Bookings

En cambio, en los agregados disponemos de esquemas en estrella, con datos precalculados a partir de la explotación de la tabla de hechos principal. A continuación, como ejemplo, la figura 41 muestra el modelo de datos del agregado Bkg_F_Bookings_Air_Agg_02.

**Figura 41: Tabla de hechos y dimensiones de un agregado de Air Bookings**

La tabla 21 muestra un resumen de los agregados creados para este módulo de datos y las aperturas de los mismos.

BKG_F_BOOKINGS_AIR					
DIMENSION	AGG_01	AGG_02	AGG_03	AGG_04	AGG_05
GDS	x	x	x	x	x
GEOGRAPHY	x	x	x	x	x
SOURCE	x	x	x	x	x
AIRLINE			x		
OFFICE	x				
ACCOUNT		x			
SEGMENT					x

Tabla 21: Tablas de hechos agregadas de Air Bookings

Respecto a las medidas numéricas de estas tablas de hechos, disponemos de Add Bookings (órdenes de venta) y Cancel Bookings (cancelaciones), tanto Active (transacciones creadas en el propio GDS) como Passive (transacciones creadas fuera del GDS pero que en algún momento del ciclo de venta pasaron a ser gestionadas por el GDS) pero la medida numérica de referencia es Net Bookings, que se calcula a partir de las anteriores.

NOMBRE DE TABLA	Bkg_F_Bookings_Air[_Agg_XX]
NOMBRE	Net Bookings
DESCRIPCIÓN	Reservas netas procedentes de los datos de actividad (CBI) o facturación (ABI)
TIPO	Entero
AGREGADO	Suma

Tabla 22: Medida Air Net Bookings



HOTEL

En el negocio de hoteles, Amadeus también actúa como intermediario entre las agencias de viajes y los hoteles.

En esta parte del negocio, aunque sean reservas de hoteles, no siempre encontramos como proveedores a cadenas hoteleras sino que también podemos encontrar aerolíneas en el caso de la reserva de un vuelo que incluya alojamiento en un hotel, gracias a un acuerdo entre el hotel y la aerolínea. Este tipo de reservas se contabilizan como reservas manuales (Manual Hotel Bookings).

Un problema que encontramos en nuestra fuente de datos es que no disponemos de un campo que nos diferencie el tipo de proveedor, como en la tabla 23:

PROVIDER TYPE	PROVIDER	PROVIDER NAME	HOTEL NET BOOKINGS	HOTEL NIGHTS	MANUAL HOTEL BOOKINGS
AI	AI	AIR INDIA LIMITED	63	90	63
HO	AI	ARMANI HOTELS	10	15	0
AI	BI	ROYAL BRUNEI AIRLINES	2	8	2
HO	BI	BACCARAT AND 1 HOTELS	80	229	0
AI	CX	CATHAY PACIFIC AIRWAYS LTD	6	9	6
HO	CX	COUNTRY INN SUITES	882	2,528	0

Tabla 23: Reservas de Hotel por tipo de proveedor

A veces, un mismo código se corresponde con una cadena hotelera y una compañía aérea por lo que no tenemos un dato por proveedor fiable, como en el ejemplo de la tabla 24:

PROVIDER	HOTEL NET BOOKINGS	HOTEL NIGHTS	MANUAL HOTEL BOOKINGS
AI	73	105	63
BI	82	237	2
CX	888	2,537	6

Tabla 24: Extracto de ABI Hotel Bookings

Afortunadamente, el interés de nuestros usuarios de hoteles no reside en los proveedores, sino en las agencias de viajes por lo que, por el momento, nuestro esquema en estrella no dispone de una dimensión proveedor aceptable. No obstante, en caso necesario, se puede calcular este dato a partir del código de proveedor y las reservas manuales, asumiendo que éstas serán las asociadas a la aerolínea.

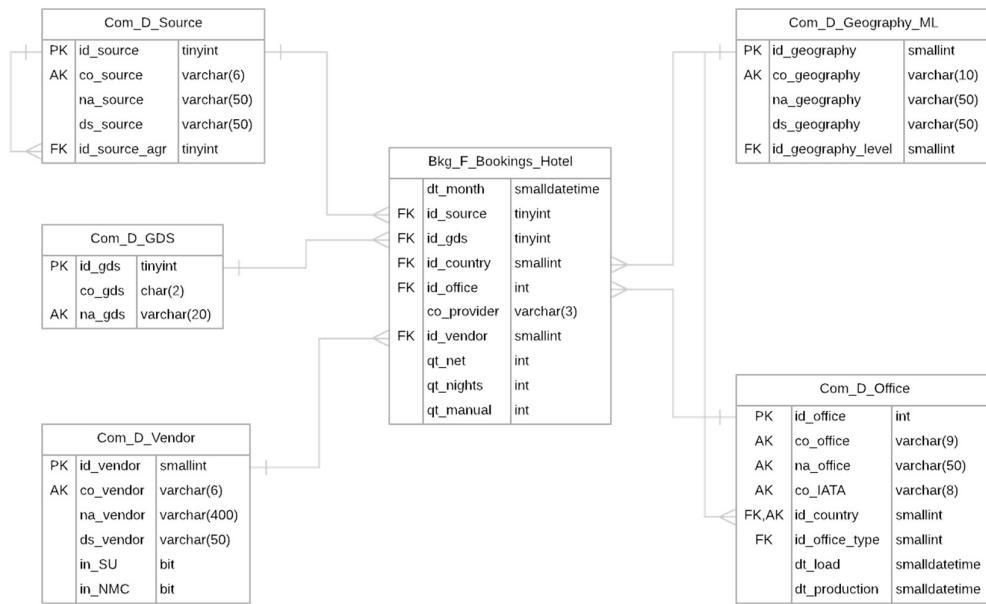


Figura 42: Tabla de hechos y dimensiones de Hotel Bookings

En cuanto a las medidas numéricas utilizadas en Hotel, tenemos:

NOMBRE DE TABLA	Bkg_F_Bookings_Hotel
NOMBRE	Hotel Net Bookings
DESCRIPCIÓN	Reservas netas de hotel facturadas
TIPO	Entero
AGREGADO	Suma

Tabla 25: Medida Hotel Net Bookings

NOMBRE DE TABLA	Bkg_F_Bookings_Hotel
NOMBRE	Manual Hotel Bookings
DESCRIPCIÓN	Reservas de hotel procedentes de una reserva aérea
TIPO	Entero
AGREGADO	Suma

Tabla 26: Medida Manual Hotel Bookings

A pesar de disponer de un dato de noches de hotel reservadas, a día de hoy no se explota porque los incentivos están basados en la cantidad de reservas, independientemente del número de días asociadas a las mismas. Sin embargo, se ha decidido almacenar este dato por si toma importancia en el futuro aunque sería importante tener un dato por cada reserva, para poder hacer algún tipo de análisis del comportamiento de los clientes.



CAR

El volumen de datos de Car es muy inferior a Air. Por esta razón, solo se dispone de una tabla de hechos conectada con las dimensiones comunes (figura 43), es decir, no hay necesidad de crear agregados.

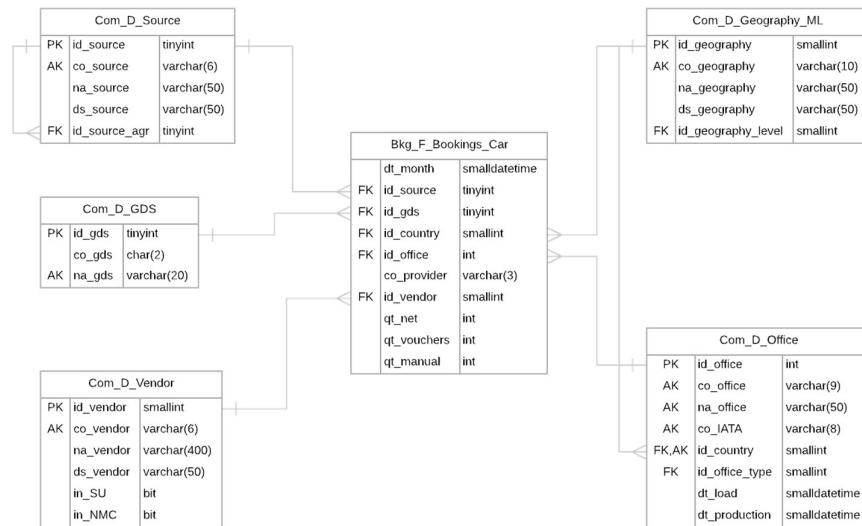


Figura 43: Tabla de hechos y dimensiones de Car Bookings

Aunque se dispone de un código de proveedor, tenemos el mismo problema que en Hotel pero los usuarios de estos datos no lo consideran relevante ya que ellos tratan con la agencia de viajes. No obstante, al tener una medida con los bookings manuales, se podría desagregar el dato en caso necesario.

Respecto a las medidas numéricas, también las reservas netas son el principal indicador de rendimiento.

NOMBRE DE TABLA	Bkg_F_Bookings_Car
NOMBRE	Car Net Bookings
DESCRIPCIÓN	Número de reservas netas de coche
TIPO	Entero
AGREGADO	Suma

Tabla 27: Medida Car Net Bookings

NOMBRE DE TABLA	Bkg_F_Bookings_Car
NOMBRE	Car Manual Bookings
DESCRIPCIÓN	Reservas de coche procedentes de una reserva aérea
TIPO	Entero
AGREGADO	Suma

Tabla 28: Medida Car Manual Bookings



RAIL

La actividad ferroviaria disponible concierne solamente a unos pocos mercados, Francia e Inglaterra, entre otros. Son reservas realizadas a través del sistema de Amadeus, donde no se encuentran más de 4 proveedores, principalmente, SNCF (el equivalente a RENFE en nuestro país) y Eurostar International Limited (presta servicios ferroviarios de alta velocidad en el túnel del Canal de la Mancha). Próximamente se incorporarán datos de otros sistemas ferroviarios, como RENFE.

Debido a esta escasez de proveedores, la apertura por proveedor no resulta relevante a día de hoy por lo que no se ha creado ninguna dimensión, aunque se ha almacenado el dato directamente en la tabla de hechos:

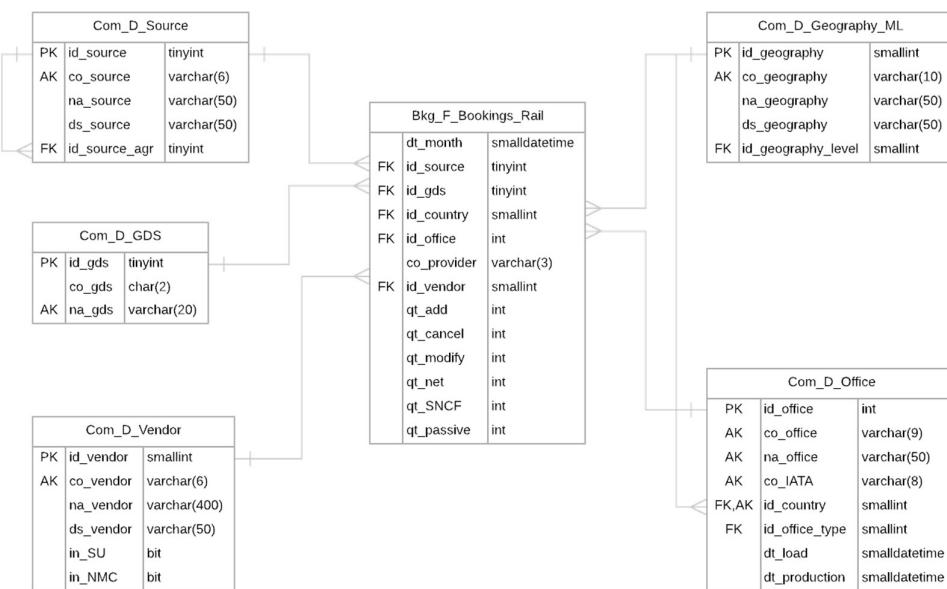


Figura 44: Tabla de hechos y dimensiones de Rail Bookings

Respecto a las medidas numéricas, las reservas netas son el principal indicador de rendimiento. Por otra parte, como estos datos proceden en mayor medida del sistema ferroviario francés, toma cierta relevancia el número de bookings producidos en el sistema de SNCF.

NOMBRE DE TABLA	Bkg_F_Bookings_Rail
NOMBRE	Rail Net Bookings
DESCRIPCIÓN	Número de reservas netas de ferrocarril
TIPO	Entero
AGREGADO	Suma

Tabla 29: Medida Rail Net Bookings



NOMBRE DE TABLA	Bkg_F_Bookings_Rail
NOMBRE	Rail SNCF Bookings
DESCRIPCIÓN	Número de reservas producidas en el sistema SNCF
TIPO	Entero
AGREGADO	Suma

Tabla 30: Medida Rail SNCF Bookings

3.2.4. TICKETING

El módulo de Ticketing es muy similar al de Bookings en términos de dimensiones. Se ha tomado como referencia el anterior, a pesar de que la información de ticketing a día de hoy es solo para Amadeus y se recibe en un único fichero. Por tanto, no se requeriría la identificación del GDS ni comprobaciones de datos de países duplicados, pero se ha decidido mantener el mismo esquema (figura 45) para reutilizar lo máximo posible el material disponible: scripts de creación de tablas de hechos, procesos ETL, queries SQL de explotación...

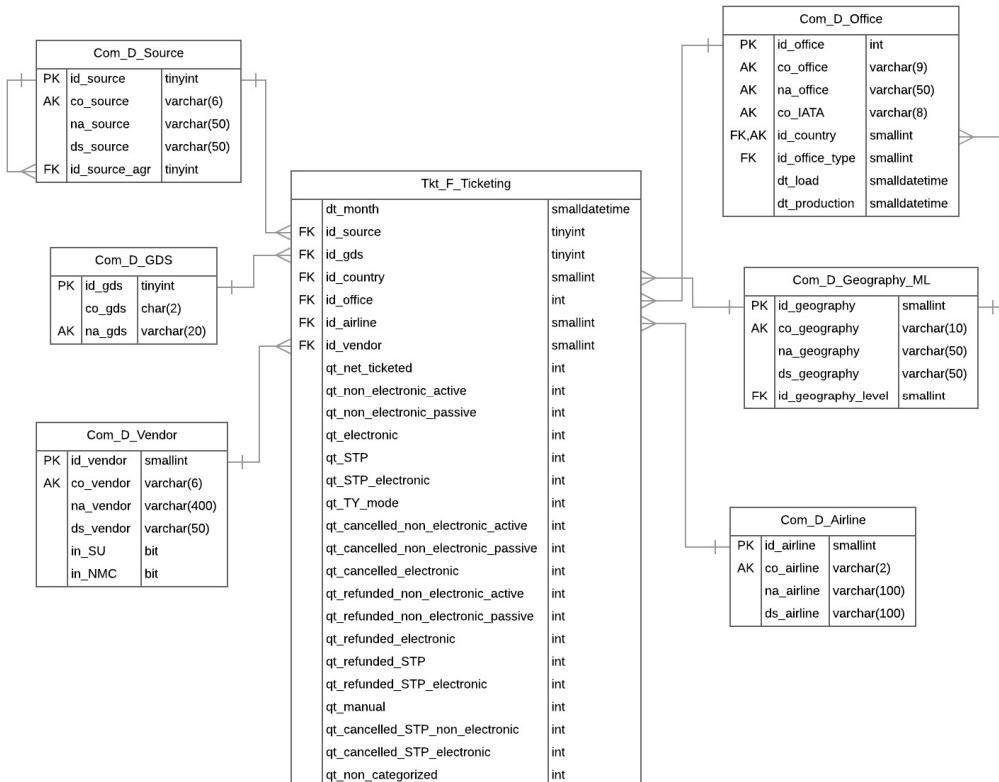


Figura 45: Tabla de hechos y dimensiones de Ticketing



Respecto a las medidas numéricas, en esta fuente de datos recibimos un gran número de indicadores, aunque el número de reservas emitidas (ticketed) son el principal indicador de rendimiento, puesto que la comparación de este valor con el número de reservas netas puede otorgarnos información de mucho valor en términos de fraude; una agencia de viajes puede realizar muchas reservas de viajes, por las que recibe un incentivo, pero luego nunca son emitidas y finalmente las cancela.

NOMBRE DE TABLA	Tkt_F_Ticketing[_Agg]
NOMBRE	Net Ticketed
DESCRIPCIÓN	Número de reservas emitidas
TIPO	Entero
AGREGADO	Suma

Tabla 31: Medida Net Ticketed

3.3. PROCESOS DE CARGA

Los procesos ETL de este proyecto se han desarrollado en TOL, usando el framework de Bayes para este entorno, TOL Process Manager. No obstante, también hemos recurrido a PDI, por ejemplo, para cargar el contenido de los ficheros fuente, con formato CSV, en tablas temporales de nuestra base de datos.

La mayoría de las fuentes de información del proyecto se reciben mensualmente, excepto los datos procedentes de CRM One View CX, que están disponibles diariamente. Teniendo en cuenta esta periodicidad, se han definido dos procesos ETL globales automáticos:

- Mensual: su ejecución está programada para el día 5 de cada mes, a las 12 de la mañana, puesto que el día 4 se deberían de haber recibido los datos de ABI, CBI y Net Ticketed Segment. No obstante, en caso de haber una incidencia que suponga un retraso en la recepción de los datos, podemos modificar esta programación e incluso lanzarlo manualmente.

En este proceso se actualizan las tablas de detalle de Bookings y Ticketing. Dicho proceso de actualización leerá la última fecha insertada en cada tabla y la comparará con la última fecha disponible en la fuente de información correspondiente. Si observa que hay datos nuevos, realizará la inserción de los meses que no tenga para ponerse al día.

