

# Proyecto Business Intelligence para el análisis de factores No Conformidad y Costes Evitables en procesos industriales.

## **Estudiante**

Antonio M. Villa Moraga  
Grado de Ingeniería Informática  
Business Intelligence

## **Nombre Consultor/a**

Xavier Martinez Fonte

## **Nombre Profesor/a responsable de la asignatura**

Atanasi Daradoumis Haralabus

13/06/2019



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

Copyright © 2019 Antonio M. Villa Moraga.  
Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.  
A copy of the license is included in the section entitled "GNU Free Documentation License".

© (Antonio M. Villa Moraga)

Reservados todos los derechos. Está prohibido la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la impresión, la reprografía, el microfilme, el tratamiento informático o cualquier otro sistema, así como la distribución de ejemplares mediante alquiler y préstamo, sin la autorización escrita del autor o de los límites que autorice la Ley de Propiedad Intelectual.

Título del trabajo:	Proyecto Business Intelligence para el análisis de factores No Conformidad y Costes Evitables en procesos industriales
Nombre del autor:	Antonio Manuel Villa Moraga
Nombre del consultor/a:	Xavier Martinez Fonte
Nombre del PRA:	Atanasi Daradoumis Haralabus
Fecha de entrega (mm/aaaa):	06/2019
Titulación:	Grado de Ingeniería Informática
Área del Trabajo Final:	Business Intelligence
Idioma del trabajo:	Español
Palabras clave	Índice de Desempeño, Data Warehouse y ETL.

**Resumen del Trabajo (máximo 250 palabras):** Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.

Este trabajo pretende indagar en la casuística de las tecnologías y disciplinas involucradas en la construcción de un producto BI. Dicho producto se ha conformado como un dashboard con el objetivo de estudiar los Costes Evitables y las No Conformidades dados en un contexto industrial. Un segundo objetivo pretende ensayar el reto que supone llevar a cabo un producto BI en cuanto a la realidad de un Sistema de Información en explotación. En otra medida de intenciones, la iniciativa incorpora un trasfondo con los propósitos de aportar nuevos conceptos BI a la cultura del tejido organizacional y, además, la toma de conciencia sobre la ventaja que otorgan los datos en la generación de conocimiento.

En cuanto a la conceptualización apuntar que esta ha sido diseñada siguiendo el modelo *Bottom-up* ideado por R. Kimball, como guía en el diseño y concreción del plan que ha propiciado la ambición de pilotar la iniciativa. Cabe decir que dicho modelo casa perfectamente con el marco de referencia agile empleado en el pilotaje del proyecto.

El trabajo también recoge las vicisitudes surgidas durante la prospección de las fuentes de datos y como su resultado ha afectado en la construcción del dashboard.

El proyecto ha concluido con la entrega de un dashboard ajustado a la configuración establecida a excepción del alcance de dos KPIs. Aun así, se considera que este hecho no desmerece la calidad del producto desarrollado por las funcionalidades que ofrece y por las complejidades derivadas de las tecnologías empleadas para su desempeño.

**Abstract (in English, 250 words or less):**

This work aims to investigate the casuistry of the technologies and disciplines involved in the construction of a BI product. This product has been formed as a dashboard with the objective of studying the Avoidable Costs and Non-Conformities given in an industrial context. A second objective aims to test the challenge of carrying out a BI product in terms of the reality of an Information System in operation. In another measure of intentions, the initiative incorporates a background with the purpose of contributing new BI concepts to the culture of the organizational fabric and, in addition, the awareness of the advantage that data provide in the generation of knowledge.

The conceptualization has been designed following the Bottom-up model devised by R. Kimball, as a guide in the design and realization of the plan that has fostered the ambition to pilot the initiative. It is possible to say that this model fits perfectly with the agile reference frame used in the piloting of the project.

The work also includes the issues that arose during the prospecting of the data sources and how the result has affected the construction of the dashboard.

The project has concluded with the delivery of a dashboard adjusted to the established configuration except for the scope of two KPIs. Even so, it is considered that this fact does not detract from the quality of the developed product, given the functionalities it offers and the complexities derived from the technologies used for its performance.

## Índice

<b>1. Introducción</b>	<b>1</b>
1.1 Contexto y justificación del Trabajo	2
1.1 Contexto	2
1.2 Justificación	3
1.2 Objetivos del Trabajo	4
1.2.1 Objetivos generales	4
1.2.2 Objetivos específicos	4
1.3 Alcance	5
1.3.1 El producto	5
1.3.1 El proyecto	6
1.4 Enfoque y método seguido	6
1.5 Planificación del Trabajo	7
1.5.1 Diagrama de Gantt	9
1.5.2 Riesgos	11
1.5.3 Control y seguimiento	13
1.5.4 Valoración económica del trabajo	13
1.5.5 Herramientas transversales	13
1.6 Breve resumen de productos obtenidos	14
1.7 Breve resumen del resto de apartados	14
<b>2. Definición del producto</b>	<b>16</b>
2.1 Prospección y recogida de Información	16
2.2 Identificación de KPIs	18
2.3 Estudio de fuentes y datos	20
2.4 Diseño de la aplicación	27
2.4.1 Estudio de la arquitectura y esquemas	27
2.4.2 Diseño de modelo conceptual	29
2.4.3 Diseño de modelo lógico	30
2.5 Elección de software y pruebas	34
<b>3. Desarrollo del producto</b>	<b>40</b>
3.1 Construcción del espacio de desarrollo	40
3.2 Construcción del Data Warehouse	44
3.3 ETL (Extracción, Transformación y Carga)	45
3.3.1 Pruebas con datos	49
3.4 Confección de OLAP	51
3.5 Configuración de elementos gráficos	55
3.6 Pruebas de visualización de datos	63
<b>4. Fase de ensayo y calidad</b>	<b>65</b>
4.1 Disposición en el sistema real	65
4.2 Pruebas de calidad	65
<b>5. Conclusiones</b>	<b>67</b>
5.1 Descripción de las conclusiones del trabajo	67
5.2 Consecución de los objetivos planteados	67
5.3 Análisis crítico de la metodología del proyecto y su planificación	68
5.4 Gestión del proyecto	69
5.5 Líneas de trabajo futuro	69
<b>6. Glosario</b>	<b>70</b>
<b>7. Bibliografía</b>	<b>71</b>
<b>8. Anexos</b>	<b>72</b>
I. Log de la ejecución de los trabajos ETL	72

## Lista de ilustraciones

Ilustración 1 desglose EDT .....	9
Ilustración 2: Cronograma de los trabajos .....	10
Ilustración 3: flujo de trabajo para el desarrollo de DW .....	16
Ilustración 4 Muestra del contenido de la DDBB empleada para la gestión de las NC. ....	21
Ilustración 5 Demostración de ausencia de estructura relacional en la DDBB. ....	22
Ilustración 6: Detalle las NC y su número de registros. ....	23
Ilustración 7: NC registradas y la cantidad de registros.....	24
Ilustración 8: Registro de importados des Impromptu.....	24
Ilustración 9: registro con la trazabilidad de NC.....	25
Ilustración 10: contenido de un archivo CSV de CE de Automoción. ....	25
Ilustración 11 : Contenido de un archivo CSV de CE de Producto.....	26
Ilustración 12: relaciones que faltan entre las tablas de la BBDD de Access .....	27
Ilustración 13: modelo R. Kimball [ <a href="http://www.zentut.com/data-warehouse/data-mart">http://www.zentut.com/data-warehouse/data-mart</a> ].....	28
Ilustración 14: adaptación del modelo de R. Kimball al producto en desarrollo. ....	29
Ilustración 15: modelo conceptual del producto ideado según los requerimientos del caso .....	30
Ilustración 16: matriz de bus de cruce entre procesos .....	30
Ilustración 17: modelo lógico del Data Warehouse.....	31
Ilustración 18: tabla de hechos No Conformidades .....	31
Ilustración 19: tabla de hechos Costes Evitables.....	32
Ilustración 20: dimensión fecha común a las dos tablas de hechos .....	32
Ilustración 21: dimensión Proyectos común con a las dos tablas de hechos .....	33
Ilustración 22: dimensión Unidad de Negocio común a las dos tablas de hechos. ....	33
Ilustración 23: dimensión Tipo NC .....	33
Ilustración 24: dimensión Tipo CE.....	33
Ilustración 25: dimensión Proveedor común a las dos tablas de hechos. ....	33
Ilustración 26: Cuadro Mágico de Gartner, Inc. ....	34
Ilustración 27: localización de las aplicaciones en el modelo de R. Kimball .....	39
Ilustración 28: captura del error producido durante la conexión desde Kettle a MySQL .....	41
Ilustración 29: parametrización de la conexión en Kettle para acceder a MySQL .....	42
Ilustración 30: requisito fundamental para la instalación de MySQL. ....	42
Ilustración 31: configuración de los canales de acceso a la BBDD. ....	43
Ilustración 32: selección del mecanismo de autenticación (sin SSL). ....	43
Ilustración 33: Ingreso de credenciales de MySQL desde Power BI. ....	44
Ilustración 34: vista de los Schemas para el Data Warehouse y el laboratorio .....	44
Ilustración 35: campos, relaciones y "Primary Keys" según la modelización del DWH.....	45
Ilustración 36: JOB principal ETL para ficheros CSV.....	46
Ilustración 37: saltos para la extracción de los archivos CSV .....	46
Ilustración 38: saltos de transformación de ficheros CSV .....	47
Ilustración 39: extracción de dimensiones de los archivos CSV.....	47
Ilustración 40: JOB ETL para las fuentes MDB.....	47
Ilustración 41: saltos de transformación de los hechos NC de la fuente MDB .....	48
Ilustración 42: extracción de las dimensiones NC .....	48
Ilustración 43: saltos para la carga de las dimensiones en Data Warehouse .....	49
Ilustración 44: saltos para la carga de los hechos en el Data Wharehouse .....	49
Ilustración 46: script para la automatización de las acciones ETL sobre las fuentes MDB.....	50
Ilustración 45: script para la automatización de las tareas ETL sobre los archivos CSV .....	51
Ilustración 47: resultado de la carga de datos desde Power BI.....	52
Ilustración 48: rectificación de las relaciones defectuosas después de la carga de datos .....	53
Ilustración 49: funcionalidad de análisis automático y corrección rápida de errores.....	54
Ilustración 50: acciones rápidas sobre los campos desde Power BI .....	54
Ilustración 51: elementos de segmentación de Power BI .....	56
Ilustración 52; parametrización del elemento de visualización de las fechas.....	56
Ilustración 53: funcionalidad para la visualización de los datos de una gráfica.....	57
Ilustración 54: composición de visualizaciones para la página CE 1.....	57
Ilustración 55: métrica para el cálculo del coste por cada tipo CE .....	58
Ilustración 56: cálculo del objetivo según el tipo CE .....	58
Ilustración 57: formato condicional mediante código DAX.....	59
Ilustración 58: composición de visualizaciones para la página CE 2.....	59

Ilustración 59: composición de visualizaciones para la página CE 3.....	60
Ilustración 60: composición de visualizaciones para la página NC 1 y NC 3 .....	61
Ilustración 61: métrica para el acumulado objetivo por proveedor .....	61
Ilustración 62: métrica para el recuento de las NC según su tipo.....	61
Ilustración 63: ejemplo de búsqueda .....	62
Ilustración 64: composición de visualizaciones para la página NC 2 y NC 3 .....	62
Ilustración 65: ejemplo de uso de la página NC 2 y NC 3 .....	63
Ilustración 66: identificación de una posible anomalía.....	65
Ilustración 67: confirmación de la anomalía .....	66

# 1. Introducción

En la industria globalizada se está viviendo un cambio de época (Cuarta Revolución Industrial) la cual es exigente de nuevas tecnologías que faciliten dinamismo y agilidad en la confección de estrategias, en la toma de decisiones y en la forma de llevar a cabo los procesos. Este escenario propicia la proliferación de tecnologías como la hiper conectividad o el tratamiento y análisis masivo de datos como nutriente que favorecerá la creación de nuevas herramientas capaces de incorporar, a la genética empresarial, el atributo de innovación.

