



**UNIVERSIDAD DE CASTILLA-LA MANCHA**  
**ESCUELA SUPERIOR DE INFORMÁTICA**

**GRADO EN INGENIERÍA EN INFORMÁTICA**  
**TECNOLOGÍA ESPECÍFICA DE INGENIERÍA DEL SOFTWARE**

**TRABAJO FIN DE GRADO**

**Sistema para la toma de decisiones en la producción  
empresarial basado en Business Intelligence**

Miguel Ampuero López –Sepúlveda

Julio, 2018





**UNIVERSIDAD DE CASTILLA-LA MANCHA**

**ESCUELA SUPERIOR DE INFORMÁTICA**

**Tecnologías y Sistemas de Información**

**TECNOLOGÍA ESPECÍFICA DE INGENIERÍA DEL SOFTWARE**

**TRABAJO FIN DE GRADO**

**Sistema para la toma de decisiones en la  
producción empresarial basado en Business Intelligence**

**Autor: Miguel Ampuero López - Sepúlveda**

**Director Avanttic: Noel García Gómez**

**Director ESI: Francisco Ruiz González**

Julio, 2018



**TRIBUNAL:**

**Presidente:**

**Vocal:**

**Secretario:**

**FECHA DE DEFENSA:**

**CALIFICACIÓN:**

**PRESIDENTE**

**VOCAL**

**SECRETARIO**

Fdo.:

Fdo.:

Fdo.:



## Resumen

Hoy en día utilizar la información sobre la producción de una empresa es un factor clave para la toma de decisiones estratégicas, claves para el éxito del negocio. Utilizar como apoyo en dicha toma de decisiones un sistema basado en Inteligencia de Negocio es muy útil para poder mejorar la agilidad de la compañía. Obtener información a partir de los datos generados internamente y convertirlo en conocimiento útil en la toma de decisiones es factible gracias a las herramientas de *Bussines Intelligence*. Para tal fin los sistemas de información operacionales de la empresa deben servir como origen de datos para la creación de un *data warehouse*.

En este proyecto se ha desarrollado un sistema de *Bussines Intelligence*. Para ello, primero se ha hecho un estudio del sistema ERP de la compañía, que es donde residen los datos sobre producción a ser analizados. Posteriormente ha sido almacenado en el *data warehouse desarrollado para tal fin*, haciendo uso de una herramienta ETL comercial para, finalmente, ser explotados por una herramienta BI.



## **Abstract**

Nowadays, using the information about the production of your company is a key factor for making strategic business decisions. Using as decision support a system based on Business Intelligence is necessary for being able to improve the agility of the company. Obtaining the information from the data generated internally and converting them into useful knowledge in the decision-making process is feasible thanks to BI tools. Nevertheless, the operational information systems of the companies must serve as data sources for the creation of the data warehouse.

In this project it has been developed a Business Intelligence system. For that, firstly the ERP system of the company, where the production data is stored, has been studied. Later, that data had been stored in a data warehouse making use of one of the most potent ETL tools in the market. Finally, the data warehouse has been exploited by a BI tool.



*A mis padres, por enseñarme el valor del esfuerzo.*

*Ad Astra Per Áspera.*



## **AGRADECIMIENTOS**

Me gustaría comenzar acordándome de mis padres por el apoyo que he recibido a lo largo de todos estos años. Agradecer el esfuerzo diario que habéis realizado para brindarme la oportunidad de estudiar una carrera, disfrutar de un año Erasmus y vivir experiencias únicas. También me gustaría acordarme en estas líneas de mis abuelos Gabriel y Carmen, habéis sido parte fundamental en esta etapa de mi vida.

A Silvia y Carlos, mis hermanos, gracias por todo vuestro apoyo.

A mi tutor Francisco Ruiz, gracias por haber confiado en mí y en esta aventura de hacer un FORTE. También agradecer a todos y cada uno de los profesores de la Escuela que me han hecho crecer tanto en lo académico como en lo personal.

A avanttic Consultoría Tecnológica SL. por darme esta increíble oportunidad. Agradecer a mi tutor Noel todo su tiempo, dedicación y paciencia. He aprendido de un verdadero profesional en estos 5 meses. Tampoco olvidar a mis compañeros de BA: Paco, Basi y Diego. Y en general a toda esa gran familia que existe dentro avanttic.

A mis compañeros de FORTE Víctor, Óscar y Juan, personas con las que he vivido esta fase final de la carrera.

Por último, pero no menos importante, me gustaría acordarme de José Ángel, Diego y Ángel. Personas con las que he compartido estos 4 años de carrera, muchas horas de trabajos y proyectos pero sobre todo de buenos e inolvidables momentos.

Miguel Ampuero López-Sepúlveda.



# ÍNDICE

|       |  |    |
|-------|--|----|
| 1     | INTRODUCCIÓN.....                                  | 1  |
| 1.1   | Estructura del documento.....                      | 3  |
| 2     | OBJETIVOS DEL TFG.....                             | 5  |
| 2.1   | Objetivo principal.....                            | 5  |
| 2.2   | Objetivos de la Empresa.....                       | 5  |
| 2.3   | Objetivos Académicos.....                          | 6  |
| 3     | ANTECEDENTES, ESTADO DE LA CUESTIÓN.....           | 7  |
| 3.1   | Inteligencia de Negocio.....                       | 7  |
| 3.1.1 | Beneficios de la Inteligencia de Negocio.....      | 8  |
| 3.2   | Data Warehouse.....                                | 9  |
| 3.2.1 | Introducción.....                                  | 9  |
| 3.2.2 | Metodologías.....                                  | 10 |
| 3.2.3 | Modelo en Estrella.....                            | 14 |
| 3.2.4 | OLAP vs OLTP.....                                  | 16 |
| 3.2.5 | Procesos ETL.....                                  | 18 |
| 3.3   | Tecnología BI.....                                 | 19 |
| 3.3.1 | Ejemplos.....                                      | 19 |
| 4     | MÉTODO DE TRABAJO.....                             | 21 |
| 4.1   | Gestión del Proyecto.....                          | 21 |
| 4.1.1 | Artefactos de SCRUM.....                           | 21 |
| 4.2   | Desarrollo del Almacén de Datos.....               | 23 |
| 4.3   | Tecnologías usadas en el proyecto.....             | 25 |
| 4.3.1 | Marco tecnológico.....                             | 25 |
| 4.3.2 | Herramientas para el desarrollo del proyecto.....  | 25 |
| 4.3.3 | Herramientas para la gestión del proyecto.....     | 28 |
| 4.3.4 | Herramientas para el desarrollo de la memoria..... | 28 |

|       |  |    |
|-------|--|----|
| 5     | RESULTADOS.....  | 31 |
| 5.1   | Sprint 0: Preparación Inicial .....                        | 31 |
| 5.1.1 | Tareas.....  | 31 |
| 5.1.2 | Resultados del Sprint .....                                | 32 |
| 5.1.3 | Revisión .....   | 38 |
| 5.1.4 | Retrospectiva.....   | 39 |
| 5.2   | Sprint 1: Análisis de Requisitos y Modelado de Datos ..... | 39 |
| 5.2.1 | Tareas.....  | 39 |
| 5.2.2 | Resultados del Sprint .....                                | 40 |
| 5.2.3 | Revisión .....   | 43 |
| 5.2.4 | Retrospectiva.....   | 44 |
| 5.3   | Sprint 2: Extracción Inicial de los Datos .....            | 44 |
| 5.3.1 | Modificación de la Pila de Producto.....                   | 45 |
| 5.3.2 | Tareas.....  | 47 |
| 5.3.3 | Resultados del Sprint .....                                | 47 |
| 5.3.4 | Revisión .....   | 67 |
| 5.3.5 | Retrospectiva.....   | 67 |
| 5.4   | Sprint 3: Extracción Periódica de Datos.....               | 68 |
| 5.4.1 | Tareas.....  | 68 |
| 5.4.2 | Resultados del Sprint .....                                | 69 |
| 5.4.3 | Revisión .....   | 75 |
| 5.4.4 | Retrospectiva.....   | 76 |
| 5.5   | Sprint 4: Explotación de Datos .....                       | 76 |
| 5.5.1 | Planificación del Sprint.....                              | 76 |
| 5.5.2 | Desarrollo de Tareas.....                                  | 77 |
| 5.5.3 | Revisión .....   | 87 |
| 5.5.4 | Retrospectiva.....   | 88 |

|     |  |    |
|-----|--|----|
| 6   | CONCLUSIONES Y PROPUESTAS.....   | 89 |
| 6.1 | Cumplimiento de Objetivos .....  | 89 |
| 6.2 | Propuestas de Mejora.....  | 90 |
| 6.3 | Opinión Personal .....   | 90 |
| 7   | BIBLIOGRAFÍA.....  | 91 |
|     | ANEXOS.....  | 95 |
|     | ANEXO A – Acrónimos .....  | 95 |
|     | ANEXO B – Notas Obtenidas en las reuniones de captura de requisitos<br>..... | 97 |



## ÍNDICE DE TABLAS, FIGURAS E ILUSTRACIONES

|   |    |
|---|----|
| Tabla 1- Objetivos de Mejora del TFG .....                          | 5  |
| Tabla 2- Objetivos Académicos del TFG.....                          | 6  |
| Tabla 3- Comparación DB con DWH [10] .....                          | 10 |
| Tabla 4- Comparación ROLAP y MOLAP .....                            | 18 |
| Tabla 5- Correspondencia Sprints con Fases de Desarrollo de un DWH. | 24 |
| Tabla 6 - Planificación del Sprint 0 .....                          | 31 |
| Tabla 7- Historia de Usuario 1 .....                                | 33 |
| Tabla 8 - Historia de Usuario 2 .....                               | 34 |
| Tabla 9 - Historia de Usuario 3 .....                               | 34 |
| Tabla 10 - Historia de Usuario 4.....                               | 35 |
| Tabla 11- Plan de Proyecto en Sprint 0.....                         | 36 |
| Tabla 12 - Análisis de Riesgos .....                                | 37 |
| Tabla 13- Resumen del Sprint 0 .....                                | 38 |
| Tabla 14- Planificación del Sprint 1 .....                          | 39 |
| Tabla 15- Tablas de Hechos Identificadas .....                      | 42 |
| Tabla 16- Resumen del Sprint 1 .....                                | 43 |
| Tabla 17 - Historia de Usuario 5.....                               | 45 |
| Tabla 18 - Plan de Proyecto en el Sprint 2.....                     | 46 |
| Tabla 19- Planificación del Sprint 2 .....                          | 47 |
| Tabla 20 - Resumen del Sprint 2 .....                               | 67 |
| Tabla 21 - Planificación Sprint 3.....                              | 68 |
| Tabla 22 - Resumen Sprint 3 .....                                   | 75 |
| Tabla 23 - Planificación Sprint 4.....                              | 77 |
| Tabla 24 - Resumen Sprint 4.....                                    | 87 |
| Tabla 25 - Consecución de los Subobjetivos.....                     | 89 |
| Figura 1- Esquema del proyecto a desarrollar .....                  | 2  |
| Figura 2 - Resumen Marco Tecnológico.....                           | 25 |
| Figura 3 - Burn up Sprint 0 .....                                   | 38 |
| Figura 4 - Burn Up Sprint 1.....                                    | 44 |
| Figura 5 - Burn Up Sprint 2.....                                    | 67 |
| Figura 6 - Burn Up Sprint 3.....                                    | 76 |
| Figura 7 - Burn Up Sprint 4.....                                    | 87 |

|   |    |
|---|----|
| Ilustración 1- Información en una organización [8] .....                    | 8  |
| Ilustración 2- Metodología Top -Down .....                                  | 11 |
| Ilustración 3- Metodología Bottom-Up .....                                  | 13 |
| Ilustración 4- Ejemplo de Enterprise Bus Matrix [3] .....                   | 14 |
| Ilustración 5- Ejemplo de Modelo en Estrella [3] .....                      | 15 |
| Ilustración 6- Comparación Estrella vs Copo de Nieve.....                   | 16 |
| Ilustración 7- Resumen del proceso ETL [8] .....                            | 19 |
| Ilustración 8- Mapping con PowerCenter [15].....                            | 20 |
| Ilustración 9 - Resumen de la Metodología de Gestión Scrum [16] .....       | 22 |
| Ilustración 10 - Ciclo de Vida de DWH/BI de Kimball [17] .....              | 23 |
| Ilustración 11 - Metodología para un Proyecto BI [8].....                   | 23 |
| Ilustración 12 - Resumen del proceso ELT [18] .....                         | 26 |
| Ilustración 13 - Enterprise Bus Matrix del TFG .....                        | 41 |
| Ilustración 14 - Modelo Dimensional del DWH .....                           | 43 |
| Ilustración 15- Diseño Físico del DWH .....                                 | 49 |
| Ilustración 16 - Modelo Lógico de la capa de Staging .....                  | 51 |
| Ilustración 17- Definición del Esquema Físico de Staging.....               | 52 |
| Ilustración 18 - Esquemas físicos importados para la copia de Staging... 52 |    |
| Ilustración 19 - Modelo de Datos de Staging importado en ODI.....           | 53 |
| Ilustración 20 - Mapping en la capa de Staging .....                        | 53 |
| Ilustración 21- Plan de Carga de Staging.....                               | 54 |
| Ilustración 22- Resultado ejecución plan de carga Staging .....             | 54 |
| Ilustración 23- Esquema Físico del DWH Importado en ODI.....                | 54 |
| Ilustración 24 - Mapping Cliente (Vista lógica).....                        | 55 |
| Ilustración 25 - Mapping Cliente (Vista Física).....                        | 55 |
| Ilustración 26 - Módulos de Conocimiento Importados .....                   | 56 |
| Ilustración 27- Ejemplo Configuración Módulos de Conocimiento ODI. 56       |    |
| Ilustración 28 - Importación de una secuencia de BBDD a ODI.....            | 57 |
| Ilustración 29- Mapping DIM Empresa.....                                    | 57 |
| Ilustración 30 - Mapping DIM Tipo Actividad.....                            | 58 |
| Ilustración 31 - Mapping de DIM Área .....                                  | 58 |
| Ilustración 32- Mapping DIM Persona .....                                   | 59 |
| Ilustración 33 - Condición Look-Up DIM Persona.....                         | 59 |

|   |    |
|---|----|
| Ilustración 34 - Mapping DIM Proyecto .....                                   | 60 |
| Ilustración 35 - Mapping DIM Subproyecto .....                                | 60 |
| Ilustración 36 - Procedimiento populate DIM FECHA .....                       | 61 |
| Ilustración 37- Mapping FACT Horas (1).....                                   | 63 |
| Ilustración 38 - Mapping FACT Horas (2).....                                  | 63 |
| Ilustración 39- Mapping FACT Horas (3).....                                   | 64 |
| Ilustración 40 - Propiedades Agregación en el Mapping FACT_Actividad<br>..... | 64 |
| Ilustración 41 - Mapping FACT Horas día .....                                 | 65 |
| Ilustración 42 - Creación de Datos Genéricos.....                             | 66 |
| Ilustración 43 - Plan de Carga Inicial DWH .....                              | 66 |
| Ilustración 44 - Resultados Ejecución Plan de Carga Inicial .....             | 66 |
| Ilustración 45 - Mapping DIM Area con SCD .....                               | 69 |
| Ilustración 46 - Modificación Modelo de datos de DIM Área .....               | 70 |
| Ilustración 47 - Modificación Atributos DIM Área.....                         | 70 |
| Ilustración 48 - Cambio propiedades mapping DIM Área.....                     | 71 |
| Ilustración 49 - Cambio de propiedades mapping DIM Área en IKM .....          | 71 |
| Ilustración 50 - Carga de Datos Genéricos con tabla Auditoría.....            | 72 |
| Ilustración 51 - Consola del Agente Standalone de ODI .....                   | 73 |
| Ilustración 52 - Programación de Plan de Carga en ODI.....                    | 74 |
| Ilustración 53 - Configuración ODI para enviar mail.....                      | 75 |
| Ilustración 54 - Acceso a Oracle Business Intelligence.....                   | 78 |
| Ilustración 55- Capa Física del Repositorio BI .....                          | 79 |
| Ilustración 56 - Capa Lógica y de Presentación del Repositorio BI.....        | 79 |
| Ilustración 57 - Modelo lógico del DWH en el Repositorio BI .....             | 80 |
| Ilustración 58 - Selección de Área Temática en OBI .....                      | 80 |
| Ilustración 59 - Ejemplo de Informe con Drill Down.....                       | 81 |
| Ilustración 60 - Análisis sobre horas de empleados.....                       | 81 |
| Ilustración 61 - Informe con detalle de semanas por empleado .....            | 82 |
| Ilustración 62 - Informe Rendimiento Área .....                               | 83 |
| Ilustración 63 - Unión de datos en ODV.....                                   | 83 |
| Ilustración 64 - Mapa de Clientes en ODV .....                                | 84 |
| Ilustración 65- Cuadro de Mando de Clientes .....                             | 84 |

|   |    |
|---|----|
| Ilustración 66- Cuadro de Mando de Áreas .....      | 85 |
| Ilustración 67 - Cuadro de Mando de Empleados ..... | 86 |

# 1 INTRODUCCIÓN

En la actualidad, las empresas generan grandes cantidades de datos a nivel interno y externo. Para aumentar la competitividad y agilidad, éstas buscan explotar sus datos convirtiéndolos en información y posteriormente en conocimiento útil para la toma de decisiones a nivel táctico y estratégico. Obtener información y poder saber cómo tratarla es un factor clave hoy en día para cualquier negocio. Es aquí donde reside el concepto de Inteligencia de Negocio, *Bussines Intelligence* en inglés (en adelante BI).

Dentro del proceso que se sigue para la realización de un proyecto BI, la construcción de un almacén de datos, en inglés *Data Warehouse* (en adelante DWH), es una etapa clave para el correcto uso de los datos. J. Watson Hugh destacaba que “*BI es un proceso con dos partes: obtener los datos y sacar los datos. Obtener los datos tradicionalmente se ha definido como la creación de un almacén de datos*” [1]. Es por ello que una de las claves de este proyecto será la creación de un DWH para posteriormente ser explotado por un entorno BI, siguiendo los requisitos de los usuarios finales de la aplicación. Uno de los mayores referentes en este ámbito es Bill Inmon [2], que propuso una estructura específica para un DWH, diferente a las habituales bases de datos transaccionales. También es de destacar Ralph Kimball [3], que definió un proceso o método para desarrollar un DWH, que se seguirá en este trabajo de forma adaptada.

Aunque son muchas las posibles fuentes de datos que se pueden dar en un sistema de información, gracias a los procesos de extracción, transformación y carga (llamados ETL del inglés *Extract – Transform – Load*), éstas se pueden integrar en un mismo DWH. Siendo el Almacén de Datos diseñado con características físicas y lógicas idóneas para ser utilizado por una aplicación BI.

Este Trabajo Fin de Grado (en adelante TFG) se centra en desarrollar un *data warehouse* y utilizar una aplicación BI que explote su información, tomando como fuentes de datos el sistema ERP de la empresa Avanttic Consultoría Tecnológica SL. Éste TFG se encuadra dentro del convenio FORTE entre la mencionada empresa y la Escuela Superior de Informática. Aspectos claves del trabajo serán la captura de los requisitos y el posterior diseño del almacén de datos, así como el diseño de cuadros de mando e informes que aprovechen el DWH para la toma de decisiones por parte de los directivos de Avanttic.

El principal origen de datos tomado para la realización del TFG ha sido la Base de Datos del sistema ERP que utiliza la empresa mencionada anteriormente. Aunque son muchos los diferentes datos que son almacenados, el Trabajo de Fin de Grado consistirá en la obtención de información sobre la producción de la empresa. Se centrará en obtener la actividad de cada empleado en relación con los proyectos y clientes que tiene el negocio.

Los sistemas ERP, del inglés Enterprise Resource Planning [4], o planificación de recursos empresariales, son sistemas de información cuya principal funcionalidad es integrar y manejar los datos de los procesos de negocio que se dan en una empresa, además de los referidos a aspectos internos como operaciones de logística, inventario, procesos de producción, o contabilidad y finanzas.

La Figura 1 muestra el proceso que se ha seguido para el desarrollo del DWH y su posterior explotación con herramientas BI. Los pasos serán:

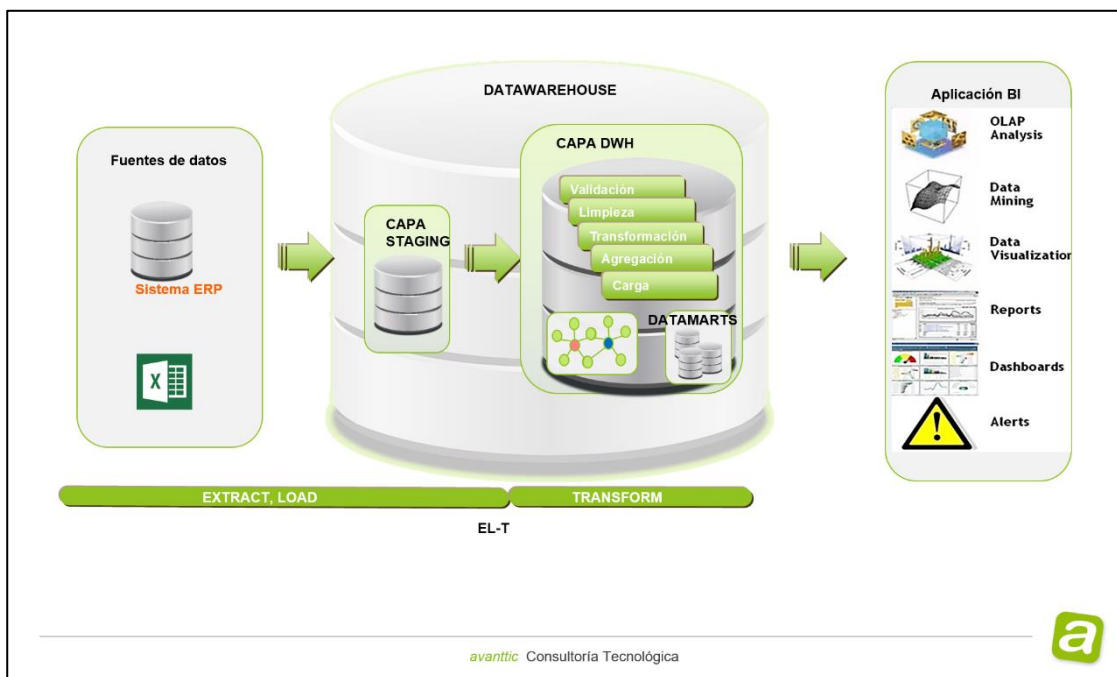


Figura 1- Esquema del proyecto a desarrollar

1. En primer lugar, se obtendrá acceso al sistema ERP, fuente de los datos para poder extraer y utilizar dichos datos. El ERP almacena los datos en una Base de Datos Oracle 11g [5].
2. Posteriormente, utilizando 'Oracle Data Integrator' (ODI en adelante) [6] se hará una copia de los datos de origen a una primera base de datos de trabajo, conocida normalmente como *Staging*. En este paso se mantiene la estructura

relacional de los datos de origen.

3. A continuación, se procederá a programar e implementar los proceso ETL (ODI suele denominarlos EL-T, ya que primero se extraen, luego se cargan y, finalmente, se transforman los datos). Para ello se hará un mapeo de los datos almacenados en *Staging* hasta el DWH final, siguiendo el modelado multidimensional, con esquema en estrella, propuesto por primera vez por Kimball [3].
4. Por último, se desarrollarán cuadros de mando e informes usando la herramienta de BI '*Oracle Bussines Intelligence*' (OBI en adelante) [7], siguiendo los requisitos establecidos para la explotación del DWH.

## **1.1 Estructura del documento**

La estructura de la memoria de este TFG se ha dividido en los siguientes capítulos con el contenido correspondiente:

- **Capítulo 1 Introducción:** Una introducción al tema tratado en el TFG, y el trabajo realizado durante el proyecto.
- **Capítulo 2 Objetivos del TFG:** Exposición del objetivo principal de este trabajo más los objetivos parciales derivados.
- **Capítulo 3 Antecedentes:** Estado de la cuestión: Presentación de los conocimientos previos necesarios para la realización de este TFG.
- **Capítulo 4 Método de Trabajo:** Explicación de la metodología de gestión y de desarrollo seleccionadas para este TFG. También se detallan las tecnologías que se han utilizado.
- **Capítulo 5 Resultados:** Se indican los resultados obtenidos tras seguir la metodología de trabajo descrita en el capítulo 4.
- **Capítulo 6 Conclusiones y Propuestas:** Comparación de los objetivos parciales descritos en el capítulo 2 con los resultados obtenidos. Además, se detallan propuestas de mejora para un futuro.
- **Capítulo 7 Bibliografía:** Listado del conjunto de libros, artículos y otro tipo de materiales bibliográficos que se han utilizado para el desarrollo de este TFG.
- **Anexos:** Material adicional.



## 2 OBJETIVOS DEL TFG

En este capítulo se define el objetivo principal del Proyecto además de los objetivos parciales y objetivos de mejora derivados.

### 2.1 Objetivo principal

Este TFG tiene como objetivo principal la creación de un almacén de datos empresariales sobre la producción de los empleados, registrando las horas destinadas a los diferentes proyectos que se realizan en la empresa, vacaciones u otras actividades para su posterior explotación *analítica*. Esto incluye todo el trabajo de captura de requisitos reales que se dan en la empresa, modelado del DWH, incorporación de los datos a partir de las fuentes, y diseño e implementación de los diferentes cuadros de mando e informes (almacenados y ad-hoc) para la explotación posterior. Las funcionalidades soportadas y los datos concretos manejados se irán definiendo con la empresa durante las etapas iniciales del proyecto.

Dado que el TFG se realiza en un ámbito empresarial real, se han considerado dos tipos de subobjetivos, los propios de la empresa, y los académicos relacionados con el TFG, y que son necesarios para alcanzar el objetivo principal.

### 2.2 Objetivos de la Empresa

Los objetivos de mejora para la empresa son:

| Objetivos de Mejora |  |
|---------------------|--|
| <b>OE1</b>          | Facilitar el proceso de creación de informes en el sistema ERP actual. |
| <b>OE2</b>          | Aumentar la calidad del análisis de los datos generados en el ERP.     |
| <b>OE3</b>          | Obtener datos históricos progresivos.                                  |

Tabla 1- Objetivos de Mejora del TFG

## 2.3 Objetivos Académicos

En cuanto a los subobjetivos propios del TFG se han identificado los siguientes:

| Subobjetivos |   |
|--------------|---|
| <b>OA1</b>   | Estudio de los diversos métodos a emplear.  |
| <b>OA2</b>   | Aprendizaje de las herramientas software que serán utilizadas, especialmente Oracle Data Integrator 12c y Oracle Business Intelligence 12c. |
| <b>OA3</b>   | Análisis y captura de requisitos.   |
| <b>OA4</b>   | Diseño del esquema de datos.  |
| <b>OA5</b>   | Realización del proceso de extracción inicial de datos.   |
| <b>OA6</b>   | Realización del proceso de actualización periódica de los datos.  |
| <b>OA7</b>   | Diseño e implementación de informes y cuadros de mando con una herramienta BI.  |

*Tabla 2- Objetivos Académicos del TFG*

### **3 ANTECEDENTES, ESTADO DE LA CUESTIÓN**

Este capítulo presenta todos los conceptos que han sido estudiados e investigados para el desarrollo del Trabajo Fin de Grado.

#### **3.1 Inteligencia de Negocio**

En el mundo empresarial actual la obtención de información a través de los datos internos del negocio sirve como soporte para la toma de decisiones críticas a nivel de organización. Conocer el estado actual de tu empresa, las posibles razones de ese comportamiento, en la forma y tiempo adecuados, además de poder prevenir futuras situaciones es parte de la motivación donde surge el concepto de Inteligencia de Negocio o *Bussines Intelligence*.

Inteligencia de Negocio es un conjunto de herramientas y procesos que sirven para el acceso a grandes cantidades de datos que apoyen a los usuarios en la toma de decisiones de negocio.

En cuanto a la información que es procesada en un sistema BI, se pueden distinguir tres tipos tal y como muestra la Ilustración 1:

- **Nivel Operacional:** Datos transaccionales, relacionados con las operaciones diarias del negocio. Suele estar consultada por el personal operativo dentro de la empresa. Los datos generados a este nivel suelen ser el origen de datos para un DWH.
- **Nivel Táctico:** Primer nivel que se podría considerar propio de BI, relacionados con los gerentes de cuentas o jefes de departamentos en un negocio.
- **Nivel Estratégico:** Nivel más alto dentro del tipo de información que se genera en una empresa, la información generada es útil para la alta dirección de la empresa.