- Diario: programado a las 4 de la madrugada, se encarga de actualizar los cambios que se hayan producido en CRM en cuanto a la segmentación comercial de las agencias de viajes y de la asignación de oficinas a agencias. Estos cambios implican la actualización de aquellas tablas de hechos agregadas que estén



afectadas por estos cambios en el cálculo de sus medidas. Por ejemplo, los agregados que tengan datos materializados del número de reservas por agencia de viajes, el número de agencias por segmento o sub-segmento, etc.

Además, se dispone de un proceso de actualización manual, en el que se pueden seleccionar los procesos ETL que se quieren actualizar de cada módulo de información. También es posible suministrar parámetros como, por ejemplo, las fechas a cargar, bien porque se ha recibido una nueva versión de los datos que requiere recargar las tablas de hechos afectadas o debido a que el proceso automático ha fallado en algún proceso ETL y existen "huecos" de meses o fuentes no informadas que el actualizador automático no consigue identificar.

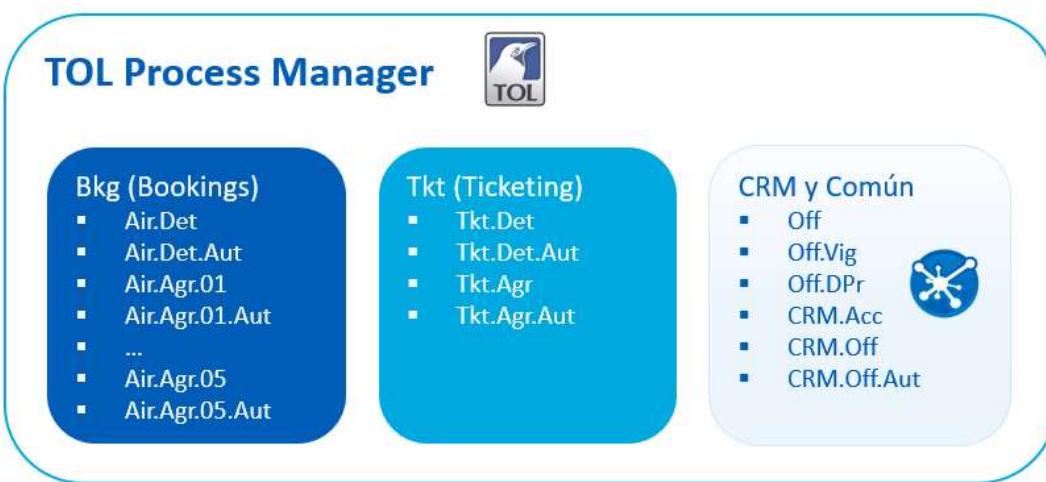


Figura 46: Esquema de procesos en TOL Process Manager

En los siguientes apartados se describen los principales procesos ETL desarrollados en el proyecto (ver esquema gráfico en la figura 46). Para ello, hemos utilizado el formato de la tabla 32:

PROCESO	DESCRIPCIÓN
Código del proceso	Descripción de las principales acciones que realiza el proceso

Tabla 32: Descripción de un proceso ETL



3.3.1. BOOKINGS

El tamaño de las tablas de hechos que contienen la información de las reservas aéreas mensuales es considerable (4 millones de registros al mes) por lo que se ha decidido crear, además de la tabla de hechos detalle, diversos agregados con el fin de poder explotar estos datos en tiempos razonables.

De esta manera, se han creado dos tipos de procesos ETL en la carga de los datos de reservas, tanto Air como Non-Air, dependiendo del tipo de tablas de hechos afectadas: detalle o agregados.

Para las tablas de hechos de detalle, se realizan las siguientes acciones:

1. La información de los ficheros fuente se carga, a través de PDI, en tablas temporales: Bkg_F_Bookings_XXX_Tmp, siendo XXX el nombre del negocio (Air, Car, Hotel, Rail).

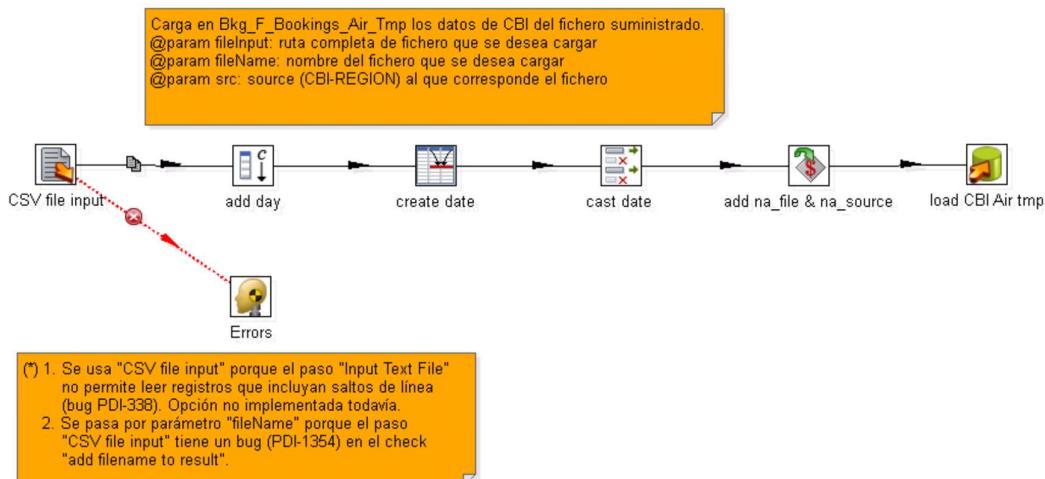


Figura 47: Transformación creada para cargar los ficheros de CBI en tablas temporales

2. Antes de realizar ningún cambio en nuestro modelo de base de datos, se realizan una serie de verificaciones de los datos cargados mediante sentencias SQL, codificadas en TOL y ejecutadas vía ODBC. Por ejemplo, se comprueba que la fecha que aparece en los ficheros se corresponde con los datos que contiene, o que los datos de un país proceden de la región correcta y no están repetidos en otro fichero.
3. Se actualizan las dimensiones afectadas por esta carga de datos, en caso de existir nuevos valores. No debería de suceder sin previo aviso pero, para que no haya problemas en la carga, se insertan nuevos valores en la dimensión GDS, Source y Geography y se traza una advertencia en el log. También se actualiza la dimensión Office, aunque este proceso es más complejo y se explica en detalle en el apartado dedicado a agencias de viajes.



4. Por último, la información del mes a cargar se inserta en la tabla de hechos principal que contiene el detalle del negocio correspondiente: Bkg_F_Bookings_XXX.

Con respecto a la carga del detalle de los datos de reservas, en los ficheros TOL del proyecto se encuentran definidos los siguientes procesos:

PROCESO	DESCRIPCIÓN
Bkg.Air.Det	Recibe como parámetro la fecha del mes que se desea cargar, así como la fuente, ABI o CBI, y ejecuta las transformaciones descritas anteriormente para cargar las reservas aéreas
Bkg.Air.Det.Aut	Busca la última fecha insertada en la tabla de hechos y comprueba en el repositorio de ficheros si existen datos más recientes. De ser así, llama al proceso Air.Det
Bkg.NAi.Det	Recibe como parámetros la fecha del mes y el nombre del negocio de Non-Air que se desea cargar y ejecuta las transformaciones descritas anteriormente para cargar las reservas del negocio correspondiente
Bkg.NAi.Det.Aut	Recibe como parámetro el nombre del negocio de Non-Air que se desea actualizar, busca la última fecha insertada en la tabla de hechos correspondiente y comprueba en el repositorio de ficheros si existen datos más recientes para ese negocio. De ser así, llama al proceso NAI.Det

Tabla 33: Procesos ETL para el detalle de reservas

Mientras que en el caso de los agregados tenemos:

PROCESO	DESCRIPCIÓN
Bkg.Air.Agr.NN (NN: 01-05)	Recibe como parámetro una fecha de inicio y una fecha de fin y lo que hace es borrar inicialmente los datos del agregado entre esas fechas e insertarlas de nuevo desde la tabla de hechos detalle
Bkg.Air.Agr.NN.Aut (NN: 01-05)	Busca la última fecha insertada en la tabla de hechos agregada y comprueba la última fecha disponible en la tabla de hechos detalle. En caso de ser más reciente el detalle, llama al proceso Air.Agr.NN

Tabla 34: Procesos ETL para los agregados de reservas



3.3.2. TICKETING

Con respecto a la información de ticketing, también se ha creado un proceso ETL para la tabla de hechos detalle y otro para el agregado.

Para la tabla de hechos que contiene el detalle, se han definido los procesos ETL en TOL recogidos en la tabla 35:

PROCESO	DESCRIPCIÓN
Tkt.Det	Recibe como parámetro la fecha del mes que se desea cargar y ejecuta las siguientes acciones: <ol style="list-style-type: none">1. Carga en una tabla temporal, Tkt_F_Ticketing_Tmp, la información mensual del fichero de Net Ticketed Segment. Esta acción se ha implementado mediante una transformación de PDI.2. Comprueba que la fecha que aparece en el fichero cargado corresponde con los datos que contiene.3. Actualiza la dimensión Geography, en caso de existir nuevos valores. No debería de suceder sin previo aviso pero, para que no haya problemas en la carga, se insertan los nuevos valores encontrados y se traza una advertencia en el log. También se actualiza la dimensión Office, aunque este proceso se explica en detalle en el apartado dedicado a agencias de viajes.4. Inserta la información del mes a cargar en la tabla de hechos principal que contiene el detalle de los datos de Ticketing: Tkt_F_Ticketing.
Tkt.Det.Aut	Busca la última fecha insertada en la tabla de hechos y comprueba en el repositorio de ficheros si existen datos más recientes. De ser así, llama al proceso Air.Det

Tabla 35: Procesos ETL para el detalle de ticketing

En el caso del agregado tenemos:

PROCESO	DESCRIPCIÓN
Tkt.Agr	Recibe como parámetro una fecha de inicio y una fecha de fin y lo que hace es borrar inicialmente los datos del agregado entre esas fechas e insertarlas de nuevo desde la tabla de hechos detalle.
Tkt.Agr.Aut	Busca la última fecha insertada en la tabla de hechos agregada y comprueba la última fecha disponible en la tabla de hechos detalle. En caso de ser más reciente el detalle, llama al proceso Tkt.Agr

Tabla 36: Procesos ETL para el agregado de ticketing



3.3.3. AGENCIAS DE VIAJES

En este apartado se explican los procesos ETL desarrollados para las dimensiones que contienen información de agencias de viajes. Nos centraremos en Office y Account.

OFFICE

Después de analizar las fuentes de datos que afectan a Office y el proceso actual de clasificación de oficinas utilizado por GSP en su actual sistema, se ha decidido considerar a la dimensión Office como una dimensión lentamente cambiante (SCD). SCD de tipo 1 para los datos procedentes de ABI, donde disponemos del campo Office Type; y SCD tipo 4, para los campos del Office Profile que solo están presentes en CBI.

A continuación, la tabla 37 describe las acciones realizadas por este proceso ETL:

PROCESO	DESCRIPCIÓN
Com.Off	Recibe como parámetros la fuente de datos y la fecha del mes que se desea cargar y ejecuta las siguientes acciones: <ol style="list-style-type: none">1. Comprueba el contenido de la tabla temporal utilizada para cargar los ficheros fuente, por ejemplo, <i>Bkg_F_Bookings_Air_Tmp</i> en el caso de reservas aéreas. Si existen nuevas oficinas con respecto a <i>Com_D_Office</i>, se insertan en la dimensión.2. Depende de la fuente de información:<ol style="list-style-type: none">a. En el caso de datos procedentes de ABI o Net Ticketed Segment, se actualizan en <i>Com_D_Office</i> el valor de <i>office_type</i> (SCD tipo 1), si procede. Es decir, si encontramos una oficina de Amadeus que no tiene informada esta columna porque se cargó en nuestra dimensión desde CBI, donde no existe esta información, sobreescrivimos este valor.b. En el caso de datos procedentes de CBI, se llama al proceso Com.Off.Vig.
Com.Off.Vig	Recibe como parámetro la fecha del mes que se desea cargar y ejecuta las siguientes acciones: <ol style="list-style-type: none">1. Compara los datos de oficina que no forman parte de la tabla <i>Com_D_Office</i> (Account Name, Trade Name, Addresses, City and Postal Code) con la tabla de vigencias, <i>Com_D_Office_Vig</i>.2. Inserta nuevos registros en caso de identificar cambios con respecto al mes anterior3. Cierra las fechas de vigencia de los registros que han dejado de estar vigentes por los cambios identificados. Por ejemplo, si una oficina ha cambiado su Trade Name con respecto al mes anterior, se insertará un nuevo registro con los nuevos valores, fecha de inicio el mes que se está cargando y fecha de fin vacía para indicar que es un registro vigente.

Tabla 37: Procesos ETL para Office

ACCOUNT

Con respecto a las agencias de viajes, Accounts en nuestro sistema, tenemos principalmente dos procesos ETL: uno para carga los datos de CRM One View CX (Account, Segment y relaciones Office-Account) y otro para asignar las oficinas que no existen en CRM.



A continuación, la tabla 38 describe el proceso que se encarga de los datos procedentes de CRM One View CX:

PROCESO	DESCRIPCIÓN
CRM.Acc	Este proceso llama a un trabajo de PDI (figura 48) que realiza las siguientes acciones: <ol style="list-style-type: none">1. Comprueba si existe el fichero procedente de CRM2. Transforma el fichero de formato XLSX a CSV por cuestiones de eficiencia en PDI.3. Ejecuta una transformación que carga el contenido de este fichero en una tabla temporal de la base de datos.4. Actualiza la dimensión Segment, que contiene la Segmentación de Marketing, con los nuevos valores recibidos, si procede.5. Ejecuta una transformación que marca las agencias de viajes globales.6. Actualiza la dimensión Account, que contiene las agencias de viajes de CRM.7. Comprueba los pares Office-Account y elimina los valores duplicados, en caso necesario. Además, hay casos en que una oficina está asignada a dos agencias durante un mes, lo que supone un solapamiento en los datos mensuales. En estos casos, se mantiene la relación más reciente.8. Actualiza la tabla de hechos sin hechos CRM_F_Account_Office_Vig con las relaciones históricas de Office-Account.

Tabla 38: Proceso ETL que carga el contenido procedente de CRM

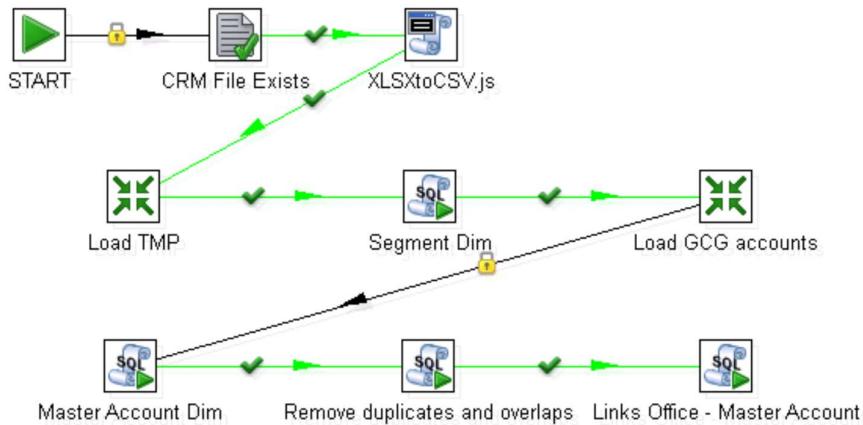


Figura 48: Trabajo que carga el contenido procedente de CRM

Respecto al proceso encargado de asignar las oficinas que no aparecen en CRM a pesar de existir en CBI, se ha decidido crear un algoritmo que empareje cada una de estas oficinas con la agencia de viajes a la que pertenece la oficina más parecida. Básicamente, consiste en “aprender” de las relaciones ya existentes en CRM_F_Account_Office_Vig.

Para efectuar esta asignación, se cruzan las oficinas recién creadas en Com_D_Office con el histórico de oficinas de esta dimensión en los últimos 2 años, en busca de similitudes.



En función de los campos coincidentes en ambos conjuntos de datos, oficinas nuevas e históricas, el algoritmo asocia un valor a este grado de similitud (% Match Type), según los valores recogidos en la tabla 39.

Además, este valor se incrementa con las correspondencias encontradas en los campos almacenados en Com_D_Office_Vig. La tabla 40 recoge todas las combinaciones posibles, junto con su Match Type asociado.

Como resultado de este cruce se obtiene, para cada nueva oficina, la oficina con mayor Match Type y, en consecuencia, se asocia la oficina recién creada con la agencia de viajes a la que está unida la oficina histórica. Es decir, se crea un registro en la tabla CRM_F_Account_Office_Vig que relaciona la oficina nueva con la agencia de viajes a la que pertenece la oficina con mayor Match Type. También se almacena el resultado del valor Match Type para que los usuarios lo validen en el módulo de mantenimiento (esto lo veremos en detalle en el apartado 3.4.4.).

Por otra parte, si en este proceso no se encuentran coincidencias con ninguna oficina ya existente, se procede a crear una nueva agencia de viajes en la dimensión Account, cuyo nombre será el valor del campo Office Name de la nueva oficina, y se le asigna la oficina en la tabla CRM_F_Account_Office_Vig. Además, el valor de la columna Match_Type será 0 para que en el interfaz de mantenimiento aparezca como una de las oficinas nuevas a revisar.