Una de estas metodologías innovadoras es la Inteligencia de Negocio. El conocimiento sobre la situación de la actividad del negocio y la de su entorno empodera a los gobiernos y estratos decisionales de la organización frente la toma de decisiones.

La iniciativa que se propone se enmarca en esta área con un enfoque dirigido a la experimentación. Su aplicación pretende ser un ejemplo de cambio de paradigma de gestión de la información en una organización, real y en activo, con un nivel de Integración Informacional e inteligencia digital desfasada respecto a sus clientes y proveedores.

Aun así, no se debe mal interpretar la intención de la iniciativa, su propósito no es iniciar un cambio tecnológico respecto al tratamiento que se hace de la información. La intención se aproxima más a la presentación de otra forma de hacer las cosas, aportando agilidad y funcionalidad. En otra medida, también procura dar visibilidad a las nuevas herramientas para la gestión de la información y, dentro de las limitaciones del conocimiento que se tiene sobre la materia, introducir nuevos conceptos en la organización.

En concreto, el producto objeto de estudio pretende analizar de los Costes Evitables<sup>1</sup> que se dan a lo largo de la vida de un proyecto y las No Conformidades<sup>2</sup> relacionadas con los proveedores.

Por último, señalar que la integración del *dashboard* en el sistema no se llevará acabo. Aun siendo un producto totalmente funcional y adaptado a la realidad de la empresa su intención es puramente académica, pese a ello, se espera la participación de integrantes del departamento de calidad como asesores en el proyecto.

---

<sup>1</sup> Costes Evitables: [ver significado en el glosario.](#)

<sup>2</sup> No Conformidades: [ver significado en el glosario.](#)



## 1.1 Contexto y justificación del Trabajo

### 1.1 Contexto

Dejando a un lado los aspectos estructurales o valoraciones sobre la operativa de la organización (en el Anexo I se presentan las dimensiones de la empresa), debemos centrar la atención sobre cómo se ha llegado al estado actual de nivel de Integración del SI en la organización.

Su naturaleza industrial -diseño y fabricación de útiles para el sector del automóvil - y disponer de una historia de más de 80 años, ha fraguado en la organización un carácter con poca iniciativa innovadora, actitud que le ha obligado a evolucionar únicamente por las exigencias de sus clientes. Este hecho se puede constatar si nos fijamos en las características del ERP en explotación (BAAN): sistema monolítico y con más de 17 años de uso y una lista interminable de adaptaciones. Como es de suponer, estos atributos han desmotivado todo intento de sustitución del sistema ERP.

Entre otras consideraciones del Sistema de Información actual también se puede destacar la falta de visibilidad de la información que ofrece a los usuarios. Las únicas representaciones de información son conformadas por listados y reportes tabulados de difícil comprensión. Además, todas aquellas funciones extra (propias de módulos CRM, SCM, PRM o BI), necesarias para el estudio de datos y requeridas por las diferentes áreas funcionales, han sido suplidas por aportaciones que los propios usuarios han realizado a lo largo del tiempo empleando hojas de cálculo, bases de datos y *queries* procedentes del ERP almacenados en formato CSV. Aun así, son muchas las ocasiones en donde las iniciativas individuales se ven frenadas por las limitaciones inherentes a las aplicaciones contenidas en el portafolio TIC.

Como ejemplo de lo dicho, y por ser objeto de estudio en el trabajo, se puede presentar el departamento de Calidad y Gestión de Medio Ambiente. Uno de sus desempeños se centra en estudiar los acontecimientos dados en los diferentes procesos que intervienen en la Cadena de Valor y la Cadena de Suministro de la organización. Su actitud proactiva ante el registro de acciones fuera de lo normal y excepciones surgió de la necesidad de defender su posición frente a los diferentes casos de No Conformidades (NC) de productos o los Costes Evitables (CE) redundantes o de gravedad. Este desempeño les propició la fama de “Asuntos Internos” dentro de la organización.

Desde dicho departamento se ha generado, a lo largo de los años, un conjunto de bases de datos empleando software de escritorio. Producto de la buena intención y una formación autodidacta en la confección de repositorios de datos, se ha generado un conglomerado de *dataset*<sup>3</sup> con un alto nivel de dispersión y una pobre, e ineficiente, vinculación entre los datos. Aun así, cabe decir que lo contenido en los *dataset*, pese a su falta de refinamiento, sí estuviese gestionado

---

<sup>3</sup> Dataset: [ver significado en el glosario.](#)

de forma correcta puede ejercer un efecto muy revelador dentro de la organización.

Por último y a propósito del origen de la iniciativa, cabe decir que la idea surge de forma proactiva por parte del que escribe, no existe una solicitud explícita de ningún componente de la organización. Dicho sea de paso, los datos sobre los que se trabajará (información real, pero manipulada con ánimo de no vulnerar la privacidad de los involucrados) son propiedad de la empresa a la que pertenece la organización mencionada.

## 1.2 Justificación

Ante tal escenario se cree necesario articular una herramienta que aporte atributos mejores a los actuales y otorgue de funcionalidad más automatizada al Sistema de Información en explotación en cuanto al tratamiento de los datos.

Cabe decir que, aun siendo un prototipo ideado para ser utilizado por un grupo restringido de usuarios, lo importante del caso es dar otra perspectiva en la gestión de la información dando a entender que su alcance puede suponer un valor más allá que la propia representación de los datos.

A groso modo algunas de las justificaciones que impulsan y acreditan la iniciativa del proyecto son:

- Unificar la visión de situación. Se da el caso que los mismos datos sobre los CE y NC son accesibles desde Calidad y Compras, con desviaciones entre las dos perspectivas.
- Búsqueda de nuevas fuentes de innovación. Poner en valor los datos recogidos en general y en particular respecto a los NC y CE.
- Estado de mejora constante. Abrir las puertas a otra forma de hacer las cosas poniendo en el debate si es posible mejorar, por ejemplo, en qué medida una herramienta BI puede influir sobre los NC y CE.
- Análisis y redefinición de las estrategias. Evaluar el grado de respuesta de los procesos frente a la necesidad de establecer nuevas estrategias en busca los objetivos fundamentado su razón mediante estudios y análisis de situación entorno a las perspectivas los CE y las NC.
- Introducir conceptos BI en el lenguaje de la organización. Según la trayectoria de la empresa se puede intuir que existe una falta de interés por parte del gobierno de la empresa en cuanto las virtudes que pueden aportar las soluciones BI y, por consiguiente, en el resto de la organización
- Showroom. Minimizar los celos del espónsor poniendo en cuanto a las tecnologías BI. Esto es posible poniendo al alcance de la capa decisonal de la organización un prototipo atractivo, low-cost, con funcionalidad real y dando la opción de experimentar en primera persona.

## 1.2 Objetivos del Trabajo

A rasgos generales, la línea de desarrollo del trabajo se centrará en la confección de un prototipo *dashboard* diseñado según las necesidades del departamento de Calidad como medio de estudio y análisis de un grupo de índices de desempeño ya preestablecidos.

Referente al enfoque otorgado al desarrollo cabe decir que es experimental, pero no exento del carácter transformador inducido por el uso de tecnologías y metodologías en constante innovación como son las pertenecientes al Business Intelligence. Es más, puesto que se desarrollará a medida conforme la realidad de un Sistema de Información en explotación, el cuadro de mando podría ser utilizada si así se decidiese.

### 1.2.1 Objetivos generales

El objetivo central del trabajo no es otro que el de ensayar la confección de una solución completa, con forma de *dashboard*, en un sistema de información en explotación. Dicho útil representará de forma gráfica aspectos relevantes del trabajo registrado y almacenado por el personal del departamento de calidad. Además, la disposición del Cuadro de Mando será accesible por todo aquel interesado e integrante a la organización.

Por otra parte, dicho objetivo clave también debe asegurar que la calidad y disponibilidad de los datos gestionados de tal modo que el conjunto de KPIs objeto de estudio sea fiable. En este sentido no se admite ningún otro resultado. Alcanzar dicho objetivo pasa por establecer, concentrar y normalizar toda la información en un solo repositorio (Data Warehouse<sup>4</sup>).

En otra medida, el trabajo también tiene como objetivo la pretensión de promover las virtudes y bondades de las soluciones BI y como estas pueden ayudar a la organización. El método previsto para conseguir este objetivo se basa en establecer un contraste entre el producto desarrollado versus las herramientas en explotación y disponibles en el portafolio actual. Aunque este objetivo va más allá de las competencias del TFG, se procurará adjuntar en el documento una primera impresión realizada por alguno de los integrantes del departamento de Calidad y Medio Ambiente.

### 1.2.2 Objetivos específicos

En concreto, el despiece del propósito principal del trabajo estará dividido en objetivos parciales. En gran medida estos objetivos específicos están vinculados entre sí y localizados a lo largo de la progresión del trabajo. Es por ello que una correcta priorización ayudará a alcanzar el logro de los sucesivos objetivos, por lo que resultará crucial en la consecución del proyecto. El conjunto de objetivos específicos más relevantes se define a continuación:

---

<sup>4</sup> Data Warehouse: [ver significado en el glosario](#).

- Elección del **software y herramientas** que ayuden en la confección de la solución BI, según el criterio que será descrito en su momento.
- Diseño de un **modelo conceptual y una versión lógica** de la solución **Data Warehouse** idóneo para la solución *dashboard* ideada. La criticidad del diseño vendrá condicionada por la disponibilidad de los datos y el acceso a las fuentes de donde parten.
- Desarrollo de una **estructura física** basada en el modelo lógico Data Warehouse diseñado y específico para el Cuadro de Mando, capaz de establecer, concentrar y normalizar toda la información en un solo repositorio.
- **Automatizar** todo lo posible las tareas involucradas en la confección de informes que llevan a cabo los usuarios, de manera que se puedan dedicar a hacer su trabajo.
- Desarrollo de un **entorno de visualización de los valores** agrupados bajo los KPIs y dispuesto en forma de *dashboard*.

## 1.3 Alcance

### 1.3.1 El producto

Tal y como se ha concretado en los objetivos, se entregará un producto que incorporará funcionalidad analítica usable por los usuarios interesados. Dicho sea de paso, la gestión de usuarios no se integrará como parte del producto.

El producto se diseñará a medida, respetando la realidad del Sistema de Información en explotación, dicho de otra forma, la solución entregada no exigirá adaptación alguna por parte del entorno en el que se ensayará. Es más, su ejecución deberá pasar desapercibida frente a los mecanismos de seguridad del sistema.

El producto se compondrá, como mínimo, de tres soluciones: la parte encargada del ETL, el almacenamiento y visualización de los datos, y su ejecución durante el ensayo será local, no será ejecutable en *cloud* ninguna de sus partes:

- Tanto los datos como las fuentes de origen de estos, serán los empleados en el departamento de Calidad. Es por ello que se creará un mecanismo ETL que automatice todo el proceso de tratamiento de los datos. También se dispondrá de mecanismos de automatización que minimice el uso de las herramientas ETL por parte del usuario.
- El producto incluirá un elemento de almacenamiento de datos estructurado, consistente y gestionable basado, preferiblemente, por un motor de DDBB SQL completa (estructurada).
- Dicho producto dispondrá una composición predefinida de elementos para la representación de datos. Los cuales ofrecerán la máxima

navegabilidad permitida por aplicación de edición de *dashboard* y que mantenga la intención de los análisis de datos.