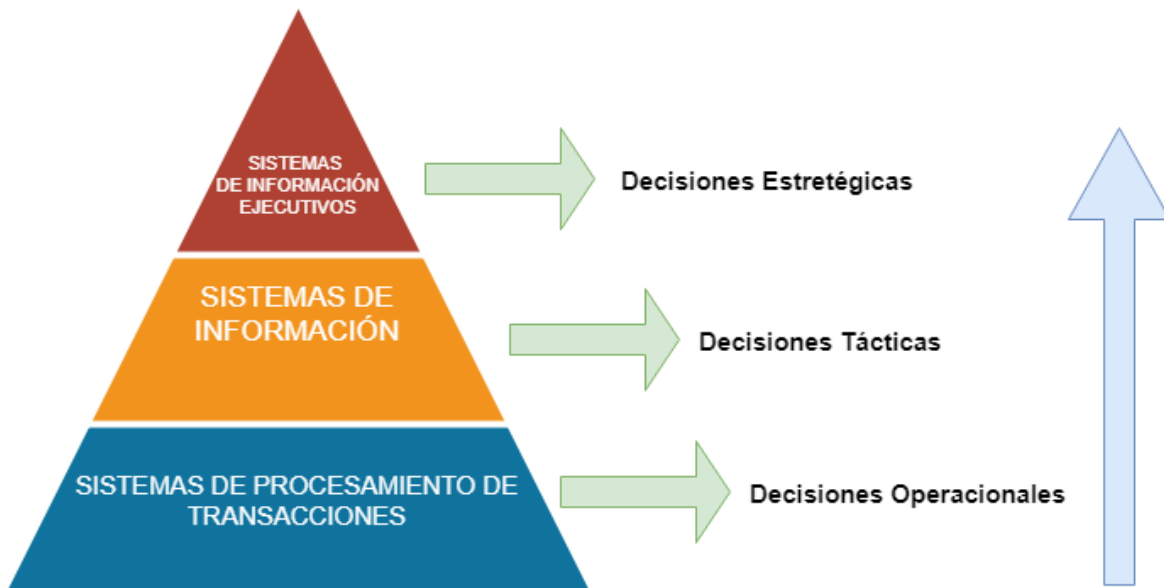


Ilustración 1- Información en una organización [8]

### 3.1.1 Beneficios de la Inteligencia de Negocio

Las empresas gracias a las herramientas y tecnologías relacionadas con BI pueden, a partir de sus sistemas operacionales, generar diferentes análisis de información que aporten valor para el negocio [9]. Las principales ventajas son las siguientes:

- **Conocer tu negocio:** Las empresas hoy en día tienen que evolucionar y cambiar en función de las necesidades de los clientes. Tener bien identificados los procesos de negocio es igual de importante que conocer su estado actual. Un sistema BI permite tener en tiempo real análisis de los datos operacionales.
- **Manejo de los costes:** Una de las principales ventajas de mantener todos los datos de tus sistemas operacionales congregados en un sistema de BI es que te permite ver el estado actual de los costes a diferentes niveles de agregación y detalle.
- **Mejor entendimiento de los clientes:** Los usuarios generaran grandes cantidades de datos que no suelen ser aprovechados. Una empresa debería transformar todas esas ‘toneladas’ de datos en información que aporte valor al negocio, mejorando la relación con los clientes.
- **Indicadores de gestión:** Gracias a la implantación de un sistema de *Bussines Intelligence* se puede acceder a la información de diferente forma. Pudiendo crear indicadores de gestión, partiendo de los datos operacionales, que sirvan de apoyo en las decisiones tácticas o estratégicas.

## 3.2 Data Warehouse

### 3.2.1 Introducción

Un *Data Warehouse* o almacén de datos en castellano, es un sistema de almacenamiento diseñado para contener datos extraídos de sistemas transaccionales y de otras fuentes externas. El almacén combinará los datos agregados para un posterior análisis y creación de informes que cumplan los requisitos del negocio.

Uno de los mayores representantes de esta temática es Bill Inmon [2]. El científico inglés, denominado por muchos como el padre del *Data Warehouse* definió 4 propiedades que caracterizan este tipo de sistemas de almacenamiento:

- **Orientado a temas:** Un DWH debe estar diseñado para poder buscar por un tema en concreto, por ejemplo “pedidos”.
- **Integrado:** Un DWH integra datos de diferentes fuentes de datos.
- **Variante en el tiempo:** Los datos históricos deben poder ser almacenados en el *Data Warehouse*.
- **No volátil:** Una vez que los datos han sido almacenados, éstos no deben ser alterados. De esta manera un DWH siempre mantiene intactos los datos históricos registrados.

Por otro lado, Ralph Kimball [3] también es un autor que ha aportado a esta temática. Kimball definió un *Data Warehouse* como: “Un almacén de datos que extrae, limpia, conforma y entrega una fuente de datos dimensional para la consulta y análisis”.

La Tabla 3 muestra una comparación entre un DWH y una base de datos convencional:

|                             | <b>Base de Datos</b>  | <i>Data Warehouse</i>   |
|-----------------------------|---|---|
| <b>Objetivo</b>             | Soporte de las actividades diarias de la empresa  | Análisis y tomas de decisiones                                    |
| <b>Tipo de datos</b>        | Enfocado a los detalles   | Centrado en los datos agregados y en distintos niveles de detalle |
| <b>Explotación</b>          | Aplicaciones específicas del negocio  | Abarca todo el negocio usando diferentes fuentes de datos         |
| <b>Perspectiva del dato</b> | Representa el dato actual   | Representan datos históricos                                      |
| <b>Rendimiento</b>          | Respuestas rápidas para información requerida a nivel de detalle o a nivel de transacción | Respuestas rápidas para información masiva                        |
| <b>Usuarios</b>             | Nivel operativo   | Nivel táctico y estratégico                                       |

Tabla 3- Comparación DB con DWH [10]

### 3.2.2 Metodologías

A continuación, se presentan las dos metodologías reconocidas para el desarrollo de un almacén de datos. Aunque ambos tienen como objetivo final lo mismo, las dos propuestas exponen diferentes formas de lograrlo.

#### 3.2.2.1 Bill Inmon

La metodología de desarrollo de un DWH creado por Bill Inmon parte de un modelo de datos corporativo. Usando este modelo, se identifican las temáticas del negocio y las entidades claves de la empresa a nivel operacional más importantes (clientes o ventas, por ejemplo). Partiendo del modelo de datos corporativo se podrá crear un modelo lógico específico para cada entidad identificada en el paso anterior. Por debajo de cada tipo de entidades se podrán identificar relaciones, claves primarias o dependencias que serán añadidas en el modelo lógico.

El aspecto clave de esta metodología, y es diferenciador con la de Ralph Kimball, es que la estructura de datos de la entidad se mantiene normalizada. De esta manera se

evita la redundancia de datos, llevando a identificar de manera más adecuada los conceptos de negocio y evitando posibles anomalías al actualizar los datos del futuro almacén. Además del diseño lógico, también en el físico se mantendrá una estructura normalizada.

Siguiendo estas pautas se crearía un *Data Warehouse* corporativo con la metodología ideada por Inmon. Gracias a este diseño normalizado, la carga de datos será menos compleja mientras que las consultas no serán simples al tener que realizar *joins* en sucesivas tablas. Para solucionar este problema en las consultas (etapa final de un proyecto BI) Inmon propone la creación de *Data Marts* específicos para cada departamento de la empresa. Estos *Data Marts* únicamente tomarán como fuente de datos las tablas del DWH. Y a diferencia de éstas, los datos en el *Data Mart* estarán sin normalizar.

Gracias a este proceso global, se asegura la integridad de los datos en la primera fase y rapidez en la creación de informes con la segunda. Esta forma de desarrollar un *data warehouse* recibe el nombre de *top-down*.

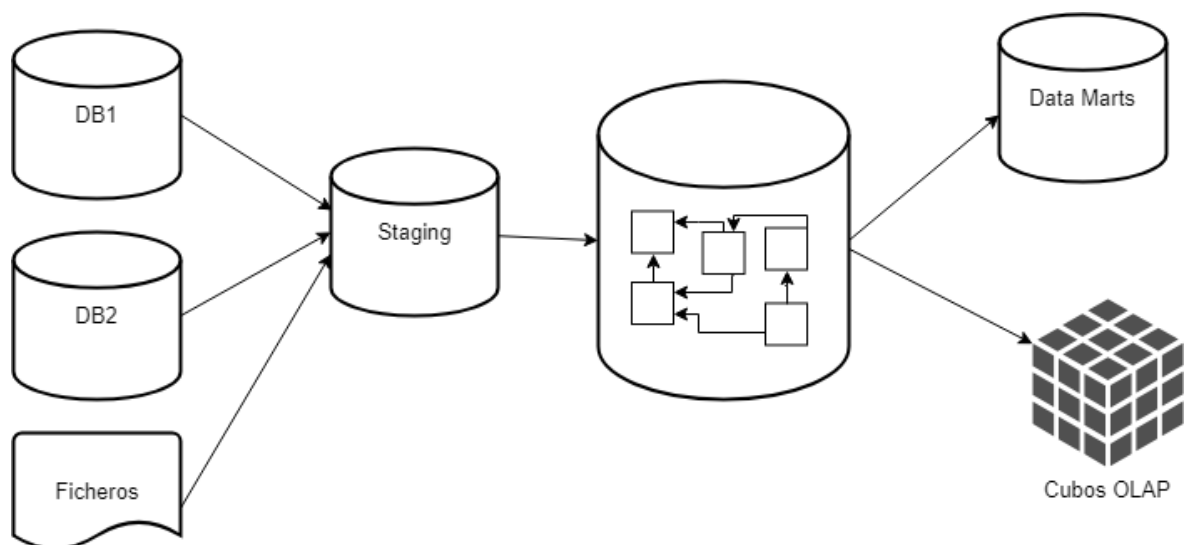


Ilustración 2- Metodología Top -Down

### 3.2.2.2 Ralph Kimball

Kimball definió otro tipo de método al establecido por Inmon. Antes de realizar el diseño físico del DWH, Kimball da importancia al desarrollo del diseño dimensional. Este diseño es la base para el desarrollo de un almacén de datos siguiendo su metodología. El método de Kimball para el desarrollo de un diseño dimensional comienza a nivel de procesos de negocio. Al identificar los procesos de negocio [11] se pueden obtener las dimensiones, que describen el contexto de los procesos. En este

paso es cuando se identifican los KPIs (*key performance indicator*), o indicador clave de rendimiento en castellano. Finalmente el resultado u *output* del proceso de negocio suele indicar la métrica correspondiente a los hechos. Kimball destaca que al pensar por departamentos o áreas de una empresa los datos que se obtienen pueden ser redundantes. Al tener una visión global, siguiendo el paradigma de procesos de negocio, obtienes la información una única vez.

Una vez definidos los procesos de negocio que se dan en la empresa y que se quieren añadir al DWH, el siguiente paso es definir la ‘granularidad’. Kimball define como granularidad de un *data warehouse* de la siguiente forma: “*el nivel de detalle que se quiere dar en una tabla de hechos*”. Este paso es fundamental al usar el método Kimball.

Tras identificar el nivel de detalle que se quiere dar en una tabla de hechos, también denominado ‘grano’ del dato, el paso posterior es identificar las dimensiones. Para identificar las dimensiones es necesario preguntarse, según Kimball, cómo las personas del negocio describen las métricas resultantes de rendimiento para cada proceso de negocio. El qué, cómo, dónde y por qué de un evento importante en un proceso de negocio.

El último paso para tener un diseño dimensional trata de la identificación de los hechos. La pregunta que lanza Kimball es: “*¿Qué está midiendo el proceso de negocio?*”. Analizando los procesos de negocio previamente identificados junto con el resto de pasos anteriores, se tendría el modelo dimensional listo. Junto con los procesos de negocio se analizan las diferentes fuentes de datos que el *Data Warehouse* puede tener. Es entonces cuando la herramienta ETL agrupa los datos de los diferentes orígenes de datos en el área de *staging*.

Una vez estén los datos listos, éstos serán cargados al modelo dimensional diseñado previamente. Es aquí donde surge la principal diferencia con el otro método expuesta en esta memoria. El modelo de *Data Warehouse* propuesto por Kimball es desnormalizado, siguiendo el modelado en Estrella (tratado en la sección 3.2.3). Una de las principales ventajas de este modelado dimensional es que el usuario final podrá hacer *drill down* sin tener que hacer uso de *joins*, uniones entre tablas, al hacer las consultas. Las operaciones de *drill down* o *drill up* son fundamentales para el análisis de datos por parte de un usuario de negocio. *Drill down*, o en castellano profundizar, significa añadir un atributo de una dimensión a la operación de agrupación en una

consulta SQL ya existente. Este parámetro de agrupación puede venir de cualquier dimensión que esté relacionada con la tabla de hechos consultada [3].

Al tener identificados diferentes procesos de negocio y sus preguntas asociadas, será necesaria la creación de múltiples estrellas que cumplan con los requisitos definidos. Para verificar la consistencia de los datos, habrá ciertas dimensiones que se consideren conformadas, dimensiones utilizadas por varias tablas de hechos, asegurando que un único registro será utilizado por todas las tablas de hechos (los conceptos de tablas de hechos y dimensiones serán tratados en el apartado 3.2.3).

La metodología de desarrollo ideada por Kimball recibe el nombre de *Bottom-Up*. Este esquema resume la estructura de un *data warehouse* siguiendo la paradigma de Kimball:

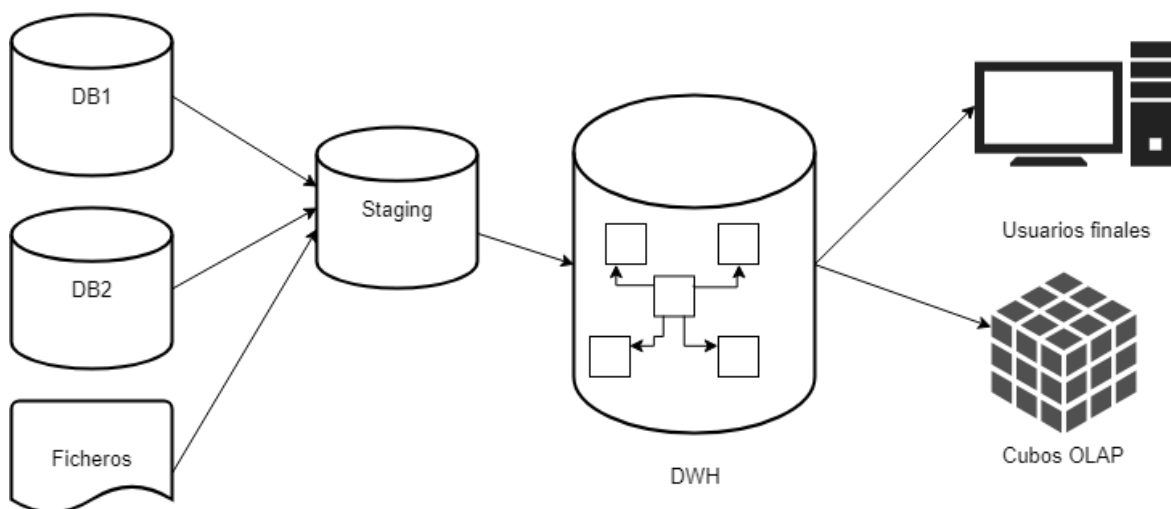


Ilustración 3- Metodología Bottom-Up

La principal diferencia entre la Ilustración 2 e Ilustración 3 se puede apreciar en la organización interna del DWH.

Una de las técnicas empleadas en el método de Kimball es la ‘*Enterprise Bus Matrix*’. Este tipo de artefacto registra, en forma de matriz, las relaciones entre los hechos (vertical) y las dimensiones (horizontal). Siguiendo los requisitos marcados por la empresa, se marcará una coincidencia cuando una tabla de hechos dependa de una

dimensión. La matriz final (Ilustración 4) servirá como documento de especificación de requisitos para el modelado en estrella.

| Business Processes     | Atomic Granularity             | Metrics                                 | Date | Product | Vendor | Contract Terms | Employee | Warehouse | Carrier |
|------------------------|--------------------------------|---|------|---------|--------|----------------|----------|-----------|---------|
| Purchase Requisitions  | 1 row per requisition line     | Requisition Quantity & Dollars          | X    | X       | X      | X              | X        |           |         |
| Purchase Orders        | 1 row per PO line              | PO Quantity & Dollars                   | X    | X       | X      | X              | X        | X         | X       |
| Shipping Notifications | 1 row per shipping notice line | Shipped Quantity                        | X    | X       | X      |                | X        | X         | X       |
| Warehouse Receipts     | 1 row per receipt line         | Received Quantity                       | X    | X       | X      |                | X        | X         | X       |
| Vendor Invoices        | 1 row per invoice line         | Invoice Quantity & Dollars              | X    | X       | X      | X              | X        | X         |         |
| Vendor Payments        | 1 row per payment              | Invoice, Discount & Net Payment Dollars | X    | X       | X      | X              |          | X         |         |

Ilustración 4- Ejemplo de Enterprise Bus Matrix [3]

### 3.2.2.3 Conclusión

Aunque las dos metodologías tienen diferencias sustanciales, ambos tienen ventajas y desventajas. Es por ello que muchos arquitectos optan por hacer un modelo híbrido. Este paradigma híbrido parte de la construcción de DWH descrita por Inmon pero desarrollando los *data marts* siguiendo el modelo en estrella.

En numerosos proyectos de BI se comienza siguiendo la metodología descrita por Kimball ya que permite desarrollarse de manera más rápida. En caso de crecimiento de la organización o del proyecto BI es conveniente evolucionar el modelado al descrito por Inmon ya que se mejoraría la eficiencia y el mantenimiento del proceso ETL.

### 3.2.3 Modelo en Estrella

El modelo en estrella es un tipo de diseño de bases de datos orientado a un *data warehouse*. Su principal característica es la distinción entre dimensiones y hechos seguido en el método de Kimball, ya que proporciona una visión de alto nivel de los diferentes elementos que se almacenan en una base de datos dimensional. El resultado del diseño dimensional dividirá los datos en dos tipos: dimensiones y hechos.

Las tablas de hechos, *fact tables* en inglés, son los datos que serán utilizados para análisis. Representan las medidas del negocio que se quieren analizar para tomar las decisiones. Por el contrario, las dimensiones, o *dimensions* en inglés representan aspectos de interés, de los cuales se podrá filtrar por ellos. Su principal función es

proveer un contexto lógico de negocio, cosa que las tablas de hechos no hacen. Las *Fact tables* están conformadas de datos cuantitativos y de claves foráneas de las dimensiones relacionadas. La Ilustración 5 muestra cómo al estar las tablas de hechos formadas de datos cuantitativos y de claves foráneas de las dimensiones relacionadas, el resultante diagrama queda en forma de estrella.

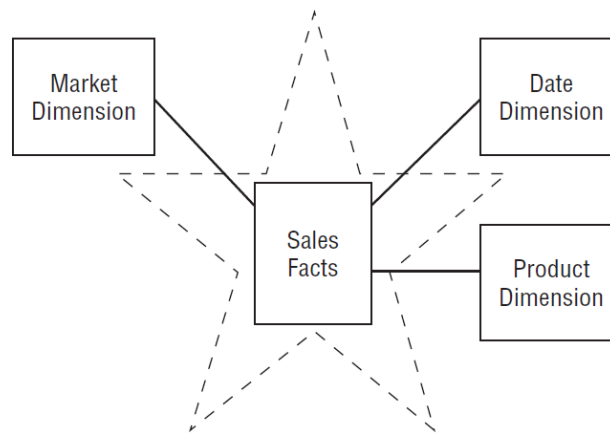


Ilustración 5- Ejemplo de Modelo en Estrella [3]

En cuanto a las tablas de hechos, se pueden destacar varios tipos:

- **Transaccionales:** Un evento concreto por fila en la tabla de hechos.
- **Instantánea Periódica:** Con múltiples filas, se toma una instantánea al final de un periodo de tiempo determinado.
- **Instantánea Acumulativa:** Múltiples filas pero con modificación del estado de una fila durante un periodo de tiempo.

Dentro de las posibles opciones de diseño para las tablas de dimensiones, existen un tipo denominadas lentamente cambiantes o SCD, del inglés *slowly changing dimension*. Las dimensiones SCD pueden ser de diferentes tipos:

- **Tipo 1:** Las filas de las dimensiones se sobrescriben en caso de cambio.
- **Tipo 2:** Se añade una nueva línea en caso de que tener que actualizar un registro previamente insertado.
- **Tipo 3:** Se incluye una columna que conserve el valor original en el anterior estado.
- **Tipo 4:** Se parte de las SCD de tipo 1 pero registrando en una tabla auxiliar los registros históricos.
- **Tipo 6:** Se denominan de tipo 6 al ser una combinación de los tipos 1, 2 y 3.

Aunque para construir un *data warehouse* es necesario desnormalizar todos los datos que partan como fuente, es posible que para cumplir con los requisitos del diseño del DWH sea necesario mantener cierta normalización. Es por ello que existe el concepto de Copo de Nieve o *snowflake* que normaliza la jerarquía de ciertas dimensiones para reducir la redundancia de datos. Esta solución es válida para el método de Kimball. La Ilustración 6 muestra una comparación entre un modelado en estrella y la variante en copo de nieve.

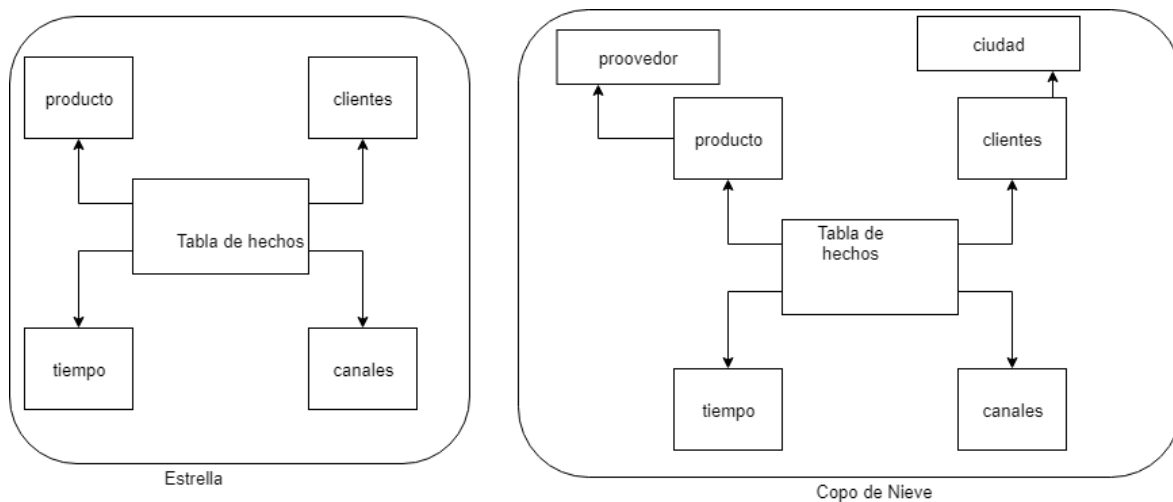


Ilustración 6- Comparación Estrella vs Copo de Nieve

### 3.2.4 OLAP vs OLTP

#### 3.2.4.1 OLTP

*Online Transactional Processing* son bases de datos relacionales orientadas a transacciones. Una transacción, dentro del contexto de gestores de bases de datos (en adelante RDBMS), es una secuencia de operaciones puestas en funcionamiento por una base de datos de forma atómica. En un RDBMS, las operaciones básicas son de cuatro tipos: insertar, actualizar, eliminar, y seleccionar. En cuanto al estado e integridad de los datos, todas estas operaciones son transaccionales, por lo que si una operación se ha finalizado se denomina *commit* y si en caso contrario *rollback* (al haber fallo, se vuelve al estado antes de la transacción).

Para que una base de datos relacional pueda asegurar un estado correcto se deben cumplir las siguientes características en cada transacción realizada sobre ella:

- **Atomicidad:** de esta forma se cumple que una transacción se ha realizado completamente, despejando la posibilidad de que no se realice de forma

completa.

- **Consistencia:** cumplir con la integración de la base de datos. Todas las transacciones que se realicen deben respetar las reglas definidas previamente.
- **Aislamiento:** dos operaciones que estén dirigidas a la misma información en una base de datos nunca generarán errores o condiciones de *deadlock* ya que trabajarán de forma independiente.
- **Durabilidad:** Las operaciones una vez finalizadas serán almacenadas en la base de datos de forma persistente.

### 3.2.4.2 OLAP

**OLAP** son las siglas de *On-Line Analytical Processing* y se definen como bases de datos destinadas al procesamiento analítico. Este tipo de bases de datos están diseñadas para soportar análisis de grandes cantidades de datos que aporten información de valor para el negocio.

Los cubos OLAP [12], que parten de un diseño en estrella, son vectores de varias dimensiones que son almacenados en una base de datos multidimensional. Siguiendo el paradigma relacional, un cubo es una tabla de hechos que tiene como columnas indicadores o dimensiones. Los indicadores suelen aportar información cuantitativa mientras que las columnas de tipo dimensional son características cualitativas. Este segundo tipo de columna suele tener claves foráneas que apuntan a las tablas de dimensión.

Aunque generalmente los cubos OLAP suelen estar almacenados en bases de datos multidimensionales, dependiendo del tipo de persistencia se pueden definir dos tipos de OLAP:

- **ROLAP:** Utilización de sistemas de administración de bases de datos relacionales, RDBMS, para almacenar los cubos siguiendo el modelado en estrella, o su variación en forma de copo de nieve, y usando tablas con datos agregados. Las consultas que ejecuten los usuarios en una herramienta de *reporting BI* serán transformadas en SQL para ser respondidas por bases de datos relacionales.
- **MOLAP:** Tecnología OLAP utilizando como persistencia bases de datos multidimensionales, en adelante MDDB. La carga de datos en el MDDB se

realiza por fracciones, de forma periódica. Los datos agregados se calculan una vez los datos elementales han sido almacenados en la base de datos multidimensional.

- **HOLAP:** Tecnología recién desarrollada que busca una solución híbrida. Esta tecnología almacena en sistemas RDBMS los registros de detalle y un MDDDB los datos agregados.

La Tabla 4 resume las ventajas y desventajas que ofrecen las soluciones ROLAP y MOLAP:

|                              | ROLAP   | MOLAP  |
|------------------------------|---|--|
| <b>Cambios en el diseño</b>  | Cubos generados automáticamente al ser lanzadas las consultas.                          | En caso de ser requerido un cambio en algún cubo, éste tiene que ser recalculado para que los cambios estén reflejados en las consultas. |
| <b>Velocidad de consulta</b> | Cada vez que una consulta es lanzada los datos del cubo ROLAP tienen que ser calculados | Los datos están precalculados, por lo que en tiempo de ejecución las consultas son más rápidas.  |

Tabla 4- Comparación ROLAP y MOLAP

### 3.2.5 Procesos ETL

Son los procesos que engloban la extracción de datos de múltiples fuentes para, tras una transformación posterior, ser almacenados en un *data warehouse*. El nombre de ETL viene de las palabras en inglés: *extract*, *transform* y *load* (extracción, transformación y carga). Una vez los datos corporativos han sido extraídos de múltiples fuentes, éstos quedan almacenados siguiendo el modelo de negocio definido para el DWH. Las etapas mencionadas anteriormente consisten en lo siguiente:

- **Extraer datos:** Esta información puede estar en diferentes formatos y sistemas de origen.
- **Transformación de los datos:** Primeramente se debe comprobar la calidad del dato, limpiarlos, normalización de las métricas y unificación de conceptos en

las diferentes fuentes de datos. En esta fase se aplican las reglas de negocio previamente definidas para el almacén de datos.

- **Carga de los datos:** Por último, los datos son almacenados en el DWH siguiendo el modelo en estrella diseñado previamente, utilizando tablas de hechos y dimensiones.

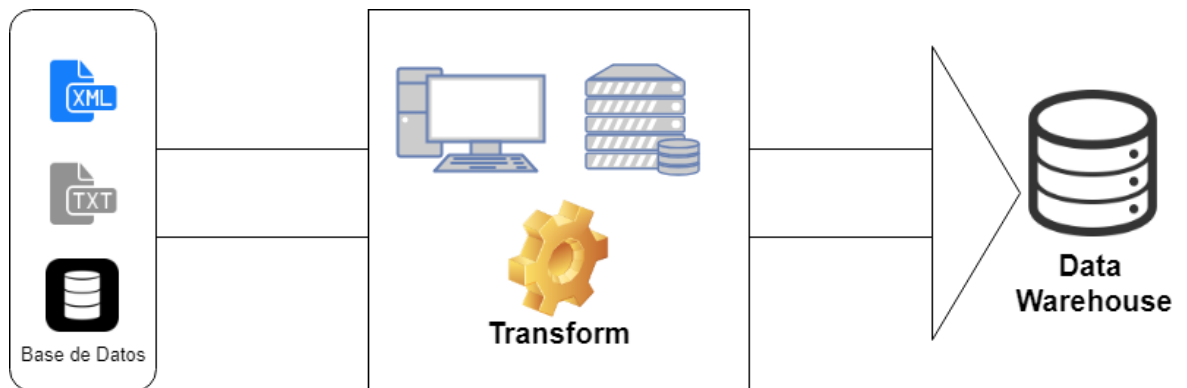


Ilustración 7- Resumen del proceso ETL [8]

### 3.3 Tecnología BI

Para el desarrollo de un proyecto *Bussines Intelligence* es conveniente el uso de un software genérico que provea funcionalidad para prácticamente todas las fases de un proyecto BI. En su defecto, será necesario recurrir a varios que cubran distintas funcionalidades o necesidades. Las herramientas específicas se pueden diferenciar en función de la etapa en la que tengan utilidad:

- Herramientas genéricas para BI.
- Herramientas para ETL.
- Herramientas para generación de consultas e informes.
- Herramientas para la generación de cuadros de mando.
- Sistema de Soporte a la Decisión.
- Sistema de Información Ejecutivo.

#### 3.3.1 Ejemplos

- **Pentaho:** Esta herramienta genérica para Inteligencia de Negocio contiene diversas utilidades que cubren todas las etapas de un proyecto BI. Desde herramientas para minería de datos o ETL, hasta software para generar

informes [13].

- **PowerCenter:** Es una de las herramientas más utilizadas para el proceso ETL según el informe de Gartner [14]. La Ilustración 8 muestra un ejemplo del software.