COM_D_OFFICE					% MATCH TYPE	CAMPOS COINCIDENTES
OFFICE ID	AMADEUS	OFFICE NAME	IATA NUMBER	OFFICE TYPE		
1	1	1	0	1	96	OfficeID (Amadeus), Office Type y Office Name
1	1	1	0	0	92	OfficeID (Amadeus) y Office Name
1	0	1	0	0	88	OfficeID (Competencia) y Office Name
1	1	0	1	1	84	OfficeID (Amadeus), Office Type y IATA Number
1	1	0	1	0	80	OfficeID (Amadeus) y IATA Number
0	N/A	1	1	N/A	75	Office Name y IATA Number
1	0	0	1	0	70	OfficeID (Competencia) y IATA Number
0	N/A	1	0	N/A	65	Office Name
1	1	0	0	1	60	OfficeID (Amadeus y Office Type)
1	1	0	0	0	55	OfficeID (Amadeus)
1	0	0	0	0	50	OfficeID (Competencia)

Tabla 39: Lista de posibles valores del proceso "Match Type" (1)



COM_D_OFFICE_VIG					% MATC H TYPE	CAMPOS COINCIDENTES
AGENCY NAME	ADDRESS	TRADE NAME	CITY NAME	POSTAL CODE		
0	0	0	0	0	0	
0	0	0	0	1	0.1	Postal Code
0	0	0	1	0	0.2	City Name
0	0	0	1	1	0.3	City Name, Postal Code
0	0	1	0	0	0.4	Trade Name
0	0	1	0	1	0.5	Trade Name, Postal Code
0	0	1	1	0	0.6	Trade Name, City Name
0	0	1	1	1	0.7	Trade Name, City Name, Postal Code
0	1	0	0	0	0.8	Some Address(es)
0	1	0	0	1	0.9	Some Address(es), Postal Code
0	1	0	1	0	1	Some Address(es), City Name
0	1	0	1	1	1.1	Some Address(es), City Name, Postal Code
0	1	1	0	0	1.2	Some Address(es), Trade Name
0	1	1	0	1	1.3	Some Address(es), Trade Name, Postal Code
0	1	1	1	0	1.4	Some Address(es), Trade Name, City Name
0	1	1	1	1	1.5	Some Address(es), Trade Name, City Name, Postal Code
1	0	0	0	0	1.6	Agency Name
1	0	0	0	1	1.7	Agency Name, Postal Code
1	0	0	1	0	1.8	Agency Name, City Name
1	0	0	1	1	1.9	Agency Name, City Name, Postal Code
1	0	1	0	0	2	Agency Name, Trade Name
1	0	1	0	1	2.1	Agency Name, Trade Name, Postal Code
1	0	1	1	0	2.2	Agency Name, Trade Name, City Name
1	0	1	1	1	2.3	Agency Name, Trade Name, City Name, Postal Code
1	1	0	0	0	2.4	Agency Name, some Address(es)
1	1	0	0	1	2.5	Agency Name, some Address(es), Postal Code
1	1	0	1	0	2.6	Agency Name, some Address(es), City Name
1	1	0	1	1	2.7	Agency Name, some Address(es), City Name, Postal Code
1	1	1	0	0	2.8	Agency Name, some Address(es), Trade Name
1	1	1	0	1	2.9	Agency Name, some Address(es), Trade Name, Postal Code
1	1	1	1	0	3	Agency Name, some Address(es), Trade Name, City Name
1	1	1	1	1	3.1	Agency Name, some Address(es), Trade Name, City Name, Postal Code

Tabla 40: Lista de posibles valores del proceso "Match Type" (2)



3.4. APPLICACIÓN WEB

Marketing Intelligence Tracking (MIT) es el nombre de la aplicación web desarrollada para Amadeus por diferentes trabajadores de la empresa Bayes Forecast, entre los que se encontraba el alumno.

Tal y como se comentó en el capítulo 1, se trata de una herramienta basada en un producto ya existente en Bayes Forecast, la BWS, personalizable en función de cada cliente.

Las tecnologías empleadas para desarrollar MIT son las siguientes:

- Back-end: programación en el lado del servidor.
- Está escrito en Java, utilizando el framework Spring. El alumno no ha participado en la implementación de esta parte, sino que se ha limitado a usar los servicios proporcionados.
- Front-end: programación en el lado del cliente.
- Se han utilizado los lenguajes HTML, CSS y JavaScript (JS). El framework Sencha Ext JS es la base de este lado de la aplicación. Además, para los gráficos se ha empleado una extensión de la librería HighCharts, disponible en Ext JS.

Como resultado, tenemos una aplicación web que dispone de tres módulos de explotación: Dashboard, Reports, Data Explorer; y uno de mantenimiento, Maintenance. Todos ellos serán explicados en los próximos apartados de este capítulo.

Asimismo, se dispone de una base de datos interna con el siguiente contenido:

- Módulos (BWS_D_Module): nombre de los módulos implementados.
- Roles (BWS_D_Role): perfiles creados para administrar el acceso a los módulos. Esto permite aplicar las mismas reglas de acceso a un grupo de usuarios. Cada rol está asociado a uno o varios módulos, con unos determinados permisos.
- Usuarios (BWS_D_User y BWS_D_User_Attributes): lista de usuarios del sistema. Un usuario se caracteriza por los roles que tiene asignados y una serie de atributos que determinan a qué datos geográficos puede acceder y si puede generar informes en el repositorio compartido.
- Datasets (BWS_D_Dataset): lista de consultas SQL parametrizadas.

Esta base de datos está alojada en la misma instancia Microsoft SQL Server en la que tenemos el DM de MIT. La figura 49 muestra su modelo de datos:

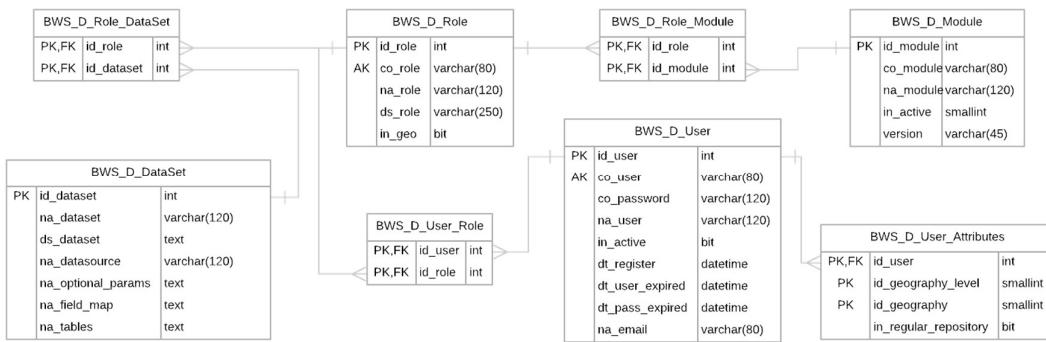


Figura 49: Modelo de datos de la base de datos interna de MIT

La tabla de datasets resulta esencial para algunos módulos del front-end, como Dashboard y Maintenance, ya que simplemente consumen información procedente de la base de datos del negocio, el DM de MIT, a través de llamadas al back-end que devuelven el resultado en formato JSON. Para ello, en necesario realizar lo siguiente:

1. Crear una consulta SQL que devuelva los datos que se desean mostrar.
2. Insertar en la tabla de datasets, BWS_D_Dataset, esta consulta junto con el nombre de los parámetros que contenga: fecha, country, account, segment...
3. Crear en el front-end, normalmente en algún fichero JS que usa el framework Ext JS, una llamada a un servlet del back-end, llamado GET, pasándole el nombre del dataset y los valores de sus parámetros, si los tiene. Este servlet es un controlador que gestiona las peticiones a la base de datos. Con el nombre del dataset y los valores de los parámetros, extrae de la tabla de datasets la consulta SQL que se lanzará contra el DM, reemplazando las variables por los valores pasados en los parámetros. Finalmente, ejecuta la consulta SQL contra el DM y devuelve al front-end un JSON con el resultado.
4. Recoger los datos enviados por el back-end en formato JSON y enlazarlos con los gráficos correspondientes mediante las estructuras de datos que ofrece Ext JS.

En resumen, teniendo en cuenta este modelo de programación, podemos decir que el trabajo del alumno en los módulos Dashboard y Maintenance se limita a crear las consultas SQL que generen los datos que se desean visualizar, introducir las llamadas al back-end en los ficheros del front-end y mostrar de forma gráfica los datos recibidos en formato JSON.

La aplicación MIT se distribuye en formato WAR y en este proyecto se ejecuta en el contenedor web Apache Tomcat, una vez desplegada. Además, en este entorno, tenemos otras dos aplicaciones:



- BIRT Engine versión 3.7.0. Es el motor que usamos para ejecutar en MIT los informes creados con BIRT Report Designer. Pueden encontrarse más detalles de esta tecnología en el apartado 2.3.4.
- Pentaho Business Analytics Server versión 5.4 CE. Plataforma BI que contiene el motor OLAP Mondrian, descrito en el apartado 2.3.3.

MIT consume los servicios de estas dos aplicaciones a través de los módulos Reports y Data Explorer, respectivamente. La figura 50 muestra un esquema de los componentes de MIT mencionados y su interacción en la aplicación:

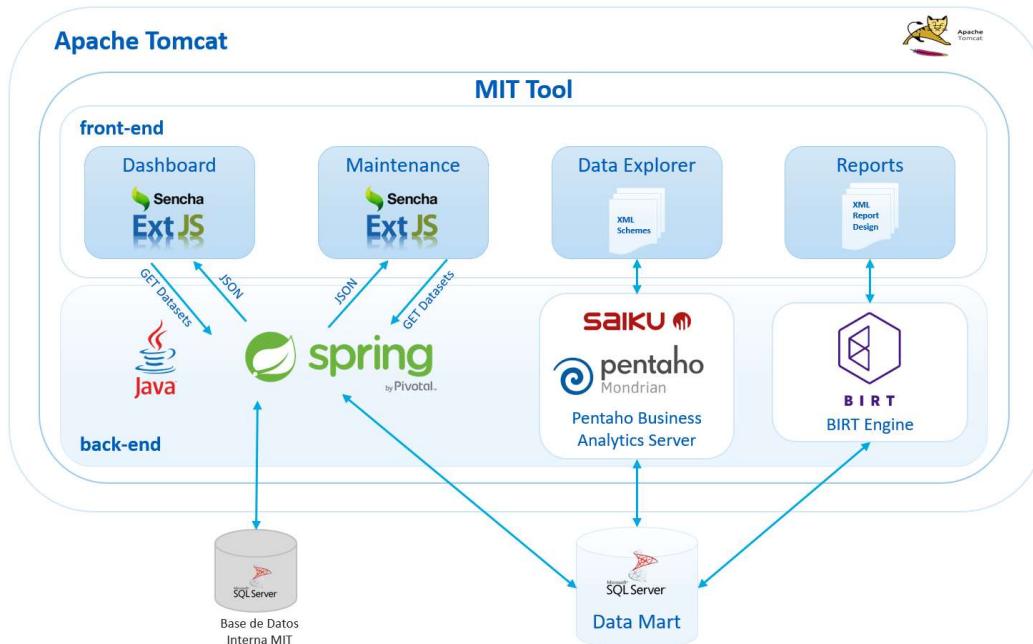


Figura 50: Componentes de MIT

3.4.1. DASHBOARD

Este módulo consiste en un cuadro de mando que ofrece una visualización intuitiva de la información almacenada en el DW del proyecto. Permite a los usuarios analizar las tendencias del mercado y de los clientes de Amadeus y sus principales competidores, de forma que puedan interpretar lo que está ocurriendo y decidir si se deben tomar medidas de mejora y/o promover nuevas acciones comerciales.

Cada GDS cuenta entre sus clientes con proveedores (aerolíneas, cadenas hoteleras...) y agencias de viajes, aunque la audiencia del Dashboard de MIT está más interesada en visualizar los KPI de la actividad de las agencias de viajes.



El principal objetivo de este módulo es proporcionar, de una manera sencilla y fácil de usar, elementos clave para una planificación, análisis, segmentación y seguimiento de ventas adecuados, con el fin de maximizar los resultados de estas actividades. Las visualizaciones creadas ayudan al usuario a:

- Medir la cuota de mercado de Amadeus en un mercado o segmento así como el rendimiento del equipo de ventas
- Identificar agencias de viajes que se están yendo a la competencia antes de que se materialice la pérdida del cliente y de las reservas que hace con Amadeus
- Encontrar oportunidades de crecimiento, en base a las tendencias del mercado y la evolución de los segmentos y los principales clientes de la competencia
- Identificar las brechas entre los objetivos y los resultados
- Alinear tácticas a corto plazo con la estrategia corporativa
- Ponerse en el lugar de un competidor, viendo en qué cuentas está creciendo y/o maximizando sus beneficios

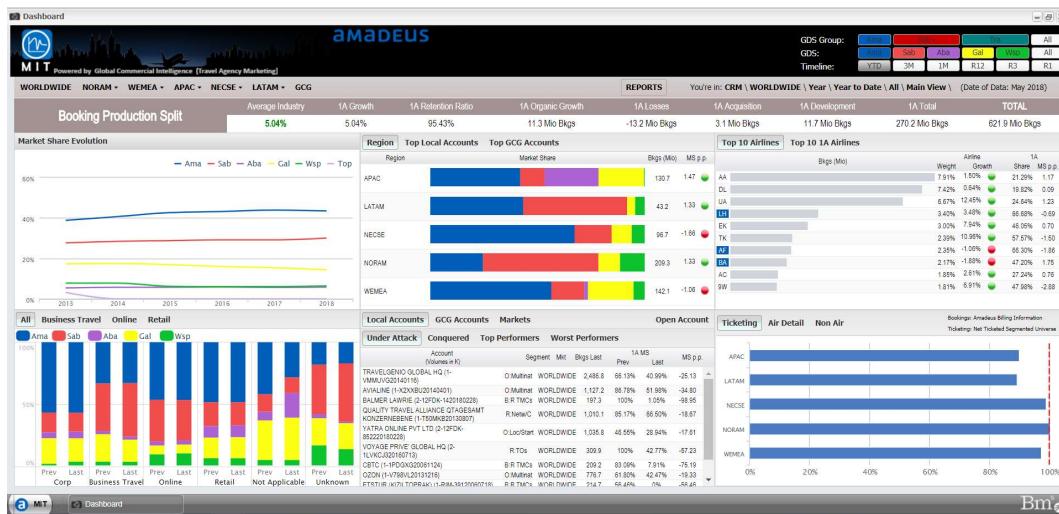


Figura 51: Portada del Dashboard

La figura 51 muestra la pantalla principal del Dashboard. La interfaz de usuario tiene tres vistas, con diferentes niveles de granularidad:

- Main View: contiene una visión general de la selección geográfica
- Account View: contiene información para una cuenta determinada
- Segment View: contiene información para un segmento dado

Las secciones posteriores están dedicadas a estos tres componentes pero antes de explicarlos en detalle, vamos a describir los menús y selectores comunes.



En la parte superior izquierda se dispone de un menú geográfico (ver figura 52) adaptado a los permisos del usuario, es decir, solamente muestra la jerarquía geográfica de los países a los que tiene acceso.



Figura 52: Menú geográfico del Dashboard

En la esquina superior derecha se encuentran los selectores de GDS y Timeline (figura 53):

GDS Group:	Amra	Sab	Tra	All
GDS:	Amra	Sab	Aba	Gal
Timeline:	YTD	3M	1M	R12

Figura 53: Selectores del Dashboard

Se puede seleccionar cada GDS individualmente o el grupo. Cada uno tiene su color corporativo. La selección predeterminada es "All". Si se selecciona un competidor de Amadeus, se obtiene una vista de ese competidor contra Amadeus.

Respecto a Timeline, la mayoría de visualizaciones muestran comparativas entre dos períodos de tiempo, por lo que este selector permite al usuario elegir la cantidad de meses (12, 3 o 1) a comparar con respecto al período actual. El valor por defecto YTD. En caso de duda, el usuario puede colocar el cursor sobre un botón y aparecerá un mensaje especificando los períodos comparados. Veamos el significado de cada uno:

- YTD: compara las cifras del año en curso con respecto al mismo período del año anterior
- 3M: compara los últimos 3 meses disponibles con los mismos 3 meses del año anterior
- 1M: compara el último mes disponible con el mismo mes del año anterior
- R12: compara los últimos 12 meses disponibles con los 12 meses anteriores
- R3: compara los últimos 3 meses disponibles con los 3 meses anteriores
- R1: compara el último mes disponible con el anterior

Por otra parte, justo debajo de estos selectores y a la altura del menú geográfico (ver figura 54), se muestra la fecha de los últimos datos disponibles y los filtros aplicados.



También hay un botón llamado “Reports” para descargar los informes disponibles, de acuerdo con la opción elegida en el menú geográfico y los selectores GDS y Timeline.



Figura 54: Botón “Reports” y mensajes informativos del Dashboard

MAIN VIEW

La figura 55 muestra el contenido de esta vista. Por simplicidad, hemos numerado los recuadros existentes, de 0 a 6. En la parte superior, tenemos el recuadro 0, mientras que el resto de la vista son 6 recuadros distribuidos en una cuadrícula de 3 columnas por 2 filas. Todos tienen el mismo tamaño pero contienen diferentes visualizaciones.