El conjunto de la solución no debe suponer un consumo de recursos más allá del desarrollo de este trabajo.

Y, por último, se deja claro que no se ofrecerá a los interesados la opción de cambiar la apariencia de las representaciones ni los datos representados. Tampoco se ofrecerá la posibilidad de parametrizar el comportamiento del *dashboard* (no confundir con la navegabilidad del *dashboard* o la variabilidad de la granularidad en alguna dimensión).

### 1.3.1 El proyecto

El proyecto tiene el fin de idear, diseñar, implementar e integrar una solución *dashboard* en un sistema de Información en explotación. Alcanzar este propósito requerirá la aplicación de acciones multidisciplinares tales como: pilotaje de proyecto, análisis del caso, implementación de la solución configurando y adaptando empleando productos con diferentes usos (motores de DDBB, aplicaciones para el tratamiento de los datos o soluciones que faciliten la visualización de información).

El proyecto finalizará en el momento que el *dashboard* sea operativo en el sistema de información para ser ensayado por algún componente del departamento de Calidad y Medio Ambiente con la intención de documentar impresiones.

Puesto que la naturaleza de la iniciativa es académica y no tiene patrocinio de terceros, el proyecto no incluirá definiciones de mantenimientos ni formación o gestión de garantías. Dicho de otra forma, el proyecto sólo incluirá una demostración en la empresa en la que sólo intervendrá algún integrante del departamento de calidad. Además, esta demostración *in situ* no implica despliegue alguno o procesos de integración completo el Sistema de Información, lo cual excluye el diseño, planificación o ejecución de la gestión de permisos, integridad de los datos, alcance a todos los interesados o documentación.

## 1.4 Enfoque y método seguido

El trabajo está orientado a la obtención de un producto (prototipo) útil en un ámbito laboral. Aplicar este enfoque exigirá tener conciencia de la forma de trabajar en la organización y conocimiento de los procesos en los que el *dashboard* puede resultar de interés. En consecuencia, una iniciativa BI como la presentada debe contemplar aspectos más allá de las exigencias técnicas de desarrollo y los conocimientos de gestión del proyecto, los cuales son los aspectos esperados desde la perspectiva académica.

En este sentido, respecto a la perspectiva académica, el enfoque que se pretende aplicar es el de la experimentación del pilotaje de un proyecto

puramente BI, mediante la construcción de una solución BI y empleando datos de una organización real. Para ello se ideará y pilotará un proyecto capaz de cubrir determinadas necesidades dadas en una organización.

Dicho proyecto abarcará, también, el análisis y la toma de decisión de todos los aspectos que configuren el modelo sobre el que se basará el prototipo de *dashboard*. Una vez ejecutada la parte más teórica, se construirá un laboratorio virtual en donde se recreará el escenario compuesto por herramientas que facilitarán el desarrollo del proyecto.

Con el fin de que la gestión del proyecto sea posible, tanto desde la perspectiva empresarial como académica, se decide seguir las indicaciones recogidas por el marco de referencia SCRUM<sup>5</sup> -metodología *agile* conocida, principalmente, entre los grupos de desarrollo de software-. No se opta por otras recomendaciones más exhaustivas y complejas, del estilo PMBOK<sup>6</sup>, por considerar tanto el volumen del proyecto como el número de participantes lo bastante reducido como para poner en riesgo el desempeño del proyecto o el grado de calidad del producto.

Como parte de las indicaciones del marco de referencia elegido, se agruparán las tareas definiendo una serie de *Sprints* con una duración en el tiempo máximo de dos semanas. Transcurrido dicho periodo se realizará una entrega a modo de revisión de lo realizado (demostración y retrospectiva). La intención de esta acción es la de corregir desviaciones en el contenido por el consultor (o Scrum Master).

## 1.5 Planificación del Trabajo

Más abajo se detalla la descomposición en tareas planificadas del trabajo. En concreto, se presenta un desglose hasta una profundidad de tercer nivel. Las tareas descritas incluyen la fecha de comienzo y fin, además del tiempo estimado de su duración en unidades de días, entendiendo que una jornada son 4h.

Con intención de que el trabajo sea más completo, se adjunta a la distribución de trabajos la columna con los roles responsables de cada momento. Por otro lado, se debe advertir del echo que todos los roles han sido llevados a cabo por el que escribe.

Por último, aclarar por qué no se han incluido las tareas subyacentes de documentación del trabajo. El motivo principal no es otro que el de representar los trabajos propios del proyecto o que sean relevantes, pero se entiende que se irán haciendo tareas de documentación a la par que se ejecutan las tareas descritas en el plan. La única referencia a una tarea de documentación aparece durante la confección de la memoria por su importancia por el refinamiento del contenido y las acciones de formalización.

---

<sup>5</sup> Scrum (o melé) adoptado de la terminología de los jugadores de Rugby.

<sup>6</sup> Project Management Book Of Knowledge

Orden	Nombre de tarea	Duración	Comienzo	Fin	Pre.	Nombres de los recursos
	<b>TFG BI 2018/19 (2)</b>	<b>76 días</b>	<b>lun 25/02/19</b>	<b>lun 17/06/19</b>		
1	<b>1. Plan de trabajo</b>	<b>14 días</b>	<b>lun 25/02/19</b>	<b>vie 15/03/19</b>		<b>Jefe de proyecto</b>
2	Prospección y elección de TFG	3 días	lun 25/02/19	mié 27/02/19		"
3	Teleconferencia	0 días	lun 04/03/19	lun 04/03/19		"
4	Descripción y justificación	2 días	jue 28/02/19	vie 01/03/19	2	"
5	Objetivos y alcance	1 día	lun 04/03/19	lun 04/03/19	4	"
6	Plan de trabajo	2 días	mar 05/03/19	mié 06/03/19	5	"
7	Cronograma, Riesgos	3 días	jue 07/03/19	lun 11/03/19	6	"
8	Control y seguimiento	1 día	mié 13/03/19	mié 13/03/19	7	"
9	Reajustes de tiempos y contenidos	1 día	jue 14/03/19	vie 15/03/19	8	"
10	<b>PEC 1</b>	<b>0 días</b>	<b>vie 15/03/19</b>	<b>vie 15/03/19</b>		<b>Jefe de proyecto</b>
11	<b>2. Fase de definición de producto</b>	<b>9 días</b>	<b>lun 18/03/19</b>	<b>vie 29/03/19</b>	<b>1</b>	
12	Prospección y recogida de información	1 día	lun 18/03/19	mar 19/03/19		Jefe de proyecto
13	Identificación de KPIs	1 día	mar 19/03/19	mié 20/03/19	12	Jefe de proyecto
14	Estudio de fuentes y datos	2 días	mié 20/03/19	vie 22/03/19	13	Arquitecto de software
15	<b>Diseño de la aplicación</b>	<b>3 días</b>	<b>vie 22/03/19</b>	<b>mié 27/03/19</b>	<b>14</b>	<b>Arquitecto de software</b>
19	Elección de software y pruebas	2 días	mié 27/03/19	vie 29/03/19	15	"
20	<b>3. Fase de desarrollo del producto</b>	<b>29 días</b>	<b>vie 29/03/19</b>	<b>mié 15/05/19</b>	<b>11</b>	<b>Técnico de software</b>
21	Construcción del espacio de desarrollo	2 días	vie 29/03/19	mar 02/04/19	19	"
22	Construcción del Data Warehouse	7 días	mar 02/04/19	jue 11/04/19	21	"
23	Construcción del ETL	7 días	jue 11/04/19	jue 25/04/19	22	"
24	<b>PEC 2</b>	<b>0 días</b>	<b>lun 22/04/19</b>	<b>lun 22/04/19</b>		<b>Jefe de proyecto</b>
25	Prueba con datos	1 día	jue 25/04/19	vie 26/04/19	23	Jefe de proyecto
26	Confección OLAP	6 días	vie 26/04/19	lun 06/05/19		Técnico de software
27	Configuración de elementos gráficos	4 días	mar 07/05/19	lun 13/05/19	25	"
28	Pruebas con datos	1 día	mar 14/05/19	mié 15/05/19	27	Jefe de proyecto
29	<b>4. Fase de ensayo y calidad</b>	<b>4 días</b>	<b>mié 15/05/19</b>	<b>mar 21/05/19</b>	<b>20</b>	<b>Técnico de software</b>
30	Disponer la solución en el SI	2 días	mié 15/05/19	vie 17/05/19	28	"
31	Documentación del ensayo	2 días	vie 17/05/19	mar 21/05/19	30	"

32	Pruebas de calidad	1 día	lun 20/05/19	lun 20/05/19	30	Jefe de proyecto
33	<b>PEC 3</b>	<b>0 días</b>	<b>lun 20/05/19</b>	<b>lun 20/05/19</b>		<b>Jefe de proyecto</b>
34	<b>Entrega final</b>	<b>18 días</b>	<b>mar 21/05/19</b>	<b>lun 17/06/19</b>	<b>29</b>	<b>Jefe de proyecto</b>
35	Revisión y redacción de memoria	7 días	mar 21/05/19	vie 30/05/19	31	"
36	Confección de presentación	7 días	vie 31/05/19	mar 11/06/19	35	"
37	Confección de autoinforme	2 días	mar 11/06/19	jue 13/06/19	36	"
38	Última revisión de entregables	2 días	vie 14/06/19	lun 17/06/19	37	"
39	<b>Liquidación de entregables finales</b>	<b>0 días</b>	<b>lun 17/06/19</b>	<b>lun 17/06/19</b>		"

Ilustración 1 desglose EDT

### 1.5.1 Diagrama de Gantt

Respecto a la distribución de las cargas de trabajo en el tiempo, cabe apuntar que se ha tenido en cuenta las fechas no vinculadas con el proyecto del tipo festivo o no laborables y fechas personales programadas. Dicho sea de paso, concretamente, los días laborales festivos son los recogidos en el calendario laboral de Barcelona:

- 18 y 22 abril
- 1 de mayo
- 10 y 24 de junio



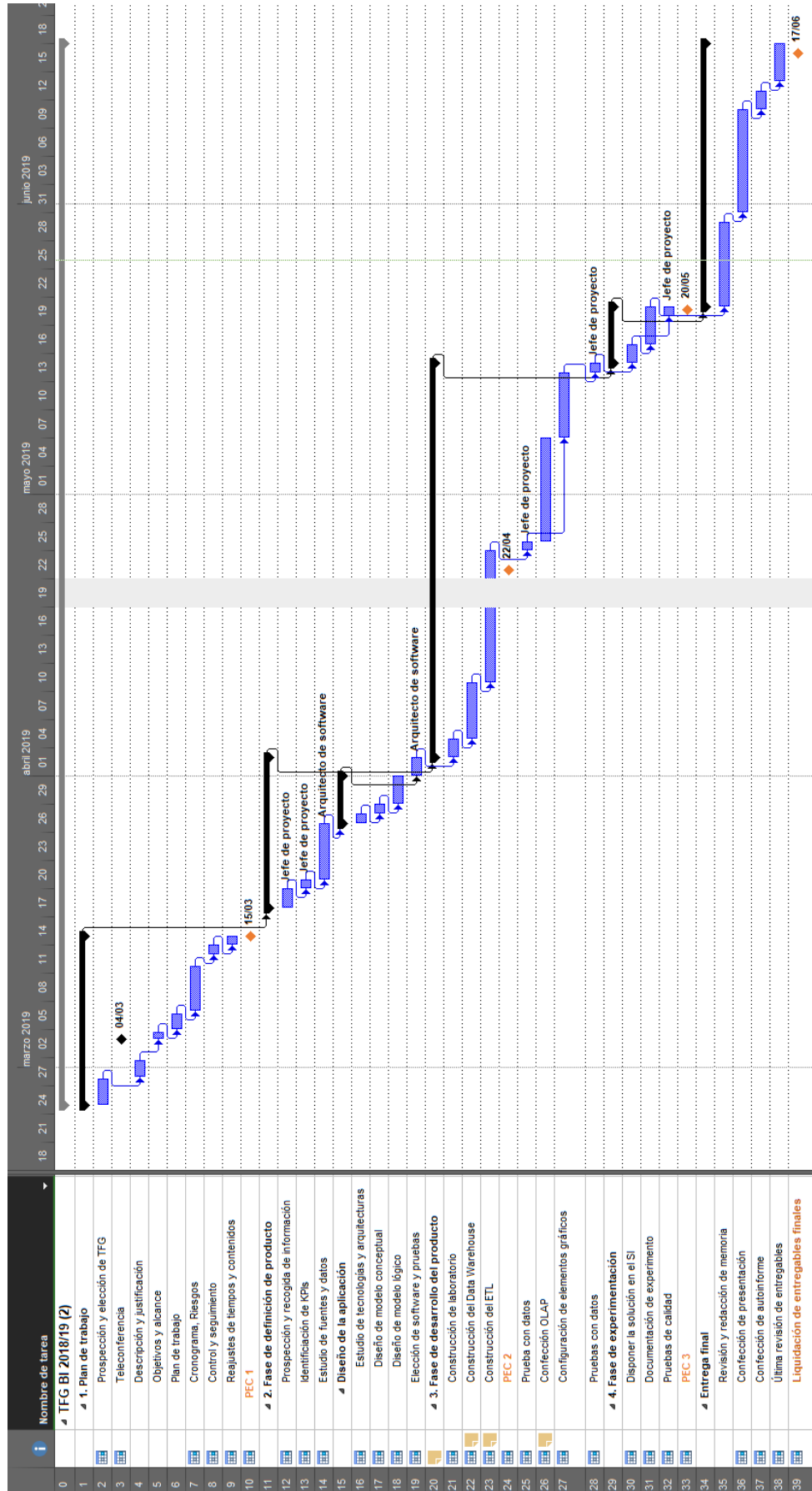


Ilustración 2: Cronograma de los trabajos

### 1.5.2 Riesgos

Ante las complejidades que supone la confección de un producto empleando tecnologías desconocidas, desde la posición del que escribe, es necesario dejar claro los riesgos esperados y como se plantea su gestión para mitigar los efectos.

Los riesgos previstos son:

<b>Saber interpretar las necesidades de los involucrados y de los posibles interesados.</b>	
Causa	Cotejar y cruzar la información con varias fuentes o perspectivas dentro del departamento.
Consecuencia	Que la aplicación no resulte de utilidad para los usuarios o peor aún, que el resultado confunda a los usuarios con el desacierto.
Probabilidad	Baja
Impacto	Alto
Acción	Cotejar y cruzar la información con varias fuentes o perspectivas dentro del departamento

<b>OLAP<sup>7</sup> poco funcional</b>	
Causa	Desconocimiento sobre la operativa y funcionalidad de las herramientas para la confección de cubos de explotación de datos durante las últimas etapas de desarrollo del cuadro de mandos.
Consecuencia	Contención de información errónea, análisis de datos complejos o de difícil representación.
Probabilidad	Media.
Impacto	Medio.
Acción	Procurar la identificación correcta de los análisis conocidos y empleados dentro del departamento de calidad. Procurar el uso de soluciones conocidas y con documentación disponible que permitan integrar los análisis en el dashboard.

<b>Dimensionar los tiempos de forma acertada.</b>	
Causa	lidar con tecnologías desconocidas hasta el momento o no disponer de los recursos físicos suficientes pueden provocar

<sup>7</sup> OLAP: [ver significado en glosario.](#)

	desvíos en el curso del proyecto. Estas circunstancias han obligado a la asignación de tiempos más elevados a trabajos clave de desarrollo y ajustando la asignación de recursos a la confección de otros desarrollos como el cubo OLAP o la recreación del escenario de ensayo.
Consecuencia	En un primer lugar, que el cubo, por ser el último desarrollo, no pueda ser creado como se espera y, en segundo lugar, cualquier contratiempo en la creación del escenario afectará directamente sobre el cronograma desde un principio de la fase de desarrollo.
Probabilidad	Alta
Impacto	Medio
Acción	Respecto a la creación del cubo, se buscará una herramienta que facilite su creación (fácil manipulación y comprensión). En el caso que se produzcan retrasos por este motivo, se ha previsto que su creación no afecte en la consecución del resto de tareas del proyecto en el caso que se decida abandonar su confección. Por otro lado, respecto a la creación del escenario de ensayo, al desarrollarse durante los primeros días del proyecto es posible remontar las desviaciones que se puedan producir y minimizar sus consecuencias.

**Valorar acertadamente de las herramientas que permitan la confección del Dashboard y ELT preparadas para facilitar el desarrollo del producto objetivo.**

Causa	No alcanzar a tiempo los conocimientos sobre la operativa, funcionamiento y parametrización del software necesario que puedan resolver las complejidades que faciliten la solución BI ideada.
Consecuencia	Efecto directo sobre el cronograma por la dependencia directa de las tareas precedentes durante el proceso de desarrollo. Por otro lado, una interpretación desviada de concepto o funcionalidad de alguna de las soluciones afectará directamente sobre el rendimiento producto y su desarrollo en etapas predecesoras.
Probabilidad	Media.
Impacto	Alto.
Acción	Procurar el uso de herramientas conocidas y reconocidas, en activo y con mucho tiempo de vida en el mercado. Asignación extra de tiempo para la familiarización con los interfaces y operativa de las aplicaciones.

### 1.5.3 Control y seguimiento

Además del mecanismo basado en entregables indicado en el plan docente del TFG, se propone hacer entregas parciales con ánimo de favorecer el seguimiento, gestión de desviaciones y aceptación de lo propuesto en la Planificación. Según se ha comentado, las entregas parciales se programarán en intervalos de tiempo no superiores a las dos semanas naturales, a expensas de imprevistos y procurando realizar los avisos pertinentes.

Por último, cabe apuntar la sustitución de las herramientas de gestión recomendadas dentro del marco de referencia SCRUM (pizarra) por un diario de seguimiento. Esta decisión se ha tomado por simplificar el seguimiento de la evolución del trabajo y porque no es necesaria la coordinación de más personal que el que escribe. Por otro lado, dicho documento también se incluirá el acta generada durante *Scrum Daily Meeting* (aunque su actualización no será estrictamente diaria) con los últimos *Sprint Tasks* realizados, las actividades del día y los impedimentos encontrados durante el abordaje de las tareas. En este sentido hay que decir que la figura del *Scrum Master* será adoptada por el consultor del TFG y su interacción se ejecutará en diferido, aun así, no se considera que esto sea un impedimento en la fluidez del proyecto.

### 1.5.4 Valoración económica del trabajo

A expensas de juicios y decisiones que se tomarán en los apartados iniciales del trabajo, y teniendo presente que se trata de un desarrollo experimental llevado a la práctica dentro de un entorno académico, se puede vaticinar que los costes derivados del mismo no supondrán un factor a tener en cuenta. Dicho de otra forma, al no ser un proyecto patrocinado y al tratarse de una iniciativa con la principal intención de demostrar los conocimientos sobre la materia BI, se procurará hacer uso de productos *Open Source* y el empleo de equipos propios o prestados.

Respecto a la posible intervención de personal experto externo cabe decir que no se dará el caso, en su conjunto, el trabajo, se llevará a la práctica por el que escribe. Aun así, se debe aclarar que los perfiles expertos incluidos en el *grid* de Plan de Trabajo se han empleado con intención informativa, por lo cual se debe entender su uso como una demostración de enriquecimiento del contenido.

### 1.5.5 Herramientas transversales

Al margen del software específico requerido para el desempeño del producto BI, el resto de aplicaciones transversales y de intención documental, o de intercambio de información, se limitan a productos estándar y ampliamente conocidos. En la siguiente lista se nombran los de más relevancias por alguno sus atributos:

- Suit Google (Docs, Sheets y Drive): Aspectos colaborativos y disponibilidad de la información gestionada o generada.

- Microsoft Word: Como herramientas de refinamiento de la documentación.
- Microsoft Project: Gestión del tiempo y representación gráfica de los tiempos.

## 1.6 Breve resumen de productos obtenidos

Al final el proyecto se espera haber obtenido un producto compuesto por varios sub productos complementarios entre sí:

- El principal componente del producto final es un dashboard confeccionado con alguna herramienta BI en forma de cuadro de mando. Dentro de lo posible, se procurará el acceso a la parte de visualización de datos analizados. Para tal propósito se habilitará un acceso a la versión publicada en el servicio *cloud* que permitirá experimentar en primera persona toda la operativa habilitada. También se entregará junto la memoria los archivos con la secuencia de sentencias ETL. Cabe decir que por necesidades de confidencialidad el conjunto de datos no será el original sobre el que se desarrollará el producto, de hecho, la información incrustada en la versión publicada será alterada con el fin de ocultar identidades o referencias a productos, personas o empresas.

Por último, existirá una tercera parte del producto con la estructura de datos (Data Warehouse), pero que no será entregada con los datos, en compensación se incluirá en el entregable un archivo con el código de construcción de la estructura.

- Una memoria que recapitulará toda la información clave concerniente al trabajo. Los aspectos de trabajo recogidos en el producto son: la planificación del proyecto, diseño del producto, implementación del producto y tareas de integración (con ánimo de experimentación y no tanto de poner en explotación).
- Una presentación grabada en video que se empleará en la defensa de la memoria ante el tribunal de revisión.
- Una presentación mediante diapositivas como apoyo a la presentación grabada en video, que recogerá resúmenes y detalles clave de la memoria.

## 1.7 Breve resumen del resto de apartados

El trabajo se compone de un conjunto de apartados ordenados de forma arbitraria con la intención de documentare el desarrollo de un proyecto Business Intelligence, de principio a fin. Según el orden de aparición, los apartados y subapartados siguientes son:

- Capítulo 2. Definición del producto. Capítulo dedicado a la concreción y justificación del diseño del producto, por lo que contendrá toda la parte más teórica de construcción del producto. El contenido del capítulo documentará todas las acciones de prospección de información sobre el caso, definición de los modelos de datos, identificación de métricas y

KPIs, terminando con la elección de software requerido por la solución perseguida.

- Capítulo 3. Desarrollo del producto. Será en este capítulo en donde se recogerá la evolución de construcción del producto y recorrido del proyecto ideado en el capítulo anterior. Los principales contenidos detallarán la construcción de la capa ETL, del Data Warehouse y la aplicación que configurará el *dashboard*.
- Capítulo 4. Fase de ensayo y calidad. El último capítulo se ha destinado a la revisión de la calidad del producto por parte de un integrante del departamento de calidad involucrado en el trabajo. Pero antes de dicha prueba se deberá trasladar el escenario de ensayo que permitirá la ejecución de los experimentos.

## 2. Definición del producto

Antes de entrar en pormenores del caso es oportuno dar unas pinceladas sobre la metodología seguida a lo largo del proyecto.

Tal y como se apuntó al inicio del trabajo, la orientación de cómo construir el repositorio de datos que alimenta la solución BI se enmarca en la metodología descrita por R. Kimball. Dichas pautas se basan en el Ciclo de Vida Dimensional del Negocio [*pie de pg. Business Dimensional Lifecycle*] el cual defiende la construcción paulatina del Data Warehouse recorriendo el conjunto de procesos o áreas funcionales de la empresa. Dicho de otra manera, a diferencia del modelo definido por Bill Inmon, que resalta la necesidad de hacer una copia de los datos total e integral (por ejemplo, todo el ERP), la idea propuesta por R. Kimball se centra en la idea de ir escogiendo las fuentes de datos según la intención del análisis para luego volcarlos sobre una estructura simple, construida con antelación y diseñada a propósito.