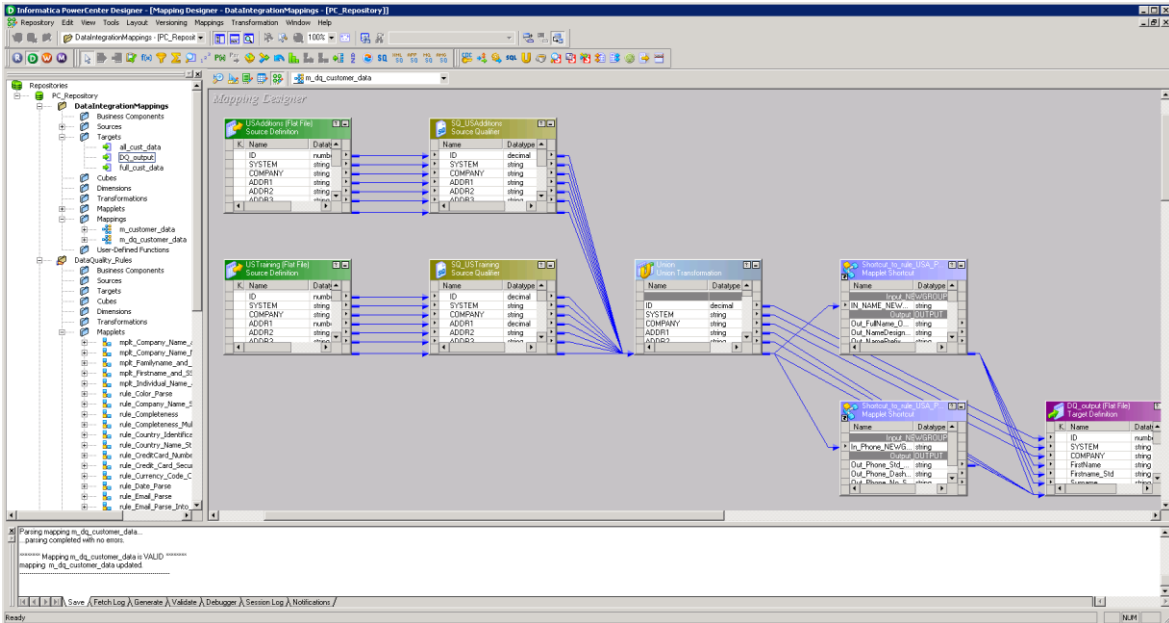


Ilustración 8- Mapping con PowerCenter [15]

- **BIRT:** Es una herramienta de *reporting* de código libre desarrollada por la fundación de Eclipse. Los informes que permite crear esta herramienta pueden ser integrados fácilmente en aplicaciones web.
- **Oracle:** Este fabricante ofrece herramientas específicas para la integración de datos y para *reporting*. Debido a que el TFG se realiza bajo el convenio FORTE con Avanttic, las herramientas de Oracle han sido las seleccionadas para el desarrollo del *Data Warehouse*, de los informes y cuadros de mando. El software correspondiente será detallado en la sección 4.3.2.

## 4 MÉTODO DE TRABAJO

En este capítulo se presentan las metodologías de trabajo que se han usado para la gestión y desarrollo de este proyecto. También se presentan las tecnologías que han sido utilizadas.

### 4.1 Gestión del Proyecto

En cuanto al método de gestión del proyecto, se ha utilizado SCRUM. Esta metodología ágil es adecuada para el contexto del TFG, ya que proporciona buenas prácticas basadas en realizar pequeñas iteraciones, y donde el contacto entre las partes interesadas se da al final de las mismas. Dadas las características de este TFG, realizado en *Avanttic Consultoría Tecnológica*, se ha utilizado una adaptación de SCRUM donde se encuentran los siguientes roles:

- **Product Owner:** Representa al interesado del producto. Es la persona encargada de lograr el mayor valor de producto para los clientes, usuarios y el resto de implicados. En este TFG el encargado de representar este rol ha sido Noel García Gómez (Tutor por parte de la empresa en este proyecto).
- **Scrum Manager:** Responsable del funcionamiento de SCRUM en la organización. En este caso es el autor, Miguel Ampuero López-Sepúlveda, quien ha ejercido este rol.
- **Scrum Team – Equipo de desarrollo:** Responsables de transformar la pila de cada iteración en un incremento. Es decir, son los encargados de desarrollar el sistema. En este caso solo habrá un miembro, Miguel Ampuero (autor del TFG).

#### 4.1.1 Artefactos de SCRUM

Dentro de la adaptación de SCRUM al TFG, se destacan los siguientes artefactos:

- **Sprint:** Nombre que recibe las iteraciones. La duración de los mismos oscila entre 1 a 4 semanas. Al final de cada iteración o *sprint*, se debería obtener un incremento sobre el producto.

- **Historia de Usuario:** Definición de una funcionalidad que debe cumplir el producto resultante.
- **Pila de Producto:** Conjunto de historias de usuario que en conjunto definen todos los requisitos del proyecto por parte del *product owner*.
- **Pila de Sprint:** Subconjunto de Pila de producto que definen las historias de usuario que van a ser desarrolladas por el *SCRUM Team* en un *sprint* dado.
- **Reuniones:** Dentro del ciclo de vida del proyecto se dan diferentes reuniones:
  - **Reunión de planificación:** En la que están presentes todos los roles definidos previamente en esta memoria.
  - **Reunión diaria:** Participa el *SCRUM Team*, realizándose al inicio de cada jornada.
  - **Reunión de retrospectiva:** Se realiza con todos los roles identificados previamente. Este encuentro se realiza al final de cada *sprint* para revisar que las historias de usuario incluidas en la pila de *sprint* han sido cumplidas, y por tanto se obtiene un incremento sobre el producto.
- **Gráficas de resultado (*Burn-up o Burn-down*):** Gráficas que representan el avance del desarrollo sobre el tiempo. En este TFG se usarán gráficos de tipo *burn-up* (estimación inicial vs estimación real).

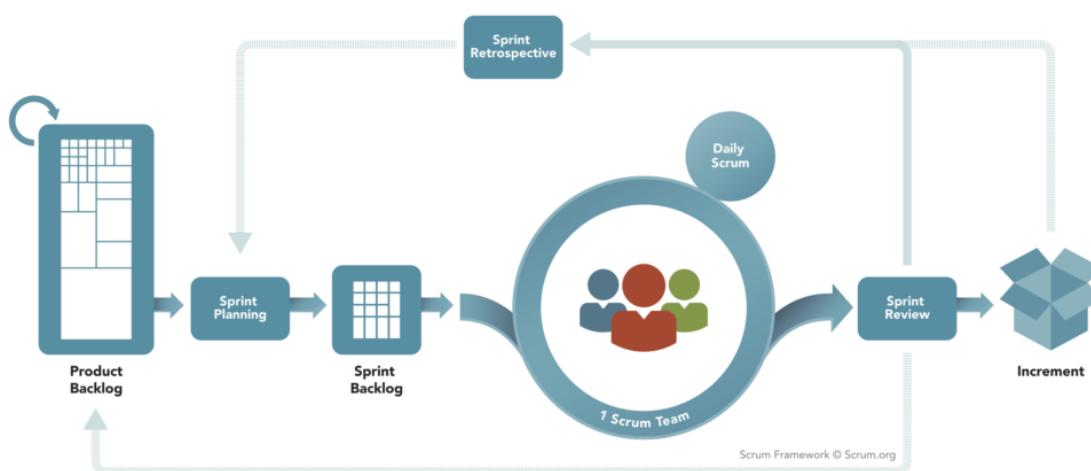


Ilustración 9 - Resumen de la Metodología de Gestión Scrum [16]

## 4.2 Desarrollo del Almacén de Datos

Para desarrollar el DWH se ha seguido la metodología propuesta en la asignatura Ingeniería de Negocio (en adelante INEG) de la Escuela Superior de Informática de Ciudad Real [8], que a su vez es una simplificación de la propuesta por Ralph Kimball, denominada *Business Dimensional Lifecycle* [3] y que ha sido detallada en la sección 3.2.2.2. La Ilustración 10 muestra el ciclo de vida que debe tener un proyecto para desarrollar un DWH y el uso de una aplicación BI. Aunque todas estas fases han sido completadas en el TFG, la metodología utilizada simplifica lo propuesto por Kimball en 5 etapas. La Ilustración 11 muestra las fases que tiene un proyecto de BI siguiendo la metodología propuesta en la asignatura Ingeniería de Negocio.

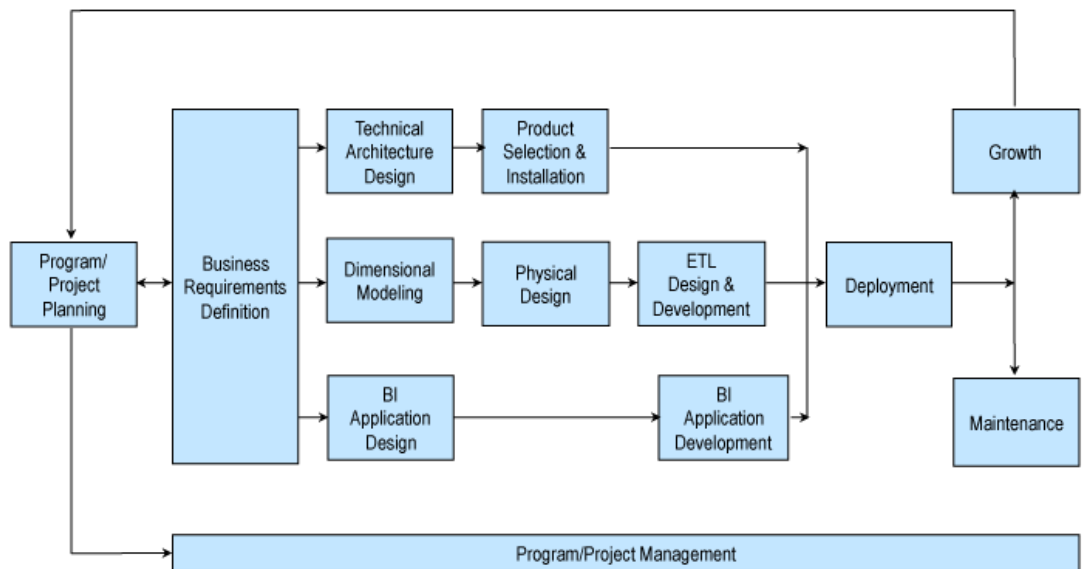


Ilustración 10 - Ciclo de Vida de DWH/BI de Kimball [17]

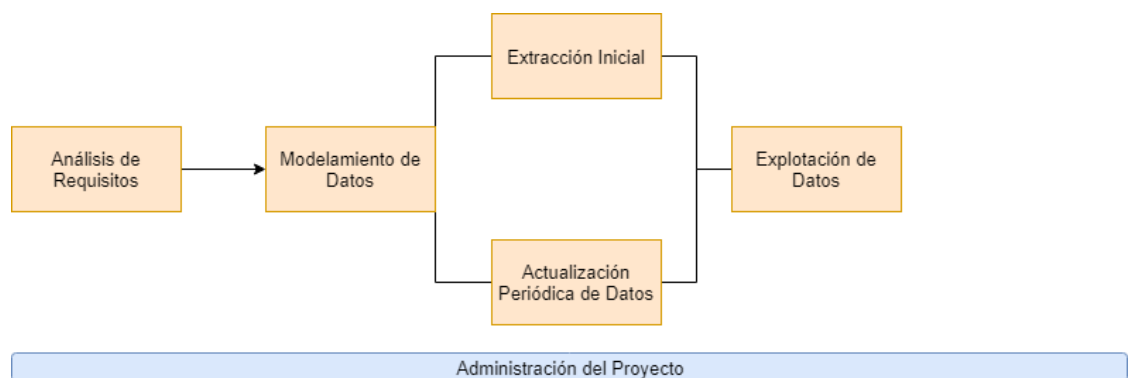


Ilustración 11 - Metodología para un Proyecto BI [8]

Así, el plan del proyecto se ha realizado en 5 fases:

1. **Análisis de Requisitos:** Se ha hecho una captura de requisitos de usuarios de la aplicación final (representantes de departamento en Avanttic Consultoría Informática S.L.). Para un correcto entendimiento de los datos que son útiles para los usuarios, se hará uso de la matriz de arquitectura bus desarrollada por Kimball [3].
2. **Modelamiento de Datos:** Tras el análisis inicial de necesidades se realizó una transformación de los datos recogidos en el ERP siguiendo un diseño de almacén de datos multidimensional, propuesta por Kimball [3], con la técnica de esquema en estrella o copo de nieve.
3. **Extracción Inicial de Datos:** Para esta tercera fase se necesita tener acceso a las fuentes de los datos y tener realizado el modelado previamente. Para esta etapa se ha utilizado la herramienta Oracle Data Integrator [6].
4. **Actualización Periódica de Datos:** Utilizando la misma herramienta que en la fase 3, se programaron las cargas periódicas al DWH.
5. **Explotación de Datos:** Por último, se han creado Cuadros de Mando e informes (fase final de un proyecto BI) utilizando la herramienta OBI [7].

Dadas estas 5 fases obtenidas tras aplicar la adaptación de Ralph Kimball, el Trabajo Fin de Grado ha sido planteado en 5 Sprints. Correspondiendo con las fases de desarrollo de la siguiente manera:

| Sprint | Fase de Desarrollo                              |
|--------|---|
| 0      | Planificación Inicial                           |
| 1      | Análisis de Requisitos y Modelamiento de Datos. |
| 2      | Extracción Inicial de los Datos.                |
| 3      | Extracción Periódica de los Datos.              |
| 4      | Explotación de Datos.                           |

Tabla 5- Correspondencia Sprints con Fases de Desarrollo de un DWH.

### 4.3 Tecnologías usadas en el proyecto

En esta sección se especifican las tecnologías utilizadas agrupadas por su utilidad.

#### 4.3.1 Marco tecnológico

La Figura 2 resume las herramientas que han sido utilizadas en el TFG:



Figura 2 - Resumen Marco Tecnológico

#### 4.3.2 Herramientas para el desarrollo del proyecto

A continuación se detallan las herramientas que se han utilizado para el desarrollo tecnológico del proyecto.

##### 4.3.2.1 Oracle Data Integrator 12c

ODI 12c es una de las herramientas principales de este TFG, ya que es donde se han implementado las cargas iniciales y la automatización de las cargas periódicas para la población de datos en el DWH.

Oracle Data Integrator es una plataforma de integración de datos que permite aplicar procesos ETL a grandes volúmenes de datos manteniendo un alto rendimiento. A diferencia de otras herramientas del sector, Oracle Data Integrator [6] (también conocida como ODI) realiza el proceso ETL en diferente orden. Oracle define el

proceso como E-LT, primero extrae los datos y los carga en destino, ya en destino es donde realiza las operaciones de transformación.

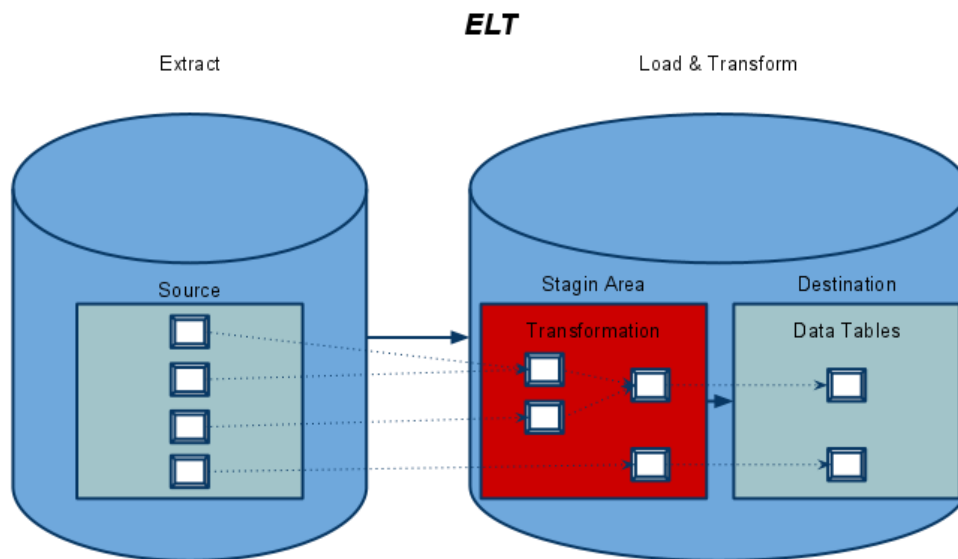


Ilustración 12 - Resumen del proceso ELT [18]

#### 4.3.2.2 Oracle Bussines Intelligence 12c

Esta tecnología ha sido utilizada para la explotación de los datos que previamente han sido insertados en el DWH. Todos los análisis y cuadros de mando resultantes de este TFG han sido desarrollados con esta herramienta.

Oracle Bussines Intelligence [7] es una herramienta visual de Inteligencia de Negocio desarrollada por Oracle. Esta plataforma de BI permite desarrollar análisis y presentación de datos de forma atractiva para su posterior explotación en cuadros de mando.

#### 4.3.2.3 Oracle Data Visualization

Oracle Data Visualizaton [19] Herramienta de análisis de datos visual desarrollada por Oracle. Este software se encuentra disponible dentro de la suite de Oracle Bussines Intelligence 12c Enterprise Edition (también conocida como OBIEE12c). A diferencia de OBIEE12c, Oracle Data Visulization permite al usuario final crear análisis a partir de un DWH o de archivos con formato .csv o .xls.

Esta solución de Oracle ha sido utilizada para la creación de análisis que posteriormente quedan integrados en cuadros de mando de OBIEE12c.

#### **4.3.2.4 PL/SQL**

Para este proyecto ha sido utilizado en la creación de la dimensión fecha en el DWH. Además se han desarrollado ciertos procedimientos que han servido a lo largo del desarrollo de la fase de exportación inicial.

*Procedural Language / Structured Query Language*, también denominado PL/SQL [20] es un lenguaje de programación definido para Bases de Datos Oracle que permite crear procedimientos estructurados.

#### **4.3.2.5 Oracle Database 12c**

Oracle Database 12c ha sido utilizado para el alojamiento de la capa de *Staging* y del *Data Warehouse*.

Oracle Database 12c [21] es un sistema de gestión de base de datos de tipo objeto-relacional, también denominado *ORDBMS*, que fue desarrollado por Oracle Corporation. La versión 12c es la última que ha lanzado el fabricante, destacando el soporte multiplataforma y la escalabilidad.

#### **4.3.2.6 SQL Developer**

Esta herramienta ha sido utilizada para la creación de los esquemas de bases de datos que han sido requeridos en el proyecto.

SQL Developer [22] es un IDE desarrollado por Oracle para el desarrollo de Bases de Datos Oracle trabajando con SQL. Gracias a su interfaz gráfica, este software permite de manera visual editar sentencias SQL, procedimientos, y la creación de objetos de bases de datos.

#### **4.3.2.7 SQL Developer Data Modeler**

Este software ha sido utilizado para la creación del modelo lógico y relacional de los distintos esquemas creados en el TFG.

SQL Developer Data Modeler [23] Herramienta gráfica desarrollada por Oracle que permite de manera visual la creación de modelos relacionales para base de datos. Esta herramienta permite la coordinación con SQL Developer, pudiendo exportar un modelo relacional a un nuevo esquema de base de datos Oracle.

### **4.3.3 Herramientas para la gestión del proyecto**

Para gestionar el proyecto, siguiendo la metodología resumida en este mismo capítulo, se ha hecho uso de las siguientes herramientas.

#### **4.3.3.1 Trello**

Para este TFG Trello [24] ha sido utilizado como herramienta de soporte para la gestión del proyecto ya que se adapta a la metodología ágil utilizada. Trello, es una herramienta *online* de gestión de proyectos que permite crear tableros para la gestión de tareas, de forma individual o cooperativa.

#### **4.3.3.2 Microsoft Excel**

Esta herramienta ha sido utilizada para anotar el tiempo dedicado en cada sprint del proyecto y la posterior creación de gráficas *Burn-up* al final de cada Sprint. Microsoft Excel [25] es un programa incluido dentro del paquete Microsoft Office que permite la creación de hojas de cálculo.

### **4.3.4 Herramientas para el desarrollo de la memoria**

Para la preparación, almacenamiento y diseño de material relacionado con la memoria del Trabajo Fin de Grado se han sido utilizadas las herramientas listadas.

#### **4.3.4.1 Dropbox**

Dropbox [26] ha sido utilizado en este proyecto como servidor de alojamiento de todos los ficheros relacionados con el TFG además de ser compartidos con el tutor académico del mismo. Dropbox es un servicio de alojamiento de archivos en la nube. Este servicio permite a los usuarios alojar ficheros en línea estando disponibles en otras plataformas y pudiendo ser compartidos.

#### **4.3.4.2 Draw.io**

Draw.io ha sido utilizada en el TFG para la creación de todos los diagramas y figuras mostradas a lo largo de la memoria. Esta herramienta desarrollada por Google permite la creación de todo tipo de diagramas, figuras o esquemas de manera *online* pudiéndose sincronizar con servicios de alojamiento en la nube.

#### **4.3.4.3 Microsoft Word**

Procesador de texto seleccionado para el desarrollo de la memoria del TFG. Microsoft Word [27] es un procesador de texto que se encuentra disponible en el paquete Microsoft Office propiedad de Microsoft.

#### **4.3.4.4 Mendeley**

En este TFG, Mendeley [28] ha sido requerido para la notación de las referencias de la memoria. Es una herramienta de gestión de referencias bibliográficas. Este software desarrollado por Mendeley Ltd. permite la gestión de citas bibliográficas de diferentes fuentes y la integración de éstas en procesadores de texto.



## 5 RESULTADOS

Este capítulo muestra los resultados obtenidos tras aplicar los métodos y tecnologías detallados en el capítulo 4, con el fin de cumplir los objetivos definidos en el capítulo 2.

### 5.1 Sprint 0: Preparación Inicial

Este primer Sprint se considera como fase previa para la preparación del proyecto. El Sprint 0 tiene como objetivo principal la identificación de las historias de usuario que van a formar parte de la pila de producto. Se ha realizado la estimación de las historias de usuario y un plan de proyecto. Como última tarea se ha detallado un plan de gestión de riesgos.

#### 5.1.1 Tareas

En el Sprint 0 se van a realizar las tareas resumidas en la Tabla 6. Estas tareas no pertenecen a la pila de producto ya que este Sprint es considerado una fase previa al desarrollo. Las tareas correspondientes son:

| SPRINT                    | LISTA DE TAREAS                              | HORAS EST. |
|---------------------------|--|------------|
| 0                         | T1: Definición del Proyecto                  | 8          |
|                           | T2: Listar Historias de Usuario              | 8          |
|                           | T3: Estimación de las Historias de Usuario   | 5          |
|                           | T4: Desarrollo de Plan de Gestión de Riesgos | 8          |
|                           | T5: Desarrollo del Anteproyecto              | 21         |
| TOTAL DE HORAS ESTIMADAS: |  | 50         |

Tabla 6 - Planificación del Sprint 0

Para la tarea T1 se puede encontrar en la sección Definición del proyecto 5.1.2.1 donde se detalla la definición del proyecto. Las tareas 2 y 3 han dado como resultado final la Pila de Producto, disponible en la sección 5.1.2.2 además del plan de proyecto inicial del TFG. En cuanto a la tarea 4, el plan de gestión de riesgos se puede

encontrar en la sección 5.1.2.4. Por último, se ha añadido como tarea la redacción del anteproyecto, como parte inicial del TFG.

## 5.1.2 Resultados del Sprint

### 5.1.2.1 Definición del proyecto

Durante la primera semana de prácticas bajo el convenio FORTE, se han dado diversas reuniones con el propietario del producto. Gracias a estas primeras reuniones se ha podido definir el alcance del proyecto, las herramientas que se van a utilizar además de la principal fuente de datos para el DWH.

Aunque este Sprint 0 se considera como fase previa al desarrollo del proyecto, en el Sprint 1 se realizarán de nuevo entrevistas específicas a los responsables del departamento de producción de la empresa. Estas segundas reuniones forman del proceso de creación de un *Data Warehouse* como parte de análisis de las necesidades del negocio, siguiendo la metodología de desarrollo explicada en la sección 4.2.

### 5.1.2.2 Pila de Producto

Tras las primeras reuniones con las partes interesadas del proyecto, se han obtenido las siguientes historias de usuario que serán añadidas a la Pila de Producto. Aunque hay historias de usuario que tienen un nivel de prioridad mayor, se van a desarrollar en el orden que marca la metodología de desarrollo utilizada en el TFG, al estar utilizando una adaptación de SCRUM en el proyecto, la escala de prioridad seleccionada va de 0 a 1000.

| HU1 Modelado Dimensional   |                           |
|--|---------------------------|
| <b>Prioridad de Negocio:</b> 600   | <b>Sprint Asignado:</b> 1 |
| <b>Responsable:</b> Miguel Ampuero López-Sepúlveda   |                           |
| <b>Descripción:</b><br>Generar, tras reuniones con los usuarios finales, un modelado dimensional de los datos que serán almacenados en el <i>Data Warehouse</i> y usados en la aplicación BI para hacer consultas. |                           |
| <b>Precondición:</b><br>No hay ninguna precondición.   |                           |

**Postcondición:**

Tras la consecución de la historia de usuario, el modelo debe ser consistente y poder cubrir las necesidades impuestas por el usuario final.

**Actividades:**

- Entrevistas con Responsable de Producción.
- Documentación de las entrevistas.
- Identificación de Dimensiones y Hechos.
- Creación de *Enterprise Bus Matrix*.
- Desarrollo del Modelo en Estrella.

**Artefactos generados:**

- *Enterprise Bus Matrix* correspondiente a lo hablado con los usuarios finales.
- Modelo de datos del DWH.

Tabla 7- Historia de Usuario 1

| HU2 Desarrollo de Cargas Iniciales  |                           |
|---|---------------------------|
| <b>Prioridad de Negocio:</b> 750  | <b>Sprint Asignado:</b> 2 |
| <b>Responsable:</b> Miguel Ampuero López-Sepúlveda  |                           |
| <b>Descripción:</b><br>Tener en el <i>Data Warehouse</i> disponibles los datos actuales de la base de datos de origen.  |                           |
| <b>Precondición:</b><br>Modelo dimensional del DWH.   |                           |
| <b>Postcondición:</b><br>Tras la consecución de esta historia de usuario, el <i>product owner</i> puede consultar los datos actuales en el DWH siguiendo el modelo de datos definido en la HU1.   |                           |
| <b>Actividades:</b> <ul style="list-style-type: none"> <li>• Diseño Físico del DWH con SQL Data Modeler.</li> <li>• Exportación del diseño físico a SQL Developer.</li> <li>• Diseño Físico de la capa de <i>Staging</i> con SQL Data Modeler.</li> <li>• Exportación del modelo de <i>Staging</i> a SQL Developer.</li> <li>• Creación de Repositorio en ODI.</li> <li>• Diseño las cargas de <i>Staging</i>.</li> </ul> |                           |

- Creación de Plan de Carga para *Staging*.
- Diseño de los mapeos para dimensiones y hechos del DWH.
- Creación del Plan de Carga inicial del DWH.
- Ejecución de los planes de carga.

**Artefacto generado:**

- *Data Warehouse* con los datos iniciales.

Tabla 8 - Historia de Usuario 2

| HU3 Configuración de Cargas Periódicas   |                           |
|--|---------------------------|
| <b>Prioridad de Negocio:</b> 750   | <b>Sprint Asignado:</b> 3 |
| <b>Responsable:</b> Miguel Ampuero López-Sepúlveda   |                           |
| <b>Descripción:</b><br>La herramienta para la creación y gestión de los procesos ETL debe ejecutar cargas al DWH de forma periódica.   |                           |
| <b>Precondición:</b><br>Tener la Historia de Usuario 2 completada.   |                           |
| <b>Postcondición:</b><br>Tras la consecución de esta historia de usuario, el <i>product owner</i> podrá comprobar como de forma periódica el DWH almacena nuevos registros. Será notificado por email cuando una carga se ha realizado correctamente.  |                           |
| <b>Actividades:</b> <ul style="list-style-type: none"> <li>• Modificación de los mapeos de las dimensiones y hechos.</li> <li>• Instalación de Agente <i>Standalone</i> de ODI.</li> <li>• Configuración de agente en proyecto de ODI.</li> <li>• Creación de paquete para configuración del servicio de notificación email.</li> <li>• Creación de tablas auxiliares DWH.</li> <li>• Configuración de las tablas auxiliares con los planes de carga.</li> </ul> |                           |
| <b>Artefacto o producto generado:</b> <ul style="list-style-type: none"> <li>• El <i>Data Warehouse</i> almacenará de forma periódica los nuevos datos insertados en las fuentes.</li> <li>• Tablas de auditoría para consultar posibles errores durante las cargas.</li> <li>• Servicio de notificación por mail.</li> </ul>  |                           |

Tabla 9 - Historia de Usuario 3

| HU4  |  | Preparación Entorno BI    |  |
|--|--|---------------------------|--|
| <b>Prioridad de Negocio:</b> 800   |  | <b>Sprint Asignado:</b> 4 |  |
| <b>Responsable:</b> Miguel Ampuero López-Sepúlveda   |  |                           |  |
| <b>Descripción:</b><br>Instalación de la infraestructura y configuración de Oracle Business Intelligence.  |  |                           |  |
| <b>Precondición:</b><br>Tener disponible el <i>Data Warehouse</i> poblado de datos.  |  |                           |  |
| <b>Postcondición:</b><br>Tras la consecución de esta historia de usuario, el <i>product owner</i> puede ver como las herramientas de Oracle para el análisis de datos están instaladas y tienen acceso a los datos (sin tener creados informes o cuadros de mando).  |  |                           |  |
| <b>Actividades:</b> <ul style="list-style-type: none"> <li>• Instalación de WebLogic Server.</li> <li>• Instalación de Oracle Bussines Intelligence.</li> <li>• Configuración de OBI con WebLogic.</li> <li>• Creación de repositorio para Oracle BI Server.</li> <li>• Conexión desde el repositorio al DWH.</li> </ul> |  |                           |  |
| <b>Artefacto o producto generado:</b> <ul style="list-style-type: none"> <li>• <i>Oracle Business Intelligence</i> instalado en el entorno del cliente.</li> <li>• Acceso a los datos del <i>Data Warehouse</i> desde <i>Oracle Business Intelligence</i>.</li> </ul>  |  |                           |  |

Tabla 10 - Historia de Usuario 4

### 5.1.2.3 Plan de Proyecto

Una vez reunidas las partes interesadas en este proyecto (equipo de desarrollo, propietario de producto y *SCRUM Master*) se han agrupado las historias de usuario en los diferentes sprints de desarrollo. El resultante Plan de Proyecto se resume en la Tabla 11.

| Sprint                    | HU | Tareas   | Duración<br>(Semanas) |
|---------------------------|----|--|-----------------------|
| 0                         | -  | T1: Definición del proyecto.<br>T2: Listar historias de usuario.<br>T3: Estimación de las historias de usuario.<br>T4: Desarrollo del plan de gestión de riesgos.<br>T5: Desarrollo del Anteproyecto.  | 2                     |
| 1                         | 1  | T1: Entrevistas con Responsable de Producción.<br>T2: Identificación de Dimensiones y Hechos.<br>T3: Creación de un <i>Enterprise Bus Matrix</i> .<br>T4: Desarrollo del Modelo en Estrella.   | 2                     |
| 2                         | 2  | T1: Creación de esquemas del DWH.<br>T2: Desarrollo de las cargas de la BBDD de origen a <i>Staging</i> .<br>T3: Desarrollo de los <i>mapping</i> iniciales del DWH.   | 4                     |
| 3                         | 3  | T1: Modificación de los <i>mapping</i> para ser ejecutados periódicamente.<br>T2: Instalación de Agente <i>Standalone</i> de ODI.<br>T3: Configuración de servicio por notificación de Email.<br>T4: Creación de tablas para generar auditorías en las cargas del DWH. | 4                     |
| 4                         | 4  | T1: Instalación de Infraestructura de Oracle para OBI.<br>T2: Instalación de Oracle Business Intelligence.<br>T3: Importación del modelo de datos del DWH para creación de repositorio de BI.  | 2                     |
| TOTAL DURACIÓN (SEMANAS): |    |  | 14                    |

Tabla 11- Plan de Proyecto en Sprint 0

#### 5.1.2.4 Plan de Gestión de Riesgos

Para la creación de un plan de gestión de riesgos se ha recurrido a una lista con los riesgos más comunes en proyectos software [29]. Tras un análisis de los 111 riesgos detallados en la lista, se ha hecho una selección de los que potencialmente pueden afectar al proyecto. La Tabla 12 muestra el resultado final del análisis:

| Riesgo   | Probabilidad | Impacto en días |
|--|--------------|-----------------|
| <b>A: Elaboración de la Planificación</b>  |              |                 |
| <b>A2: Planificación optimista, <i>mejor caso</i> (en lugar de realista, <i>caso esperado</i>).</b>          | 30%          | 4               |
| <b>A11: Un retraso en una tarea produce retrasos en cascada en las tareas dependientes.</b>                  | 60%          | 6               |
| <b>B: Organización y Gestión</b>   |              |                 |
| <b>B2: El proyecto languidece demasiado en el inicio difuso.</b>   | 10%          | 5               |
| <b>E: Cliente</b>  |              |                 |
| <b>E1: El Cliente insiste en nuevos requisitos.</b>  | 20%          | 2               |
| <b>G: Requisitos</b>   |              |                 |
| <b>G2: Los requisitos no se han definido correctamente y su redefinición aumenta el ámbito del proyecto.</b> | 15%          | 4               |

Tabla 12 - Análisis de Riesgos

El riesgo considerado con mayor probabilidad es el A11. La principal razón es que el equipo de desarrollo está conformado por una única persona, por lo que un retraso en cualquier tarea de desarrollo provocaría un retraso en cascada para el proyecto. Para mitigar este posible riesgo, el equipo de desarrollo intentará estimar las tareas con un ligero margen de tiempo para poder solventar los retrasos que puedan suceder.