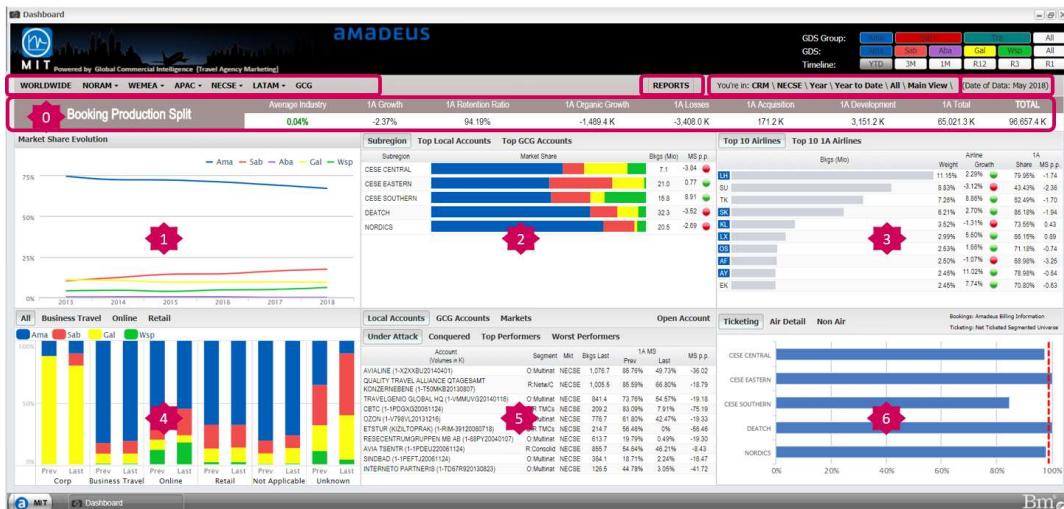


Figura 55: Dashboard - Main View

Box 0: Booking Production Split

Es una tabla resumen que contiene los siguientes KPI:

- Average Industry: es el crecimiento del mercado seleccionado, fruto de la comparación entre las reservas del periodo actual y el anterior. Estos periodos vienen determinados por el valor seleccionado en Timeline.
- GDS Growth: refleja el crecimiento del GDS seleccionado, por defecto Amadeus.
- Retention Ratio: muestra los volúmenes mantenidos en porcentaje, para el período y GDS seleccionados.
- Organic growth: se calcula aplicando el crecimiento de la cuenta a los volúmenes de GDS en el período anterior.



- Losses: representa el número total de reservas perdidas. Contiene las reservas de aquellas cuentas que tenían producción en el GDS seleccionado en el periodo anterior pero en el actual han dejado de producir reservas con este GDS.
- Acquisition: incluye adquisiciones de cuentas que aparecen por primera vez en el mercado y adquisición de cuentas de la competencia.
- Development: contiene las reservas procedentes de cuentas en las que el GDS seleccionado ha aumentado su presencia en términos de cuota de mercado.
- GDS Total: es el número de reservas en el período actual del GDS seleccionado.
- Total: es el número de reservas de todos los GDS en el período actual.

Box 1: Market Share Evolution

Muestra un gráfico de líneas con la evolución de la cuota de mercado de todos los GDS. No obstante, se pueden filtrar en la parte superior del cuadro. Al pasar el ratón por encima de las líneas del gráfico, aparece un mensaje con la cuota de mercado y el número de reservas netas, tal y como se muestra en la figura 56. Además, el usuario puede centrarse en un período de tiempo concreto haciendo clic y arrastrando un área dentro del gráfico. Para deshacer la ampliación basta con pulsar el botón “Reset zoom”.



Figura 56: Zoom en Market Share Evolution



Box 2: Regions and Top Accounts

Este cuadro consta de 3 pestañas que contienen el mismo formato, una tabla con un gráfico de barras apiladas incrustado, pero con datos diferentes.

En las figuras 57, 58 y 59 se puede ver que el gráfico está basado en la cuota de mercado de cada GDS para el período más reciente, según el valor seleccionado en Timeline. Al pasar el puntero del ratón por encima del gráfico se muestran el porcentaje de cuota de mercado y el número de reservas.

La columna de Bkgs contiene el total de reservas de todos los GDS, mientras que la última columna, MS p.p., es la evolución de la cuota de mercado, respecto al período anterior, del GDS seleccionado en el selector de la aplicación. Por defecto, se muestran los valores de Amadeus. Además de mostrar el valor numérico, tenemos un semáforo con colores condicionados al crecimiento de la cuota de mercado.

La primera pestaña toma los nombres Region, Subregion o Market, dependiendo del valor seleccionado en la jerarquía del menú geográfico. Asimismo, esta selección impacta en los elementos geográficos que mostrados. Por ejemplo, si se selecciona una región, aparece el listado de subregiones que pertenecen a la misma.

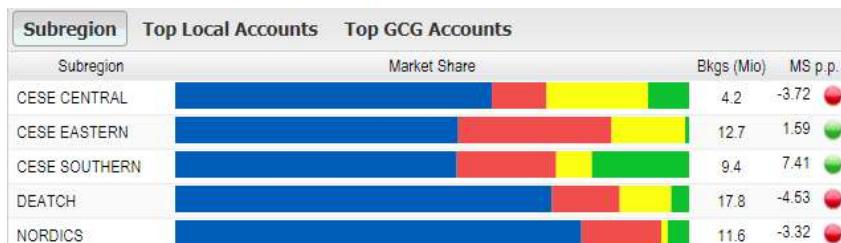


Figura 57: Main View - Box 2.1

Subregion	Top Local Accounts	Top GCG Accounts	Ranking	
Account	Mkt	Market Share	Bkgs (K)	MS p.p.
TRIPSTA (1-YSJSKU20140710)	NECSE		1,926.9	20.19
E-TRAVEL (1-VXJBK020140203)	NECSE		1,887.4	28.29
VIP SERVICE (1-1PEYI020061124)	NECSE		1,649.4	22.40
DER TOURISTIK GROUP (1-1QDLP20041214)	NECSE		1,490.0	-0.07
TICKET LEISURE TRAVEL (1-WZ4WG120140326)	NECSE		932.7	0
BERG-HANSEN REISEBUREAU AS (1-5Y1-18220031006)	NECSE		659.0	-0.61

Figura 58: Main View - Box 2.2

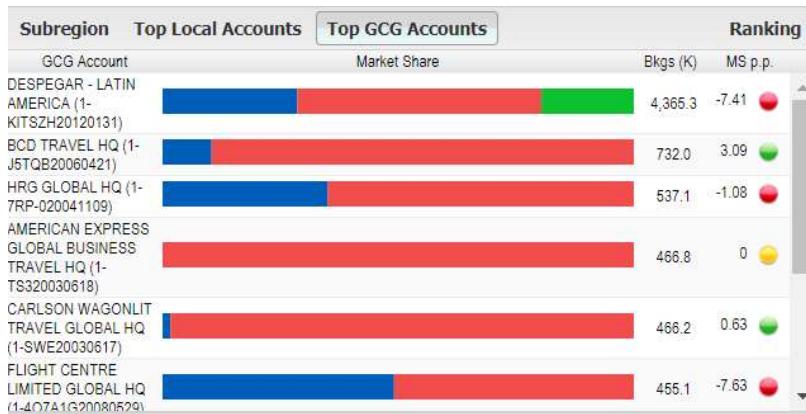


Figura 59: Main View - Box 2.3

Las otras dos pestañas muestran un ranking basado en bookings de las 10 mejores agencias de viajes, locales y globales (GCG), respectivamente. Además, si se desea consultar un número mayor de agencias, el usuario puede descargar un informe llamado “Ranking” pulsando en el botón correspondiente. El contenido de este informe está definido en el apartado 3.4.3.

Si se selecciona un GDS, la lista cambiará de acuerdo con las 10 mejores para el GDS dado.

Al hacer clic en el nombre de una agencia, se accede a la vista por cuenta, Account View, que se explica más adelante.

Box 3: Top 10 Airlines

Este cuadro, referenciado en la figura 60, dispone de 2 pestañas que muestran un ranking de aerolíneas basado en los volúmenes totales y del GDS dado, respectivamente, en el área geográfica y el período de tiempo seleccionado.

La columna Weight se refiere al porcentaje de la aerolínea sobre el mercado total, mientras que en Growth se muestra el crecimiento de la aerolínea respecto al período anterior.

Las dos últimas columnas muestran la cuota de mercado del GDS seleccionado en la aerolínea, así como la evolución de este KPI, es decir, MS p.p.



		Top 10 Airlines			
		Top 10 1A Airlines			
		Bkgs (K)	Airline	1A	
LA			Weight	Growth	Share MS p.p.
AV		17.42%	35.44%	40.56%	-1.08
AM		15.76%	-2.13%	45.18%	-0.67
CM		9.87%	11.39%	21.18%	2.90
AR		6.79%	-2.31%	49.70%	2.70
AA		6.69%	-5.48%	45.18%	-4.34
JJ		5.06%	2.85%	32.60%	-1.38
UA		3.11%	-53.37%	39.87%	2.45
IB		3.00%	8.71%	36.25%	-0.29
DL		2.65%	4.07%	46.46%	-2.55
		2.33%	-17.35%	38.75%	0.46

Figura 60: Main View - Box 3

Cuando se pasa el ratón sobre las barras, aparecen el número de reservas aéreas de la aerolínea, la cantidad de esas reservas que se han realizado con el GDS seleccionado, por defecto Amadeus, y el nombre del producto “Altea Res”, si es un cliente de Amadeus. Este último dato, también se indica mediante el color azul en el código de la aerolínea.

Además, si se sitúa el cursor sobre el código de la aerolínea, se mostrará el nombre completo del transportista.

Box 4: Segment Distribution

Contiene información basada en la Segmentación de Marketing para agencias de viajes.

En la figura 61 se puede ver el aspecto de este cuadro. La información se muestra en cuatro pestañas, una principal con todos los segmentos y el resto con los sub-segmentos de los segmentos principales: Business Travel, Online y Retail.

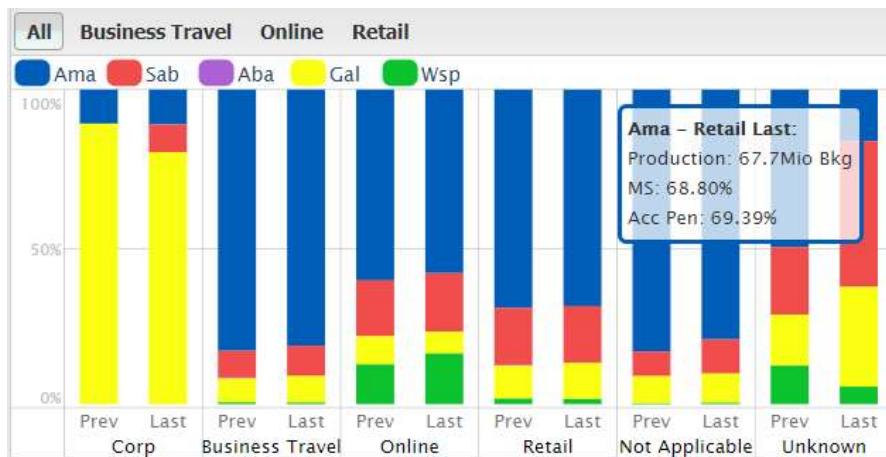


Figura 61: Main View - Box 4



Consiste en un gráfico de barras agrupadas que refleja los valores del período actual y previo para cada segmento. Cada barra se distribuye según la cuota de mercado de los GDS, permitiendo comparar su evolución en cada segmento.

Para visualizar la producción, la cuota de mercado y la penetración del GDS en el segmento (porcentaje de agencias de viajes en que está presente) basta con pasar el puntero del ratón sobre el segmento a analizar.

Por último, al hacer clic en un segmento o sub-segmento determinado, se accede a la vista por segmento, Segment View, que se explica más adelante.

Box 5: Account and Market Information

Este es el único cuadro que no contiene gráficos, ver figura 62. Consta de 3 pestañas para mercados, cuentas locales y cuentas globales que, a su vez, disponen de 4 tablas de datos en las que se aplican los siguientes criterios:

- Under Attack: agencias de viajes en las que el GDS seleccionado está perdiendo cuota de mercado.
- Conquered: agencias de viajes en las que el GDS seleccionado está ganando cuota de mercado.
- Top Performers: agencias de viajes, ya sean clientes del GDS seleccionado o no, cuyo porcentaje de crecimiento es superior a la media.
- Worst Performers: agencias de viajes, independientemente del GDS con el que trabajen, donde el crecimiento de la cuenta es mínimo.

Under Attack	Conquered	Top Performers	Worst Performers				
Account (Volumes in K)		Segment	Mkt	Bkgs Last	1A MS Prev.	Last	MS p.p.
E-TRAVEL (1-VXJBK020140203)	O:Multinat	NECSE	3,162.8	19.62%	55.42%	35.80	
TRIPSTA (1-YJSJSKU20140710)	O:Multinat	NECSE	3,243.4	11.47%	29.86%	18.39	
VIP SERVICE (1-1PEYI020081124)	R:Consolid	NECSE	2,727.8	23.60%	43.11%	19.51	
ALTAIR TRAVEL (1-RSE-16020060720)	O:Multinat	NECSE	961.9	17.52%	36.69%	19.17	
KIWI.COM S.R.O. (1-ZAUNKV20140811)	O:Multinat	NECSE	372.9	59.07%	76.45%	17.38	
ESKY (1-TD67MX20130823)	O:Multinat	NECSE	281.2	24.11%	39.48%	15.38	
RBT HOLDING (1-7P4VAD20090518)	B:R TMCs	NECSE	134.3	0.07%	32.27%	32.20	
E-TICKETS SERVICE (1-W4FHY020140214)	R:Consolid	NECSE	758.3	82.30%	85.14%	2.84	
ONETWOTRIP (1-TD67RS20130823)	O:Multinat	NECSE	605.8	14.99%	18.33%	3.34	
OTI HOLDING (1-10ZAGN20141015)	R:KAs	NECSE	81.9	21.31%	43.42%	22.10	

Figura 62: Main View - Box 5

En estos listados se muestra el segmento al que pertenece la cuenta, los volúmenes producidos en el último período y la cuota de mercado en el período anterior y actual, así como la evolución de este valor.

Por otra parte, en este cuadro también se dispone de un formulario con un botón, “Open Account”, donde el usuario puede escribir parte del nombre o parte del Master Account



Number de la agencia que está buscando. Como resultado, se visualizan en una tabla las agencias que cumplen los criterios de búsqueda. Desde ahí, el usuario puede hacer clic en cualquiera de ellas y acceder a la vista Account View. La figura 63 contiene un ejemplo de esta búsqueda.

The screenshot shows two identical search interface windows stacked vertically. Both windows have tabs for 'Local Accounts', 'GCG Accounts', and 'Markets', with 'Local Accounts' selected. There are two radio buttons: 'CRM' (selected) and 'GCG'. A search input field contains 'satguru' and a 'Search' button. An 'Open Account' button is located in the top right corner of each window. The results table has columns: Account, Mkt, Segment, and GCG. The data is as follows:

Account	Mkt	Segment	GCG
SATGURU ENTERPRISES (2-12FDK-1082020180228)	NECSE	R:SME Mass	NOT GCG
SATGURU GROUP (2-3JACKM520171023)	NECSE	R:SME Mass	NOT GCG
SATGURU HOLIDAYS INDIA PVT LTD (2-12FDK-24820180228)	NECSE	Unknown	NOT GCG
SATGURU INTERNATIONAL TRAVELS (2-12FDK-431120180228)	NECSE	R:SME Mass	NOT GCG
SATGURU TRAVEL & TOUR KHARTOUM BRANCH (1-7FOIA20111123)	NECSE	R:SME U	NOT GCG
SATGURU TRAVEL & TOUR SERVICES (2-3LJLB20150330)	NECSE	R:SME Mass	NOT GCG
SATGURU TRAVEL & TOURISM GLOBAL			

Figura 63: Main View – Buscador de cuentas

Box 6: Amadeus Bookings Analysis

Este cuadro tiene 2 pestañas: Ticketing y Non-Air. Por tanto, solamente contiene información de Amadeus procedente de ABI y Net Ticketed Segment, cuyos valores cambian según la selección realizada en el menú geográfico y el selector de Timeline.

La pestaña de Ticketing muestra la proporción entre los tickets emitidos y las reservas facturadas procedentes de ABI. Si el número de mercados a mostrar es muy grande, este gráfico consta de varias páginas, como en la figura 64. Además, muestra mediante una línea discontinua de color rojo la media de esta medida en el área geográfica seleccionado. Por el contrario, si los datos a mostrar conciernen a un solo mercado, el gráfico se transforma en un gráfico circular, como el de la figura 65.