Mediante la gráfica que se acompaña (ilustración 3) se puede apreciar, con más detalle, como el flujo de desempeño secuencia una serie de actividades enfocadas a la gestión individual de cada análisis y de forma cíclica hasta completar el desarrollo del DW.

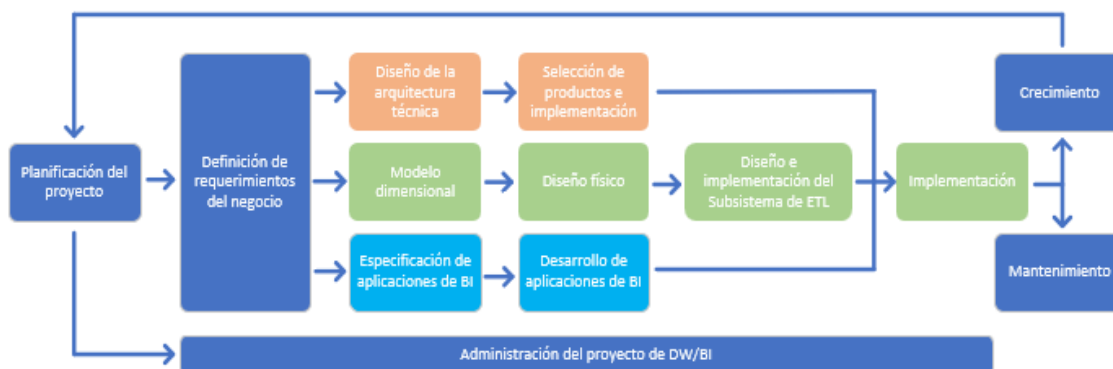


Ilustración 3: flujo de trabajo para el desarrollo de DW

Del conjunto de bloques representados se debe destacar el propósito del bloque “Planificación del Proyecto” por ser el punto en el que se divide en partes el producto final, *dashboard* definiendo así el número de iteraciones necesarias del resto de trabajos.

Por último, se debe aclarar que en trabajo no aparecerán de forma explícita todas las fases indicadas en el diagrama. Como se ha dicho antes, el seguimiento de la metodología Kimball es ajustado a las necesidades del TFG.

### 2.1 Prospección y recogida de Información

Las **acciones que se han llevado** a cabo durante esta fase comprenden esos trabajos que facilitan una perspectiva del caso y sus particularidades. En concreto, los trabajos realizados son: entrevista con personal relevante del departamento, revisión *in situ* de la operativa de las soluciones y un paseo por los repositorios de trabajo y *dataset*.

Gracias a las pesquisas se ha conseguido recrear el escenario actual dentro del laboratorio virtual y concretar objetivos de interés para el personal del departamento y para los interesados pertenecientes al gobierno de la empresa.

También se ha podido ver, en primera persona, el desarrollo de algunas de las tareas de documentación y gestión de los datos almacenados y en donde se almacenan.

Uno de los objetivos generales de la prospección ha sido la identificación de los procesos de negocio a estudiar, departamentos o áreas funcionales de la organización. En este sentido, algunos de los desempeños que se llevan a cabo desde el departamento y que han levantado interés son: gestión de las No Conformidades<sup>8</sup> y los Costes Evitables<sup>9</sup>. Dicho sea de paso, ambas perspectivas otorgan una visión crítica de la actividad en la organización, es por ello que los resultados de las métricas que las dimensionan suelen ser desencadenantes de Propuestas de Mejora, principalmente, en procesos relacionados con la Cadena de Valor y la Cadena de Suministro.

Por otro lado, y pese a la criticidad de ambos puntos de vista, cabe decir que ha sorprendido identificar los motivos que frenan el uso de la información que desprenden sus indicadores de desempeño. A continuación, se señalan algunos de los hechos principales:

- Alto grado de dispersión de los conjuntos de datos y pobre vinculación entre ellos.
- Conjunto de datos intratable por volumen y falta de unificación en el criterio de tipo de datos.
- Variedad de aplicaciones utilizadas para el tratamiento de los contenidos.
- Reutilización de aplicaciones en circunstancias que no son para las que se han diseñado.
- Repositorios inalcanzables bien sea por las restricciones de permisos de usuario o porque están localizadas en las estaciones de trabajo de los propios usuarios.
- Falta de unificación de conciencia tanto en los términos empleados como del lenguaje relacionado al contenido (emplear todos los mismos conceptos y sus referencias).
- No existe un mecanismo que promocióne y divulgue la existencia de los dataset y sus contenidos a la vez que los hace accesible.
- Una pobre representación de los contenidos o difícil comprensión de los números.

---

<sup>8</sup> No Conformidades o NC: ver significado en glosario

<sup>9</sup> Costes Evitables o CE: ver significado en glosario



En general, la propuesta que se ofrece como solución al conjunto de resistencias detectadas se centran en torno a la integración de una configuración de soluciones que, en última instancia, automatizarán la visualización de los datos de forma amigable.

A efectos de uso, durante la visualización de los datos por parte del usuario, la propuesta resultará ser un único entorno que incluirá una preconfiguración de las visualizaciones de los datos, pero que además ofrecerá la capacidad de modificar o aplicar el tratamiento de la información. Es por ello que se considera posible salvar todos los escollos relacionados con: localizaciones de los datos, premisos de acceso, estructuración de los datos, simplificar del software empleado, un único lenguaje y, por supuesto, mejorar la visualización de la información en cualquiera de sus intenciones (monitorización de evolución, métricas o KPIs).

## 2.2 Identificación de KPIs

Una de las intenciones de la prospección tenía como objetivo identificar las fuentes de información en uso por los componentes del departamento de calidad, en general, el catálogo de métricas y de KPIs. Dicho sea de paso, esta serie de análisis se tipifican dentro del conjunto descriptivo, no se da ningún caso en donde el estudio sea del tipo predictivo y mucho menos prescriptivo.

Volviendo a los resultados de la prospección, según las fuentes y datos identificados se han concretado cuales pueden ser los KPIs más apropiados relativos a las No Conformidades y los Costes Evitables. En su conjunto estas son métricas empleadas en el propio departamento de Calidad, por lo cual, su concreción es aportada por el propio departamento.

La lista presentada recoge la selección las métricas y objetivos:

CE 1	
Objeto	Conocer que tipos de Costes Evitables y el coste en el tiempo según cada Unidad de Negocio
Objetivo	contraste según el umbral marcado desde la dirección técnica según lo conocido
Dimensiones	tipo de Costes Evitables; UN, fecha
Métrica	acumulado del coste un tipo de Costes Evitables en el tiempo según la UN
Granularidad	variable en el tiempo y la Unidad de Negocio
Valor óptimo	tiende a cero (0)
Fuente de datos	<archivos>.csv

<b>CE 2</b>	
Objeto	Conocer el número de veces que aparece un tipo de Coste Evitable que se reproducen lo largo de un proyecto
Objetivo	contraste según el umbral marcado desde la dirección técnica según lo conocido.
Dimensiones	tipo de Costes Evitables; proyecto; Unidad de Negocio
Métrica	acumulado de Costes Evitables recurrentes por proyecto.
Granularidad	Unidad de Negocio y proyecto
Valor óptimo	tiende a cero (0)
Fuente de datos	<archivos>.csv

<b>CE 3</b>	
Objeto	Conocer las consecuencias de un tipo de Coste Evitable en un proyecto
Objetivo	contraste según el umbral marcado desde la dirección técnica según lo conocido en cuento el tipo de CE y proyecto
Dimensiones	tipo de Costes Evitables; proyectos; UN
Métrica	número veces que un Coste Evitable se repite y el coste acumulado del total por tipo
Granularidad	Unidad de Negocio y proyecto
Valor óptimo	tiende a cero (0)
Fuente de datos	<archivos>.csv

<b>NC 1</b>	
Objeto	conocer cuanto recurrente es una No Conformidad por un mismo proveedor en un periodo de tiempo.
Objetivo	contraste según el umbral marcado desde Calidad según lo conocido.
Dimensiones	tipo de No Conformidad; proveedor, fecha.
Métrica	suma de todas las conformidades de un mismo proveedor, pero sólo interesan si el número es mayor que 1.
Granularidad	variable en la Unidad de Negocio y el tiempo.
Valor óptimo	tiende a uno (1).
Fuente de datos	TRABILIAD 2003.MDB

<b>NC 2</b>	
Objeto	conocer cuantas No Conformidades se producen por proyecto.
Objetivo	contraste según el umbral marcado desde Calidad según lo conocido.
Dimensiones	tipo de No Conformidad; proyecto; tiempo; UN.
Métrica	suma de todas la No Conformidades de un mismo proyecto.
Granularidad	variable en la Unidad de Negocio y el tiempo.
Valor óptimo	tiende a cero (0).
Fuente de datos	TRABILIAD 2003.MDB

<b>NC 3</b>	
Objeto	conocer qué proporción de No Conformidades se dan de un mismo tipo respecto el total.
Objetivo	contraste según el umbral marcado desde Calidad según lo conocido
Dimensiones	tipo de No Conformidad; tiempo; UN.
Métrica	acumulado de un tipo de conformidad entre el total de conformidades que se dan dentro de un mismo periodo de tiempo.
Granularidad	variable en el tiempo y Unidad de Negocio
Valor óptimo	tiende a cero (0).
Fuente de datos	TRABILIAD 2003.MDB

### 2.3 Estudio de fuentes y datos

Otra de las intenciones de la prospección era alcanzar las fuentes capaces de alimentar la solución de análisis ideada. Como resultado se han identificado dos tipos de fuentes bases de datos MDB y archivos CSV, cada uno con información referente a los NC y los CE, respectivamente.

Cabe decir que localizar el origen de los datos resultó fácil, el verdadero problema se ha detectado en la pobre organización estructural de los contenedores como de la calidad los datos que estas albergan. Se cree que este último hecho se traducirá como un sobre esfuerzo en el diseño de los trabajos ETL.

A modo ilustrativo se incluyen dos capturas en donde se puede apreciar el contenido de un del archivo MDB para la gestión de los registros de No Conformidad.

The screenshot displays a database application interface with a sidebar on the left containing a tree view of tables. The main area shows several data tables:

- 002 DATOS IMPROMPTU:** Columns include ID, CODIGO PROVEEDOR, NUMERO ORDEN, PROYECTO, FECHA RECEPCION, CANT APROBADA, CANTIDAD RECHAZADA, MOTIVO RECHAZO, IMPORTE, IMPORTE FACTURADO, ARTICULO, ESTATUS EN, and NO INCIDEN.
- DATOS ENCABEZADO:** Columns include ID, COD\_PROYECTO, FECHA\_ETIQUETADO, UN, RESPONSABLE, and various quality control indicators (crítica, alto, medio, bajo).
- HOJA TRAZABILIDAD:** Columns include ID, COD\_PROYECTO, FECHA\_ETIQUETA, N\_PEDIDO HTB, NC COMPRAS, RECHAZADA, CALIDAD PRODUCTO, SUBTIPO NC, TIPO DE RECEPCION, REFERENCIA, ID AUTOMA, riesgo, Cer, Mat, rez, cer.
- INFORME:** Columns include NINFORME, ID, PRODUCTO, FECHA, PROVEEDOR, QF, TIEMPO1, TIEMPO2, PROYECTO, UNIDAD DE NEGOCIO, PLANO, evaluado, TOTAL PIEZAS, and ADMITIDAS.
- TIPO REFERENCIAS:** Columns include N INFORME SUB, PROYECTO, N PEDIDO TR, CODIGO NC1, CODIGO NC2, PLANO, DESTINO, TOTAL PIEZAS, RECHAZADAS, COSTE UINIUNITARIO, RES ENTRADA REC, FECHA IN REC, and RES SALID.
- CODIGO-1:** Columns include NINFORME, coste horas, CODIGO, HORAS, PROVEEDOR, fecha, HORAS DPTO, DPTO, GRAVEDAD, HORAS DPTO1, DPTO1, GRAVEDAD1, and HORAS DPTO2.