En cuanto al riesgo A2, al ser una de las primeras veces donde el equipo de desarrollo se enfrenta a estimar un proyecto completo, puede que se falle a la hora de crear las estimaciones al no tener en cuenta todos los aspectos que conlleva las tareas.

Sobre el riesgo E1, al ser desarrollado el TFG bajo una adaptación de la metodología ágil SCRUM, los cambios de requisitos están aceptados si se hacen en las reuniones de planificación de sprint y se mantienen dentro de la definición inicial del proyecto. Algo similar sucede con el riesgo G2. Al estar realizándose con una

adaptación a SCRUM, las reuniones de revisión al finalizar los sprints servirán para verificar que los requisitos iniciales se han definido de forma correcta.

### 5.1.3 Revisión

| SPRINT                           | LISTA DE TAREAS                              | HORAS EST. | HORAS REALES |
|----------------------------------|--|------------|--------------|
| 0                                | T1: Definición del Proyecto                  | 8          | 10           |
|                                  | T2: Listar Historias de Usuario              | 8          | 8            |
|                                  | T3: Estimación de las Historias de Usuario   | 5          | 5            |
|                                  | T4: Desarrollo de Plan de Gestión de Riesgos | 8          | 6            |
|                                  | T5: Desarrollo del Anteproyecto              | 21         | 30           |
| <b>TOTAL DE HORAS ESTIMADAS:</b> |  | <b>50</b>  | <b>59</b>    |

Tabla 13- Resumen del Sprint 0

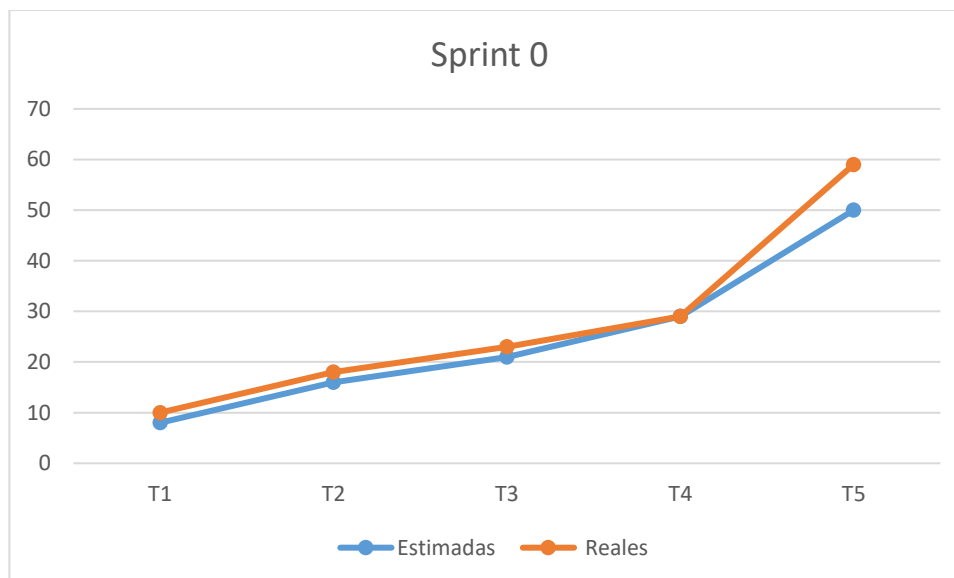


Figura 3 - Burn up Sprint 0

#### 5.1.4 Retrospectiva

El Sprint 0 ha durado ligeramente algo más de lo estimado al comienzo de él. Al ser el primer sprint del TFG, se ha dedicado especial importancia a la definición del proyecto. Donde se ha dedicado mucho más tiempo de lo esperado ha sido en la redacción del anteproyecto, donde el equipo de desarrollo erró en la estimación inicial.

#### 5.2 Sprint 1: Análisis de Requisitos y Modelado de Datos

Una vez ha sido completado el Sprint 0, el de planificación, para esta nueva iteración se comenzará a utilizar la metodología de desarrollo del DWH. Sin embargo, este primer Sprint servirá para hacer una captura de los procesos de negocio de la compañía, conocer sus fuentes de datos y sus correspondientes necesidades de información. Finalmente se transformarán todas esas necesidades a un documento que resuma el modelado dimensional del DWH. Es por ello que al final de este Sprint se tendrá que modificar parte de la planificación final del proyecto, ya que se conocerán más detalles sobre la última fase (explotación de la información). Aun así, esto no es un problema para el transcurso del TFG ya que se optó por ser gestionado con una adaptación de la metodología ágil de SCRUM, adecuación continua a las circunstancias de la evolución del proyecto.

##### 5.2.1 Tareas

Tras la reunión de planificación en el Sprint 0, se decidió que la Historia de Usuario 1 será desarrollada y completada en esta iteración, por lo que debe ser añadida en la pila del Sprint. Posteriormente, el equipo de desarrollo realizó la estimación de las tareas utilizando *Planning Poker* [30]. La Tabla 14 muestra el resultado de la misma:

| SPRINT                    | LISTA DE TAREAS                                  | HORAS EST. |
|---------------------------|--|------------|
| 1                         | T1: Entrevistas con el responsable de Producción | 8          |
|                           | T2: Identificación de Dimensiones y Hechos       | 13         |
|                           | T3: Desarrollo del Modelado Dimensional          | 21         |
| TOTAL DE HORAS ESTIMADAS: |  | 42         |

Tabla 14- Planificación del Sprint 1

Como resultado de las tareas incluidas en el Sprint 1, se ha realizado un análisis del proceso de negocio a incluir en el DWH tras las entrevistas con el responsable de producción (5.2.2.1). Posteriormente, se han obtenido las dimensiones y hechos que se usarán en el DWH documentada en una *Enterprise bus matrix* (5.2.2.2) además de un documento que muestra el modelo dimensional que se implementará en el siguiente Sprint (5.2.2.3).

## **5.2.2 Resultados del Sprint**

### **5.2.2.1 Análisis de los Procesos de Negocio**

Como primera tarea para el desarrollo de un DWH es necesario conocer los procesos de negocio y las necesidades de información de la empresa. Para ello se han realizado capturas de requisitos con el responsable del departamento del cual se analizarán sus datos. La técnica utilizada ha sido la de entrevistas con el responsable del departamento de producción de Avanttic.

Como resultado de esta tarea, se ha obtenido el principal proceso de negocio a analizar y a almacenar en el *data warehouse*. Desde el departamento de producción se quiere analizar la producción que tienen los empleados en las actividades diarias, registrado en el sistema ERP de Avanttic. Los consultores de la empresa imputan actividades a subproyectos. A su vez, estos subproyectos están asociados a un proyecto. Las necesidades de información son las siguientes:

- **Conocer la situación con los clientes:** Analizar la cantidad de horas que se facturan a los clientes en los diferentes proyectos y subproyectos que tengan con la empresa.
- **Análisis del estado actual y pasado de los departamentos de la empresa:** El responsable de producción debe ser capaz de ver el rendimiento de cada área en los diferentes ejercicios de la empresa.
- **Horas facturables por departamento:** Desde el área de producción se desea poder ver de manera rápida y visual la cantidad de horas que son facturadas a clientes en proyectos y las horas que los consultores de un área no facturan.
- **Control de empleados:** Es necesario poder visualizar y analizar las horas que una persona dedica cada día a los subproyectos y proyectos, y en cómputo global de forma diaria y mensual. Además, se podrán comparar con las horas estipuladas para cada empleado.

### 5.2.2.2 Identificación de Dimensiones y Hechos

La siguiente tarea consiste en la identificación de las dimensiones y hechos. Para ello se han seguido los pasos establecidos por Ralph Kimball [3]. Kimball define cuatro pasos para poder completar el modelado dimensional que han sido seguidos en el TFG:

1. **Elegir el proceso de negocio:** Al ser un *data warehouse* con datos sobre la producción empresarial, el proceso de negocio seleccionado es el de la actividad que un empleado registra en el ERP. Como también se requiere analizar la actividad planificada a un empleado, también será seleccionado el proceso de negocio de creación de un calendario personal de un empleado.
2. **Establecer un nivel de granularidad:** En este caso se ha optado por agrupar las actividades, de un mismo tipo, que un empleado realiza cada día. Gracias a tener este nivel de detalle, se podrá ver en detalle las horas que una persona dedica a subproyectos, proyectos y por tanto clientes.
3. **Elegir las dimensiones:** Tras realizar diferentes reuniones con el responsable de producción, las dimensiones identificadas se pueden ver en la Ilustración 13 que representa la *Enterprise Bus Matrix* generada tras la captura de requisitos.
4. **Identificar las tabla de hechos y medidas:** En este último paso se han identificado las dos tablas de hechos que formaran parte en el *data warehouse*. Las tablas de hechos definidas se pueden observar en la Tabla 15. La principal medida de las tablas de hechos son las horas imputadas por un consultor y las horas que fueron asignadas por parte del departamento de producción.

| Proceso de Negocio                           | Dimensiones |          |             |       |         |      |                     |                |
|--|-------------|----------|-------------|-------|---------|------|---------------------|----------------|
|  | Cliente     | Proyecto | Subproyecto | Fecha | Persona | Área | Calendario Personal | Tipo Actividad |
| Imputar horas a una actividad de un proyecto |             |          |             |       |         |      |                     |                |
| Generar Calendario del Personal              |             |          |             |       |         |      |                     |                |

Ilustración 13 - Enterprise Bus Matrix del TFG

| TABLA DE HECHOS | ATRIBUTOS   |
|-----------------|---|
| ACTIVIDAD       | <ul style="list-style-type: none"> <li>• Empresa.</li> <li>• Cliente.</li> <li>• Proyecto.</li> <li>• Subproyecto.</li> <li>• Persona.</li> <li>• Área.</li> <li>• Tipo de Actividad.</li> <li>• Fecha.</li> <li>• Cantidad Horas.</li> </ul> |
| HORAS_DÍA       | <ul style="list-style-type: none"> <li>• Persona.</li> <li>• Fecha.</li> <li>• Horas Calendario Personal.</li> <li>• Horas Realizadas.</li> </ul>   |

Tabla 15- Tablas de Hechos Identificadas

### 5.2.2.3 Modelo Dimensional

Tras la identificación de las dimensiones y hechos, el siguiente paso es el desarrollar el modelo dimensional que será implementado en el *data warehouse*. Aunque el modelado de datos más conocido es el de Estrella, tratado en la sección 3.2.3, el equipo de desarrollo ha optado por usar la variante de copo de nieve. La principal razón de utilizar este diseño es la posibilidad de crear una jerarquía entre las dimensiones de: Empresa, Cliente, Proyecto y Subproyecto. Siguiendo el modelo en estrella se perdería la normalización de estas tres dimensiones. Sin embargo, tras analizarse las necesidades de información de la empresa es más eficiente mantener la jerarquía en esas 3 dimensiones y por tanto seguir el modelado en copo de nieve. La Ilustración 14 muestra el modelo dimensional resultante.

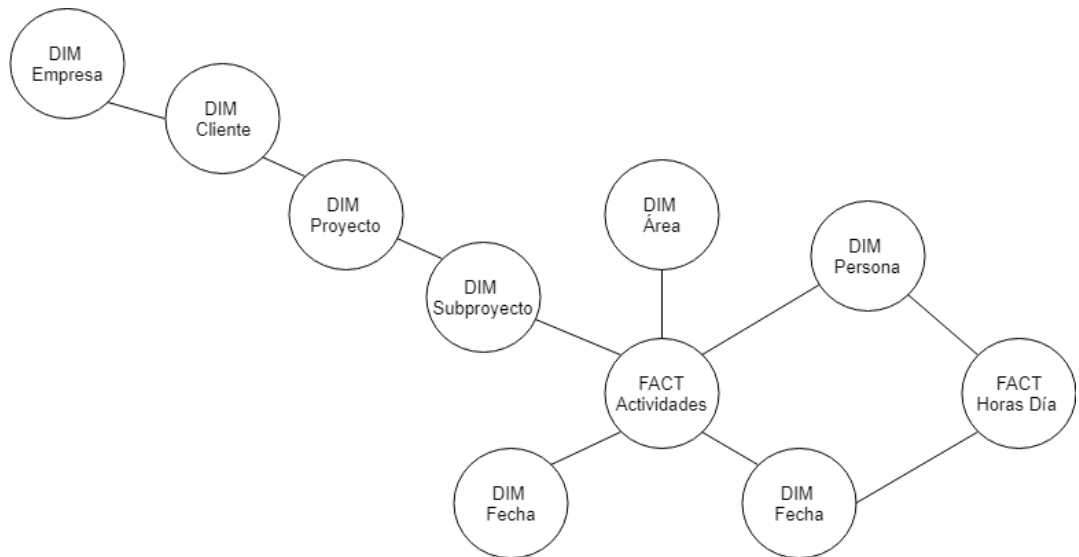


Ilustración 14 - Modelo Dimensional del DWH

### 5.2.3 Revisión

| SPRINT                           | LISTA DE TAREAS                             | HORAS EST. | HORAS REALES |
|----------------------------------|---|------------|--------------|
| 1                                | T1: Entrevistas con el responsable de Prod. | 8          | 10           |
|                                  | T2: Identificación de Dimensiones y Hechos  | 13         | 12           |
|                                  | T3: Desarrollo del Modelado Dimensional     | 21         | 30           |
| <b>TOTAL DE HORAS ESTIMADAS:</b> |   | 42         | 52           |

Tabla 16- Resumen del Sprint 1

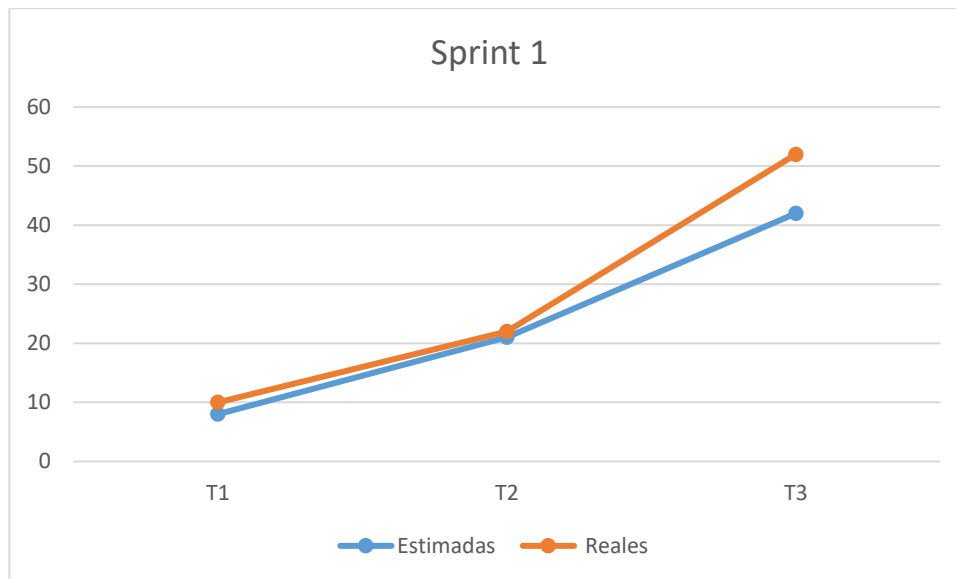


Figura 4 - Burn Up Sprint 1

#### 5.2.4 Retrospectiva

En el segundo sprint del TFG, primero de desarrollo, también se ha sido optimista en la planificación. Aunque la reunión con el responsable de producción duró ligeramente más de lo planificado, el principal error vino con el modelado dimensional. Y es que esta tarea era clave para el devenir del resto del proyecto, es por ello que el equipo de desarrollo dedicó más tiempo del planificado en el inicio con el propietario de producto para comprobar que el artefacto final de la tarea cumpla con los requisitos iniciales.

#### 5.3 Sprint 2: Extracción Inicial de los Datos

Una vez obtenidos los requisitos del almacén de datos en el Sprint anterior, ahora el proyecto se centra en comenzar con el desarrollo del DWH. Primeramente se ha decidido en qué sistema de Base de Datos se optará para almacenar el DWH. Seguidamente, ha sido necesaria la creación de los esquemas físicos en el sistema de BBDD elegido antes. Finalmente se ha realizado el diseño de las cargas de *Staging* y la configuración de las cargas iniciales del DWH.

Como resultado del Sprint anterior, el equipo de desarrollo ya conoce en detalle que informes y cuadros de mando se van a realizar en la fase de explotación BI. Por lo que en la reunión de planificación del Sprint 2, se ha añadido una nueva historia de usuario a la pila de producto. Esta nueva historia de usuario ha especificado

todo lo relacionado con el desarrollo de los informes y cuadros de mando. Como consecuencia, se ha rediseñado el plan de proyecto detallado en el Sprint 0.

### 5.3.1 Modificación de la Pila de Producto

| HU5 Desarrollo Cuadros de Mando  |                           |
|--|---------------------------|
| <b>Prioridad de Negocio:</b> 850   | <b>Sprint Asignado:</b> 4 |
| <b>Responsable:</b> Miguel Ampuero López-Sepúlveda   |                           |
| <b>Descripción:</b><br>Crear informes y Cuadros de Mando para visualizar datos sobre la producción empresarial.  |                           |
| <b>Precondición:</b><br>Tener disponible el <i>Data Warehouse</i> poblado de datos y el entorno de BI instalado y configurado.   |                           |
| <b>Postcondición:</b><br>Tras la consecución de esta historia de usuario, el <i>product owner</i> podrá tener acceso a los Informes y Cuadros de Mando creados.  |                           |
| <b>Actividades:</b> <ul style="list-style-type: none"> <li>• Desarrollo tabla dinámica para mostrar clientes y proyectos.</li> <li>• Desarrollo informe sobre empleados y resumen de horas.</li> <li>• Configuración informe dinámico sobre horas diarias de empleados.</li> <li>• Desarrollo de informes para áreas.</li> <li>• Creación de Mapa interactivo de Clientes.</li> <li>• Configuración de Cuadro de Mando de Clientes.</li> <li>• Configuración de Cuadro de Mando de Empleados.</li> <li>• Configuración de Cuadro de Mando de Áreas.</li> </ul> |                           |
| <b>Artefacto o producto generado:</b> <ul style="list-style-type: none"> <li>• Cuadros de Mando sobre la producción empresarial (clientes, empleados y áreas).</li> </ul>  |                           |

Tabla 17 - Historia de Usuario 5

| Sprint                           | HU | Tareas   | Duración<br>(Semanas) |
|----------------------------------|----|--|-----------------------|
| 0                                | -  | T1: Definición del proyecto.<br>T2: Listar historias de usuario.<br>T3: Estimación de las historias de usuario.<br>T4: Desarrollo del plan de gestión de riesgos.<br>T5: Desarrollo del Anteproyecto.  | 2                     |
| 1                                | 1  | T1: Entrevistas con Responsable de Producción.<br>T2: Identificación de Dimensiones y Hechos.<br>T3: Creación de un <i>Enterprise Bus Matrix</i> .<br>T4: Desarrollo del Modelo en Estrella.   | 2                     |
| 2                                | 2  | T1: Creación de esquemas del DWH.<br>T2: Desarrollo de las cargas de la BBDD de origen a <i>Staging</i> .<br>T3: Desarrollo de los <i>mapping</i> iniciales del DWH.   | 4                     |
| 3                                | 3  | T1: Modificación de los <i>mapping</i> para ser ejecutados periódicamente.<br>T2: Instalación de Agente <i>Standalone</i> de ODI.<br>T3: Configuración de servicio por notificación de Email.<br>T4: Creación de tablas para generar auditorías en las cargas del DWH. | 4                     |
| 4                                | 4  | T1: Instalación de Infraestructura de Oracle para OBI.<br>T2: Instalación de Oracle Business Intelligence.<br>T3: Importación del modelo de datos del DWH para creación de repositorio de BI.  | 2                     |
|                                  | 5  | T1: Desarrollo de Informes con OBI.<br>T2: Desarrollo de Informes con ODV.<br>T3: Configuración de Cuadro de Mando.  | 2                     |
| <b>TOTAL DURACIÓN (SEMANAS):</b> |    |  | <b>16</b>             |

Tabla 18 - Plan de Proyecto en el Sprint 2

### 5.3.2 Tareas

| SPRINT                    | LISTA DE TAREAS  | HORAS EST. |
|---------------------------|--|------------|
| 2                         | T1: Tomar decisión sobre el tipo de BBDD donde almacenar el DWH. | 3          |
|                           | T2: Diseño Físico del DWH.                                       | 13         |
|                           | T3: Creación de esquemas del DWH.                                | 13         |
|                           | T4: Creación de capa de <i>Staging</i> .                         | 21         |
|                           | T5: Desarrollo de mapeos de la capa de <i>Staging</i> .          | 21         |
|                           | T6: Desarrollo de los mapeos iniciales del DWH.                  | 30         |
| TOTAL DE HORAS ESTIMADAS: |  | 101        |

Tabla 19- Planificación del Sprint 2

Como resultado de este Sprint, se ha generado un documento que muestra el esquema físico que ha sido utilizado en el DWH, resumido en la sección 5.3.3.3 . Tras la creación de los esquemas del DWH y la capa de *Staging*, se han realizado las relaciones de datos para hacer la inserción inicial de los datos (5.3.3.6).

### 5.3.3 Resultados del Sprint

#### 5.3.3.1 Elección de Tecnología para el DWH

Una vez realizada la planificación del proyecto y la posterior captura de necesidades de información de la empresa, el siguiente paso fue elegir el motor de base de datos para almacenar el DWH. Debido a que el TFG se está realizando bajo el convenio de prácticas FORTE en la empresa Avanttic, se ha decidido utilizar la base de datos *Oracle 12c*.

Por otro lado, para la creación de los cubos se ha optado por utilizar cubos ROLAP (detallados en la sección 3.2.4.2). Esta decisión se ha tomado en base al tiempo y recursos que se va a dedicar al TFG. Ya que el uso de cubos ROLAP es el más adecuado para proyectos de corta duración y de bajo coste, no es necesario el uso de bases de datos multidimensionales. Se hará uso de un motor de Base de Datos

relacional como es el de *Oracle 12c* para almacenar el DWH y el área de *staging*. Esta decisión de diseño cumple con los requisitos pactados con el propietario de producto.

### 5.3.3.2 Modelo Físico del DWH

Para la creación de las tablas que formarán parte del DWH se ha realizado con la herramienta SQL Data Modeler, realizando primero un diagrama lógico y posteriormente un diseño físico. Tras un estudio de la bbdd de origen y unido a las necesidades de información de la empresa, se han creado los atributos para cada tabla de dimensión y de hechos. Para las tablas de dimensiones se ha optado por utilizar claves subrogadas.

En origen, las claves principales de las tablas son de tipo *VARCHAR* y de cara al diseño del DWH aumenta el rendimiento de respuesta a las consultas si el tipo de datos es un numérico. Es por ello que todas las tablas de dimensiones no utilizarán la clave primaria de origen como clave principal si no una secuencia numérica única para cada fila de la tabla. Posteriormente, en la tabla de hechos se añadirán como claves foráneas las claves subrogadas de las dimensiones. La Ilustración 15 muestra el resultado final del diseño físico del DWH:



### 5.3.3.3 Creación de esquemas del DWH

Antes de crear el esquema correspondiente al *data warehouse*, se ha estipulado un cálculo del espacio de la base de datos para que cumpla con las necesidades del cliente. Para el desarrollo de esta tarea, se ha hecho uso de la herramienta SQL Developer de Oracle.

Para las tablas de dimensiones y hechos, SQL Developer permite importar un diseño físico del Data Modeler e insertarlo en el esquema deseado. Por tanto se ha utilizado el diseño realizado en la tarea anterior para ser importado.

Sobre las claves subrogadas que tienen las dimensiones como claves principales, se ha optado por crear una secuencia para cada tabla de la base de datos. El Código 1 muestra la sentencia SQL ejecutada para crear la secuencia de la dimensión área.

```
CREATE SEQUENCE "TFG_DWH"."SQ_DIM_AREA" MINVALUE 1 MAXVALUE 2147483647  
INCREMENT BY 1 START WITH 17 NOCACHE NOORDER NOCYCLE NOPARTITION;
```

*Código 1 - Sentencia SQL para crear secuencia en DWH*

Tras la creación de la secuencia ha sido necesaria la creación de *triggers* o disparadores para cada secuencia creada. El Código 2 muestra como se ha realizado para la dimensión área:

```
create or replace  
trigger TRIGGER1  
BEFORE INSERT ON DIM_AREA  
FOR EACH ROW  
DECLARE V_NEWVAL NUMBER(12) := 0;  
BEGIN  
  V_NEWVAL := TFG_DWH.SQ_DIM_AREA.NEXTVAL;  
  :new.DWH_AREA := V_NEWVAL;  
END;
```

*Código 2 - Creación de Trigger para DWH*

### 5.3.3.4 Creación de la capa de Staging

Como la fuente de datos es la BBDD del sistema ERP de la empresa, es conveniente desarrollar una capa de *staging*. El área de *staging* es una duplicación de las tablas o estructuras de datos del origen, de carácter temporal. Esta área del DWH permite evitar que los procesos posteriores de carga afecten a la disponibilidad de las fuentes de datos operacionales.

La Ilustración 16 muestra el modelo lógico que mantiene la capa de *Staging*. Esta capa ha sido creada a partir del modelo de BBDD del sistema ERP de la empresa

y han sido seleccionadas las tablas que son utilizadas en las cargas del almacén de datos.

Para la realización de esta tarea, primeramente se ha creado el usuario y esquema para almacenar las tablas correspondientes a la capa de *staging*. Al igual que con las tablas del *data warehouse*, se ha realizado un diseño físico con la herramienta SQL Data Modeler y ha sido importado en SQL Developer. Como se ha definido previamente que datos se van a necesitar para la población, no es necesario replicar toda la estructura del sistema ERP en esta capa. Simplemente se han importado las tablas que servirán como origen para las dimensiones y hechos.