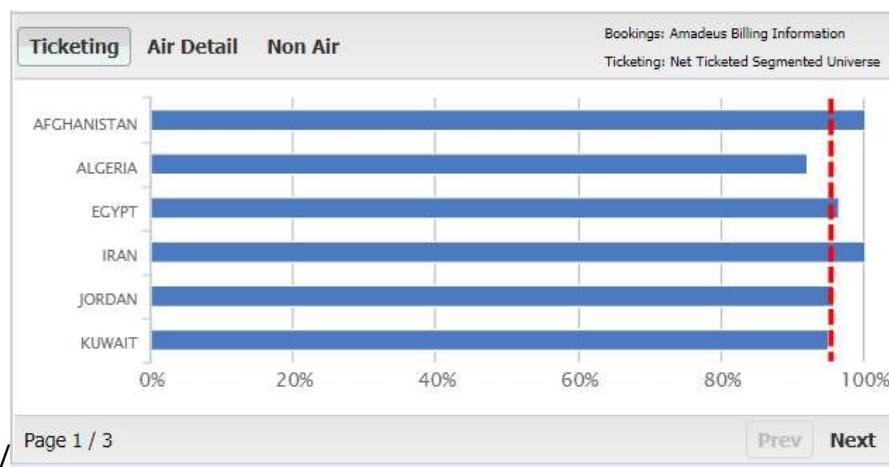


Figura 64: Main View - Box 6.1

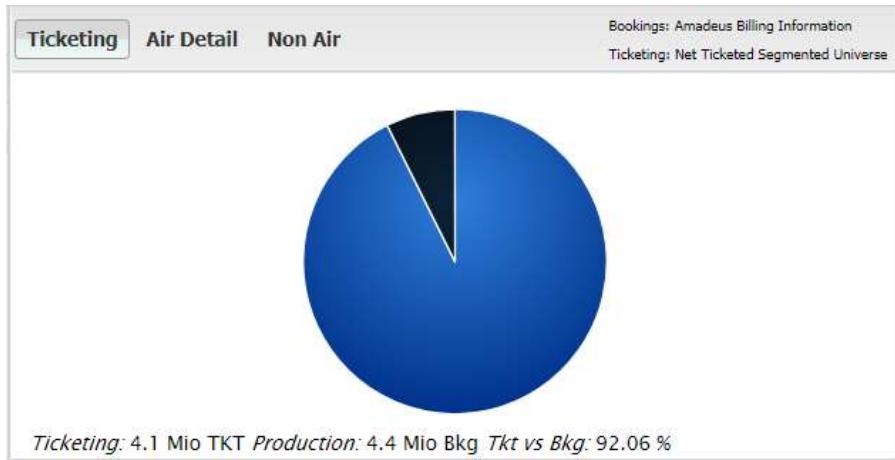


Figura 65: Main View - Box 6.2

La pestaña de Non-Air está destinada a mostrar una comparación entre las reservas aéreas y el resto: Car, Hotel y Rail.

Al igual que sucede con Ticketing, disponemos de varias páginas en caso de haber muchos mercados y existen 2 tipos de visualizaciones, figuras 66 y 67, dependiendo del número de mercados a mostrar.

Además, en el caso del gráfico de barras agrupadas, figura 66, hay que tener en cuenta que, dependiendo del área seleccionada y debido a los enormes volúmenes de aire a nivel regional, subregional y de mercado, el eje X comienza en un porcentaje muy alto para que el resto de negocios se puedan visualizar claramente.

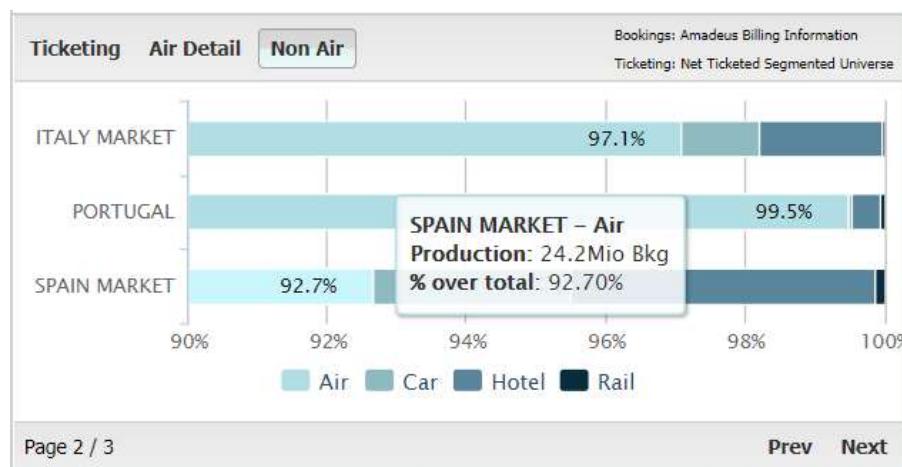


Figura 66: Main View - Box 6.3

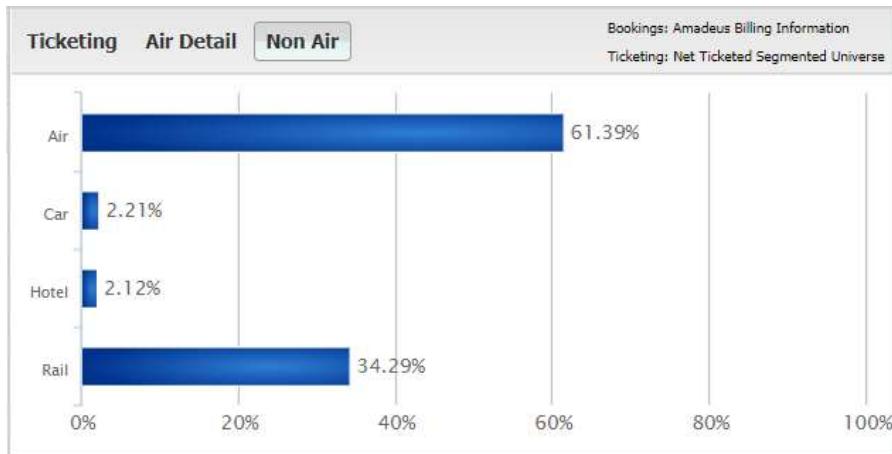


Figura 67: Main View - Box 6.4

ACCOUNT VIEW

Esta vista permite profundizar en el análisis de una agencia de viajes.

En la figura 68 podemos ver que la pantalla está dividida bajo el mismo formato que Main View y se mantienen algunas de las visualizaciones existentes pero con los datos de la agencia seleccionada:

- En la parte superior se muestra el cuadro de Bookings Production Split.
- En los cuadros 1 y 3 se mantienen las visualizaciones de Market Share Evolution y Top 10 Airlines, respectivamente.
- Los cuadros 4 y 5 contienen los datos de Ticketing y Non-Air, respectivamente. Se corresponden con las pestañas presentes en el cuadro 6 de Main View.



La barra informativa situada en la esquina superior derecha, debajo del selector de GDS y Timeline, muestra los identificadores de la agencia de viajes (Master Account Number y Name).

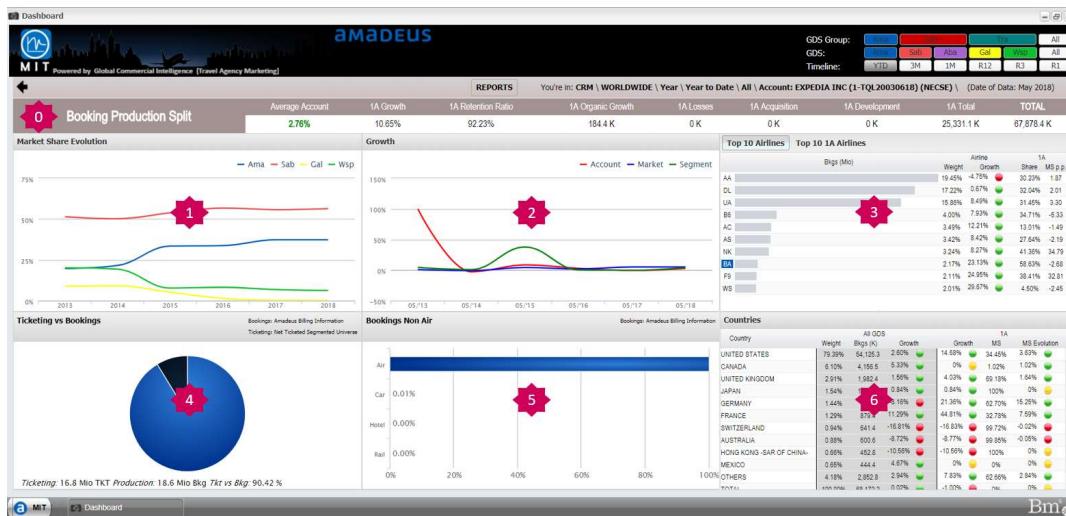


Figura 68: Account View

Las principales novedades que encontramos en esta vista son los cuadros 2 y 6:

- El cuadro 2, Growth, muestra un gráfico de líneas con la evolución del crecimiento de la agencia en el contexto del área geográfica filtrada y el resto de agencias que pertenecen al mismo segmento. Es decir, muestra la evolución del crecimiento de la agencia frente al mercado y el segmento.
- La figura 69 contiene un ejemplo de esta visualización. Al pasar el puntero del ratón por encima de una línea aparece un mensaje con la producción y el crecimiento del elemento asociado.

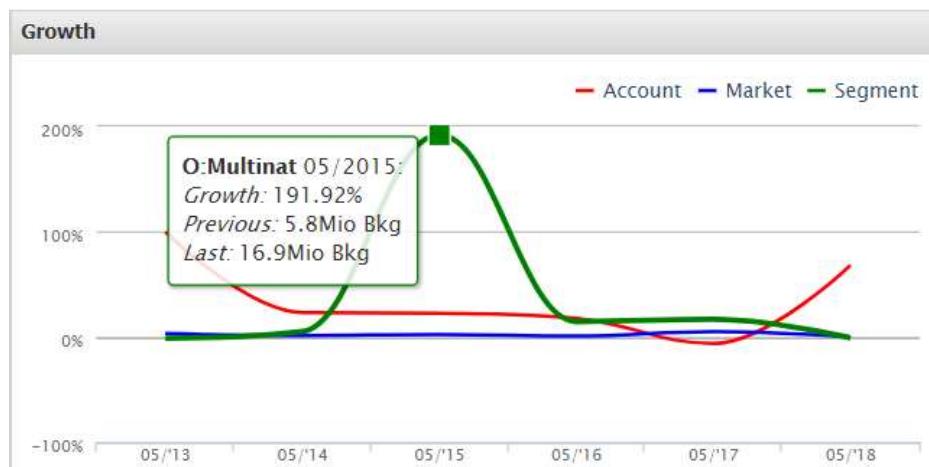


Figura 69: Account View – Box 2



- El cuadro 6, Countries, es una tabla con el listado de países en que la agencia de viajes está presente. Por defecto, estos países aparecen ordenados en función del total de bookings de la agencia en cada país. Si se selecciona un GDS, la clasificación se basará en el número de bookings de la cuenta con ese GDS.
- Esta tabla muestra 6 columnas, 3 con valores para todos los GDS y 3 para el GDS seleccionado que, por defecto, es Amadeus. Las primeras 3 columnas indican el peso del país en la cuenta, el número de reservas y la evolución del crecimiento. Las restantes son el crecimiento del GDS en la cuenta para ese mercado, la cuota de mercado y la evolución de esta cuota de mercado respecto al período anterior. La figura 70 muestra un ejemplo de esta visualización.

Countries							
Country	All GDS			IA			MS Evolution
	Weight	Books (#)	Growth	Growth	MS	MS	
UNITED STATES	79.34%	82,312,6	-2.07% ●	2.91% ●	32.89%	1.59% ●	
CANADA	5.87%	6,087,5	10.85% ●	0% ●	0%	0% ●	
UNITED KINGDOM	2.81%	2,919,6	-8.50% ●	5.09% ●	67.87%	8.78% ●	
JAPAN	1.64%	1,697,7	18.13% ●	18.13% ●	100%	0% ●	
GERMANY	1.52%	1,576,2	-4.66% ●	-15.26% ●	48.95%	-6.14% ●	
FRANCE	1.13%	1,168,0	-14.44% ●	3.31% ●	25.63%	4.41% ●	
SWITZERLAND	1.05%	1,087,5	-2.68% ●	-2.64% ●	99.78%	0.05% ●	
AUSTRALIA	1.01%	1,051,8	5.39% ●	5.37% ●	99.90%	-0.02% ●	
HONG KONG-SAR OF CHINA-	0.75%	777,3	2.31% ●	2.31% ●	100%	0% ●	
MEXICO	0.69%	718,9	41.21% ●	0% ●	0%	0% ●	
OTHERS	4.20%	4,355,4	4.58% ●	-0.82% ●	59.98%	-3.27% ●	
TOTAL	100.00%	103,752,6	-0.01% ●	-1.00% ●	0%	0% ●	

Figura 70: Account View – Box 6

SEGMENT VIEW

Esta vista permite profundizar en el análisis de un segmento de marketing.

En la figura 71 podemos ver que la pantalla está dividida bajo el mismo formato que Main View, excepto que no se dispone del recuadro de Booking Production Split. También se mantienen algunas de las visualizaciones existentes:

- Los cuadros 1 y 3 contienen exactamente las mismas visualizaciones, Market Share Evolution y Top 10 Airlines, respectivamente. Simplemente aplican los filtros del segmento o sub-segmento seleccionado.
- Mientras que los listados de agencias de los cuadros 2 y 5 unifican las cuentas locales y globales en vez de tener una pestaña para cada tipo. Además, haciendo clic en cualquier agencia, se accede a la vista por cuenta.

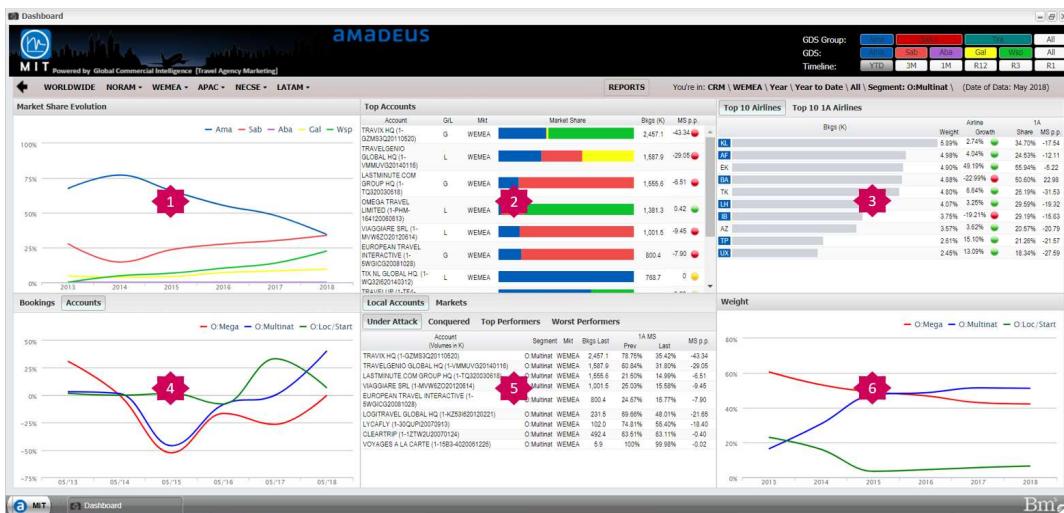


Figura 71: Segment View

Las principales diferencias residen en los cuadros 4 y 6:

- El cuadro 4, Segment Growth Evolution, tiene 2 pestañas, Bookings y Accounts, que contienen gráficos de líneas con la evolución del crecimiento de todos los segmentos, en caso de filtrar por un segmento, o de todos los sub-segmentos de un segmento, en caso de seleccionar un sub-segmento. Por ejemplo, si se selecciona Online, se mostraran líneas para Online, Business Travel, Retail, etc., mientras que si se selecciona un sub-segmento de Online, el gráfico contendrá una línea por cada sub-segmento de Online.

La figura 72 muestra un ejemplo de esta visualización. Si se sitúa el ratón sobre una línea, aparece un mensaje con el crecimiento del segmento y los valores en el periodo actual y anterior, en términos de reservas o número de cuentas, dependiendo de la pestaña en la que estemos.

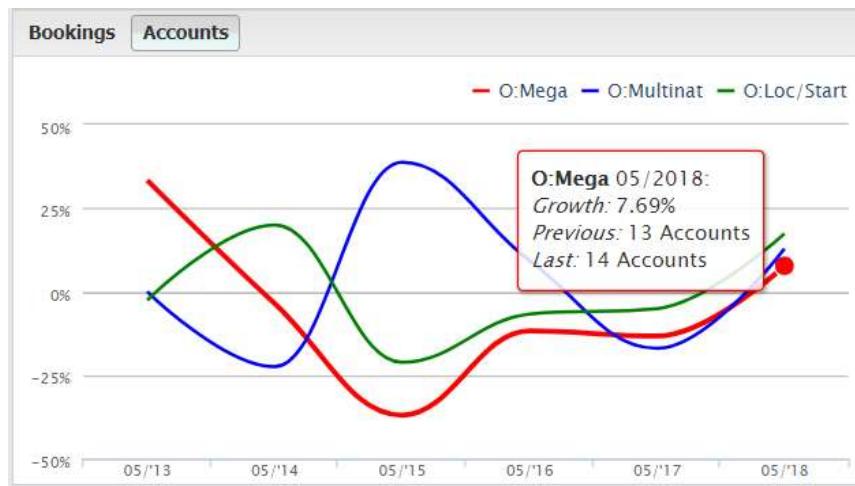


Figura 72: Segment View – Box 4

- El cuadro 6, Segment Weight Evolution, es un gráfico de líneas basado en el peso de cada segmento o sub-segmento. La figura 73 muestra un ejemplo de esta visualización.