Ilustración 4 Muestra del contenido de la DDBB empleada para la gestión de las NC.

En la siguiente captura se puede apreciar la ausencia de relaciones entre dichas tablas y consecuencia de ellos se puede vaticinar el coste que supondrá su transformación.

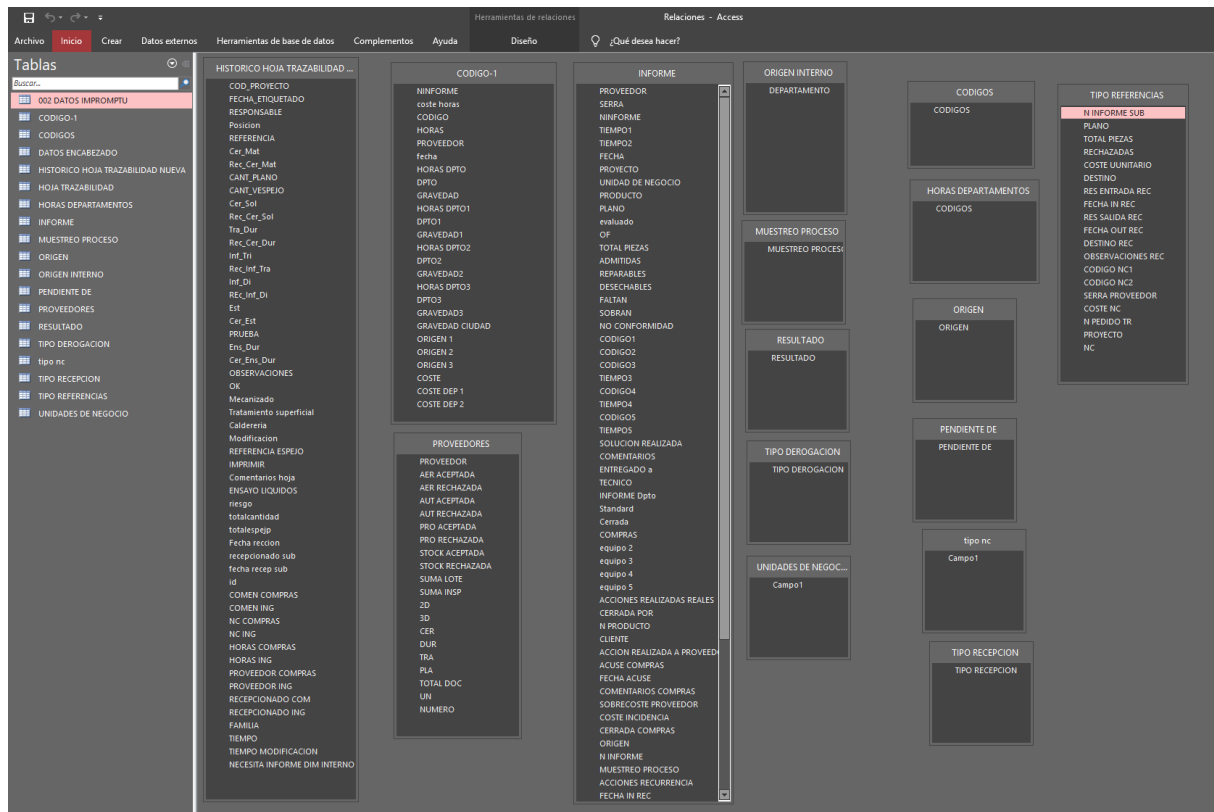


Ilustración 5 Demostración de ausencia de estructura relacional en la DDBB.

En su favor se debe resaltar que gracias a que es una misma aplicación la que gestiona el contenido de las DDBB, Microsoft Access, tratamiento del tipo de dato está unificado en el conjunto de las tablas.

Antes de desgranar las DDBB identificadas como válidas se debe apuntar un hecho: pese a que habría sido lo deseado, los datos no se extraerán atacando las DDBB del ERP en explotación, Baan, en su lugar se emplearán los datos que se almacenan en los MDB y los archivos CSV. Dichos repositorios son cargados por los usuarios con los datos extraídos del ERP mediante una aplicación dedicada para la confección de listados y reportes. En la medida de lo posible, esta posibilidad se valorará como una mejora en futuras ampliaciones.

Volviendo a las fuentes de interés localizadas, en la lista de a continuación se indican principales tablas de consulta y algunos de sus atributos:

INFORME	
Fuente	TRAZABILIDAD_2003.mdb
Objetivo	<i>back-end</i> <sup>10</sup> que centraliza los datos que configuran los informes fruto de cada No Conformidad

<sup>10</sup> Back-end: [ver significado en el glosario](#).

Estructura de datos	Datos tabulados embebidos en DDBB y sin relaciones ni PK.
Origen de los datos	los datos son introducidos por todo el personal de ingeniería y el departamento de Calidad a través del <i>front-end</i> <sup>11</sup> .
Frecuencia de actualización	su uso es diario y dependiendo de los procesos de recepción de productos
Cantidad de registros	en el momento de su estudio 3678 con un incremento anual entorno a los 800.

NINFORME	ID	PRODUCTO	FECHA	TIEMPO1	TIEMPO2	PROYECTO	UNIDAD DE NEGOCIO	PLANO
3092	18673	LATERAL CUERPO DCHO./	14/02/2017	0	0	ALM	PRODUCTO	
3091	18671	SOPORTE CON CALA	13/02/2017	0	0	90253	PRODUCTO	
3090	18668	PERFIL APOYO	13/02/2017	0	0	90253	PRODUCTO	
3089	18667	ESTRUCTURA ARMARIOS	13/02/2017	0	0	90253	PRODUCTO	

Ilustración 6: Detalle las NC y su número de registros.

## HOJA DE TRAZABILIDAD

Fuente	TRAZABILIDAD_2003.mdb
Objetivo	<i>back end</i> que centraliza los datos de seguimiento de las No Conformidad, desde su creación hasta su cierre y manteniéndolas como históricos
Estructura de datos	Datos tabulados embebidos en DDBB y sin relaciones ni PK.
Origen de los datos	recoge los datos del <i>front end</i> los cuales son introducidos por los usuarios involucrados en el seguimiento de la No Conformidad y de la tabla 002 DATOS IMPROMPTU.
Frecuencia de actualización	su uso es diario y dependiendo de los procesos de negocio
Cantidad de registros	en el momento de su estudio 3.8701 con un incremento anual entorno a los 800

<sup>11</sup> Front-end: [ver significado en el glosario](#).

id	COD_PROYECTO	FECHA_ETIQUET	N_PEDIDO_HTB	NC_COMPRAS	NC_ING	SUBTIPO_NC	TIPO_DE_RECEPCION	REFERENCIA	ID_AUTOMA	riesgo
21170	31129	27/11/2018		RECHAZADA	CALIDAD PRODUCTO		DOCUMENTACION + VISUAL	31129_01_M_145_1_03_018	38926	ALTO
21170	31129	27/11/2018		RECHAZADA	CALIDAD PRODUCTO		DOCUMENTACION + VISUAL	31129_01_M_145_1_03_017	38925	ALTO
21169	ALMACEN	27/11/2018	660662	FALTA MATERIAL			SOLO DOCUMENTACION	109728		38924
21168	90276	27/11/2018	879206	FALTA MATERIAL			SOLO DOCUMENTACION	90276_0750_014		38921
21168	940211	27/11/2018	879187	FALTA MATERIAL			SOLO DOCUMENTACION	D681021142		38920
21168	940211	27/11/2018	879187	FALTA MATERIAL			SOLO DOCUMENTACION	D681021141		38919

Ilustración 7: NC registradas y la cantidad de registros

002 DATOS IMPROMPTU	
Fuente	TRAZABILIDAD_2003.mdb
Objetivo	a modo de <i>staging</i> se emplea de repositorio intermedio de donde recogen datos el resto de tablas.
Estructura de datos	Datos tabulados embebidos en DDBB y sin relaciones ni PK.
Origen de los datos	contiene los datos extraídos del ERP mediante la aplicación Impromptu
Frecuencia de actualización	su uso es diario y dependiendo de los procesos de negocio.
Cantidad de registros	45714 registros con un incremento anual de 15000 registros brutos.

id	CODIGO_PROVEEDOR	NUMERO_ORDEN	PROYECTO	FECHA_RECEPCION	CANT_APROBADA	CANTIDAD_RECHAZADA	MOTIVO_RECHAZO	IMPORTE	IMPORTE_FACTURADO	ARTICULO	ESTATUS_EN
2443066	E01545	879045	90304	23/11/2018	5	0		76		0 78369-01-160-015	1
2444207	E00100	878886	90304	23/11/2018	1	0		912		0 78370-01-130-001	1
2444206	E00100	878886	90304	23/11/2018	2	0		1240		0 78369-01-140-001	1
2444205	E00100	878886	90304	23/11/2018	1	0		912		0 78369-01-130-001	1
2444204	E00388	660537		23/11/2018	0	-2		-69		-69 ABONO	3
2443076	E03442	879131	90307	23/11/2018	2	0		55		0 78017.01.329.025	1

Ilustración 8: Registro de importados des Impromptu

DATOS ENCABEZADO	
Fuente	TRAZABILIDAD_2003.mdb
Objetivo	contiene los datos que definen cada caso de trazabilidad, de cada registro cuelgan un grupo de entradas de la tabla HOJA DE TRAZABILIDAD.
Estructura de datos	Datos tabulados embebidos en DDBB y sin relaciones ni PK.
Origen de los datos	roge los datos del <i>front end</i> que son introducidos por los usuarios involucrados en el seguimiento de la No Conformidad y de la tabla 002 DATOS IMPROMPTU

Frecuencia de actualización	su uso es diario y dependiendo de los procesos de negocio
Cantidad de registros	en el momento de su estudio 3.8701 con un incremento anual entorno a los 800.

Ilustración 9: registro con la trazabilidad de NC.

REGISTRO CE DE AUTOMOCIÓN	
Fuente	<mes_proyectos>.CSV
Objetivo	recoger la relación entre las Ordenes de Fabricación y los casos de Costes Evitables que se dan durante un solo mes en la Unidad de Negocio de Producto
Estructura de datos	Datos separados por “;” y sin relaciones ni PK.
Origen de los datos	los datos son importaciones, manuales, de los listados de la aplicación Impromptu la cual visualiza información de las tablas del ERP
Frecuencia de actualización	su uso es diario y dependiendo de los procesos de negocio.
Cantidad de registros	no se puede concretar hasta que no realice el tratamiento de los datos, pero se estima que oscila el número de registros entre 75.000 a 150.000 transacciones en bruto, por cada mes.

Ilustración 10: contenido de un archivo CSV de CE de Automoción.