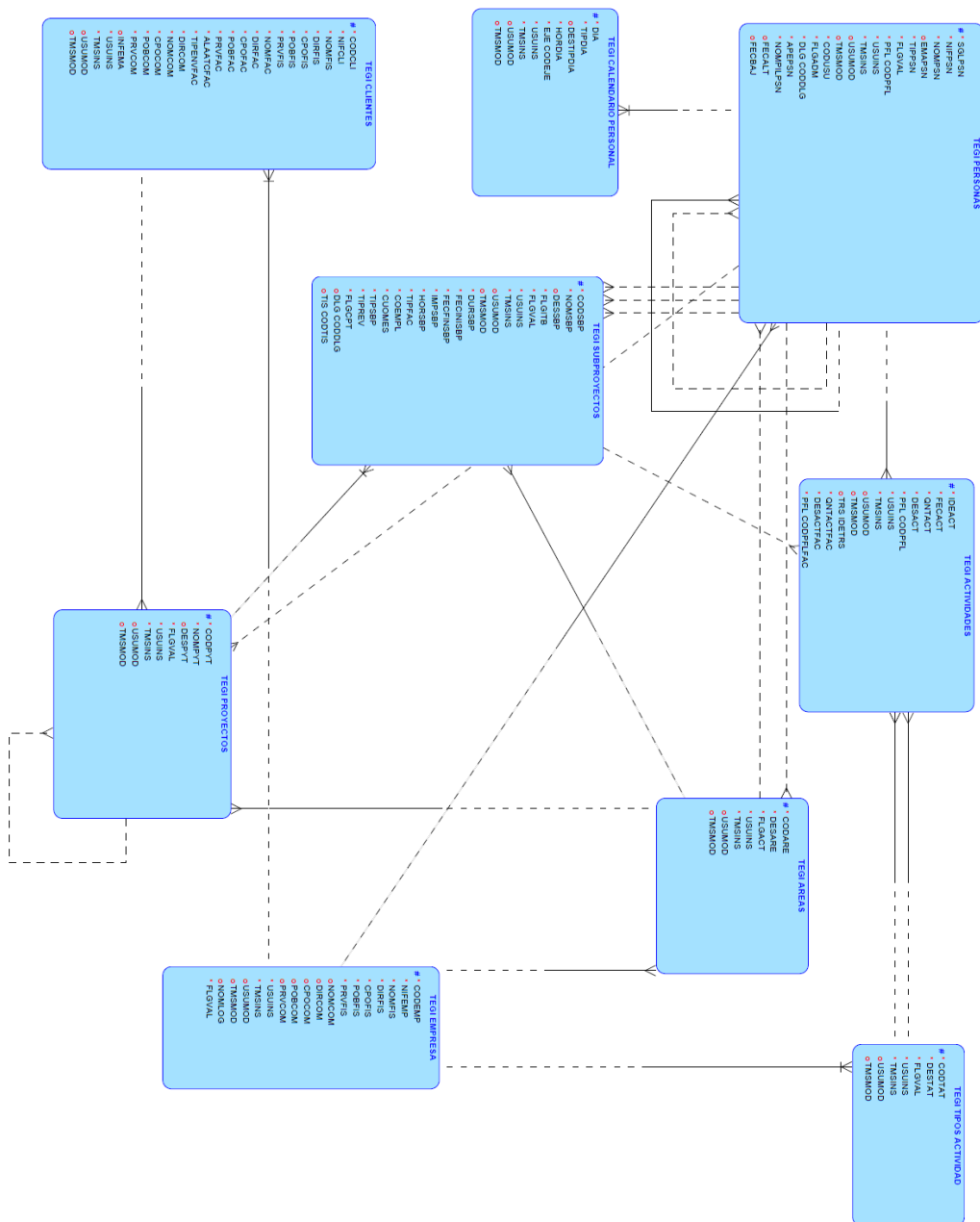


Ilustración 16 - Modelo Lógico de la capa de Staging

### 5.3.3.5 Mapeos en la capa Staging

Para la realización de los procesos ETL, se ha utilizado la herramienta Oracle Data Integrator. Esta herramienta precisa de una base de datos para crear un repositorio donde almacenar los proyectos ETL que se vayan a desarrollar. Por lo que antes de realizar cualquier diseño, ha sido necesario la creación de un repositorio en una base de datos. ODI almacena en el repositorio metadatos de las aplicaciones y los datos involucrados en las transformaciones y procesos de integración además de la información generada al crear ejecuciones. El equipo de desarrollo ha optado por utilizar la misma base de datos donde se está almacenando el DWH para crear el repositorio.

Una vez el repositorio ha sido creado, es necesario especificar en la herramienta las fuentes de datos que van a ser utilizados a lo largo del proceso ETL. Se ha importado la fuente principal de datos del TFG, la base de datos donde reside el sistema ERP de la empresa, y el esquema del área de *staging*. La Ilustración 17 resume como se importa un esquema físico, en este caso la BBDD correspondiente a la capa de *staging*. Una vez definido el origen de datos, en ODI los esquemas importados son los mostrados en la Ilustración 18.

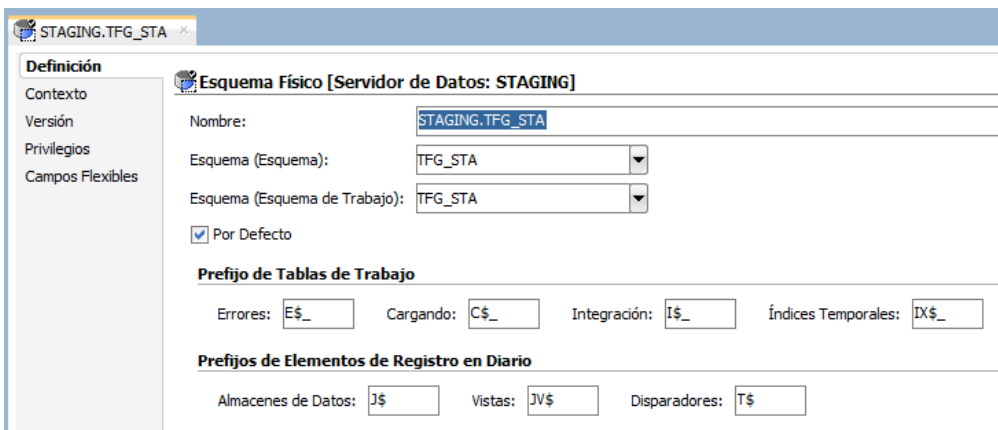


Ilustración 17- Definición del Esquema Físico de Staging

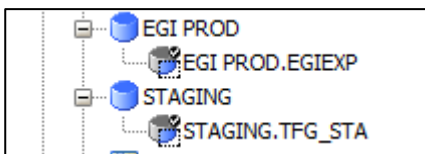


Ilustración 18 - Esquemas físicos importados para la copia de Staging

Una vez definidos los esquemas físicos, y sus correspondientes lógicos, en Oracle Data Integrator es necesario definir diferentes modelos de datos para cada fuente. Y es que la herramienta de Oracle es capaz de realizar ingeniería inversa, tal y como se muestra en la Ilustración 19, de la fuente de datos si se ha definido un esquema lógico.

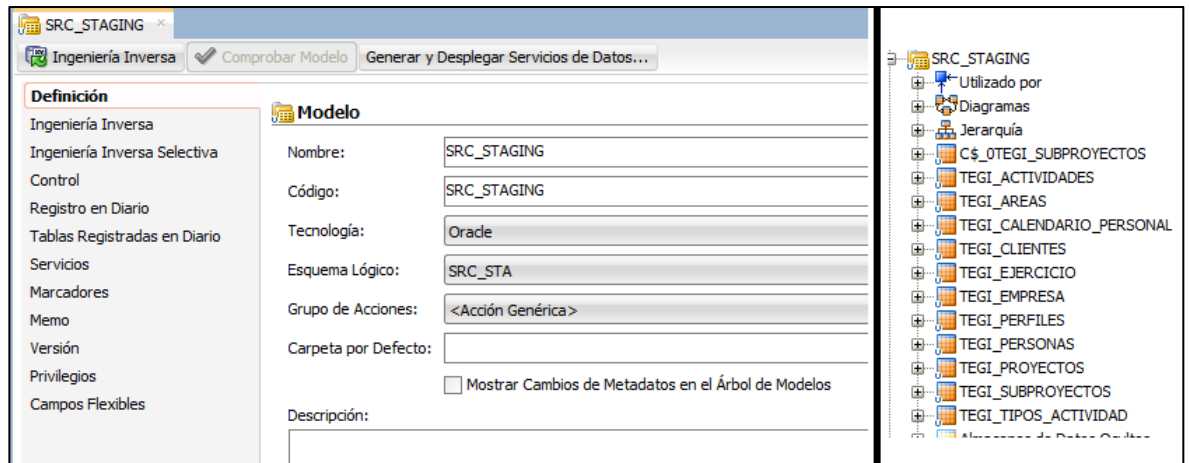


Ilustración 19 - Modelo de Datos de Staging importado en ODI

Tras importar los esquemas, se han realizado los diseños de los mapeos de la fuente de datos a las tablas de *staging*. Debido a que se están haciendo copias de tablas idénticas, no se explicará en detalle los mapeos realizados en esta tarea. La Ilustración 20 muestra el ejemplo de la tabla TEGI\_PROYECTOS.

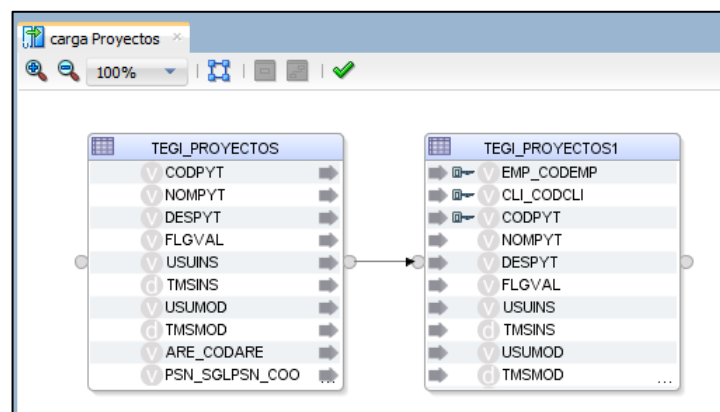


Ilustración 20 - Mapping en la capa de Staging

Una vez los mapeos han sido diseñados, se ha generado un escenario (código compilado para cada tarea ETL diseñada) para cada flujo de datos y se han añadido a un plan de carga, secuencia de escenarios a ejecutar. En los planes de carga de ODI se pueden indicar el orden de ejecución de los diferentes escenarios,

pudiéndose ejecutar de forma paralela algunos de ellos. La Ilustración 21 muestra el plan de carga diseñado, y la Ilustración 22 el resultado tras su ejecución.

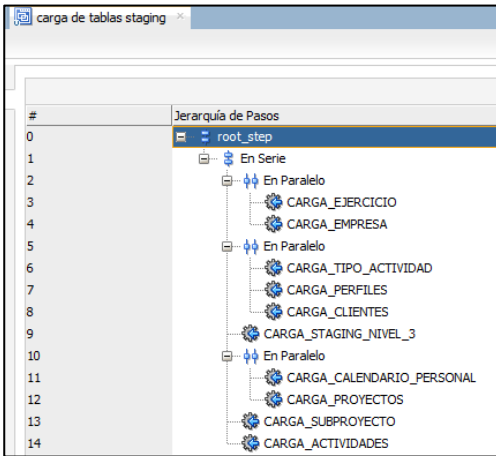


Ilustración 21- Plan de Carga de Staging

| #  | Jerarquía de Pasos        | Estado | Duración | Inicio   | Fin      |
|----|---------------------------|--------|----------|----------|----------|
| 0  | root_step                 | ✓      | 01:43    | 13:29:34 | 13:31:17 |
| 1  | En Serie                  | ✓      | 01:43    | 13:29:34 | 13:31:17 |
| 2  | En Paralelo               | ✓      | 00:02    | 13:29:34 | 13:29:36 |
| 3  | CARGA_EJERCICIO           | ✓      | 00:00    | 13:29:35 | 13:29:35 |
| 4  | CARGA_EMPRESA             | ✓      | 00:00    | 13:29:35 | 13:29:35 |
| 5  | En Paralelo               | ✓      | 00:02    | 13:29:36 | 13:29:38 |
| 6  | CARGA_TIPO_ACTIVIDAD      | ✓      | 00:01    | 13:29:36 | 13:29:37 |
| 7  | CARGA_PERFILES            | ✓      | 00:01    | 13:29:36 | 13:29:37 |
| 8  | CARGA_CLIENTES            | ✓      | 00:01    | 13:29:36 | 13:29:37 |
| 9  | CARGA_STAGING_NIVEL_3     | ✓      | 00:02    | 13:29:38 | 13:29:40 |
| 10 | En Paralelo               | ✓      | 00:25    | 13:29:41 | 13:30:06 |
| 11 | CARGA_CALENDARIO_PERSONAL | ✓      | 00:24    | 13:29:41 | 13:30:05 |
| 12 | CARGA_PROYECTOS           | ✓      | 00:01    | 13:29:41 | 13:29:42 |
| 13 | CARGA_SUBPROYECTO         | ✓      | 00:01    | 13:30:06 | 13:30:07 |
| 14 | CARGA_ACTIVIDADES         | ✓      | 01:09    | 13:30:08 | 13:31:17 |

Ilustración 22- Resultado ejecución plan de carga Staging

### 5.3.3.6 DWH poblado con los datos iniciales

Como se ha explicado en la tarea anterior, antes de proceder con el desarrollo de los mapeos es necesario importar el esquema del DWH. Por lo que se añade al esquema de *staging* y la base de datos del sistema ERP que ya estaban previamente configurados.

Ilustración 23- Esquema Físico del DWH Importado en ODI

Al igual que se ha explicado en la tarea anterior, se debe crear un modelo para la arquitectura importada del DWH. Una vez el modelo de datos ha sido importado, los mapeos se han diseñado para posteriormente añadirlos a un plan de carga (carga inicial del DWH).

A continuación se detallan los mapeos de las dimensiones del DWH. En el primer *mapping*, el de la tabla DIM\_CLIENTES, se detalla todas las funcionalidades que ofrece la herramienta ETL utilizada en el TFG. Para el resto de mapeos se han destacado las particularidades de cada diseño.

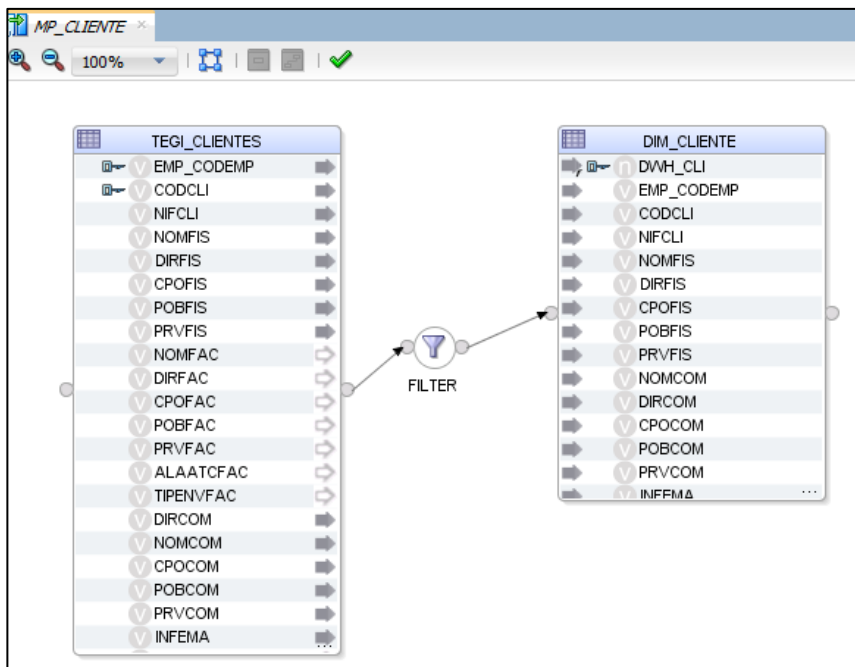


Ilustración 24 - Mapping Cliente (Vista lógica)

La Ilustración 24 muestra el resultado del *mapping* de la dimensión cliente. La herramienta ETL ofrece dos puntos de vista a la hora de diseñarlo, siendo esta primera la vista lógica. Tras diseñar como se va a realizar la transformación de los datos, en este caso se ha hecho un filtro por el código de la empresa, es necesario especificar físicamente como se va a realizar la transformación.

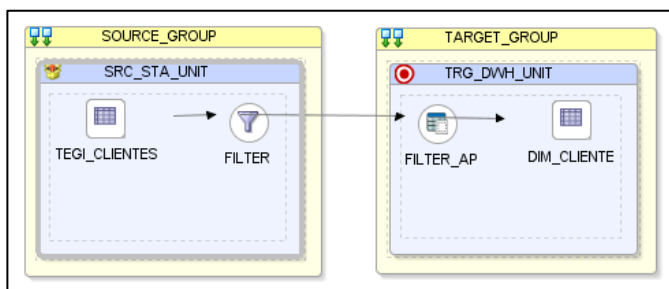


Ilustración 25 - Mapping Cliente (Vista Física)

La Ilustración 25 muestra de manera física cómo se va a gestionar la extracción, carga y transformación de los datos. Para gestionar las diferentes fuentes de datos y sus transformaciones, ODI desarrolla en concepto de módulo de conocimiento. Estos módulos implementan los flujos de los datos para definir plantillas de generación de código a través de los diferentes sistemas que están involucrados en cada proceso.

Aunque son varios los tipos de módulo de conocimiento que tiene la herramienta, este TFG ha hecho uso de LKMs (*Load knowledge modules*), IKMs (*Integration Knowledge Modules*) y CKMs (*Comprobatión Knowledge Modules*) específicos para las fuentes de datos utilizadas, mostrados en la Ilustración 26.

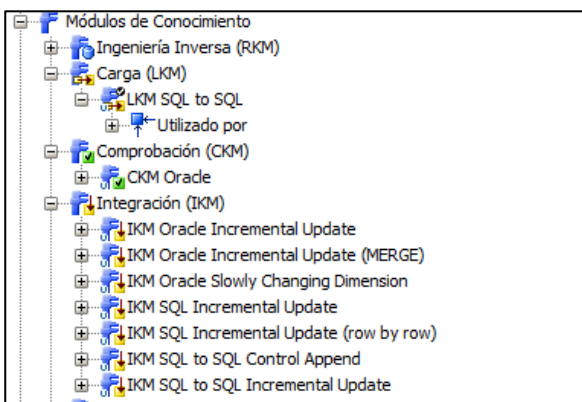


Ilustración 26 - Módulos de Conocimiento Importados

Siguiendo con el ejemplo de la carga de la dimensión Cliente, la Ilustración 27 muestra cómo se han configurado los IKM y LKM.

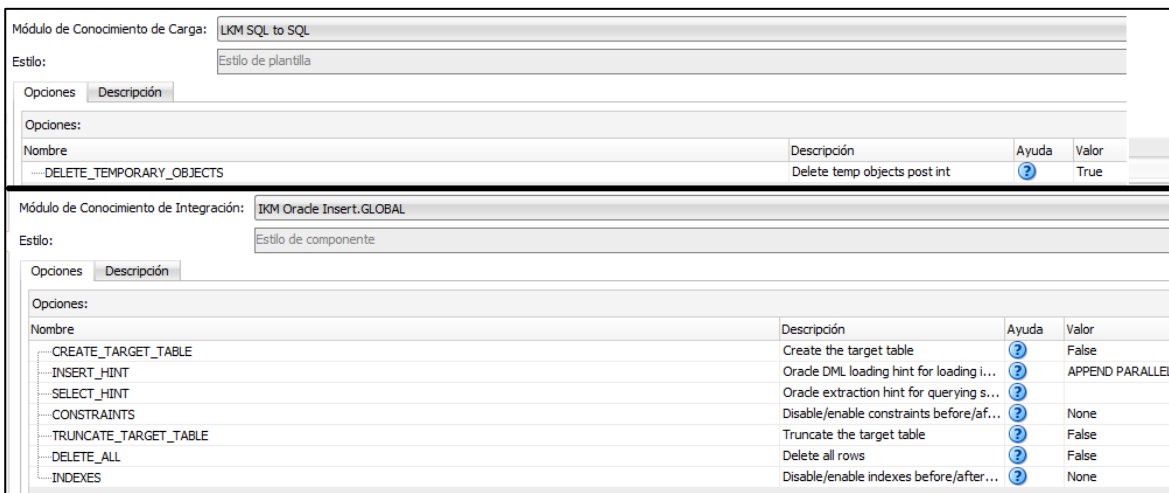


Ilustración 27- Ejemplo Configuración Módulos de Conocimiento ODI

Oracle Data Integrator también permite generar secuencias en BBDD o importarlas. Para el diseño de las cargas se han importado las secuencias previamente creadas y se ha asignado directamente a las claves subrogadas de cada tabla. En caso de cambiar de herramienta de ETL, la numeración de la secuencia no se perdería al estar almacenada en el propio DWH y no en la herramienta ETL. La Ilustración 28 muestra la secuencia para la dimensión área.

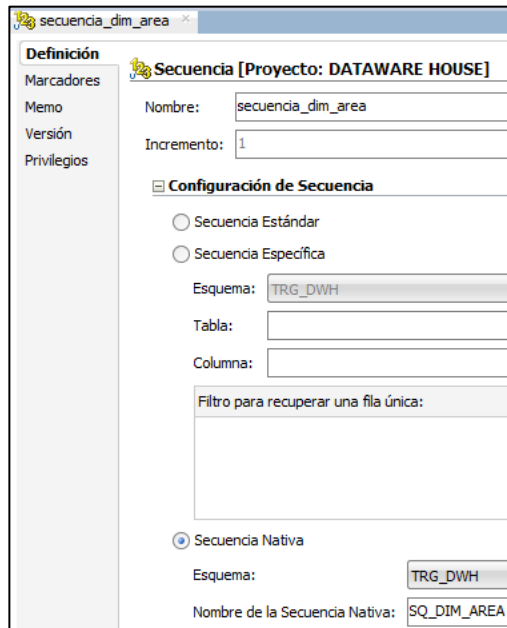


Ilustración 28 - Importación de una secuencia de BBDD a ODI

A continuación se detallan todos los mapeos que se han diseñado para el resto de dimensiones en la fase de carga inicial.

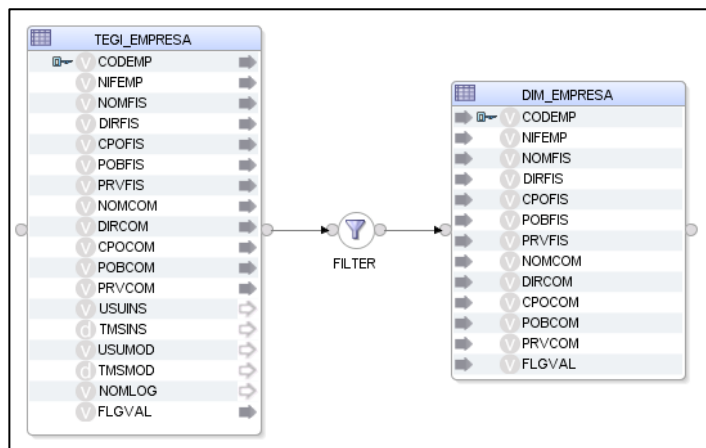


Ilustración 29- Mapping DIM Empresa

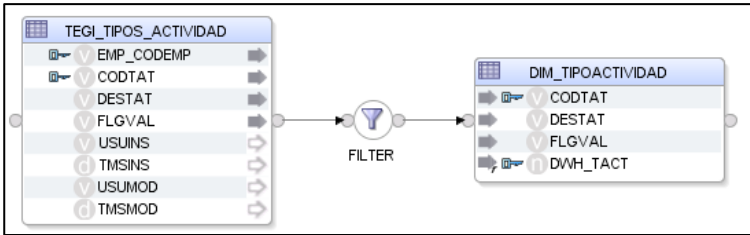


Ilustración 30 - Mapping DIM Tipo Actividad

La Ilustración 29 e Ilustración 30 muestran los *mappings* de las dimensiones empresa y tipo de actividad respectivamente. Estos dos flujos de datos son sencillos ya que son similares a los datos que vienen de origen. Sobre la dimensión empresa, no se ha generado clave subrogada ya que en el almacén de datos tan solo se tendrá un registro. Sin embargo, para la dimensión tipo de actividad si se ha utilizado una clave subrogada (DWH\_TACT).

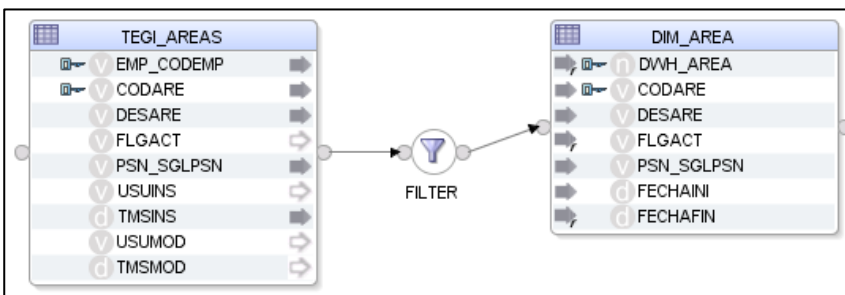


Ilustración 31 - Mapping de DIM Área

También en el mapeo para la dimensión Área (Ilustración 31) se ha realizado como los anteriores, en este caso creando también una clave subrogada.

Como se está utilizando un modelado en copo de nieve, será necesario utilizar las claves subrogadas de ciertas dimensiones como atributos en dimensiones de jerarquía inferior, pudiendo así facilitar operaciones de *drill down* posteriormente. Tal y como se puede apreciar en la Ilustración 32 que corresponde al *mapping* de DIM Persona, se hace un *look up* con la tabla recién cargada DIM ÁREA para obtener su nueva clave subrogada comparando con el código de área utilizado en origen (Ilustración 33).

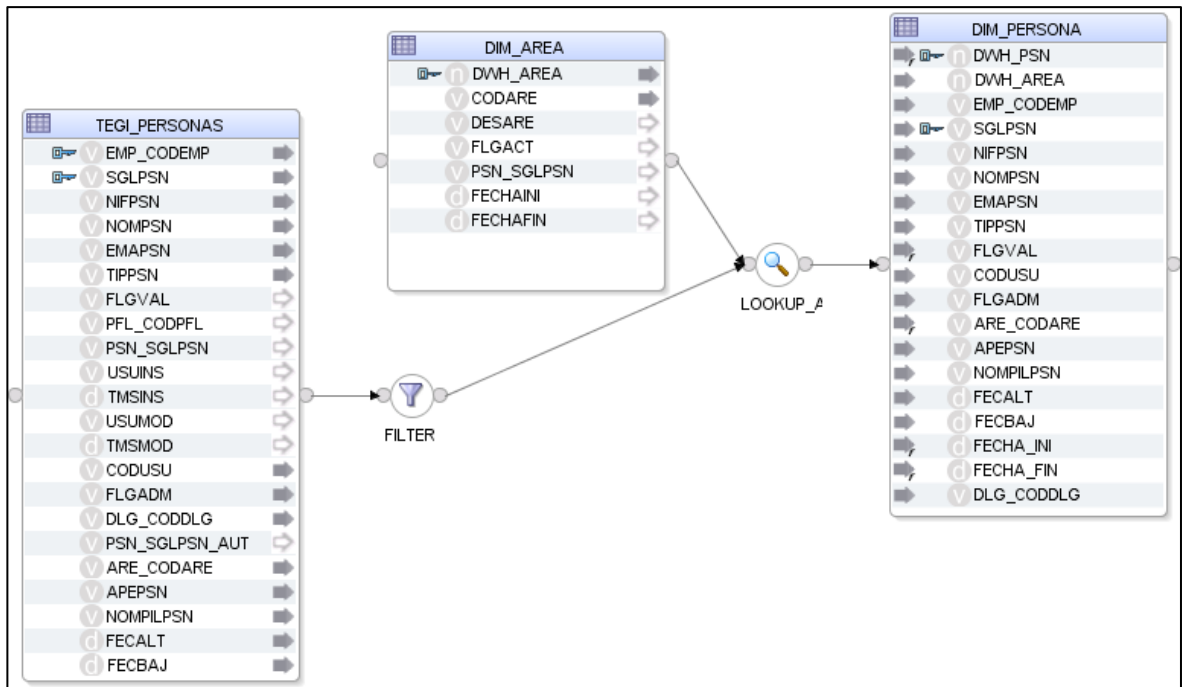


Ilustración 32- Mapping DIM Persona

| LOOKUP_ARE - Propiedades            |   |
|-------------------------------------|---|
| <input type="text" value="Buscar"/> |   |
| <b>Condición</b>                    |   |
| Reglas de Filas Coincidentes        | Condición de Consulta: <code>TEGI_PERSONAS.ARE_CODARE = DIM_AREA.CODARE</code>                                  |
| General                             | Descripción Técnica: <code>[TEGI_PERSONAS (TEGI_PERSONAS)].[ARE_CODARE] = [DIM_AREA (DIM_AREA)].[CODARE]</code> |
| Puntos de Conector                  | Indicación de Ejecución: <code>Ninguna Indicación</code>  |

Ilustración 33 - Condición Look-Up DIM Persona

Para los mapeos de las dimensiones de proyecto (Ilustración 34) y subproyecto (Ilustración 35) se seguirá un diseño similar, haciendo uso de la operación *look-up* para insertar las claves subrogadas de las dimensiones relacionadas.