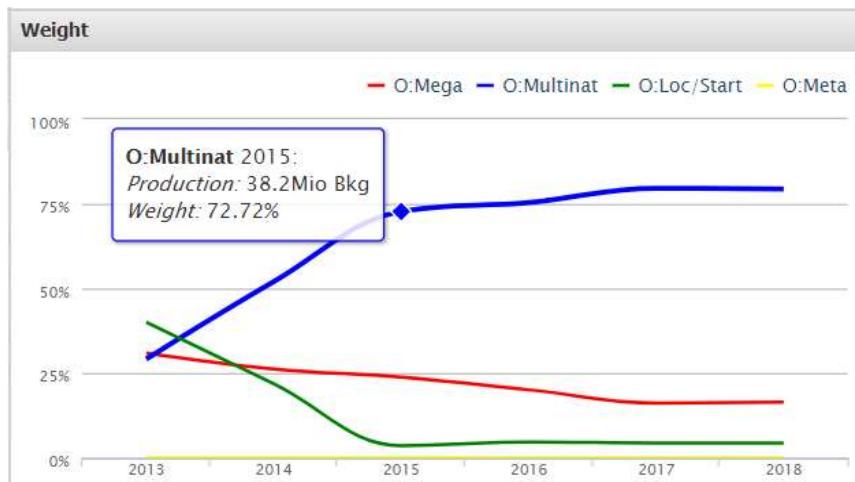


Figura 73: Segment View – Box 6

3.4.2. DATA EXPLORER

Data Explorer es el nombre del módulo de MIT que permite consultar, visualizar y analizar la información de nuestro DW, a través del motor de cubos OLAP Mondrian, descrito en el apartado 2.3.3., sin necesidad de tener conocimientos de MDX ni de bases de datos.



En este módulo, los usuarios de negocio pueden interactuar con las dimensiones y medidas de los cubos definidos en los esquemas XML. Pueden arrastrar y soltar estos elementos como columnas, filas, filtros o dentro del propio cubo de acuerdo con la visualización deseada.

Principalmente, se ha definido un cubo con información de reservas aéreas con todas las dimensiones disponibles en el modelo de datos. Mondrian genera varias consultas SQL que devuelven los datos contenidos en las tablas de hechos de reservas. En función de las dimensiones seleccionadas, estas consultas utilizan un agregado u otro, aunque esto es transparente para los usuarios.

De esta manera, un usuario puede generar sus propios informes, ejecutar las consultas que estime oportunas y realizar sus análisis de manera independiente.

Las principales funcionalidades del Data Explorer son:

- Visualización de los datos en forma de tabla dinámica y/o gráficos
- Exportación de datos en formato Excel, CVS y PDF
- Restricción del acceso a los datos dependiendo del perfil de los usuarios

Al inicio de este capítulo, se comentó que MIT utiliza la versión 5.4 de Pentaho CE. A pesar del potencial de la plataforma de Pentaho, MIT solamente utiliza Mondrian. Es más, simplemente embebe alguno de los front-end disponibles para este motor OLAP en un frame del módulo en cuestión, por lo que no está integrado como tal en nuestra aplicación, sino que simplemente se ejecutan en el mismo contenedor web (Apache Tomcat).

Por tanto, el funcionamiento del resto de módulos de MIT no se ve afectado por el funcionamiento de Pentaho.

Debido al alto coste de las licencias de Pentaho Data Analyzer, se han probado diversos plugins de la plataforma Pentaho para este módulo, manteniendo Mondrian como motor OLAP, ya que la mayoría de plugins Open Source disponibles son compatibles con las versiones 5.x y superiores.

Finalmente, se ha optado por Saiku Analytics, un proyecto Open Source fundado en 2008 por Tom Barber y Paul Stoellberger, que ofrece, al igual que Pentaho, versiones CE y EE. La principal diferencia entre ellas es que la suite EE incluye, además del visor de cubos OLAP, un módulo para crear dashboards. Amadeus ha optado por utilizar la versión CE.

Saiku ofrece una solución analítica fácil de usar que permite a los usuarios crear fácilmente diferentes análisis de los cubos OLAP existentes en Mondrian, permitiendo visualizar y analizar la información en forma de tabla dinámica (figura 74) o mediante gráficos simples (figura 75). Además, dispone de funciones de exportación a Excel 2007-2010 (XLSX), CSV y PDF.



En general, tiene buenas prestaciones aunque inferiores a las de Pentaho Data Analyzer.

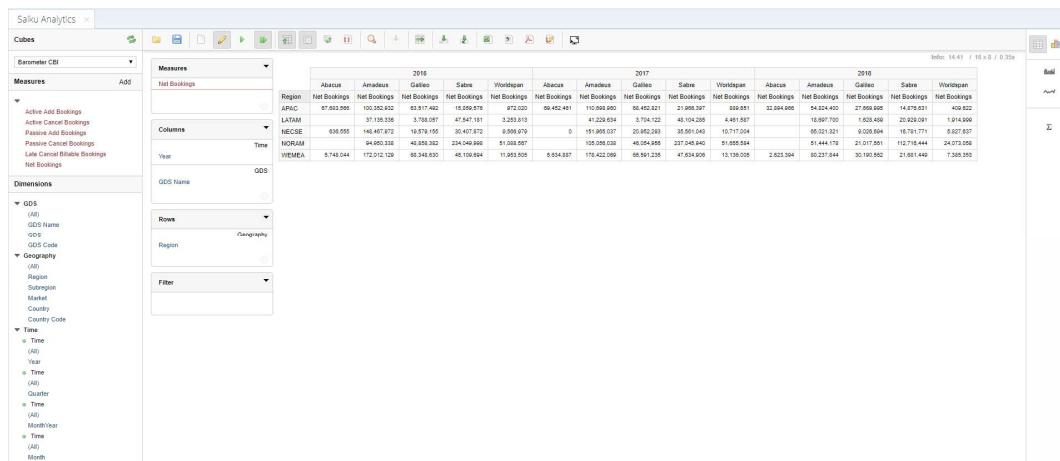


Figura 74: Visualización de un cubo en forma de tabla dinámica

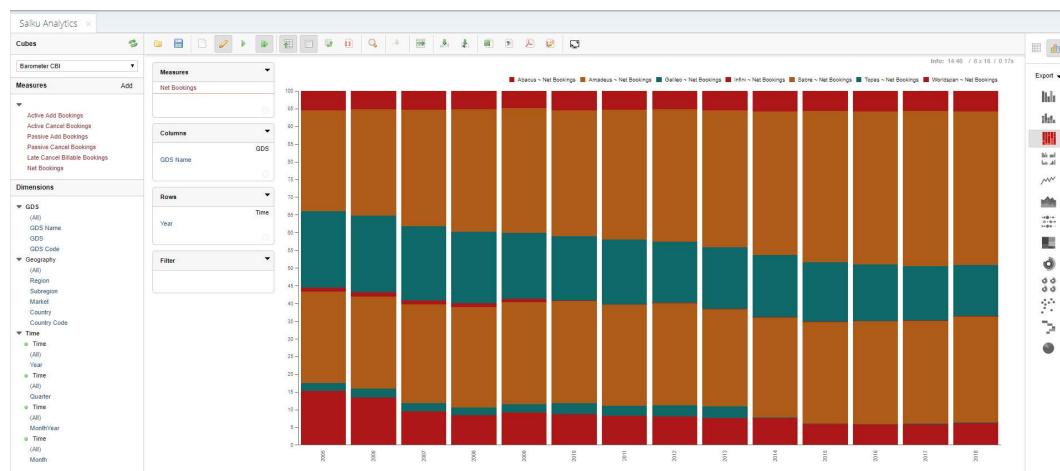


Figura 75: Visualización de un cubo en forma de gráfico de barras apiladas

Embeber este nuevo visor en MIT ha requerido modificar el código fuente del front-end, para que se muestre este visor al acceder al módulo *Data Explorer* aunque esta tarea no la ha llevado a cabo el alumno. El alumno ha sido partícipe de la construcción de los cubos para que el usuario final pueda ejecutar sus propias consultas en función de sus necesidades específicas, por ejemplo, vista geográfica, cuentas principales, etc.

Por último, hemos de reconocer que el rendimiento de estos cubos no ha sido óptimo por lo que no han resultado muy útiles y tampoco se han desarrollado en exceso.



3.4.3. REPORTS

El módulo de Reports permite generar y descargar, en formato Excel, una serie de informes predefinidos. Fueron desarrollados con BIRT Report Designer, aunque el alumno no fue partícipe de esta tarea, simplemente los mantiene y hace cambios puntuales sobre los mismos.

Se dispone de un interfaz para generar los informes, figura 76, y otra ventana con una vista en árbol, para descargarlos, figura 77.

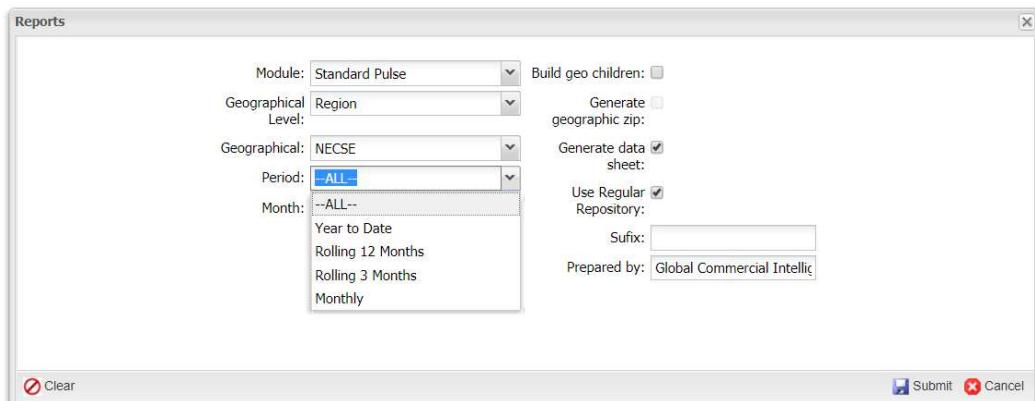


Figura 76: Generador de informes

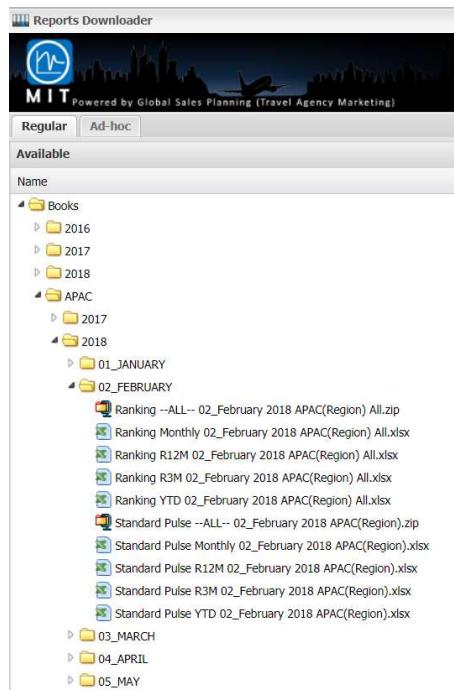


Figura 77: Descarga de informes



A lo largo del proyecto se desarrollaron varios informes pero finalmente el cliente se centró en tener en esta nueva plataforma los más utilizados: Standard Pulse, Marketing Pulse y Ranking.

STANDARD PULSE

Es un informe que compara dos períodos de tiempo, previous y last. La mayoría de datos proceden de CBI y se encuentran agrupados a nivel del cliente CRM. El informe consiste en una serie de hojas con tablas resumen, ver figuras 78 y 79, que proporcionan KPI claros y métricas para la gestión del mercado y las ACO.



WORLDWIDE Standard Pulse: WorldWide view

Prev: January 2016 to October 2016

Last: January 2017 to October 2017

Prepared By: Global Commercial Intelligence

Segment	January 2016 to October 2016						January 2017 to October 2017					
	Amadeus	Abacus	Galileo	Sabre	Worldspan	TOTAL	Amadeus	Abacus	Galileo	Sabre	Worldspan	TOTAL
Grand TOTAL	469,765	62,923	175,097	319,306	65,468	1,092,559	499,348	64,108	176,013	334,665	69,786	1,143,920
Corp	207	7	118	86	56	474	212	5	96	102	28	444
Managed Business-Corporation	207	7	118	86	56	474	212	5	96	102	28	444
Business Travel	89,120	11,599	70,257	114,035	8,956	293,967	96,278	11,293	66,528	121,241	8,958	304,299
Managed Business-Global TMCs	40,837	4,873	24,156	72,663	859	143,388	42,201	4,943	23,703	75,009	856	145,711
Managed Business-Regional TMCs	25,028	3,498	23,058	30,931	7,352	89,866	28,169	3,358	23,531	32,824	7,370	95,253
Managed Business-Local TMCs	9,863	2,932	4,269	6,177	627	23,886	9,716	2,612	4,377	6,324	629	23,657
Managed Business-Specialty	10,582	34	17,998	3,393	113	32,120	13,573	35	13,556	6,193	101	33,457
Managed Business-Network	2,810	262	777	871	4	4,725	2,618	346	1,362	891	3	5,220
Online	139,387	3,975	20,311	107,429	29,390	300,492	151,360	4,022	21,557	110,371	30,072	317,382
Online-Mega OTAs	73,606	36	1,989	79,214	19,654	174,498	76,386	169	229	78,424	17,063	172,271
Online-Multinational OTAs	59,011	560	8,825	24,058	8,159	100,613	66,793	497	9,636	27,754	11,476	116,147
Online-Local OTA/Start-ups	6,736	3,379	9,497	4,157	1,577	25,347	8,154	3,366	11,633	4,193	1,533	28,938
Online-Metasearch	34	-	-	-	-	34	26	-	-	-	-	26
Online-Gatekeepers	0	-	-	-	-	0	-	-	-	-	-	-
Retail	239,315	46,831	79,296	90,016	23,839	479,297	249,621	45,724	80,781	92,480	26,544	495,151
Retail-Tour Operators	31,467	3,683	5,578	10,490	4,125	55,343	33,242	3,938	5,529	10,888	4,163	57,525
Retail-Consolidators	29,774	3,957	9,324	11,316	2,710	56,981	34,908	3,778	10,147	13,323	2,555	64,711
Retail-Network / Consortia	21,577	560	2,232	4,602	164	29,195	21,732	553	2,636	4,591	186	29,698
Retail-Key Accounts	50,259	11,000	16,689	18,527	8,113	104,588	55,216	11,637	18,295	19,731	11,840	116,719
Retail-SME Upper	19,236	7,775	7,100	10,500	2,255	46,866	19,885	7,094	7,387	10,803	2,047	47,217
Retail-SME Mass	63,618	14,120	26,637	25,990	4,833	135,399	62,330	13,430	25,668	24,751	4,309	130,499
Managed Business-SME	23,394	5,835	11,677	8,592	1,639	51,126	22,308	5,294	11,355	8,392	1,445	48,793
Not Applicable	604	121	455	647	329	2,156	637	151	600	440	568	2,396
Managed Business-Partners	21	-	0	0	0	21	18	-	2	(0)	-	19
Online-Partners	1	-	-	0	(0)	1	0	-	0	-	-	0
Retail-Partners	0	-	-	1	-	1	0	-	-	0	-	1
Retail-TO Partners	195	82	281	161	12	731	323	44	415	145	10	937
Retail-Providers	17	2	-	0	0	19	23	1	-	0	0	24
Transversal-Partners	0	-	1	0	97	99	0	-	1	0	57	58
-	-	-	-	-	-	-	-	-	-	-	-	-

Figura 78: Standard Pulse

Dependiendo del periodo de tiempo comparado, tenemos diferentes versiones del documento:

- Rolling 12 (R12): compara los últimos 12 meses contra los 12 anteriores.
- Rolling 3 (R3): compara los últimos 3 meses contra los 3 anteriores.
- Month: compara las cifras de un mes con respecto al mismo mes del año anterior.
- YoY (Year on Year): compara las cifras de un año a otro frente al mismo período del año anterior.