REGISTRO CE DE PRODUCTO	
Fuente	<mes_proyectos>.csv
Objetivo	recoger la relación entre las Ordenes de Fabricación y los casos de Costes Evitables que se dan durante un solo mes en la Unidad de Negocio de Producto.
Estructura de datos	Datos tabulados embebidos en DDBB y sin relaciones ni PK.
Origen de los datos	los datos son importaciones, manuales, de los listados de la aplicación Impromptu la cual visualiza información de las tablas del ERP.
Frecuencia de actualización	su uso es diario y dependiendo de los procesos de negocio
Cantidad de registros	no se puede concretar hasta que no realice el tratamiento de los datos, pero se estima que oscila el número de registros entre 75.000 a 150.000 transacciones en, cada mes

```

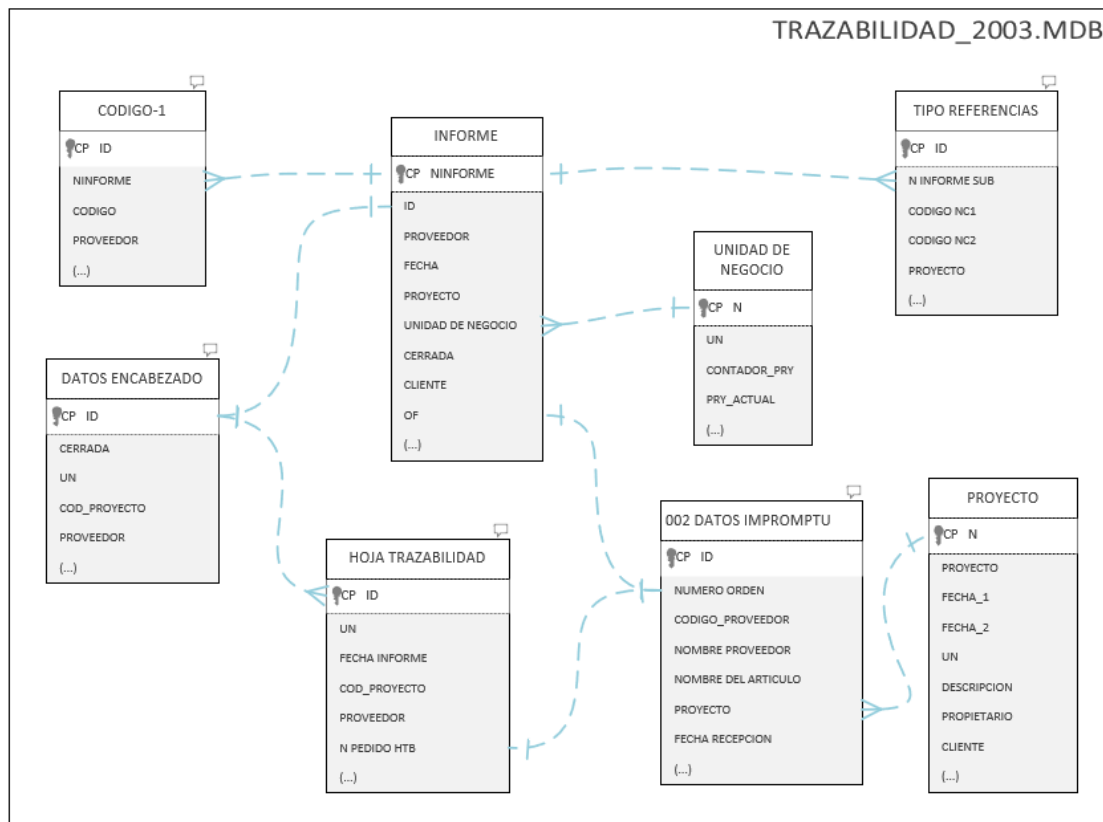
1 | Proj., "OF Base", "Tscos", "Tsonno", "Tspono", "Tscost", "Lin.", "Tscst", "Tsunid", "Tsamnt", "Tscurs", "C.Coste", "Pos.", "Tipo", "Linea", "Tknno3", "Tknno4", "Tknno5", "Tknno6", "Tknno7", "Tknno8", "Tknno9", "Tknno10", "Tknno11", "Tknno12", "Tknno13", "Tknno14", "Tknno15", "Tknno16", "Tknno17", "Tknno18", "Tknno19", "Tknno20", "Tknno21", "Tknno22", "Tknno23", "Tknno24", "Tknno25", "Tknno26", "Tknno27", "Tknno28", "Tknno29", "Tknno30", "Tknno31", "Tknno32", "Tknno33", "Tknno34", "Tknno35", "Tknno36", "Tknno37", "Tknno38", "Tknno39", "Tknno40", "Tknno41", "Tknno42", "Tknno43", "Tknno44", "Tknno45", "Tknno46", "Tknno47", "Tknno48", "Tknno49", "Tknno50", "Tknno51", "Tknno52", "Tknno53", "Tknno54", "Tknno55", "Tknno56", "Tknno57", "Tknno58", "Tknno59", "Tknno60", "Tknno61", "Tknno62", "Tknno63", "Tknno64", "Tknno65", "Tknno66", "Tknno67", "Tknno68", "Tknno69", "Tknno70", "Tknno71", "Tknno72", "Tknno73", "Tknno74", "Tknno75", "Tknno76", "Tknno77", "Tknno78", "Tknno79", "Tknno80", "Tknno81", "Tknno82", "Tknno83", "Tknno84", "Tknno85", "Tknno86", "Tknno87", "Tknno88", "Tknno89", "Tknno90", "Tknno91", "Tknno92", "Tknno93", "Tknno94", "Tknno95", "Tknno96", "Tknno97", "Tknno98", "Tknno99", "Tknno100"
2 | "910077",610006,1,0,0,340737,270,0,1,5,8,0,"01",10,"Compra",10,270,0,0,2099.95,"Materiales","PEDIDO ORIGINAL","79633",11239.07,2018-08-27 12:00
3 | "910077",610006,1,0,0,340737,300,0,2,60,41,0,"01",10,"Compra",10,300,0,0,2099.95,"Materiales","PEDIDO ORIGINAL","79633",11239.07,2018-08-27 12:00
4 | "910077",610006,1,0,0,340737,210,0,1,3,16,0,"01",10,"Compra",10,210,0,0,2099.95,"Materiales","PEDIDO ORIGINAL","79633",11239.07,2018-08-27 12:00
5 | "910077",610006,1,0,0,340737,390,0,2,229,11,0,"01",10,"Compra",10,390,0,0,2099.95,"Materiales","PEDIDO ORIGINAL","79633",11239.07,2018-08-27 12:00
6 | "910077",610006,1,0,0,340737,380,0,1,8,28,0,"01",10,"Compra",10,380,0,0,2099.95,"Materiales","PEDIDO ORIGINAL","79633",11239.07,2018-08-27 12:00
7 | "910077",610006,1,0,0,340737,340,0,1,208,0,"01",10,"Compra",10,340,0,0,2099.95,"Materiales","PEDIDO ORIGINAL","79633",11239.07,2018-08-27 12:00
8 | "910077",610006,6,871186,5,340737,230,0,2,64,0,"01",10,"Compra",10,230,0,0,2099.95,"Materiales","PEDIDO ORIGINAL","79633",11239.07,2018-08-27 12:00
9 | "910077",610006,1,0,0,340737,220,0,1,1,48,0,"01",10,"Compra",10,220,0,0,2099.95,"Materiales","PEDIDO ORIGINAL","79633",11239.07,2018-08-27 12:00
10 | "910077",610006,1,0,0,340737,290,0,1,110,5,0,"01",10,"Compra",10,290,0,0,2099.95,"Materiales","PEDIDO ORIGINAL","79633",11239.07,2018-08-27 12:00
11 | "910077",610006,1,0,0,340737,370,0,1,1,65,0,"01",10,"Compra",10,370,0,0,2099.95,"Materiales","PEDIDO ORIGINAL","79633",11239.07,2018-08-27 12:00

```

Ilustración 11 : Contenido de un archivo CSV de CE de Producto

Para finalizar, como resultado del estudio de las fuentes se ha concluido cómo deberían ser las relaciones entre las tablas de cada mdb y los archivos csv, como parte del *back end*. En el caso de las ddbb, la implementación del *front end* en Microsoft Access<sup>12</sup>, desde donde se recogen los datos. El diseño de la gestión de las tablas no se ha basado en relaciones, simplemente se hacen búsquedas tabla por tabla cada vez que se ejecuta una *query* desde los informes.

<sup>12</sup> En la actualidad la versión de Microsoft Access es la incorporada en la suit Office 2007 Professional y se está pendiente de migración de dicho paquete a la versión 2013.



*Ilustración 12: relaciones que faltan entre las tablas de la BBDD de Access*

## 2.4 Diseño de la aplicación

### 2.4.1 Estudio de la arquitectura y esquemas

No se entrará en contrastes con otros modelos (Modelo de Inmon, Hefesto, The SAS Rapid Data Warehouse Methodology o Mastering Data Warehouse Design) [2][3], pero sí se aporta unas propiedades teóricas sobre la arquitectura seguida antes de entrar en el caso particular.

El modelo de DW desarrollado por R. Kimball se enfoca hacia una estructura interna de datos *bottom-up*, es decir, la estructura de almacenamiento se inicia desarrollando particiones de la estructura general según el objetivo que debe cumplir cada subestructura hasta alcanzar el esquema completo, dando como resultado una constelación de Data Marts interconectados.

Este mecanismo de desarrollo facilita la configuración y escalabilidad del DW centrando su intención en el tratamiento particular de cada caso. Tanto es así que se pueden gestionar cada datamarts por cada caso de análisis en base a las dimensiones recuperadas de unos datos específicos.

La estructura de almacenamiento del modelo de R. Kimball, orientada a las dimensiones y de contenido no normalizado, se diseña al margen del origen de los datos y es configurada en base a las tablas que recogen las dimensiones, por un lado, y tablas con los hechos (*facts tables*), por otro lado.

Otro elemento relevante del DW es el *bus structure* como mecanismo que interconecta entre sí los repositorios internos, el cual permite la ejecución de *queries* sobre el conjunto de datos para resolver las consultas de las soluciones de visualización.

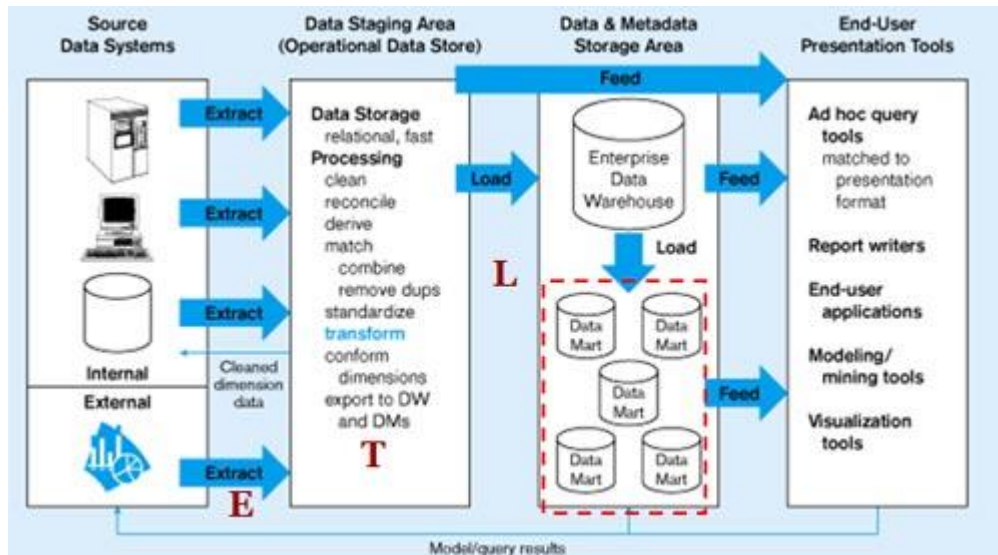


Ilustración 13: modelo R. Kimball [<http://www.zentut.com/data-warehouse/data-mart/>]

Respecto a los datamarts, que componen la base de la estructura, comentar que se pueden encontrar dos tipos diferenciados según su jerarquía relacional. Las dos configuraciones básicas son: esquema en estrella -la tabla de hechos hace referencia a todas las tablas de con las dimensiones- o esquema en copo de nieve -la tabla de hechos no centraliza las referencias de todas las tablas de dimensiones, también se pueden referencias entre ellas-.