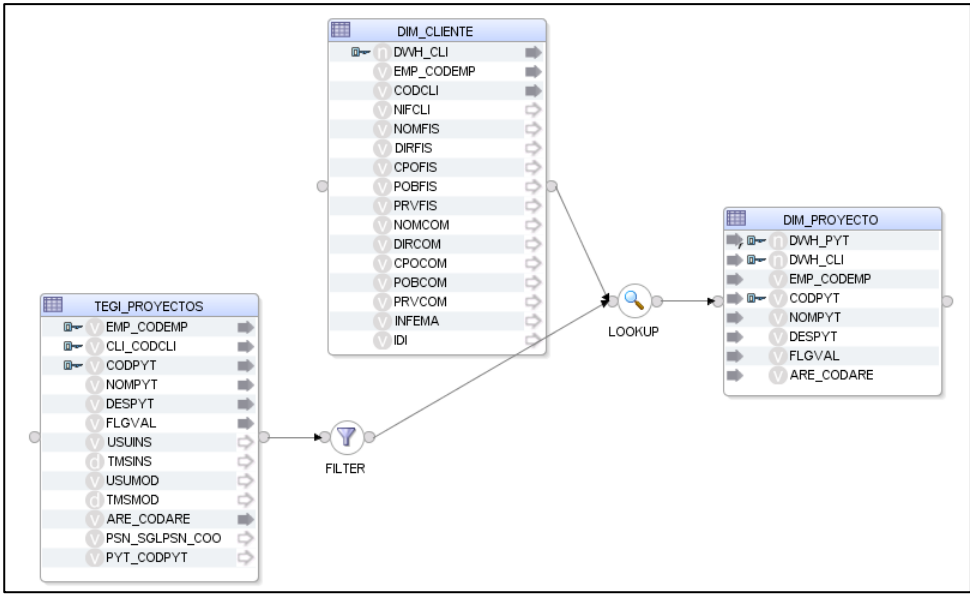


Ilustración 34 - Mapping DIM Proyecto

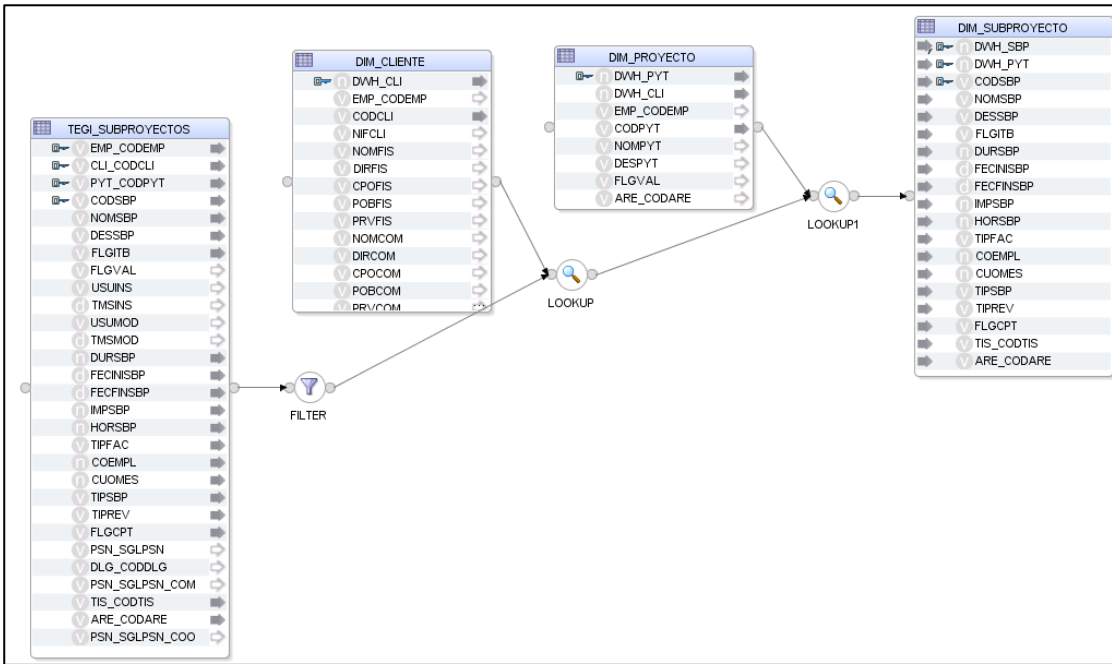


Ilustración 35 - Mapping DIM Subproyecto

Para la creación de la dimensión tiempo, se ha utilizado un script (Código 3) desarrollado en PL/SQL [31] que genera los registros necesarios en una dimensión de este tipo siguiendo los parámetros especificados. Este código, que ha sido almacenado en el DWH, es ejecutado en Oracle Data Integrator a través de un procedimiento. Un procedimiento en ODI detalla las sentencias que se pueden ejecutar en las fuentes de datos o en destino, en función de lo deseado. En este caso el procedimiento (Ilustración 36) consta de dos llamadas, la primera de ellas ejecuta un truncado de la dimensión en caso de que haya datos anteriores, y seguidamente ejecuta el procedimiento SQL almacenado en el *data warehouse* que genera la tabla de dimensión tiempo.

| Nombre de la Tarea | Limpieza                 | Ignorar Er...            | Opciones ...                        | Comando de Destino                       | Comando de Origen | Tecnología de Destino |
|--------------------|--------------------------|--------------------------|-------------------------------------|--|-------------------|-----------------------|
| truncate table     | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | TRUNCATE TABLE TFG_DWH.DIM_FECHA cascade |                   | Oracle                |
| populate table     | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | beginCARGADIMIEMPO;end;                  |                   | Oracle                |

Ilustración 36 - Procedimiento populate DIM FECHA

```

create or replace
PROCEDURE CARGADIMIEMPO IS
tmpVar NUMBER;
FechaDesde date;
FechaHasta date;
FechaDesdeStr VARCHAR2(8);
err_num NUMBER;
err_msg VARCHAR2(255);
BEGIN
    tmpVar := 0;
    FechaDesde := TO_DATE('19941231','YYYYMMDD');
    FechaHasta := TO_DATE('20181231','YYYYMMDD');
    WHILE FechaDesde <= FechaHasta LOOP
    FechaDesdeStr := to_char( FechaDesde, 'YYYYMMDD' ) ;
    INSERT INTO DIM_FECHA
    (ID_FECHA, FECHA, AÑO, TRIMESTRE, MES, SEMANA, DIA, DIASEMANA, NTRIMESTRE, NMES, NMES3L,
    NSEMANA, NDIA, NDIASEMANA, FECHACORTE, FECHACORTEANT
    )
    VALUES
    (
    to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ) , 'YYYYMMDD'),
    FechaDesde,
    to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'YYYY'),
    to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'Q'),
    to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'MM'),
    to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'WW'),
    to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'DD'),
    to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'D'),
    'T'||to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'Q')||'/'||to_char(TO_DATE(
    FechaDesdeStr,'YYYYMMDD' ), 'YY'),
    LTRIM(RTRIM(to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'MONTH' ))),
    to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'MON'),
    'Sem '||to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'WW')||'/'||to_char(TO_DATE(
    FechaDesdeStr,'YYYYMMDD' ), 'YY'),
    to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'DD MON'),
    to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'DAY'),
    CASE
    WHEN to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'Q') = 1 THEN TO_DATE('31/03/' ||
    to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'YYYY'), 'DD/MM/YYYY')
    WHEN to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'Q') = 2 THEN TO_DATE('30/06/' ||
    to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'YYYY'), 'DD/MM/YYYY')
    WHEN to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'Q') = 3 THEN TO_DATE('30/09/' ||
    to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'YYYY'), 'DD/MM/YYYY')
    WHEN to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'Q') = 4 THEN TO_DATE('31/12/' ||
    to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'YYYY'), 'DD/MM/YYYY')
    END,
    CASE
    WHEN to_char(to_date(FechaDesdeStr,'YYYYMMDD' ), 'Q') = 1 THEN TO_DATE('31/12/' ||
    to_char(to_number(to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'YYYY'))-1), 'DD/MM/YYYY')
    WHEN to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'Q') = 2 THEN TO_DATE('31/03/' ||
    to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'YYYY'), 'DD/MM/YYYY')
    WHEN to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'Q') = 3 THEN TO_DATE('30/06/' ||
    to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'YYYY'), 'DD/MM/YYYY')
    WHEN to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'Q') = 4 THEN TO_DATE('30/09/' ||
    to_char(TO_DATE( FechaDesdeStr,'YYYYMMDD' ), 'YYYY'), 'DD/MM/YYYY')
    END
    );

    commit ;
    FechaDesde := FechaDesde + 1;
    END LOOP;
END CARGADIMIEMPO;

```

*Código 3 - Script generación DIM Tiempo*

Tras las dimensiones, se han diseñado las cargas correspondientes a los hechos.

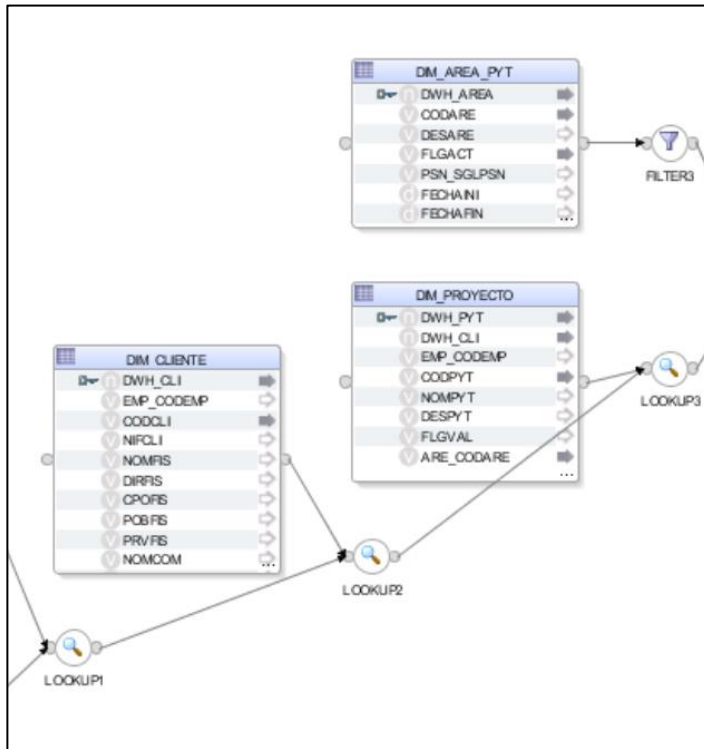


Ilustración 37- Mapping FACT Horas (1)

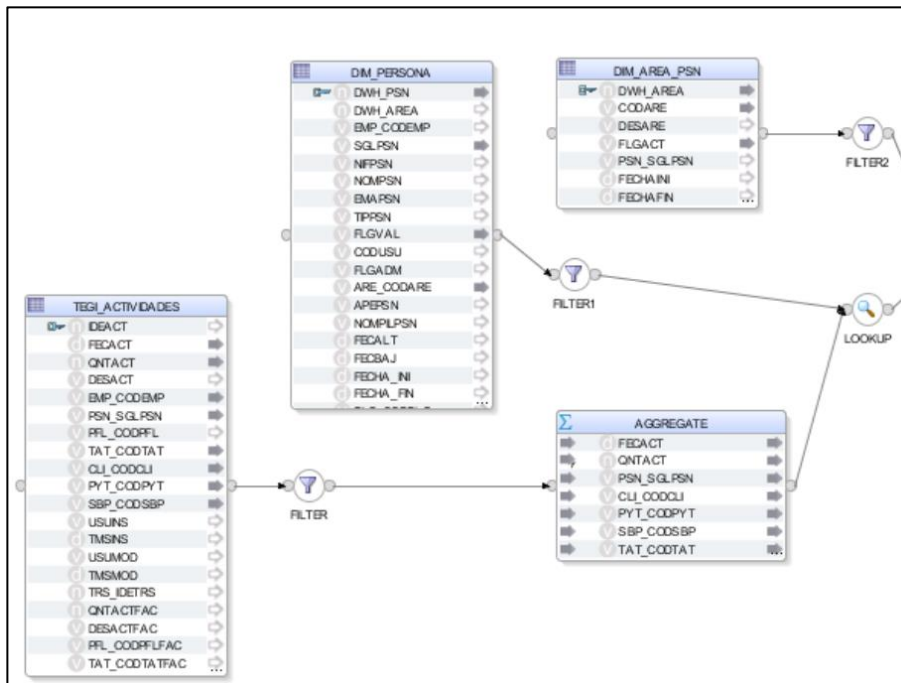


Ilustración 38 - Mapping FACT Horas (2)

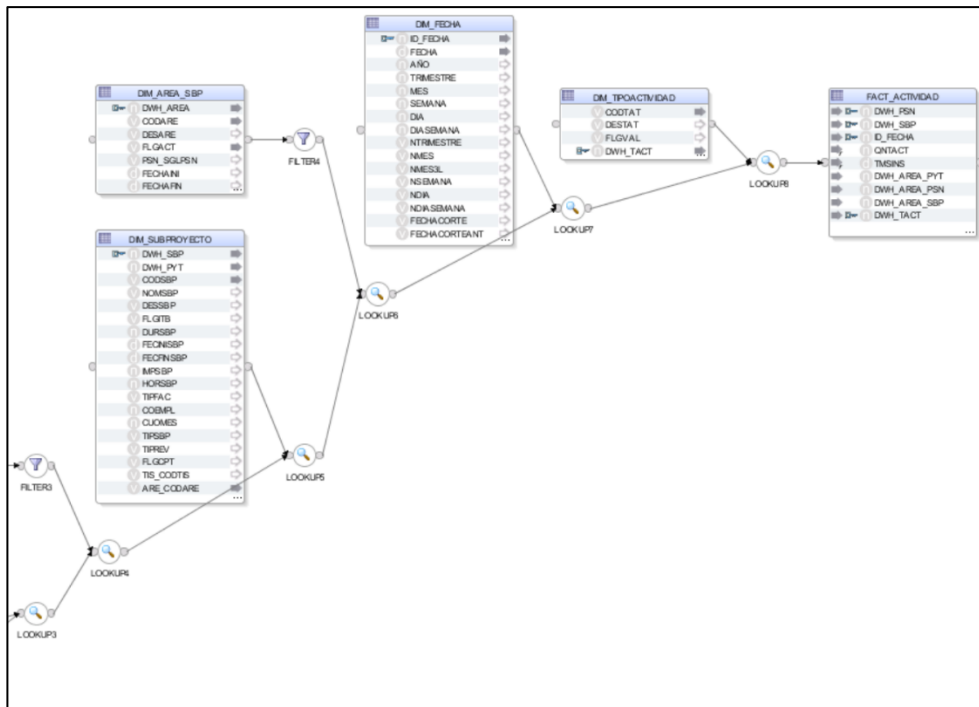


Ilustración 39- Mapping FACT Horas (3)

La Ilustración 37, Ilustración 38 e Ilustración 39 muestran el *mapping* diseñado para la tabla de hechos de actividad. Para añadir todas las claves subrogadas de las diferentes dimensiones ha sido necesario el uso de *look-ups* para comparar las claves naturales y obtener la secuencia creada en el momento de la inserción. En la parte izquierda de la imagen, se puede observar cómo se realiza una operación de agregación. En la tabla de origen se inserta una fila por cada actividad imputada en un día concreto a un subproyecto por parte de un empleado. La operación (Ilustración 40), agrupa todas las actividades de una persona en un mismo día, a un cliente con un subproyecto y que sea del mismo tipo de actividad. De esta forma se cumple con el nivel de detalle del dato definido en el sprint anterior.

| des        |              |          |        |             |                              |                         |                             |             |
|------------|--------------|----------|--------|-------------|------------------------------|-------------------------|-----------------------------|-------------|
| Atributos: |              |          |        |             |                              |                         |                             |             |
| Nombre     | Tipo de Dato | Longitud | Escala | Data format | Expresión                    | Indicación de Ejecución | Ubicación de Ejecución Fija | Agrupar por |
| FECTACT    | DATE         |          |        |             | TEGI_ACTIVIDADES.FECTACT     | Ninguna Indicación      |                             | Sí          |
| QNTACT     | NUMERIC      | 5        | 0      |             | SUM(TEGI_ACTIVIDADES.QNTACT) | Ninguna Indicación      |                             | No          |
| PSN_SGLPSN | VARCHAR      | 3        |        |             | TEGI_ACTIVIDADES.PSN_SGLPSN  | Ninguna Indicación      |                             | Sí          |
| CLI_CODCLI | VARCHAR      | 9        |        |             | TEGI_ACTIVIDADES.CLI_CODCLI  | Ninguna Indicación      |                             | Sí          |
| PYT_CODPYT | VARCHAR      | 9        |        |             | TEGI_ACTIVIDADES.PYT_CODPYT  | Ninguna Indicación      |                             | Sí          |
| SBP_CODSBP | VARCHAR      | 14       |        |             | TEGI_ACTIVIDADES.SBP_CODSBP  | Ninguna Indicación      |                             | Sí          |
| TAT_CODTAT | VARCHAR      | 9        |        |             | TEGI_ACTIVIDADES.TAT_CODTAT  | Ninguna Indicación      |                             | Sí          |

Ilustración 40 - Propiedades Agregación en el Mapping FACT\_Actividad

Finalmente, se hace una transformación del dato que contiene la cantidad de actividad. En origen esa información viene dada en segundos pero en el *data warehouse* es transformada a horas.

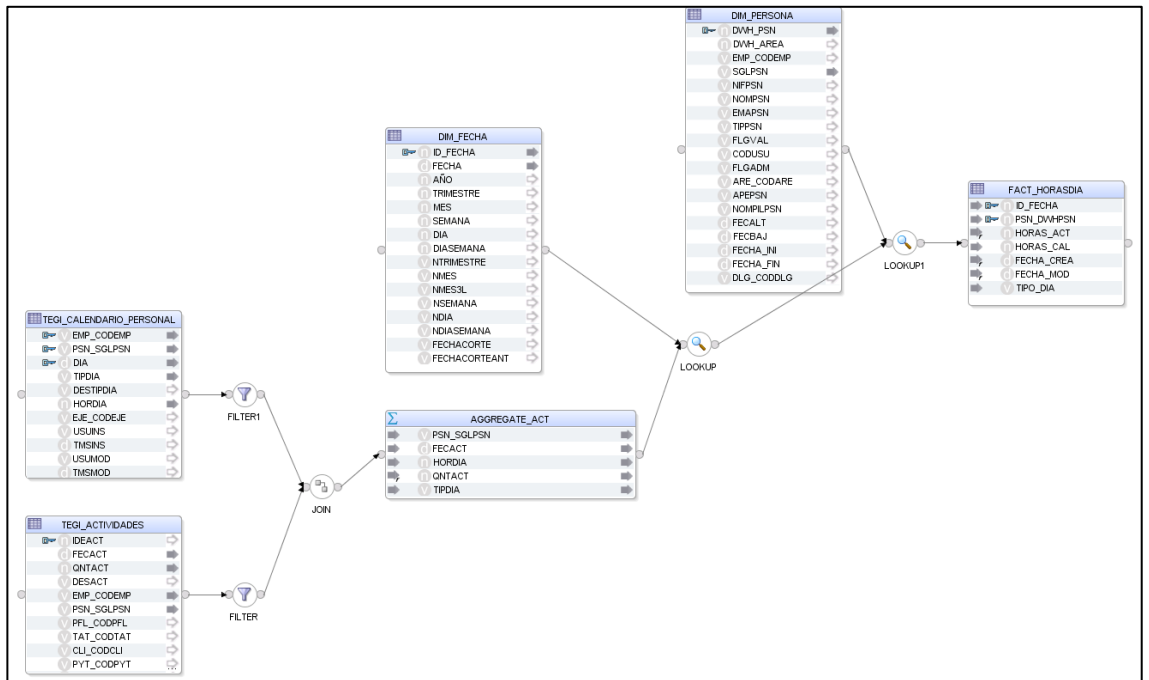


Ilustración 41 - Mapping FACT Horas día

Por último, la Ilustración 41 detalla el flujo de datos que se realiza para la carga de la tabla de hechos de horas día. En este caso parte de dos tablas de origen y realiza una agregación para una misma persona en un día obtener las horas planificadas en su calendario y las horas imputadas en el sistema ERP, además del tipo de día. Posteriormente se realizan *look-ups* con las dos dimensiones que tiene relación esta tabla de hechos.

Como en las cargas iniciales hay ciertos datos que están incompletos, se ha creado un proceso de carga que inserta en cada dimensión un registro que representa un dato genérico o desconocido. Así, cuando estos datos sean utilizados en la herramienta de BI, el usuario final verá un 'desconocido'. De esta forma, el equipo de desarrollo se asegura que la información presentada en los informes y cuadros de mando sea completa.

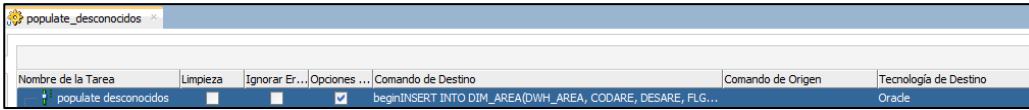


Ilustración 42 - Creación de Datos Genéricos

Tras el diseño de todos los flujos de datos, se ha repetido el mismo proceso que en la tarea anterior. Un escenario por cada *mapping* ha sido generado y se añadió a un plan de carga en ODI. La Ilustración 43 muestra el plan de carga, y la Ilustración 44 el resultado tras su ejecución.

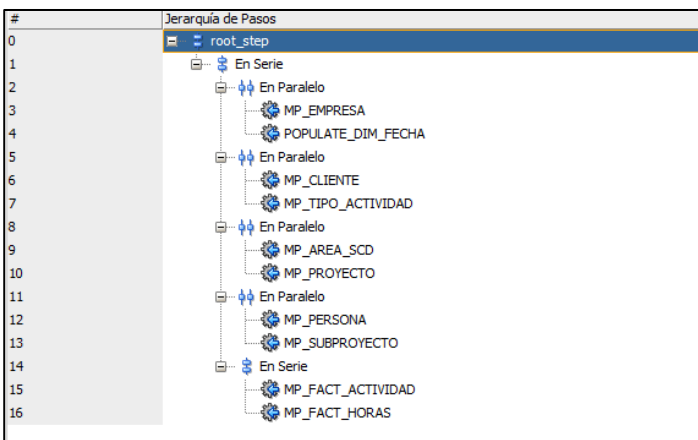


Ilustración 43 - Plan de Carga Inicial DWH

| #  | Jerarquía de Pasos                       | Estado | Duración |
|----|--|--------|----------|
| 0  | root_step                                | ✓      | 00:31    |
| 1  | En Serie                                 | ✓      | 00:31    |
| 2  | En Paralelo                              | ✓      | 00:05    |
| 3  | MP_EMPRESA                               | ✓      | 00:01    |
| 4  | POPULATE_DIM_FECHA                       | ✓      | 00:04    |
| 5  | En Paralelo                              | ✓      | 00:02    |
| 6  | MP_CLIENTE                               | ✓      | 00:01    |
| 7  | MP_TIPO_ACTIVIDAD                        | ✓      | 00:01    |
| 8  | Variable CASE: GLOBAL.Id_datos_genericos | ✓      | 00:00    |
| 9  | En Paralelo                              | ✓      | 00:03    |
| 10 | MP_AREA_SCD                              | ✓      | 00:02    |
| 11 | MP_PROYECTO                              | ✓      | 00:01    |
| 12 | En Paralelo                              | ✓      | 00:04    |
| 13 | MP_PERSONA                               | ✓      | 00:02    |
| 14 | MP_SUBPROYECTO                           | ✓      | 00:02    |
| 15 | En Serie                                 | ✓      | 00:17    |
| 16 | MP_FACT_ACTIVIDAD                        | ✓      | 00:09    |
| 17 | MP_FACT_HORAS                            | ✓      | 00:06    |

Ilustración 44 - Resultados Ejecución Plan de Carga Inicial

### 5.3.4 Revisión

| SPRINT                           | LISTA DE TAREAS  | HORAS EST. | HORAS REALES |
|----------------------------------|--|------------|--------------|
| 2                                | T1: Tomar decisión sobre el tipo de BBDD donde almacenar el DWH. | 3          | 1            |
|                                  | T2: Diseño Físico del DWH.                                       | 13         | 15           |
|                                  | T3: Creación de esquemas del DWH.                                | 13         | 8            |
|                                  | T4: Creación de capa de <i>Staging</i> .                         | 21         | 26           |
|                                  | T5: Desarrollo de mapeos de la capa de <i>Staging</i> .          | 21         | 21           |
|                                  | T6: Desarrollo de los mapeos iniciales del DWH.                  | 30         | 40           |
| <b>TOTAL DE HORAS ESTIMADAS:</b> |  | <b>101</b> | <b>111</b>   |

Tabla 20 - Resumen del Sprint 2

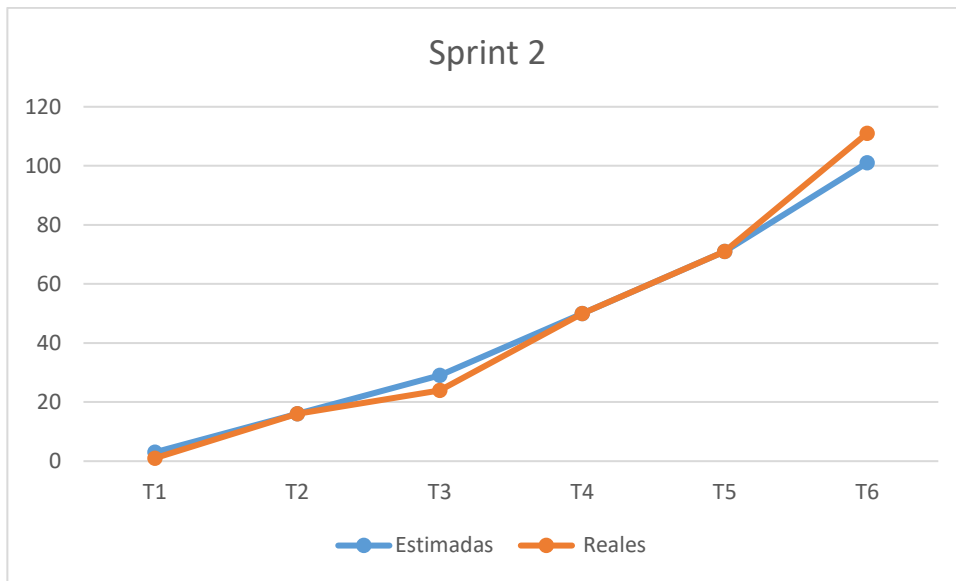


Figura 5 - Burn Up Sprint 2

### 5.3.5 Retrospectiva

Aunque este segundo Sprint ha sido optimista en la planificación, la última tarea es donde peor se ha estimado. Al ser la primera vez que se realizaba una planificación con respecto a desarrollos ETL por parte del equipo de desarrollo, se ha notado la falta de experiencia en el mismo. Se espera que para el próximo Sprint las estimaciones sean más exactas.

## 5.4 Sprint 3: Extracción Periódica de Datos

En este momento del trabajo, el DWH ya tiene su carga inicial completada, todos los datos que hasta el día de la carga se encontraban en el sistema ERP han sido añadidos. El siguiente paso ha sido configurar que las cargas se realicen de manera periódica, en función de las necesidades del propietario del producto. Además, ciertas modificaciones en los mapeos se han realizado en este sprint para que las necesidades de información dadas por la empresa sean cumplidas.

### 5.4.1 Tareas

| SPRINT                           | LISTA DE TAREAS   | HORAS EST. |
|----------------------------------|---|------------|
| 3                                | T1: Modificación de los <i>mappings</i> para ser ejecutados periódicamente. | 30         |
|                                  | T2: Creación de las tablas para generar auditorias en las cargas del DWH.   | 21         |
|                                  | T3: Instalación y Configuración de Agente <i>Standalone</i> de ODI.         | 13         |
|                                  | T4: Configuración de servicio de notificación por Email.                    | 13         |
| <b>TOTAL DE HORAS ESTIMADAS:</b> |   | <b>77</b>  |

Tabla 21 - Planificación Sprint 3

Tras la realización de las tareas, el DWH está configurado para recibir cargas de datos de forma periódica realizando las tareas 1 y 3 (5.4.2.1). Además, han sido creadas tablas auxiliares de bases de datos, en la tarea 2, para que los administradores del DWH puedan hacer un seguimiento de las cargas realizadas (5.4.2.2). Finalmente, con la consecución de la tarea 4, los usuarios relacionados con el DWH serán notificados por email cada vez que un plan de carga se haya ejecutado correctamente.

## 5.4.2 Resultados del Sprint

### 5.4.2.1 Cargas Periódicas de Datos

Dentro del proceso de automatización de las cargas en el *data warehouse*, hay ciertas dimensiones que podrán tener modificaciones en sus datos a lo largo del tiempo. Dentro de los varios tipos de dimensiones, detallados en la sección 3.2.3 de este TFG, hay algunas que serán del tipo 2 (se añade una nueva fila por cada registro que se haya modificado en origen). Si se pone el caso práctico de que un área cambia de responsable, el almacén de datos deberá tener constancia del cambio. Por ello sería interesante para el negocio poder ver para una misma área, la actividad que tuvo con un responsable y con el nuevo. Manteniendo un registro para cada configuración de las áreas, personas, proyectos, subproyectos o clientes el negocio podrá obtener información más detallada sobre su producción.

Por tanto, para cumplir con esta necesidad de información se deben cambiar ligeramente los mapeos de ciertas dimensiones. La Ilustración 45 muestra cómo se haría el nuevo *mapping* de la dimensión área. Lo que este nuevo flujo de datos hace es comparar los registros que ya están guardados en la dimensión con los que vienen de origen. Al estar los datos de origen en la capa de *Staging*, ambos están en una base de datos Oracle y es posible hacer la operación *minus*. Tras esta operación, se obtienen los registros que tienen alguna diferencia con la tabla de dimensión, ya sea por alguna modificación o por ser nuevos. En caso de ser nuevos, éstos se insertan directamente en la dimensión. Si son registros modificados, también se insertarán en la dimensión, pero con el *join* que se puede observar en la parte superior de la imagen se cierra el registro anterior que se encuentra activo en la dimensión. De esta forma se inserta una nueva fila con los nuevos datos y se deja inactivo, pero persistiendo, el anterior.