WORLDWIDE Standard Pulse: Amadeus top wins

Prev: January 2016 to October 2016

Last: January 2017 to October 2017

Prepared By: Global Commercial Intelligence

Rk#	Country	Account	Segment	Subsegment	January 2016 to October 2016						January 2017 to October 2017							
					Amadeus	Abacus	Gattee	Sale	Worldspan	TOTAL	Amadeus	Abacus	Gattee	Sale	Worldspan	TOTAL		
1	US, CA, GB, JP, DE, FR, AU, CH, MK, IT and 3 others	EXPEDIA INC (I-TGL200368)	Online	Online-Mega OTAs	42,283,243	-	1,981,043	70,724,720	10,381,029	125,870,135	44,186,421	-	221,768	69,454,325	9,076,345	122,951,457		
2	FR, DE, NL, BE, ES, PT, US, MK, FI, SE and 3 others	CGOGEON (I-NERKA201222)	Online	Online-Mega OTAs	19,476,909	-	-	20,268	-	19,507,077	19,300,252	-	-	2	-	-	19,300,254	
3	SE, DE, ND, NO, FI, NL, PR, CH, GB, ES and 3 others	EUROPEAN TRAVEL INTERACTIVE (I-SVWICG200102)	Online	Online-Multinational OTAs	6,940,576	-	-	2,734,420	-	11,676,396	10,417,087	-	-	2,678,508	-	-	13,093,593	
4	US, CA, GB, MX	FAPEPORTAL (I-CQDEGX2005118)	Online	Online-Mega OTAs	9,259,256	-	6,663	7,778,491	-	16,395,410	9,284,312	-	6,648	8,007,187	-	-	17,298,147	
5	US, DE, GB, NL, MX, SE, BE, FR, CA, AU and 3 others	BCD TRAVEL HQ (I-HSTC2006042)	Business Travel	Managed Business- Global TMCs	8,645,187	174,747	4,068,405	12,943,424	74,940	26,564,723	8,747,943	186,969	4,225,257	12,946,434	74,954,74	26,698,377		
6	US, FR, NL, DE, CA, MK, AT, SE, BE, AMERICAN EXPRESS GLOBAL BUSINESS TRAVEL HQ (I-TS200368)	AMERICAN EXPRESS GLOBAL BUSINESS TRAVEL HQ (I-TS200368)	Business Travel	Managed Business- Global TMCs	8,046,455	521,437	6,838,351	26,342,058	9,573	41,751,874	8,450,457	467,497	6,111,833	28,394,582	9,295	43,433,625		
7	US, CA, DE, NL, CH, MK, SE, BE, FR, DK, SE, PT and 3 others	VOYAGES A LA CARTE (I-2K57Kw02070302)	Online	Online-Multinational OTAs	6,226,953	-	71,450	23,318	-	6,321,761	8,222,401	-	10,324	303	-	-	8,233,028	
8	US, NO, FR, SE, MX, DE, DK, GB, AU, BE and 3 others	EGENICA GLOBAL HQ (I-1X00L204128)	Business Travel	Managed Business- Global TMCs	7,051,284	1,642	12,456	5,852,621	8,189	12,326,261	7,694,074	1,116	98	6,595,721	9,666	-	14,291,958	
9	US, DE, FR, ES, AU, CA, FI, CH, SE and 3 others	CARLSON VACATION TRAVEL GLOBAL HQ (I-SWE300087)	Business Travel	Managed Business- Global TMCs	7,047,034	76,388	7,354,854	21,403,111	6,043	36,520,259	7,620,954	7,059	7,388,024	21,640,188	4,731	-	36,971,989	
10	DE, US, NL, AT, ES, FR, CH, IT, GB, TR and 3 others	ACERTICHE IMFT (I-2B56BD0L2070123)	Retail	Retail-Consolidators	8,153,173	-	388,965	1741,881	28,513	7,350,742	8,954,538	-	268,383	1,703,527	107,314	-	8,933,762	
11	NL, DE, SE, CH, BE, SG, FR, GB, AT, ES and 3 others	RAVAKHQ (I-G2-65G2010202)	Online	Online-Multinational OTAs	6,851,119	-	14	-	88,369	6,741,502	6,768,371	-	16,762	6	2,027,375	8,814,454	-	
12	DE, FR, NL, CH, AT, BE, ES, GB, SE, PT and 3 others	TU AG (I-TG72003068)	Retail	Retail-Tour Operators	6,532,988	-	84,535	363,770	801	6,982,024	6,447,391	-	99,912	305,266	1,957	6,854,426		
13	AU, GB, NZ, US, CA, ZA, AR, BR, SG, AT and 3 others	FLIGHT CENTRE LIMITED GLOBAL HQ (I-407A62008952)	Business Travel	Managed Business- Specialty	3,180,560	3,155	16,295,153	2,967,334	341	22,446,554	5,961,977	4,094	11,971,886	5,790,671	62	-	23,728,082	
14	IN, BH, US	MAKEMYTRIP	Online	Online-Multinational OTAs	5,282,370	-	-	-	-	5,282,370	5,686,672	-	-	-	-	-	5,686,672	
15	CA, UK, DE, NO, SE, BR, CH, BO, FI and 3 others	HKG GLOBAL HQ (I-7TP-20204109)	Business Travel	Managed Business- Global TMCs	5,027,748	123,926	3,766,273	5,569,827	-	14,497,774	5,053,176	146,737	3,320,125	5,001,905	-	-	10,521,943	
16	DE, CH, BE	DER TOURISTIK GROUP (I-1QCLP2004124)	Retail	Retail-Key Accounts	4,802,114	-	-	358,451	19,432	47,126	5,107,183	4,906,286	-	304,035	12,357	115	-	5,222,793
17	ES, MX, CL, PT, CO, PE, AR, FR, PA, EC and 3 others	GRUPO EL CORTE INGLEZ (I-GIG0442007057)	Retail	Retail-Key Accounts	4,298,674	-	-	16	-	4,298,674	4,369,291	-	-	-	-	-	4,368,291	
18	AR, BR, CO, CL, MX, PE, EC, US, CR, PA and 3 others	DESPRAR - LATAMERICA (I-KT1TS20120101)	Online	Online-Multinational OTAs	2,954,417	-	-	5,230,622	1,352,321	9,577,368	4,196,441	-	-	5,683,041	2,075,341	-	12,054,823	
19	SA, EG, AE	AL TAYYAR TRAVEL GROUP-MENA (I-20LN9N020103)	Business Travel	Managed Business- Region TMCs	1,545,778	-	-	845,400	168,889	95,848	2,653,881	3,574,953	-	744,827	230,657	85,212	4,534,849	
20	FR, GR, BE	SELECTOUR (I-16-19-2010101)	Retail	Concierge	3,476,763	-	202,420	186,279	-	3,953,461	3,438,505	-	205,317	261,392	-	-	3,905,104	
21	DE, FR, ES, IT, DK, SE, AT, GB, FI, PT and 3 others	TRAVELGOAL HQ (I-YMUVG20101016)	Online	Online-Multinational OTAs	2,931,134	-	119,368	24	-	4,051,138	3,314,246	-	1,384,781	13	-	-	4,709,040	
22	SE, NO, DK, IT, AU, NL, ES, MT, IE	TICKET LEISURE TRAVEL (I-V24WSI0041028)	Retail	Retail-Key Accounts	2,376,000	-	-	-	-	2,376,000	3,300,377	-	-	-	-	-	3,109,377	
	Query WorldWide Airlines 1A growth 1A wins 1A decrease 1A losses 1A Lost dead 1A Underdevelop 1A Development 1A Hunting (+) : 4 : 4 : 4																	

Figura 79: Standard Pulse - Top Wins

MARKETING PULSE

Este informe se libera cuatrimestralmente. Contiene un resumen mundial de la actividad de todos los GDS por segmento y región geográfica (figura 80). Al igual que el Standard Pulse, compara dos períodos de tiempo para obtener la mayoría de los KPI presentes en el informe.

A pesar de que las hojas principales del informe son datos muy agregados, nuestra audiencia desea conocer los datos a nivel cliente para poder usarlos a su gusto y crear nuevas visualizaciones e informes, en caso necesario.

RANKING

Este informe es el más sencillo, simplemente contiene una tabla con las 100 mejores agencias de viajes (figura 81). Se puede generar para toda la industria o para cada GDS, por lo que es una herramienta útil a la hora de reconocer los principales clientes de la competencia.



MARKETING INTELLIGENCE TRACKING TOOL



AMADEUS WORLDWIDE Pulse Marketing: Segment Quantification



Prev: November 2015 to October 2016

Last: November 2016 to October 2017

Prepared By: Global Commercial Intelligence

Region	Total GDS Accounts	Corp										Amadeus Growth	
		Total GDS Booking Volumes ('000)				Segment Weight			Amadeus Market Share				
		November 2016 to October 2017	November 2015 to October 2016	November 2016 to October 2017	Trend	November 2016 to October 2017	November 2016 to October 2017	Trend	November 2016 to October 2017	November 2016 to October 2017	Trend		
APAC	8	56	58		↑	0.0%	0.0%	↓	68.8%	69.1%	↑	3.7%	
LATAM	3	29	27		↓	0.0%	0.0%	↓	-	0.0%	↑	-	
NECSE	16	102	88		↓	0.0%	0.0%	↓	9.9%	8.9%	↓	(22.7%)	
NORAM	2	2	1		↓	0.0%	0.0%	↓	-	-	↓	-	
WEMEA	130	314	335		↑	0.1%	0.1%	↑	62.5%	59.5%	↓	14%	
WORLDWIDE TOTAL	159	503	509		↑	0.0%	0.0%	↓	48.7%	48.5%	↓	0.7%	

Region	Total GDS Accounts	Corp: Managed Business Corporation										Amadeus Growth	
		Total GDS Booking Volumes ('000)				Subsegment Weight			Amadeus Market Share				
		November 2016 to October 2017	November 2015 to October 2016	November 2016 to October 2017	Trend	November 2016 to October 2017	November 2016 to October 2017	Trend	November 2016 to October 2017	November 2016 to October 2017	Trend		
APAC	8	56	58		↑	0.0%	0.0%	↓	68.8%	69.1%	↑	3.7%	
LATAM	3	29	27		↓	0.0%	0.0%	↓	-	0.0%	↑	-	
NECSE	16	102	88		↓	0.0%	0.0%	↓	9.9%	8.9%	↓	(22.7%)	
NORAM	2	2	1		↓	0.0%	0.0%	↓	-	-	↓	-	
WEMEA	130	314	335		↑	0.1%	0.1%	↑	62.5%	59.5%	↓	14%	
WORLDWIDE TOTAL	159	503	509		↑	0.0%	0.0%	↓	48.7%	48.5%	↓	0.7%	

Region	Total GDS Accounts	Business Travel										Amadeus Growth	
		Total GDS Booking Volumes ('000)				Segment Weight			Amadeus Market Share				
		November 2016 to October 2017	November 2015 to October 2016	November 2016 to October 2017	Trend	November 2016 to October 2017	November 2016 to October 2017	Trend	November 2016 to October 2017	November 2016 to October 2017	Trend		
APAC	225	59,647	64,772		↑	24.4%	24.1%	↓	27.8%	28.1%	↑	9.9%	
LATAM	42	17,365	17,720		↑	19.2%	18.3%	↓	24.2%	26.0%	↑	9.6%	
NECSE	111	51,668	53,134		↑	25.0%	24.4%	↓	83.7%	82.4%	↓	13%	
NORAM	159	107,543	141,746		↑	32.3%	32.3%	↓	3.6%	3.7%	↑	7.9%	
WEMEA	219	70,669	73,861		↑	23.5%	23.8%	↑	49.1%	54.7%	↑	16.4%	

Figura 80: Marketing Pulse

AMADEUS NECSE Ranking All GDS



Prev: January 2017 to May 2017

Last: January 2018 to May 2018

Prepared By: Global Commercial Intelligence

Rank	Account	Resp	Countries	January 2017 to May 2017		January 2017 to May 2017 Total		January 2018 to May 2018		January 2018 to May 2018 Total		Growth	Amadeus MS	Amadeus MS evolution
				January 2017 to May 2017 Amadeus	January 2017 to May 2017 Total	January 2017 to May 2017 Amadeus	January 2017 to May 2017 Total	January 2018 to May 2018 Amadeus	January 2018 to May 2018 Total	Account weight (Total bkgd)	Account % growth			
				66,596,371	96,69,965	65,621,321	96,657,423	0.0%	67.3%	(17)				
1	EUROPEAN TRAVEL INTERACTIVE (15401 CG20081028)	GCG	SE, GR, NO, DK, DE, FI, CH, IL, AT, BG and 5 others	5,327,427	6,350,037	3,752,038	4,579,737	4.7%	(20.3%)	68.0%	(14.5)			
2	ABERTOUR ITALIA S.p.A. (26504 CG20170229)	GCG	DE, AT, DK, SE, PL, FI, DK, SK	2,930,144	4,143,255	2,900,969	4,160,653	4.2%	0.3%	68.9%	(11.0)			
4	TRIPSTA (15174 CG20140709)	Local	GR, MT, DE	221,023	169,445	98,304	342,073	3.4%	68.7%	23.9%	18.4			
13	E-TRAVEL (14918 CG20130331)	Local	GR, RU, CY	438,903	2,236,516	1,753,938	3,362,753	3.5%	41.4%	55.4%	26.8			
3	ECO TRAVEL HQ (14702 CG20160421)	GCG	DE, SE, DK, CH, NQ, FI, C2, AT, GR, A2 and 1 other	3,271,324	3,320,111	3,012,225	3,18,306	3.2%	(6.3%)	97.1%	(14)			
4	CARLSON WAGNER (14612 CG20160717)	GCG	DE, FI, CH, DK, SE, AT, PL, EE, HU, PR and 8 others	2,149,112	2,982,666	2,134,391	2,949,783	3.1%	(11.5%)	72.4%	0.3			
6	AMERICAN EXPRESS GLOBAL BUSINESS TRAVEL HQ (1-75203 CG20160308)	GCG	DE, SE, AT, FI, DK, CH, NO, PL, C2, GR and 5 others	2,523,701	2,895,723	2,628,525	2,902,117	3.0%	3.4%	90.6%	0.6			
5	TU AG (14702 CG20160319)	GCG	DE, CH, AT, SE, ND, PL, FI, DK, SK	2,739,763	2,933,714	2,565,029	2,799,598	2.9%	(4.9%)	92.0%	(1.5)			
7	DER TOURISTIK GROUP (14201 CG20140214)	Local	DE, CH	2,576,263	2,759,234	2,565,770	2,743,725	2.8%	(0.6%)	93.5%	0.1			
8	VIP SERVICE (14961 CG20130814)	Local	RU, AM, DE	595,309	2,922,594	1,766,049	2,277,769	2.6%	0.1%	43.1%	19.5			
10	EGENDA GLOBAL HQ (1-N-20140219)	GCG	NO, SE, DE, DK, CH, FI, PL, C2	2,503,839	2,904,635	2,561,329	2,962,272	2.7%	2.3%	100.0%	(0.0)			
9	CGEO (14961 CG20130222)	GCG	DE, NO, SE, FI, DK, AT, PL, CH	2,516,859	2,914,859	2,425,921	2,425,921	2.6%	(3.5%)	100.0%	(0.0)			
12	HKG GLOBAL HQ (1-TPP-20140119)	GCG	DE, NO, SE, CH, FI, DK, RU, PL, C2 and 3 others	2,091,721	2,307,453	2,10,551	2,343,348	2.4%	0.3%	90.2%	0.7			
11	EXPEDIA INC (1-TD-20030918)	GCG	DE, CH, FI, SE, ND, DK, AT	1,660,397	2,465,470	1,954,248	2,160,982	2.2%	(18.8%)	72.3%	5.7			
15	TICKET TO LEISURE TRAVEL (14612 CG20140203)	Local	SE, NO, DK, FI, DE, MT	1,638,081	1,638,861	1,748,182	1,748,182	1.6%	6.7%	100.0%	-			
17	LUFTHANSA CENTER INTERNATIONAL (1-930208060127)	GCG	DE, SI, CH, DK, AL	1,266,574	1,302,356	1,26,366	1,282,376	1.5%	(1.5%)	86.7%	(0.1)			
18	BEPG-HANSEN PREISEBUERAU AS (19131 CG20130305)	Local	NO	1,131,142	1,129,998	1,139,147	1,163,108	1.2%	3.0%	97.9%	(0.7)			
16	AVALINA (1X20010203040401)	Local	PL, UA, DE, AT, K2, EE, TR, BY, TJ, MD and 5 others	1,354,870	1,579,889	555,467	1,076,694	1.1%	(31.9%)	49.7%	(36.0)			
19	QUALITY TRAVEL ALLIANCE (17455 CG20130305)	Local	DE	799,299	931,963	671,721	1,005,521	1.0%	7.7%	66.9%	(10.7)			
27	ALTAIR TRAVEL (14961 CG2013020723)	Local	GR, DE	718,121	674,270	562,886	96,084	1.0%	42.7%	36.7%	19.2			
21	TRAVI-HQ (143-6320307020)	GCG	DE, CH, AT, PL	372,417	838,958	276,188	98,271	1.0%	9.5%	31.0%	(14.5)			
24	ATA TSENTRI (1-POECD-200104)	Local	PL, UA, BY	414,933	797,706	395,417	855,597	0.9%	0.9%	46.2%	(8.4)			

Figura 81: Ranking



3.4.4. MAINTENANCE

El módulo de Maintenance se utiliza para revisar la agrupación de oficinas y agencias de viajes generadas por los procesos de carga que se explicaron en el apartado 3.3.

Se utiliza principalmente para aquellos mercados en que CRM no está disponible aunque también pueden existir oficinas sin CRM Master Account asignada en los mercados con CRM. Esto suele ocurrir cuando se crean nuevas oficinas en los sistemas y los vendedores no han tenido tiempo para revisarlas. En todos estos casos, se valida el resultado del ETL que se refleja en el valor de Match Type, aprobando la asignación creada o corrigiéndola en caso necesario.

La participación del alumno en este módulo ha consistido en crear las sentencias SQL que se llaman desde el front-end y traducen las operaciones realizadas en el interfaz por los usuarios en consultas o cambios sobre la base de datos. Los cambios en los ficheros JS del front-end fueron realizados por un compañero del equipo de Productos.

La figura 82 muestra la pantalla principal de este interfaz. Se trata de una tabla con la relación entre Offices y Accounts. Cada mes, los usuarios efectúan una revisión de las oficinas no asignadas en CRM con el propósito de clasificarlas bajo la agencia de viajes que corresponda. Esta tabla proporciona información de las oficinas almacenadas en el sistema, mostrando para cada una de ellas las agencias de viajes a las que pertenecen.

The screenshot shows a software interface titled "Maintenance Interface" with the Amadeus logo at the top. The main area is a grid table with the following columns: Account ID, Account Name, Match Type, Country, Office ID, Office Name, IATA Number, Load Date, Production Date, Agency Name, Trade Name, and Address. The rows list various travel agency offices, some of which are assigned to specific agencies like "AERIVO SOLUTIONS S.L.", while others like "AFFAQ TRAVEL / PAKLINKS VIAJES SL" are listed under "UNKNOWN". A tooltip provides details for one row: "Matching for OfficeID (OID) 11A and IATA Number and Agency Name, some Address(es), Trade Name, City Name, Postal Code". At the bottom, there are navigation buttons for page numbers and a status message: "Displaying 401 - 500 of 31654".

Figura 82: Maintenance Interface

Esta tabla contiene la siguiente información:

- Agencia de viajes: Account ID y Account Name (editable).
- Oficina: Country, Office ID, Office Name, IATA Number, Office Type, Agency Name, Trade Name, Address, Address 2, Address 3, City, Postal Code.