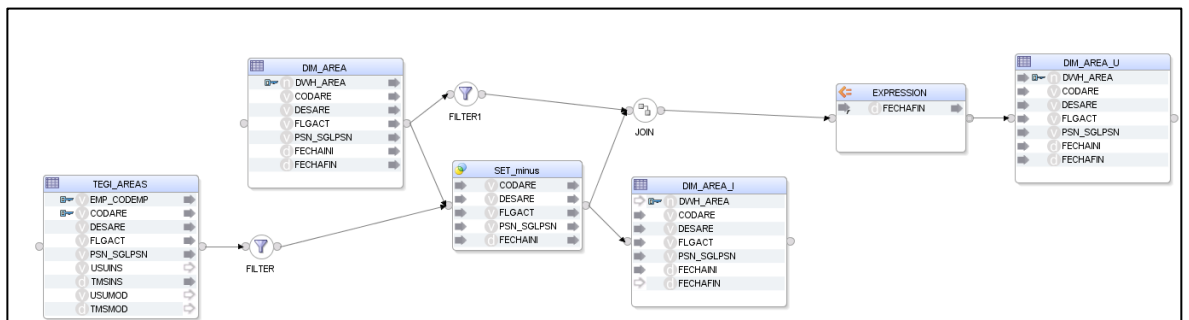


Ilustración 45 - Mapping DIM Area con SCD

Sin embargo, Oracle Data Integrator ofrece una novedad en la nueva versión de su herramienta. No será necesario modificar todos los mapeos como se ha mostrado en la Ilustración 45 ya que, modificando el modelo de datos, importado en el Sprint anterior, es posible indicar que las dimensiones son lentamente cambiantes. En la Ilustración 46 se puede observar como ODI en el modelo de las tablas importadas del DWH se pueden identificar con un tipo OLAP, en este caso dimensión de cambio lento.

| Almacén de Datos [Modelo: TRG_DWH ▶ Submodelo: Global] |          |            |                           |
|--|----------|------------|---------------------------|
| Nombre:  | DIM_AREA | Alias:     | DIM_AREA                  |
| Tipo de Almacén de Datos:                              | Tabla    | Tipo OLAP: | Dimensión de Cambio Lento |
| Nombre del Recurso:                                    | DIM_AREA |            |                           |

Ilustración 46 - Modificación Modelo de datos de DIM Área

Una vez el modelo ha sido modificado, basta con configurar los atributos de las tablas para modelar el comportamiento de la dimensión. En este caso se puede especificar la clave de sustitución, o clave subrogada, y la clave natural para hacer la comparación con origen. Además, se concreta que cambios se tendrían que realizar en caso de encontrar atributos modificados. Finalmente se detalla que variables sirven para almacenar la fecha de inicio y fin del registro, y del *flag* que muestra la validez del dato.

| ... | Nombre     | Tipo     | Lo... | ... | No Nulo                             | Comportamiento de SCD            |
|-----|------------|----------|-------|-----|-------------------------------------|----------------------------------|
| 1   | DWH_AREA   | NUMBER   |       |     | <input checked="" type="checkbox"/> | Clave de Sustitución             |
| 2   | CODARE     | VARCHAR2 | 9     |     | <input type="checkbox"/>            | Clave Natural                    |
| 3   | DESARE     | VARCHAR2 | 50    |     | <input type="checkbox"/>            | Sobrescribir en Caso de Cambio   |
| 4   | FLGACT     | VARCHAR2 | 1     |     | <input type="checkbox"/>            | Indicador de Registro Actual     |
| 5   | PSN_SGLPSN | VARCHAR2 | 3     |     | <input type="checkbox"/>            | Agregar Fila en Caso de Cambio   |
| 6   | FECHAINI   | DATE     |       |     | <input type="checkbox"/>            | Registro de Hora de Inicio       |
| 7   | FECHAFIN   | DATE     |       |     | <input type="checkbox"/>            | Registro de Hora de Finalización |

Ilustración 47 - Modificación Atributos DIM Área

Una vez todos los modelos han sido modificados, es necesario cambiar las propiedades de los mapeos para que se les atribuya el comportamiento de dimensión de cambio lento. Siguiendo con el ejemplo de la dimensión área, las nuevas propiedades son las mostradas en la Ilustración 48, donde se mantienen las asignaciones creadas en el Sprint anterior pero especificando como módulo de conocimiento de integración el creado por Oracle para dimensiones de cambio lento (Ilustración 49).

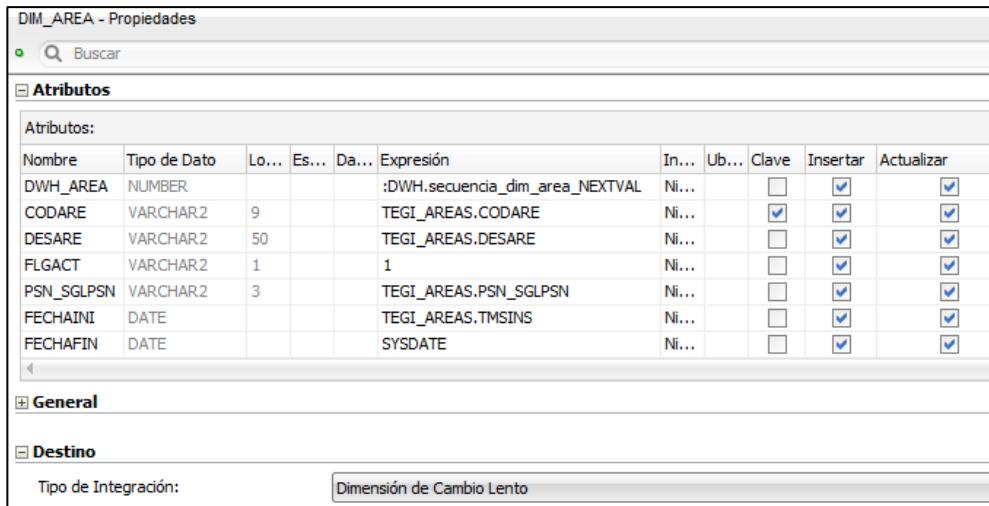


Ilustración 48 - Cambio propiedades mapping DIM Área

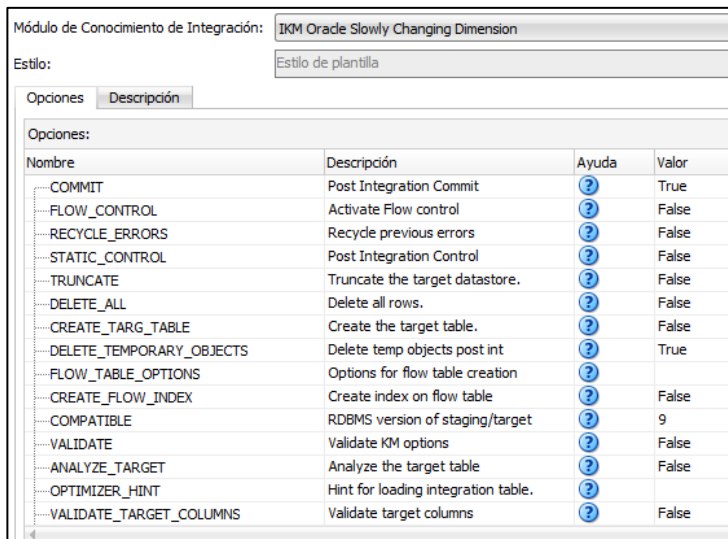


Ilustración 49 - Cambio de propiedades mapping DIM Área en IKM

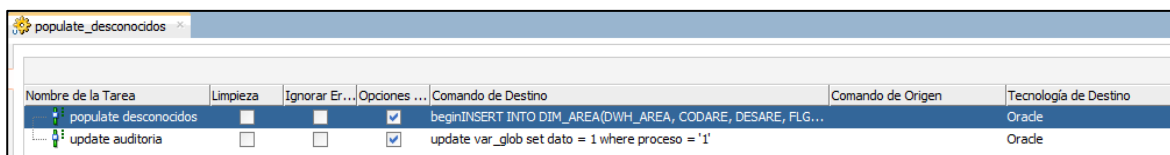
Para el resto de dimensiones se ha realizado el mismo proceso descrito. Tras la consecución de esta tarea, todas las dimensiones de cambio lento están ahora modeladas con ese comportamiento.

#### 5.4.2.2 Tablas para Generar Auditorias en las cargas del DWH

Siguiendo con la preparación del almacén de datos para las cargas periódicas, es necesaria la creación de un esquema de base de datos, guardado en el DWH, que se utilice como tablas auxiliares para un correcto mantenimiento del DWH. Para ello se ha creado un esquema donde diferentes tablas aportarán utilidad en los procesos ETL del DWH.

Se ha creado una tabla en la cual se almacenan los detalles de cada plan de carga, que ha sido ejecutado. Así, en caso de fallo se podría tener detalle de que plan de carga falló en ejecución.

Además de esta tabla de registros, se ha creado una tabla donde se almacenan los *flags* o indicadores que serán utilizados en las cargas. Como es el caso de la inserción de los datos genéricos. Ahora, cuando se ejecute el plan de carga consultará directamente a la tabla de *flags* del esquema de auditoría para saber si los datos genéricos ya han sido cargados.



| Nombre de la Tarea    | Limpieza                 | Ignorar Er...            | Opciones ...                        | Comando de Destino   | Comando de Origen | Tecnología de Destino |
|-----------------------|--------------------------|--------------------------|-------------------------------------|--|-------------------|-----------------------|
| populate desconocidos | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | beginINSERT INTO DIM_AREA(DWH_AREA, CODARE, DESARE, FLG... |                   | Oracle                |
| update auditoria      | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | update var_glob set dato = 1 where proceso = '1'           |                   | Oracle                |

Ilustración 50 - Carga de Datos Genéricos con tabla Auditoría

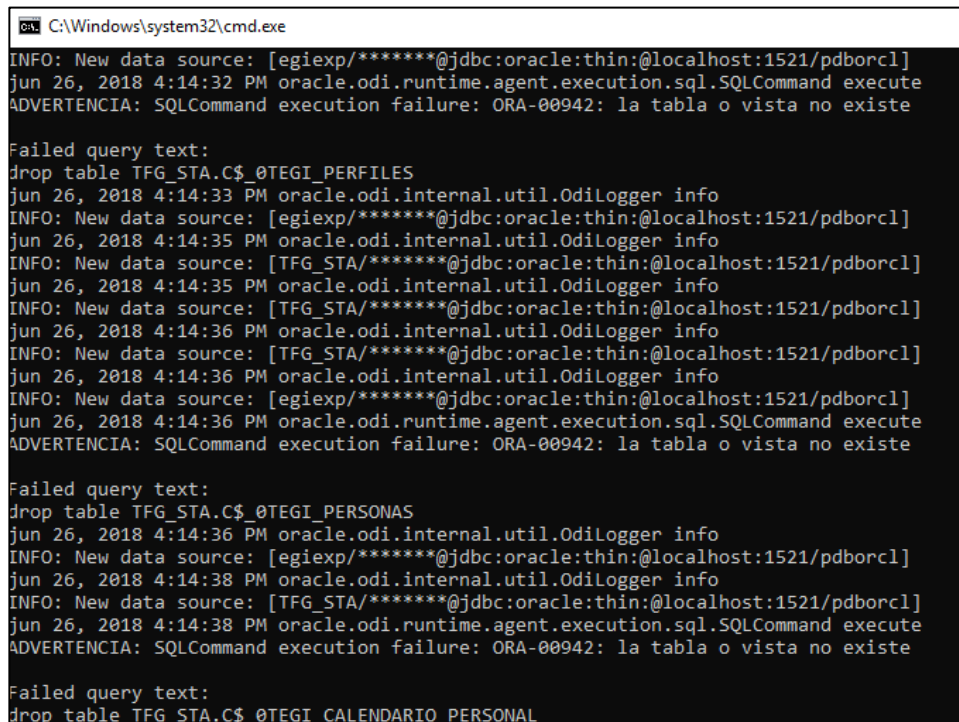
### 5.4.2.3 Agente *Standalone* de ODI configurado para ejecución de cargas

Para la ejecución de un plan de carga es necesario el uso de agentes de ODI. Un agente de ODI es un proceso de Java que es capaz de ejecutar, coordinar y orquestar los escenarios o planes de carga desarrollados en Oracle Data Integrator. Mientras se han desarrollado los mapeos, éstos se pueden ejecutar directamente sin hacer uso de agentes. Pero una vez llegado al punto de automatización de las cargas, es necesaria la instalación de un agente para que ejecute de forma periódica el plan de carga del *data warehouse*. Dentro de Oracle Data Integrator, existen varios tipos de agentes:

- ***Standalone***: Ligero, fácil de desplegar y es ejecutado en la máquina virtual de Java.
- ***Collocated***: Similar al *standalone*, ejecutado en un WebLogic Server [32] y administrado por un AdminServer [33].
- ***J2EE***: Similar al *standalone* también ejecutado en un WebLogic Server además de tener un un servicio web que publique los contextos y escenarios almacenados en el repositorio de ODI.

Para el TFG el equipo de desarrollo ha optado por la instalación de un agente *standalone* por su simplicidad y debido al contexto en el que se realiza este proyecto.

Una vez el agente ha sido instalado, se puede observar en la Ilustración 51 cómo ejecuta los escenarios del plan de carga.



```
C:\Windows\system32\cmd.exe
INFO: New data source: [egiexp/*****@jdbc:oracle:thin:@localhost:1521/pdborcl]
jun 26, 2018 4:14:32 PM oracle.odi.runtime.agent.execution.sql.SQLCommand execute
ADVERTENCIA: SQLCommand execution failure: ORA-00942: la tabla o vista no existe

Failed query text:
drop table TFG_STA.C$_0TEGI_PERFILES
jun 26, 2018 4:14:33 PM oracle.odi.internal.util.OdiLogger info
INFO: New data source: [egiexp/*****@jdbc:oracle:thin:@localhost:1521/pdborcl]
jun 26, 2018 4:14:35 PM oracle.odi.internal.util.OdiLogger info
INFO: New data source: [TFG_STA/*****@jdbc:oracle:thin:@localhost:1521/pdborcl]
jun 26, 2018 4:14:35 PM oracle.odi.internal.util.OdiLogger info
INFO: New data source: [TFG_STA/*****@jdbc:oracle:thin:@localhost:1521/pdborcl]
jun 26, 2018 4:14:36 PM oracle.odi.internal.util.OdiLogger info
INFO: New data source: [TFG_STA/*****@jdbc:oracle:thin:@localhost:1521/pdborcl]
jun 26, 2018 4:14:36 PM oracle.odi.internal.util.OdiLogger info
INFO: New data source: [egiexp/*****@jdbc:oracle:thin:@localhost:1521/pdborcl]
jun 26, 2018 4:14:36 PM oracle.odi.runtime.agent.execution.sql.SQLCommand execute
ADVERTENCIA: SQLCommand execution failure: ORA-00942: la tabla o vista no existe

Failed query text:
drop table TFG_STA.C$_0TEGI_PERSONAS
jun 26, 2018 4:14:36 PM oracle.odi.internal.util.OdiLogger info
INFO: New data source: [egiexp/*****@jdbc:oracle:thin:@localhost:1521/pdborcl]
jun 26, 2018 4:14:38 PM oracle.odi.internal.util.OdiLogger info
INFO: New data source: [TFG_STA/*****@jdbc:oracle:thin:@localhost:1521/pdborcl]
jun 26, 2018 4:14:38 PM oracle.odi.runtime.agent.execution.sql.SQLCommand execute
ADVERTENCIA: SQLCommand execution failure: ORA-00942: la tabla o vista no existe

Failed query text:
drop table TFG_STA.C$_0TEGI_CALENDARIO_PERSONAL
```

Ilustración 51 - Consola del Agente Standalone de ODI

Finalmente, ya es posible programar el agente de ODI para que ejecute los planes de carga con la frecuencia acordada con el propietario de producto. La Ilustración 52 muestra la configuración del plan de carga de la capa de *staging*.

Ilustración 52 - Programación de Plan de Carga en ODI

#### 5.4.2.4 Servicio de Notificación por Email

Por último, una de las funcionalidades de la herramienta ETL utilizada en el TFG es que permite sincronizar una cuenta de correo electrónico para notificar mediante email. Como parte de uno de los requisitos, se ha configurado el plan de carga del *data warehouse* para que una vez finalizado comunique al administrador que ha concluido las cargas periódicas. La Ilustración 53 detalla cómo se ha configurado en la herramienta ETL para que envíe emails notificando el fin del plan de carga. Este caso se ha utilizado un servidor de correo electrónico local para hacer las pruebas pertinentes. En caso de querer utilizarlo con una cuenta corporativa, es necesario tener los permisos de administrador para obtener los datos del servidor de correo.

| Nombre del Paso                       |   |
|---------------------------------------|---|
| OdiSendMail 1                         |   |
| Parámetro                             | Valor                                     |
| Servidor de Correo                    | localhost                                 |
| Puerto                                | 25  |
| Protocolo                             | smtp                                      |
| Usar Autenticación                    | No  |
| Mecanismos de Autenticación           | LOGIN                                     |
| Nombre de Usuario para Autenticación. | dbadmin@localhost.com                     |
| Contraseña para Autenticación         | *****                                     |
| De                                    | dbadmin@localhost.com                     |
| Para                                  | [info@localhost.com]                      |
| Cc                                    | <input type="checkbox"/>                  |
| CCO                                   | <input type="checkbox"/>                  |
| Asunto                                | Carga al DWH completada                   |
| Anexo                                 | <input type="checkbox"/>                  |
| Cuerpo del Mensaje                    | Carga completada <%=odiRef.getSysDate()%> |

Ilustración 53 - Configuración ODI para enviar mail

### 5.4.3 Revisión

| SPRINT                           | LISTA DE TAREAS   | HORAS     | HORAS     |
|----------------------------------|---|-----------|-----------|
|                                  |   | EST.      | REALES    |
| 3                                | T1: Modificación de los <i>mappings</i> para ser ejecutados periódicamente. | 30        | 32        |
|                                  | T2: Creación de las tablas para generar auditorias en las cargas del DWH.   | 21        | 15        |
|                                  | T3: Instalación y Configuración de Agente <i>Standalone</i> de ODI.         | 13        | 8         |
|                                  | T4: Configuración de servicio de notificación por Email.                    | 13        | 8         |
| <b>TOTAL DE HORAS ESTIMADAS:</b> |   | <b>77</b> | <b>63</b> |

Tabla 22 - Resumen Sprint 3

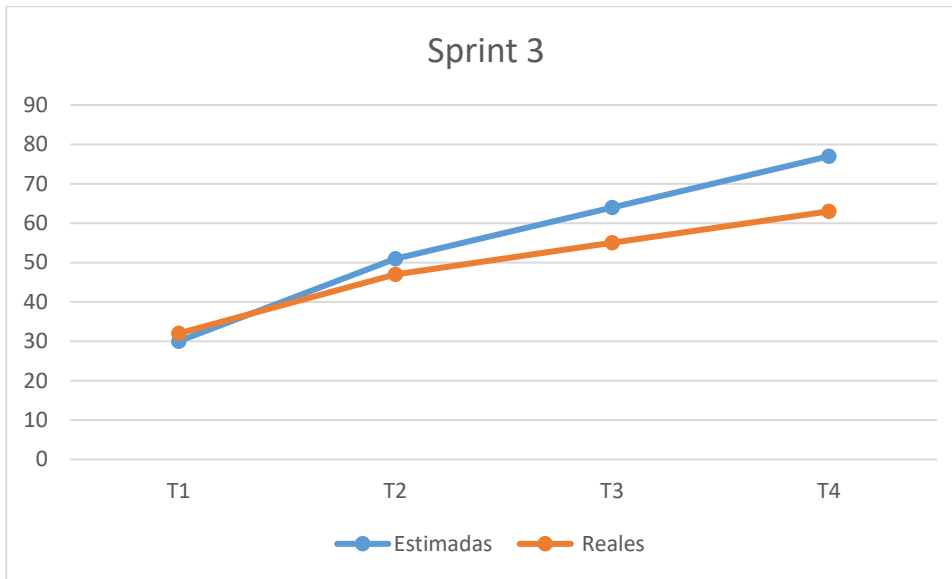


Figura 6 - Burn Up Sprint 3

#### 5.4.4 Retrospectiva

Primer Sprint dentro del desarrollo del TFG donde se ha sido pesimista en la planificación. Se ha notado la adquisición de experiencia en el control de la herramienta ETL por parte del equipo de desarrollo, ya que en la tarea 1 se ha prácticamente cumplido con el tiempo estimado inicialmente. Sin embargo el resto de tareas se han realizado en menor tiempo del estimado el comienzo del Sprint.

### 5.5 Sprint 4: Explotación de Datos

En este último sprint se ha realizado todo lo correspondiente con la capa de *Business Intelligence*. Primero se ha instalado la infraestructura necesaria y el software que será utilizado para la creación de informes y cuadros de mando. Después se han creado los análisis y cuadros de mando siguiendo las necesidades de información detalladas en el Sprint 1.

#### 5.5.1 Planificación del Sprint

Las dos últimas historias de usuario que quedaban en la pila de producto han sido añadidas a la pila del sprint 4.

| SPRINT                           | LISTA DE TAREAS   | HORAS EST. |
|----------------------------------|---|------------|
| 4                                | T4.1: Instalación de Infraestructura de Oracle para OBI.                          | 13         |
|                                  | T4.2: Importación del modelo de datos del DWH para creación de repositorio de BI. | 21         |
|                                  | T5.1: Desarrollo de Informes con OBI.   | 21         |
|                                  | T5.2: Desarrollo de Informes con ODV.   | 8          |
|                                  | T5.3: Configuración del Cuadro de Mando.  | 21         |
| <b>TOTAL DE HORAS ESTIMADAS:</b> |   | <b>84</b>  |

Tabla 23 - Planificación Sprint 4

Con la consecución de las tareas mostradas en la Tabla 23 se ha conseguido cumplir con todas las historias de usuario que se encontraban en la pila de producto. Con la tarea 4.1 se ha completado la instalación de la infraestructura necesaria para poder hacer uso de la herramienta BI (5.5.2.1). El acceso a los datos del DWH a través de OBI se ha realizado con la tarea 4.2. Finalmente, todos los análisis e informes que fueron identificados tras el Sprint 1 han sido realizados en las tareas 5.1 y 5.2. Para concluir con el proyecto, el cuadro de mando de Clientes, Empleados y Áreas (5.5.2.5) han sido generados con la tarea 5.3.

## 5.5.2 Desarrollo de Tareas

### 5.5.2.1 Instalación de Infraestructura de Oracle para OBI

Antes de poder utilizar los datos almacenados en el *data warehouse* es necesario instalar la infraestructura que requiere Oracle Business Intelligence. Para poder desplegar los servicios de OBI, Oracle Web Logic Server 12c se ha instalado en la misma máquina donde luego se ha instalado OBI 12c. Antes de poder ejecutar los servicios BI, es necesario configurar OBI con la base de datos y con el servidor de aplicaciones previamente instalado.

Al igual que con Oracle Data Integrator, es preciso crear un repositorio donde todos los metadatos relacionados con el entorno BI serán almacenados. Una vez el repositorio ha sido instalado, se podrá acceder al portal de Oracle Business Intelligence tal y como muestra la Ilustración 58.

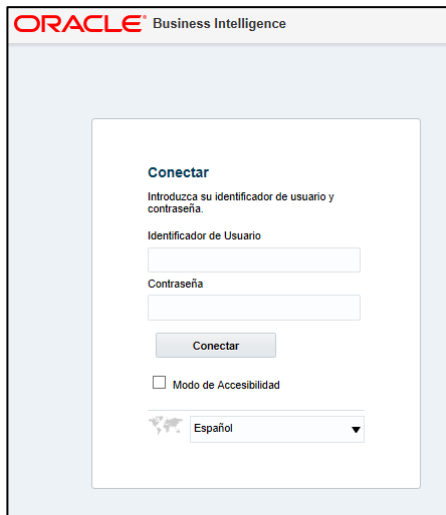


Ilustración 54 - Acceso a Oracle Business Intelligence

### 5.5.2.2 Importación del modelo de datos del DWH para creación de repositorio de BI

Antes de crear los informes, es necesario crear un repositorio de datos con el que trabajar en OBI. En este repositorio se definen las fuentes de datos con las cuales el servidor BI de Oracle va a trabajar. En el caso del TFG, la fuente de datos será el *data warehouse* creado y desarrollado en los sprints 2 y 3. Para la configuración de este repositorio, OBI cuenta con una herramienta de administración en el cual se especifican las fuentes a nivel físico, lógico y de presentación.

Para la capa física ha sido necesario especificar la conexión al almacén de datos. Una vez se ha conectado, toda la información sobre el modelo del DWH está cargada en el repositorio. La Ilustración 56 muestra como se ha añadido la conexión al esquema del DWH y se han importado las dimensiones y hechos. Como se puede apreciar en la ilustración, se han creado alias para las dimensiones y hechos siguiendo una numeración. Sin embargo, para la dimensión área se han creado 3 alias. Esto se ha hecho para remarcar la diferencia entre el área de una persona, un proyecto y un subproyecto. Posteriormente en la capa lógica, esto se modelará para cumplir con las operaciones de *drill-down* en los cuadros de mando.

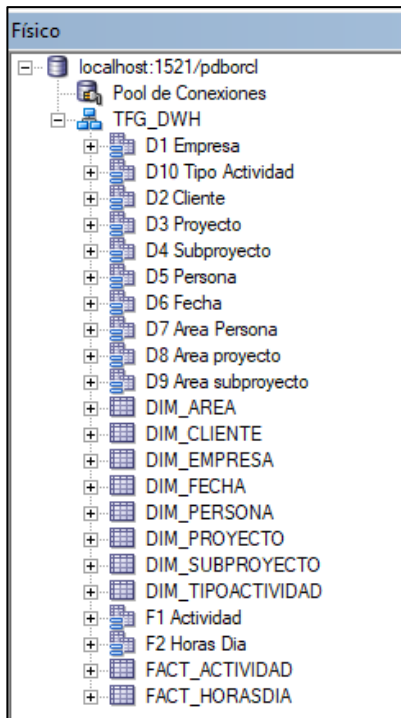


Ilustración 55- Capa Física del Repositorio BI

El principal cambio que se puede apreciar con la creación de los alias es que se han formateado los nombres de las dimensiones y hechos para que sean más legibles para los usuarios de la aplicación BI.

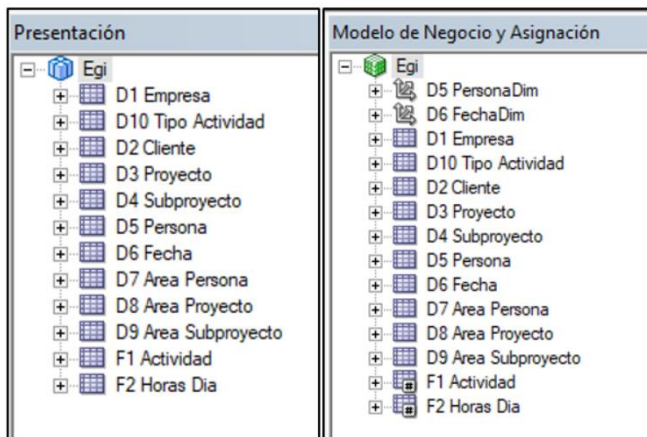


Ilustración 56 - Capa Lógica y de Presentación del Repositorio BI

En la capa de Negocio, resumida en la Ilustración 56, se han creado dos dimensiones lógicas (D5 PersonaDim y D6 FechaDim) en la cual se detallan las jerarquías por las que se podrá hacer *drill down* posteriormente en los informes.

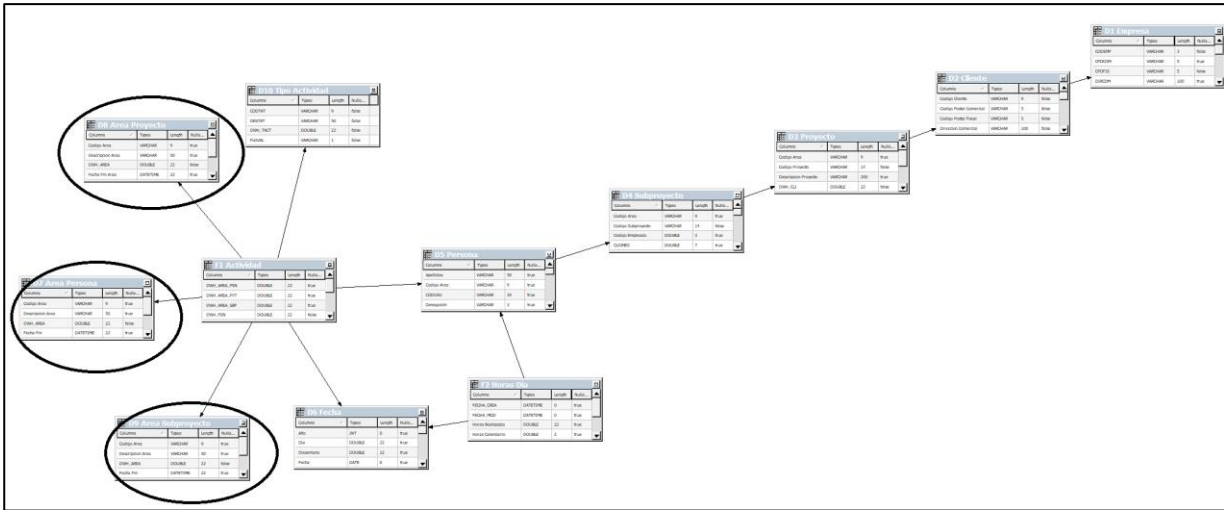


Ilustración 57 - Modelo lógico del DWH en el Repositorio BI

La Ilustración 57 muestra como se ha modelado en la capa lógica las dimensiones y hechos para que sean utilizados por el servidor BI de Oracle. Las dimensiones remarcadas representan tres copias lógicas de la dimensión área, para diferenciar entre el área de un proyecto, subproyecto y un empleado en los cuadros de mando. Este cambio con respecto al modelo original no es aplicado en el *data warehouse*, si no es orientado al repositorio BI que utiliza la herramienta de Inteligencia de Negocio.

Tras haber configurado el repositorio, éste ha sido subido al servidor BI poder crear los informes y cuadros de mando. La Ilustración 58 muestra como desde la herramienta Oracle Business Intelligence ya se tiene acceso al repositorio recién subido, en este caso llamado Egi por ser el nombre que recibe el ERP de la compañía.



Ilustración 58 - Selección de Área Temática en OBI

### 5.5.2.3 Informes creados en OBI

Siguiendo los requisitos tomados en el Sprint 1, el equipo de desarrollo ha decidido dividir el desarrollo de los informes por tres temas: clientes, empleados y áreas. Por un lado se van a desarrollar los análisis correspondientes a los clientes de la empresa y los proyectos en proceso. Por otro lado se hará un panel donde el responsable de producción podrá tener una visión general de cada empleado. Y un último panel donde ver el estado actual y pasado de las diferentes áreas de la empresa.

| Nmes  | Codigo Cliente | Codigo Proyecto | Horas |
|-------|----------------|-----------------|-------|
| ABRIL | ████           | ████████        | 6     |
|       |                | ████            | 31    |
|       |                | ████            | 2     |
|       |                | ████            | 194   |
|       |                | ████            | 2     |
|       |                | ████████        | 67    |
|       |                | ████            | 44    |
|       |                | ████            | 3     |
|       |                | ████            | 12    |
|       |                | ████            | 64    |
|       |                | ████████        | 137   |
|       |                | RES             | 24    |

| Nmes  | Codigo Cliente | Codigo Proyecto | Descripcion Subproyecto | Nombre Persona   | Horas |
|-------|----------------|-----------------|-------------------------|------------------|-------|
| ABRIL | ████           | ████            | ████████████████        | ████████████████ | 67    |

Volver - Atrás - Crear Enlace de Marcador

Ilustración 59 - Ejemplo de Informe con Drill Down

La Ilustración 59 muestra el análisis correspondiente a la parte de clientes en el cual se puede hacer *drill down*. Con este análisis, el propietario de producto puede consultar en un año y/o mes concreto las horas que han sido imputadas a cada cliente, pudiendo acceder a detalle de proyecto, subproyecto y empleado.

| Fecha      | Horas Realizadas | Horas Calendario |
|------------|------------------|------------------|
| 03/04/2017 | 8,50             | 8,00             |
| 04/04/2017 | 8,50             | 8,00             |
| 05/04/2017 | 8,50             | 8,00             |
| 06/04/2017 | 8,50             | 8,00             |
| 07/04/2017 | 6,00             | 8,00             |
| 10/04/2017 | 8,50             | 8,00             |
| 11/04/2017 | 8,50             | 8,00             |
| 12/04/2017 | 6,00             | 8,00             |
| 17/04/2017 | 8,50             | 8,00             |
| 18/04/2017 | 8,50             | 8,00             |
| 19/04/2017 | 8,50             | 8,00             |
| 20/04/2017 | 8,50             | 8,00             |
| 21/04/2017 | 6,00             | 8,00             |
| 24/04/2017 | 8,50             | 8,00             |
| 25/04/2017 | 8,50             | 8,00             |

Ilustración 60 - Análisis sobre horas de empleados

Continuando con la parte de empleados, se ha desarrollado un análisis que cubre la necesidad de conocer si un empleado está realizando las mismas horas que le han sido asignadas. Para ello, el informe (Ilustración 60) remarca en color rojo si el número de horas registradas en el sistema ERP son menores a las planificadas por el responsable de área.