- Atributos de la relación:

- Match Type: campo auxiliar utilizado por el proceso de carga, para clasificar e indicar al usuario el grado de similitud de esa oficina con otras oficinas del sistema. Si el valor es 100 significa que está clasificada en CRM o ya se ha revisado. El resto de valores indican que es una asociación creada por el ETL. En particular, cuando Match Type tiene el valor 0 es debido a que no existe en la base de datos otra oficina con características similares que puedan agilizar la clasificación de la misma. Al pasar el ratón por encima de estos valore se muestra un mensaje con la descripción asociada al valor numérico.
- Start y End Date: Fechas en que está vigente la relación Office-Account. Normalmente Start Date tomará el valor del primer mes en que la oficina apareció en el sistema (Load Date) y End Date estará vacío para indicar que la relación está vigente en la actualidad. Estas fechas, en caso de Offices pertenecientes a distintas Accounts a lo largo del tiempo, permitirán clasificar a estas Offices bajo una u otra Account en cada periodo de tiempo.
- Observations: Contiene anotaciones que hacen referencia a los cambios que ha sufrido un link (fecha y usuario que efectuó los cambios). No obstante, los usuarios pueden editar libremente su contenido y realizar sus propias observaciones.

Para facilitar las tareas de mantenimiento se dispone de varios filtros en la parte superior de la pantalla (figura 83). Por ejemplo, se pueden filtrar todas las oficinas de un país, de un mes en concreto, aquellas con match type superior o inferior a un valor concreto, etc.

The screenshot shows the maintenance interface with several filter fields at the top:

- Account Name: Write an account
- Match Type: Match Type dropdown
- Country: AR - ARGENTINA dropdown
- IATA Number: Write an IATA
- Load Date: Select begin (button), Select end (button)
- Office ID: Write an office ID
- Office Name: Write an office
- Agency Name: Write an agency
- Trade Name: Write an agency
- Production Date: Select date (button)

Below these are two expandable dropdown menus:

- Match Type: Match Type dropdown with operators <, >, = and Enter Number... buttons.
- Load Date: Select begin button, followed by a calendar grid showing months from Jan to Dec for years 2014 to 2023, with 2018 highlighted.

Figura 83: Filtros disponibles en Maintenance

Los usuarios pueden seleccionar las oficinas que deseen, pulsar el botón derecho sobre alguna de ellas y efectuar las acciones disponibles en la figura 84:

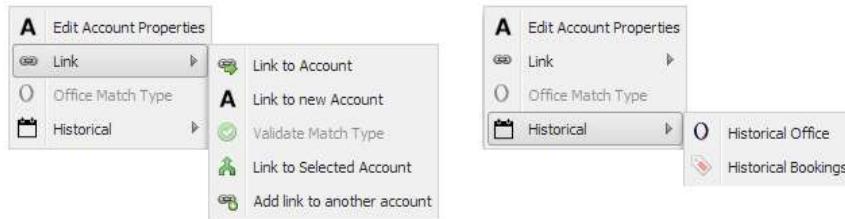


Figura 84: Operaciones disponibles en Maintenance

- A Edit Account Properties** Permite editar los atributos de una agencia de viajes (Master Account Number y Name).
- Link**
- Validate Match Type** Aplicable a los links con Match Type inferior a 100. Esta opción cambia el valor de este campo a 100, almacena el antiguo valor en Old Match y en Observations la fecha y el usuario que efectuó la operación.
- Link to Account** Permite cambiar la agencia de viajes de todas las oficinas marcadas en el interfaz.
- Link to new Account** Permite cambiar la agencia de viajes asignada por una nueva que se creará en ese instante.
- Link to Selected Account** Permite cambiar la agencia de viajes de todas las oficinas marcadas en el interfaz, por la agencia de viajes seleccionada en la pantalla con el ratón.
- Add link to another account** Permite añadir un nuevo link account-office, además de los ya existentes, de todas las oficinas marcadas en el interfaz, siempre y cuando no se solapen en el tiempo.
- Office Match Type** Abre un formulario auxiliar mostrando únicamente la oficina con la que se han encontrado campos coincidentes en el ETL. En caso de seleccionar varias oficinas, se mostrarán las oficinas emparejadas a cada una de ellas.
- Historical**
- Historical Office** Muestra la evolución de la dimensión SCD Office en la fuente CBI (cambios en Agency, Trade, Address, City, Postcode) para la oficina seleccionada.
- Historical Bookings** Muestra la producción, en bookings mensuales, de la oficina seleccionada.



4. CONCLUSIONES Y LÍNEAS FUTURAS

Este apartado contiene unos breves comentarios sobre el producto que se ha desarrollado, en base al cumplimiento de los requisitos marcados al inicio del proyecto, así como una serie de mejoras o líneas futuras que podrían aplicarse al trabajo realizado.

4.1. CONCLUSIONES

Nuestros objetivos estaban orientados a mejorar la arquitectura de la base de datos actual, los procesos de carga encargados de alimentarla, así como la explotación de esta información.

Por un lado, se ha logrado incorporar información procedente de nuevas fuentes de datos y se ha optimizado el modelo de datos de la base de datos del negocio. Para ello, por ejemplo, se ha disminuido la granularidad de los datos de oficinas de agencias de viajes a cambio de una mayor complejidad en el tratamiento de los datos. No obstante, se han creado un conjunto de procesos ETL que simplifican las cargas de datos, reduciendo las tareas manuales que hasta ahora se requerían, aunque la tecnología empleada, TOL, no está muy extendida por lo que no resulta fácil introducir mejoras ni encontrar gente con conocimientos en esta plataforma.

Por otra parte, se ha conseguido mejorar el proceso de mantenimiento de la información de agencias de viajes gracias a tres aspectos:

1. Incorporación de la información procedente de CRM.
2. Corrección de errores actuales en el proceso de asignación de cuentas. Se ha creado un proceso ETL encargado de asignar las oficinas que no están clasificadas en CRM, buscando en el histórico de la base de datos oficinas con campos coincidentes. El sistema no es perfecto al 100%, pero al menos ahora entienden la lógica aplicada en esta asignación.
3. Interfaz de mantenimiento, Maintenance, que dota al usuario de herramientas que agilizan este proceso.

Con respecto a la explotación de la información, la creación del Dashboard ha supuesto una mejora notable en el acceso a la información de toda la actividad relativa a agencias de viajes, facilitando la toma de decisiones rápidas, claras y concisas con datos que avalen las mismas.

Se ha creado una interfaz web dinámica, sencilla y fácil de usar, con visualizaciones simples donde se han introducido colores y símbolos para transmitir la información y



mostrar la evolución de ciertos KPI a través del tiempo. Gracias a este módulo, los equipos de ventas de la empresa pueden, entre otras cosas, identificar fácilmente la información más relevante y analizar los aspectos positivos o negativos de un segmento o una agencia.

En cambio, los informes creados han estado demasiado supeditados a los ya existentes, sin llegar a convertirse en algo novedoso que aporte un valor añadido. Simplemente se han migrado a la nueva plataforma, con un aspecto mejorado, pero sin introducir apenas mejoras ni visualizaciones atractivas. En la mayoría de los casos consisten en tablas enormes que dificultan toda visualización al comparar el rendimiento de dos períodos de tiempo.

Por último, en el área de cubos OLAP, se han configurado unos esquemas en Mondrian con los que cualquier usuario, sin conocimientos técnicos como el lenguaje SQL simplemente con conocer las medidas y dimensiones que ofrecen los cubos, puede extraer la información existente en la base de datos. No obstante, se han encontrado bastantes inconvenientes a la hora de explorar los cubos con esta solución, ya que el rendimiento no ha resultado óptimo, en gran medida debido a la altísima cardinalidad de las dimensiones Account y Office, ya que se ha intentado abarcar demasiada información.

En conclusión, se ha generado un sistema que mejora al anterior, con una aplicación web que facilita el trabajo diario de los usuarios y la explotación de la información.

4.2. LINEAS FUTURAS

A continuación se presentan algunas posibles evoluciones y mejoras que se pueden desarrollar a partir del proyecto construido.

En primer lugar, se debería replantear la consulta de la información con cubos OLAP en Mondrian. Este sistema no ha tenido un buen rendimiento y, finalmente, es un módulo que está cayendo en desuso. Si Amadeus apuesta por la tecnología Microsoft SQL Server, la herramienta de análisis de este proveedor sería una opción a valorar. Habría que crear estos cubos desde cero en esta plataforma pero el rendimiento sería mucho mejor que el actual.

En segundo lugar, la tecnología empleada para desarrollar la parte visual de MIT requiere conocimientos avanzados de programación y muchas de las librerías utilizadas sería conveniente actualizarlas.

Por otra parte, la apuesta tecnológica de Amadeus es muy amplia y está trabajando con sistemas como HPE Vertica para los almacenes de datos y Qlik Sense para el desarrollo



de dashboards e informes, por lo que se podría plantear la migración del sistema actual a este entorno.

La plataforma HPE Vertica es un almacén relacional y columnar, que no agrupa la información en filas, sino que en discos por columnas. El rendimiento de las consultas mejora notablemente en este sistema por lo que la evolución en este caso no pasaría por tener un datamart y crear nuevos cubos en otra plataforma sino que se migraría nuestra base de datos de Microsoft SQL Server a HPE Vertica, cuyo rendimiento nos permitiría seguir usando muchas de las actuales consultas SQL optimizadas. Además, se ha comprobado que es tan potente que podríamos manejar datos diarios en vez de mensuales por lo que la mejora en el lado de negocio también sería sustancial.

Tal y como mencionamos en las conclusiones, con el sistema creado en este proyecto, se puede ayudar a los usuarios a responder preguntas sobre el negocio conocidas de antemano (tendencias, evolución de clientes, mercados, etc.) pero después de consultar esta información en el dashboard o en algún informe pueden surgir nuevas inquietudes. Qlik Sense es una plataforma BI Self-Service (autoservicio) para el análisis de datos que Amadeus ha decidido adquirir. Con Qlik Sense, los usuarios pueden hacer descubrimientos de datos por su cuenta, explorar la información libremente y crear sus propias visualizaciones para responder sus preguntas, siguiendo su propio camino hacia la comprensión. Con esta herramienta se pueden automatizar procesos de generación de información y conocimiento, disponiendo también de posibilidades para la visualización y el análisis que nos ayudan a la toma de decisiones, todo ello trabajando con herramientas de uso sencillo para un usuario de negocio.

En definitiva, si se opta por este camino, el tiempo empleado para crear visualizaciones será inferior al actual pues Qlik Sense es un producto muy potente que permite a usuarios sin perfil técnico crear visualizaciones simples. Asimismo, un desarrollador puede crear productos web basados en Qlik, es decir, se podría combinar lo mejor de ambos lados.



BIBLIOGRAFÍA

- Abramson, I. (2002). *Data Warehouse: The Choice of Inmon versus Kimball*. Retrieved Junio 1, 2018, from ISMLL: <https://www.ismll.uni-hildesheim.de/lehre/bi-10s/script/Inmon-vs-Kimball.pdf>
- Amadeus IT Group, S.A. (2017). *Informe Global 2017*. Retrieved Junio 1, 2018, from Amadeus: <http://www.amadeus.com/msite/global-report/2017/es/pdf/amadeus-informe-global-2017.pdf>
- Bernabeu, D. (2010, Enero 4). *01 Introducción a Spoon*. Retrieved Junio 1, 2018, from Pentaho: <https://wiki.pentaho.com/pages/viewpage.action?pageId=14844841>
- Carrera, R. (2017, Julio 10). *[Modelo dimensional] Dimensiones que cambian lentamente – Parte II*. Retrieved Junio 1, 2018, from BI-Saga: <http://bisaga.com/dimensiones-que-cambian-lentamente-2/>
- Dertiano, V. (2014, Septiembre 22). *¿Qué es Business Intelligence?* Retrieved Junio 1, 2018, from BI-GEEK: <http://blog.bi-geek.com/que-es-business-intelligence/>
- Dertiano, V. (2015, Marzo 9). *Arquitectura BI (Parte II): El enfoque de William H. Inmon*. Retrieved Junio 1, 2018, from BI-GEEK: <http://blog.bi-geek.com/arquitectura-enfoque-de-william-h-inmon/>
- Dertiano, V. (2015, Abril 6). *Arquitectura BI (Parte III): El enfoque de Ralph Kimball*. Retrieved Junio 1, 2018, from BI-GEEK: <http://blog.bi-geek.com/arquitectura-el-enfoque-de-ralph-kimball/>
- Fernández, Á. (2017, Abril 10). *Business Intelligence, ETL. Introducción a Kettle*. Retrieved Junio 1, 2018, from BI-GEEK: <http://blog.bi-geek.com/kettle-parte-i-introduccion/>
- Global Mentoring. (n.d.). Curso de Java con JDBC. *Lección 1*.
- Grupo Atrium. (n.d.). Master Certificado Experto Java EE. *Módulo 1, Capítulo 13*.
- Hitachi Vantara. (n.d.). *Pentaho Business Analytics*. Retrieved Junio 1, 2018, from Hitachi Vantara: <https://www.hitachivantara.com/en-us/products/big-data-integration-analytics/pentaho-business-analytics.html>
- IBM Knowledge Center. (n.d.). *Esquemas de copo de nieve*. Retrieved Junio 1, 2018, from IBM: https://www.ibm.com/support/knowledgecenter/es/SS9UM9_9.1.2/com.ibm.datatoolsdimensional.ui.doc/topics/c_dm_snowflake_schemas.html
- IBM Knowledge Center. (n.d.). *Esquemas de estrella*. Retrieved Junio 1, 2018, from IBM:



https://www.ibm.com/support/knowledgecenter/es/SS9UM9_9.1.2/com.ibm.datatoolsdimensional.ui.doc/topics/c_dm_star_schemas.html

Imagina Formación. (n.d.). Curso de Desarrollo de Aplicaciones Java con Spring. *Tema 1.*

International Air Transport Association. (n.d.). *IATA by Region*. Retrieved Junio 1, 2018, from IATA: <http://www.iata.org/about/worldwide/Pages/index.aspx>

Kimball, R. (2008, Agosto 21). *Slowly Changing Dimensions*. Retrieved Junio 1, 2018, from Kimball Group: <https://www.kimballgroup.com/2008/08/slowly-changing-dimensions/>

Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition*. Indianapolis: Wiley.

Menasalvas Ruiz, E. (2007). Bases de Datos Deductivas. *Tema 2: Diseño Multidimensional*.

MikeRayMSFT, C. G. (2017, Mayo 24). *Editions and supported features of SQL Server 2016*. Retrieved Junio 1, 2018, from Microsoft: <https://docs.microsoft.com/en-us/sql/sql-server/editions-and-components-of-sql-server-2016?view=sql-server-2017#SSIS>

Oracle. (n.d.). *JDBC Overview*. Retrieved Junio 1, 2018, from Oracle: <http://www.oracle.com/technetwork/java/overview-141217.html>

Ramos, S. (2016). *Data Warehouse, Data Marts y Modelos Dimensionales. Un pilar fundamental para La Toma de Decisiones*. SolidQ Global, S.A.

Rangarajan, S. (2016, Septiembre 1). *Data Warehouse Design – Inmon versus Kimball*. Retrieved Junio 1, 2018, from TDAN: <http://tdan.com/data-warehouse-design-inmon-versus-kimball/20300>

Ross, M. (2013, Febrero 5). *Design Tip #152 Slowly Changing Dimension Types 0, 4, 5, 6 and 7*. Retrieved Junio 1, 2018, from Kimball Group: <https://www.kimballgroup.com/2013/02/design-tip-152-slowly-changing-dimension-types-0-4-5-6-7/>

TOL Project. (n.d.). *Welcome to TOL Project*. Retrieved Junio 1, 2018, from TOL: <https://www.tol-project.org/>

Wikipedia. (n.d.). *Apache Tomcat*. Retrieved Junio 1, 2018, from Wikipedia: https://en.wikipedia.org/wiki/Apache_Tomcat

Wikipedia. (n.d.). *Business Intelligence and Reporting Tools*. Retrieved Junio 1, 2018, from Wikipedia: https://es.wikipedia.org/wiki/Business_Intelligence_and_Reportin g_Tools



Wikipedia. (n.d.). *Cascading Style Sheets*. Retrieved Junio 1, 2018, from Wikipedia:
https://en.wikipedia.org/wiki/Cascading_Style_Sheets

Wikipedia. (n.d.). *Data warehouse*. Retrieved Junio 1, 2018, from Wikipedia:
https://en.wikipedia.org/wiki/Data_warehouse

Wikipedia. (n.d.). *Ext JS*. Retrieved Junio 1, 2018, from Wikipedia:
https://en.wikipedia.org/wiki/Ext_JS

Wikipedia. (n.d.). *Extract, transform, load*. Retrieved Junio 1, 2018, from Wikipedia:
https://en.wikipedia.org/wiki/Extract,_transform,_load

Wikipedia. (n.d.). *HTML5*. Retrieved Junio 1, 2018, from Wikipedia:
<https://en.wikipedia.org/wiki/HTML5>

Wikipedia. (n.d.). *JavaScript*. Retrieved Junio 1, 2018, from Wikipedia:
<https://en.wikipedia.org/wiki/JavaScript>

Wikipedia. (n.d.). *JavaServer Pages*. Retrieved Junio 1, 2018, from Wikipedia:
https://en.wikipedia.org/wiki/JavaServer_Pages

Wikipedia. (n.d.). *Pentaho*. Retrieved Junio 1, 2018, from Wikipedia:
<https://es.wikipedia.org/wiki/Pentaho>

Wikipedia. (n.d.). *Spring Framework*. Retrieved Junio 1, 2018, from Wikipedia:
https://en.wikipedia.org/wiki/Spring_Framework



MARKETING INTELLIGENCE TRACKING TOOL