Nmes ABRIL ▼

| Semana    | Codigo Proyecto | Codigo Subproyecto | QNTACT |
|-----------|-----------------|--------------------|--------|
| Sem 14/17 | ██████          | AUF2017            | 9      |
|           | ██████          | PVT2014            | 5      |
|           | ██████          | PVT2017            | 6      |
|           | ██████          | PHL                | 22     |
| Sem 15/17 | ██████          | AUF2017            | 11     |
|           | ██████          | FOR2017            | 2      |
|           | ██████          | SOPDESObI          | 12     |
| Sem 16/17 | ██████          | AUF2017            | 16     |
|           | ██████          | PVT2014            | 2      |
|           | ██████          | SOPDESObI          | 18     |
|           | ██████          | PHL                | 6      |
| Sem 17/17 | ██████          | AUF2017            | 18     |
|           | ██████          | PVT2014            | 19     |
|           |                 | PVT2017            | 5      |

Ilustración 61 - Informe con detalle de semanas por empleado

El informe expuesto en la Ilustración 61 detalla la dedicación de un empleado en concreto a los diferentes proyectos y subproyectos a nivel de semana, pudiéndose hacer *drill down* hasta detalle de día. Mientras que en la Ilustración 62 se muestra un análisis realizado que compara las diferentes áreas de la empresa.

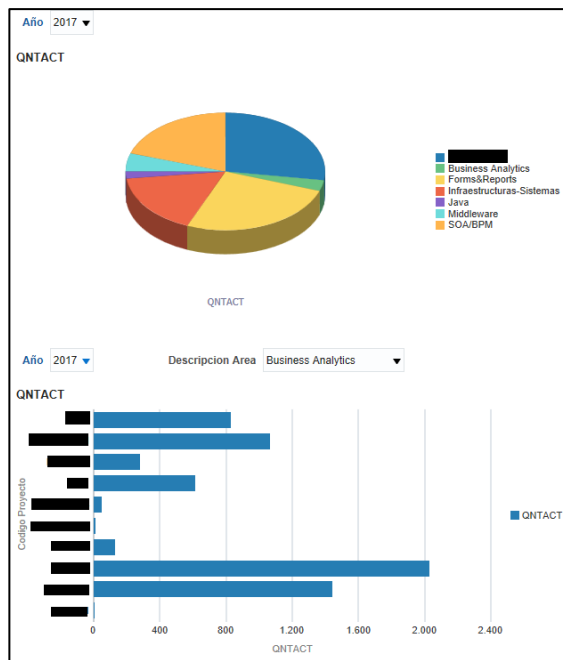


Ilustración 62 - Informe Rendimiento Área

#### 5.5.2.4 Mapa interactivo sobre Clientes

También se ha hecho uso de esta herramienta debido a la facilidad que muestra, en comparación con OBI, a la hora de crear informes. Oracle Data Visualization está dirigida a usuarios sin conocimientos avanzados de BI que quieren crear análisis sencillos y visuales de datos. Esta herramienta viene incluida dentro de la instalación de OBI y hace uso del mismo repositorio de datos del servidor BI de Oracle.

Se ha hecho uso de ODV para la creación de un mapa en el cual se situarán a los diferentes clientes de la compañía. Para poder situarlos en el mapa, se ha tenido que cargar dentro del servidor BI un archivo .xls con las coordenadas de cada municipio, haciendo una unión con la dimensión cliente. La Ilustración 63 muestra la unión de las dos fuentes de datos que se ha realizado en Oracle Data Visualization.

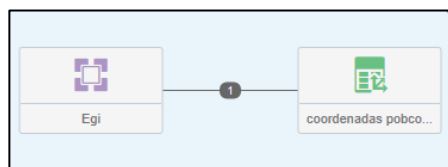


Ilustración 63 - Unión de datos en ODV

La Ilustración 64 muestra el resultado final. Los puntos azules representan los diferentes clientes situados por su dirección comercial y el tamaño del círculo son las horas facturadas a cada uno de ellos.

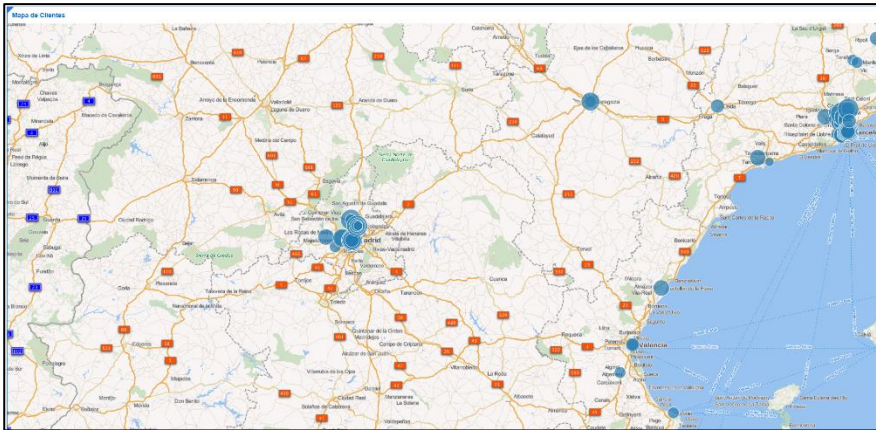


Ilustración 64 - Mapa de Clientes en ODV

### 5.5.2.5 Cuadros de Mando Finales

Tras desarrollar los diferentes análisis en OBI y ODV, se han organizado en un cuadro de mando que muestra, en 3 pestañas, la información requerida por el responsable de producción.

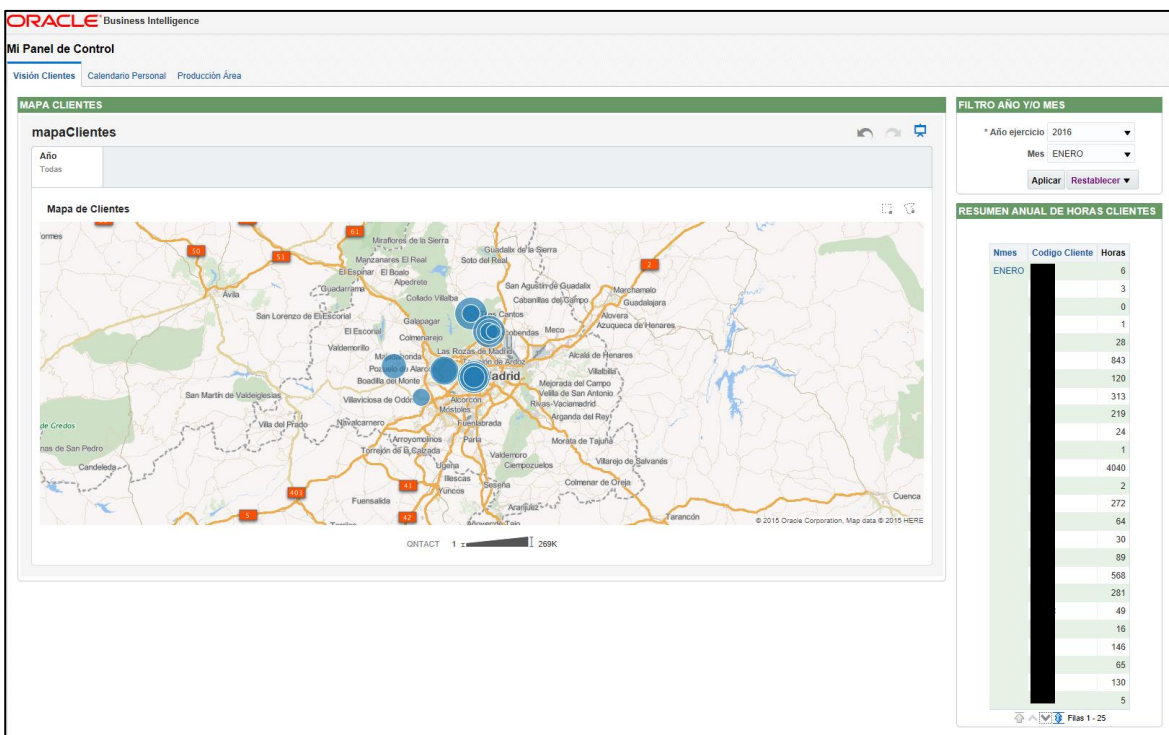


Ilustración 65- Cuadro de Mando de Clientes

Para la parte de clientes se ha desarrollado un cuadro de mando donde muestra el mapa desarrollado con ODV y un informe en el cual se puede especificar

el año y/o el mes para realizar *drill down* en función del detalle deseado por parte del usuario. Este *dashboard* está detallado en la Ilustración 65.



Ilustración 66- Cuadro de Mando de Áreas

Tras los clientes, se ha completado un *dashboard* que muestra las métricas relacionadas con las áreas de la empresa. En la parte derecha se puede observar la proporción de horas que ha hecho cada área en cada año. Justo debajo del *pie chart* se puede ver en detalle las horas de cada proyecto en función del área seleccionado en el gráfico anterior. En la parte derecha de este cuadro de mando, mostrado en la Ilustración 66, se puede ver una gráfica donde muestra las horas realizadas por año y las horas que estaban planificadas en cada departamento de la empresa. Por último, justo debajo están los indicadores desarrollados en el sprint anterior que muestran el porcentaje de horas facturables de cada área.

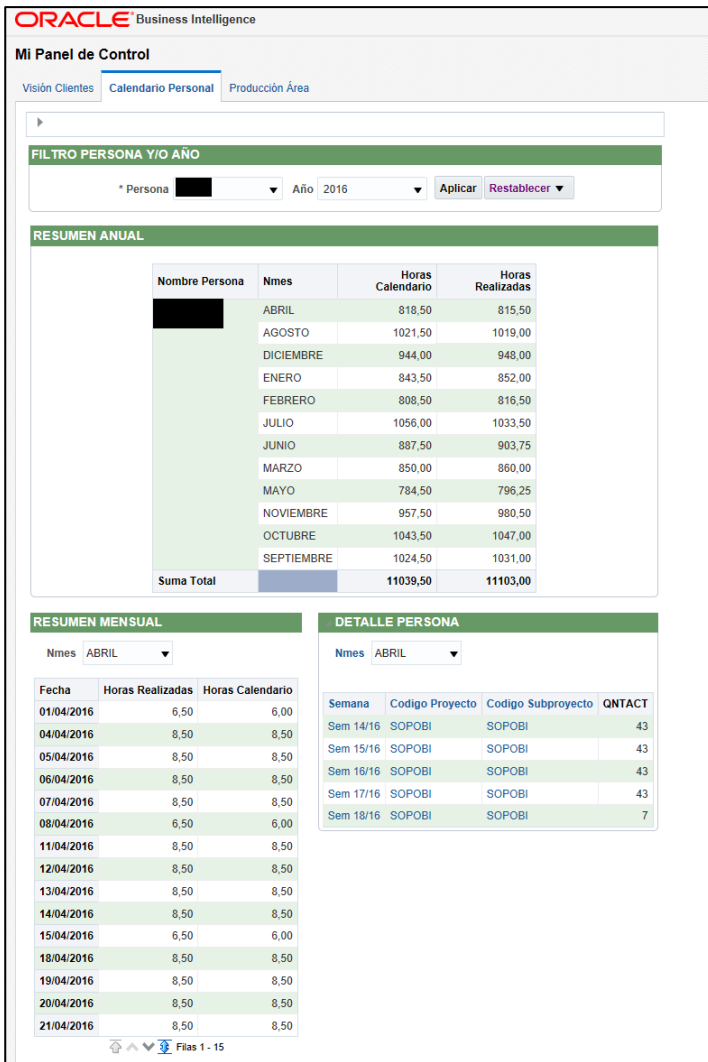


Ilustración 67 - Cuadro de Mando de Empleados

Finalmente, la Ilustración 67 muestra el cuadro de mando relacionado con los empleados. Una vez seleccionada la persona y el año, se puede ver un resumen anual de las horas realizadas. En la parte inferior izquierda se detalla las horas por día en cada mes. Por último, en la sección inferior derecha se muestra en detalle las horas dedicadas, en la misma franja de fechas, a cada proyecto.

### 5.5.3 Revisión

| SPRINT                           | LISTA DE TAREAS   | HORAS EST. | HORAS REALES |
|----------------------------------|---|------------|--------------|
| 4                                | T4.1: Instalación de Infraestructura de Oracle para OBI.                          | 13         | 8            |
|                                  | T4.2: Importación del modelo de datos del DWH para creación de repositorio de BI. | 21         | 20           |
|                                  | T5.1: Desarrollo de Informes con OBI.   | 21         | 18           |
|                                  | T5.2: Desarrollo de Informes con ODV.   | 8          | 8            |
|                                  | T5.3: Configuración del Cuadro de Mando.  | 21         | 22           |
| <b>TOTAL DE HORAS ESTIMADAS:</b> |   | <b>84</b>  | <b>76</b>    |

Tabla 24 - Resumen Sprint 4

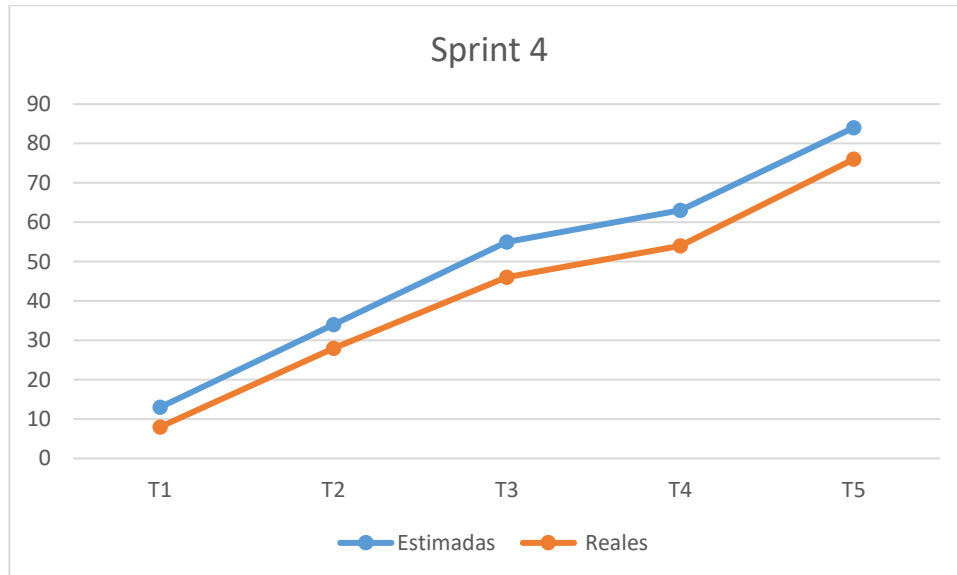


Figura 7 - Burn Up Sprint 4

#### **5.5.4 Retrospectiva**

En la última reunión de retrospectiva, el Equipo de Desarrollo se reunió para la verificación de que todas las historias incluidas en la pila del Sprint habían sido completadas. Al igual que en el Sprint anterior, este ha sido pesimista con la planificación inicial. Una de las razones fue el conocimiento de las herramientas utilizadas para el desarrollo de las historias de usuario.

## 6 CONCLUSIONES Y PROPUESTAS

En este capítulo se expone si se ha cumplido con el objetivo principal del TFG además de los objetivos parciales descritos en el capítulo 2. También se detallan propuestas de mejora futuras y la opinión personal del autor.

### 6.1 Cumplimiento de Objetivos

En el capítulo 2 se detalló cual era el objetivo principal establecido para el Trabajo Fin de Grado: la creación de un almacén de datos empresariales sobre la producción de los empleados, registrando las horas destinadas a los diferentes proyectos que se realizan en la empresa para su posterior explotación *analítica*. Además del objetivo principal, en este mismo capítulo de la memoria se han descrito objetivos académicos de la empresa y objetivos parciales.

La Tabla 25 lista los objetivos académicos además del Sprint donde se han conseguido:

|            | Descripción   | Sprint                 |
|------------|---|------------------------|
| <b>OA1</b> | Estudio de los diversos métodos a emplear.  | Sprint 0 (Fase Previa) |
| <b>OA2</b> | Aprendizaje de las herramientas software que serán utilizadas, especialmente Oracle Data Integrator 12c y Oracle Business Intelligence 12c. | Sprint 0 (Fase Previa) |
| <b>OA3</b> | Análisis y captura de requisitos.   | Sprint 1               |
| <b>OA4</b> | Diseño del esquema de datos.  | Sprint 1               |
| <b>OA5</b> | Realización del proceso de extracción inicial de datos.   | Sprint 2               |
| <b>OA6</b> | Realización del proceso de actualización periódica de los datos.  | Sprint 3               |
| <b>OA7</b> | Diseño e implementación de informes y cuadros de mando con una herramienta BI.  | Sprint 4               |

Tabla 25 - Consecución de los Subobjetivos

## **6.2 Propuestas de Mejora**

Al inicio del TFG se barajaron diferentes ideas para el desarrollo del *data warehouse* pero que finalmente no fueron incluidas en el alcance del proyecto debido a la duración del mismo.

A continuación se exponen las propuestas de mejora:

- Utilización de otros sistemas ERP para el desarrollo del *data warehouse*.
- Usos de lenguajes de programación estadísticos para la búsqueda de patrones en los datos.
- Desarrollo de nuevos esquemas en estrella utilizando el resto de información almacenada en la misma fuente de datos usada en el TFG.

## **6.3 Opinión Personal**

Personalmente, la realización del TFG me ha supuesto un aprendizaje continuo. Principalmente por haberlo realizado dentro de una empresa con tecnologías punteras en el sector como es Avanttic. Realizarlo bajo el convenio FORTE, me ha servido para aplicar todos los conocimientos que he ido adquiriendo a lo largo de la carrera, haciendo uso de ellos durante el desarrollo del TFG.

La Inteligencia de Negocio es un tema de la informática que personalmente me gusta al estar cerca de las empresas, conociendo sus necesidades de información y dando una solución útil a todo ello. Desde Avanttic se me ha brindado la oportunidad de adquirir experiencia profesional realizando un proyecto de esta índole a la par que he podido participar junto con un equipo en un proyecto real.

Para concluir, resalto la satisfacción que he obtenido tras realizar este proyecto. Tras él, me siento listo para dar comienzo a una nueva etapa profesional en mi vida.

## 7 BIBLIOGRAFÍA

- [1] H. J. Watson y B. H. Wixom, «The Current State of Business Intelligence», *Computer (Long. Beach. Calif.)*, vol. 40, n.º 9, pp. 96-99, sep. 2007.
- [2] W. H. Inmon, D. Strauss, y G. Neushloss, *DW 2.0 : the architecture for the next generation of data warehousing*. Morgan Kaufmann/Elsevier, 2008.
- [3] R. Kimball y M. Ross, *The data warehouse toolkit: the complete guide to dimensional modelling*. 2011.
- [4] Aner, «¿Qué es un ERP?» [En línea]. Disponible en: <http://www.aner.com/que-es-un-erp.html>. [Accedido: 20-mar-2018].
- [5] Oracle Corporation, «Oracle Database 11g». [En línea]. Disponible en: <http://www.oracle.com/technetwork/es/database/index.html>. [Accedido: 20-mar-2018].
- [6] Oracle Corporation, «Oracle Data Integrator». [En línea]. Disponible en: <http://www.oracle.com/technetwork/middleware/data-integrator/overview/index.html>. [Accedido: 19-feb-2018].
- [7] Oracle Corporation, «Business Intelligence (BI) | Oracle España». [En línea]. Disponible en: <https://www.oracle.com/es/solutions/business-analytics/business-intelligence/index.html>. [Accedido: 20-feb-2018].
- [8] F. Ruiz, «Material de la asignatura “Ingeniería de Negocio”. Escuela Superior de Informática, Universidad de Castilla - La Mancha.» .
- [9] Julio Castro, «¿Qué es la inteligencia de negocios y cómo beneficia a tu empresa?» [En línea]. Disponible en: <http://blog.corponet.com.mx/que-es-la-inteligencia-de-negocios>. [Accedido: 12-jun-2018].
- [10] Ana Buigues, «Diferencias entre un Data Warehouse y las bases de datos operacionales | El Blog de Ana Buigues». [En línea]. Disponible en: <http://anabuigues.com/2010/01/14/data-warehouse-y-las-bases-de-datos-operacionales/trackback/index.html>. [Accedido: 04-jun-2018].
- [11] B. Hitpass, «BPM Fundamentos y Conceptos de Implementación».
- [12] Classora Technologies, «Bases de datos multidimensionales: OLAP vs OLTP». [En línea]. Disponible en: <http://blog.classora.com/2013/06/25/bases-de-datos-multidimensionales-olap-vs-oltp/>. [Accedido: 04-jun-2018].

- [13] Hitachi Vantara, «Pentaho | Hitachi Vantara». [En línea]. Disponible en: <https://www.hitachivantara.com/go/pentaho.html?source=pentaho-redirect>. [Accedido: 14-jun-2018].
- [14] Gartner, «Data Integration Tools and Software Reviews». [En línea]. Disponible en: <https://www.gartner.com/reviews/market/data-integration-tools>. [Accedido: 14-jun-2018].
- [15] Informatica LLC, «PowerCenter: plataforma de integración de datos empresariales | Informatica España». [En línea]. Disponible en: <https://www.informatica.com/es/products/data-integration/powercenter.html>. [Accedido: 14-jun-2018].
- [16] Scrum, «About Scrum». [En línea]. Disponible en: <https://www.scrum.org/>. [Accedido: 14-jun-2018].
- [17] Kimball Group, «Kimball Group Kimball DW/BI Lifecycle Methodology». [En línea]. Disponible en: <https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dw-bi-lifecycle-method/>. [Accedido: 14-jun-2018].
- [18] Bahchi's Technologies, «ELT Drill Down». [En línea]. Disponible en: <http://www.bahchis.com/elt-drill-down/>. [Accedido: 14-jun-2018].
- [19] Oracle Corporation, «Oracle Data Visualization - Visual Analytics | What is Data Visualization? | Oracle España». [En línea]. Disponible en: <https://www.oracle.com/es/solutions/business-analytics/data-visualization.html>. [Accedido: 24-may-2018].
- [20] Oracle Corporation, «Oracle PL/SQL». [En línea]. Disponible en: <http://www.oracle.com/technetwork/database/features/plsql/index.html>. [Accedido: 24-may-2018].
- [21] Oracle Corporation, «Database 12c | Oracle España». [En línea]. Disponible en: <https://www.oracle.com/es/database/index.html>. [Accedido: 20-feb-2018].
- [22] Oracle Corporation, «Oracle SQL Developer». [En línea]. Disponible en: <http://www.oracle.com/technetwork/developer-tools/sql-developer/overview/index.html>. [Accedido: 20-feb-2018].
- [23] Oracle Corporation, «SQL Developer Data Modeler». [En línea]. Disponible en: <http://www.oracle.com/technetwork/developer-tools/datamodeler/overview/index.html>. [Accedido: 20-feb-2018].

- [24] Trello Inc., «¿Qué es Trello?» [En línea]. Disponible en: <https://trello.com/about>. [Accedido: 24-may-2018].
- [25] Microsoft Corporation, «Microsoft Excel 2016, hojas de cálculo, prueba gratuita». [En línea]. Disponible en: <https://products.office.com/es-es/excel>. [Accedido: 24-may-2018].
- [26] «Dropbox». [En línea]. Disponible en: <https://www.dropbox.com/>. [Accedido: 20-feb-2018].
- [27] Microsoft Corporation, «Microsoft Word 2016: procesamiento de textos | Office». [En línea]. Disponible en: <https://products.office.com/es-es/word>. [Accedido: 24-may-2018].
- [28] «Mendeley». [En línea]. Disponible en: <https://www.mendeley.com/reference-management/reference-manager>. [Accedido: 20-feb-2018].
- [29] C. Blanco, «Lista de comprobación de riesgos en proyectos software», pp. 1-4, 1997.
- [30] Javier Garzás, «La técnica del Planning Poker». [En línea]. Disponible en: <http://www.javiergarzas.com/2018/01/la-tecnica-del-planning-poker.html>. [Accedido: 21-jun-2018].
- [31] D. P. Carlos Fernández, «Estructura de la Dimension Tiempo y Script de carga con Oracle SQL». [En línea]. Disponible en: <http://www.dataprix.com/blogs/carlos/estructura-dimension-tiempo-script-carga-para-oracle>. [Accedido: 25-abr-2018].
- [32] Oracle, «Oracle WebLogic Server Technical Information». [En línea]. Disponible en: <http://www.oracle.com/technetwork/middleware/weblogic/overview/index.html>. [Accedido: 01-jun-2018].
- [33] Oracle, «Oracle AdminServer». [En línea]. Disponible en: <https://www.oracle.com/corporate/acquisitions/adminserver/index.html>. [Accedido: 01-jun-2018].



## **ANEXOS**

### **ANEXO A – ACRÓNIMOS**

- DWH: Data Warehouse.
- TFG: Trabajo Fin de Grado
- BD: Base de Datos.
- BI: Bussines Intelligence.
- HU: Historia de Usuario.
- OBI: Oracle Bussines Intelligence.
- ODI: Oracle Data Integration.
- ODV: Oracle Data Visualization.
- INEG: Ingeniería de Negocio.
- PL / SQL: Procedural Language/ Structured Simple Query Language.
- ETL: Extract Transform Load.
- KM: Knowledge Module.
- IKM: Integration Knowledge Module.
- LKM: Load Knowledge Module.
- ERP: Enterprise Resource Planning.
- RDBMS: Relational Database Management System
- OLTP: Online Transaction Processing.
- OLAP: Online Analytical Processing.
- ROLAP: Relational Online Analytical Processing.
- MOLAP: Multidimensional Online Analytical Processing.
- HOLAP: Hybrid Online Analytical Processing.



## **ANEXO B – NOTAS OBTENIDAS EN LAS REUNIONES DE CAPTURA DE REQUISITOS**

El sistema EGI está pensado para ser multiempresa. Es por ello que hay una tabla EMPRESA que almacena las diferentes entradas. Se podría dar el caso que un mismo empleado esté asociado a dos diferentes empresas, teniendo por tanto dos posibles logins internos. EGI no está pensado como un sistema para gestionar las tarifas, está pensado para HORAS + VACACIONES.

**Perfiles:** Asociado a una persona tenemos perfiles, pueden ser de consultor, jefe de proyecto o manager o responsable de área.

**Actividad:** la persona imputa actividades a un subproyecto que pertenece a un proyecto que es asociado a un cliente. Por lo que tener claro el concepto de que un trabajador computa horas a una ACTIVIDAD.

**Calendario:** Tenemos dos tipos, calendario laboral y calendario personal. El calendario laboral está asociado a la delegación de una empresa (toma los festivos locales y autonómicos del sitio). Aunque son dos tablas diferentes, la del empleado/persona es creado a partir del calendario laboral. Es interesante este segundo calendario ya que se almacenan sus vacaciones, las horas que han sido imputadas en un día a diferentes actividades (o una única) y las horas que debería haber realizado según contrato. Gracias a esta información se puede ver el uso de las horas al día y el balance total de horas realizadas en una semana o incluso ya en el mes final. Aunque casi todas las personas tienen el propio, no es obligatorio que una persona tenga su calendario personal (el calendario personal tiene que ser creado por el responsable de Recursos Humanos).

Este sistema está pensado como herramienta de apoyo en la gestión del departamento de producción. Por lo que es más interesante obtener información clara.

El sistema consta de dos claves principales: Conocer las horas dedicadas a los clientes y sus proyectos/subproyectos. En este cuadro de mando se podría llegar a conocer el tiempo dedicado por consultor haciendo drill-down.

El segundo bloque principal se enfoca a los consultores. Está pensado para ver y analizar las horas que tenían planteadas (de su calendario personal) y las horas que realmente han sido realizadas. Se puede ver a nivel de mes (Horas por hacer/Horas hechas) aunque donde más utilidad se puede encontrar es poniendo periodos de varios meses o incluso de un año (para ver el cómputo global). Tras ese primer cuadro de

mando simple, debajo se muestra un desglose de las horas dedicadas (en función del periodo seleccionado) a cada proyecto y por día del mes. Este segundo cuadro de mando es útil si se quiere buscar algo en concreto. Si se buscara una visión global es más útil utilizar el mencionado antes ya que es menos sobrecargada la presentación de la información.

Por último, se puede tener acceso a las horas que no son computadas. Es decir **horas no facturables**. Este aspecto es interesante para la gestión de la producción ya que se hacen previsiones de las horas que tiene que dedicar un consultor (horas facturables y no). Ahora se pueden conocer las horas que no facturadas a nivel global. Pero no tienen indicativos de ningún tipo, tan solo se muestra la información. Sería interesante introducir algún tipo de gráfico o indicador que representara la información de manera más visual.