Retomando el caso de estudio. Basándonos en la estructura conceptual de la arquitectura Kimball podemos resumir el modelo según las necesidades de caso tal y como se presenta en la gráfica.

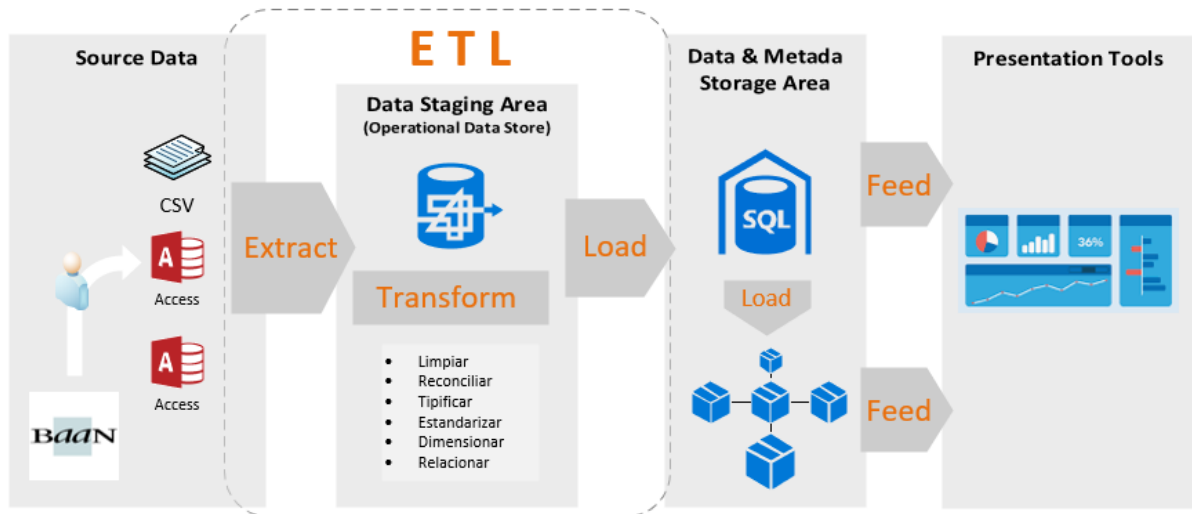


Ilustración 14: adaptación del modelo de R. Kimball al producto en desarrollo.

De la ilustración cabe destacar el conjunto de trabajos del proceso ETL por ser opcional y condicionado por las fuentes de datos. Para el caso particular del trabajo en desarrollo se ha optado por incluir una capa de almacenamiento intermedio, o *staging*, para una mejor manipulación de los datos durante su transformación. La decisión se ha tomado ante la previsión por la dispersión de los datos en pequeños archivos de base de datos (mdb). Este aspecto puede variar dependiendo de las capacidades y limitaciones del software seleccionado para tal desempeño.

Para finalizar, comentar que existen tablas transversales que serán reutilizadas en varios cruces de datos entre dimensiones y hechos, esto obliga, por motivos de optimización, compartir las dimensiones entre tablas de hechos.

#### 2.4.2 Diseño de modelo conceptual

La modelización de los procesos sobre los que se quieren actuar, pertenecientes al departamento de Calidad, se centran en la monitorización de los unos hechos relacionados con las No Conformidades y los Costes Evitables, todo ello con la intención de mejorar en agilidad, visibilidad y unificar el lenguaje en el estudio de los análisis, traduciéndose el esfuerzo en un ahorro de tiempo y desempeño del personal.

Pese a que los elementos objeto de estudio, NC y CE, pueden tener relación en algún momento dentro del ámbito del departamento de Calidad, desde la perspectiva de gestión de sus datos, en este caso, se tratarán por separado.

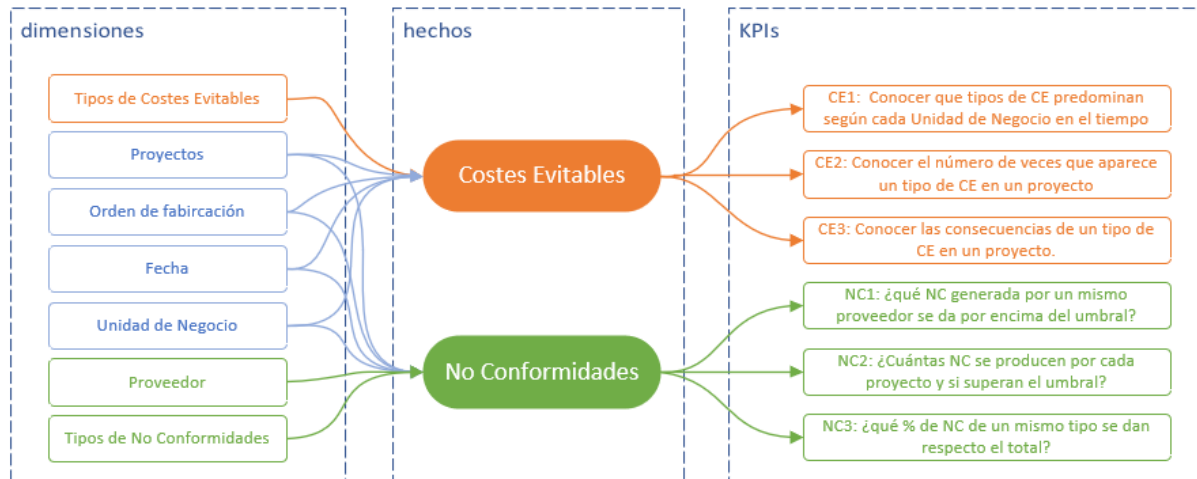


Ilustración 15: modelo conceptual del producto ideado según los requerimientos del caso

Respecto a la frecuencia de actualización de los datos del DW comentar que pese a no ser necesaria una antigüedad de días sí es crece necesaria que los datos no superen los 10 días. También es importante tener en cuenta el hecho que son varios los *dataset* involucrados con orígenes diversos. Para este factor sería conveniente disponer de un mecanismo que coordine las fechas de las diferentes de las dimensiones (buen ejemplo de la gestión de los metadatos). Tal y como ha ocurrido en otros aspectos del trabajo se considera que esta actividad superaría la intención académica y alcance del proyecto, pero que puede ser valorable en futuras mejoras del producto.

Para concluir el apartado se adjunta la matriz (Tabla 1) de bus para el cruce entre procesos y dimensiones como ayuda para el diseño del modelo lógico.

		dimensiones					
		Tipo CE	Proyecto	Fecha	Unidad de Negocio	Proveedor	Tipo NC
procesos	Costes Evitables	*	*	*	*		
	No Conformidades		*	*	*	*	*

Ilustración 16: matriz de bus de cruce entre procesos

### 2.4.3 Diseño de modelo lógico

Manteniendo en el punto de vista la modelización de los procesos objeto de estudio, CE y NC, se confecciona el modelo lógico del Data Warehouse como plataforma para el sistema de visualización de los KPIs.

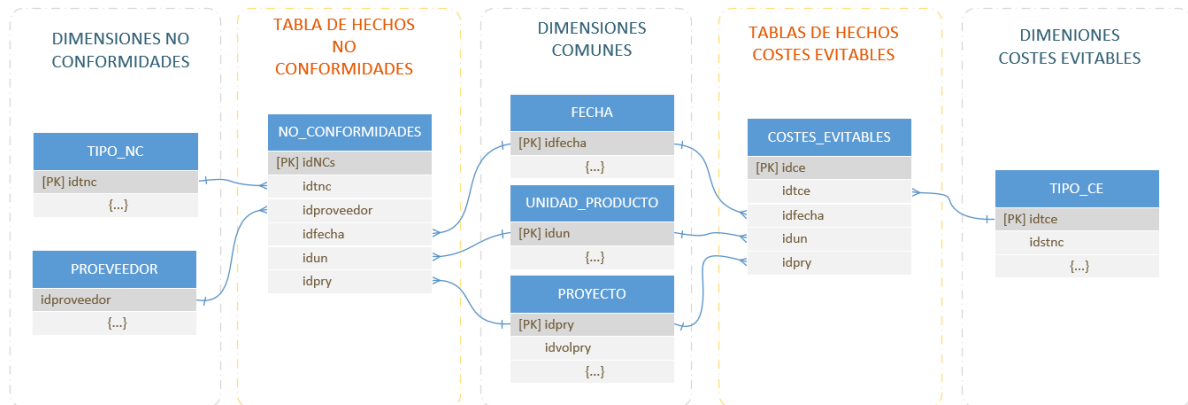


Ilustración 17: modelo lógico del Data Warehouse

En el diagrama se puede distinguir como la estructura se compone por dos subestructuras en estrella que dependen de la reutilización de dimensiones comunes. Se ha optado por esta fórmula por la agilidad y mejora en el rendimiento del *dashboard*.

### Hechos.

Este tipo de tabla contiene los indicadores de negocio que dan sentido al análisis de las dimensiones. Tal y como se ha indicado, serán dos las tablas de hechos las que compondrá el Data Warehouse.

Tabla de hechos No Conformidades			
campo	tipo	descripción	Observaciones
idncs	(FK) INT	FK del hecho No Conformidad	Indicador único del hecho NC.
idfecha	fecha	Momento en el que se da el hecho.	El momento en el que se produce el hecho
idpry	INT	FK del proyecto involucrado en el hecho	El hecho solo se da en un solo proyecto.
idun	INT	FK de la UN	El hecho solo puede tener una sola UN.
idproveedor	INT	FK del proveedor	El hecho solo puede tener un proveedor y este puede ser externo o interno.
idNC	INT	FK de los tipos de NC	El hecho define el motivo de la NC

Ilustración 18: tabla de hechos No Conformidades

Tabla de hechos Costes Evitables			
campo	tipo	descripción	Observaciones
idces	(FK) INT	FK del hecho Costes Evitables.	Indicador único del hecho CE.
idfecha	INT	FK del momento.	El momento en el que se produce el hecho.
idpry	INT	FK del proyecto involucrado en el hecho.	El hecho solo se da en un solo proyecto.
idun	INT	FK de la UN	El hecho solo puede tener una sola UN.
idprove	INT	FK del proveedor involucrado el hecho.	El hecho solo puede tener un proveedor y este puede ser externo o interno.
idce	INT	FK de los tipos de CE.	El hecho se cataloga por el tipo y subtipo.
coste	Decimal	Valor en €	Coste que supone el CE.

Ilustración 19: tabla de hechos Costes Evitables

**Claves subrogadas.** Otro aspecto contemplado en las tablas de hechos son las claves subrogadas asignadas a las dimensiones. Estas se emplearán con la intención de establecer una indización **independiente** de la existente en las fuentes originales.

Pese a lo deseado, por mucho empeño puesto en el diseño de las estructuras del tipo estrella, no es fácil el purismo en este sentido. En general, por la necesidad de mantener la jerarquía entre las tablas, sus estructuras suelen resultar con cierto grado en forma de copo de nieve. Por otro lado, esto no debe preocupar en el caso tratado en el trabajo ajustado volumen de datos tratado, se cree que no implicarán graves consecuencias sobre el rendimiento de la solución final.

### Dimensiones.

Dimensión FECHA			
campo	tipo	descripción	Observaciones
idfecha	(FK) INT	Clave foránea de tiempo	
Fecha	Date	Fecha completa	Formato: yyyy-MM-dd
anyo	INT	Solo el año	
mes_n	INT	Solo el mes en números	Número del mes
mes_corto	Varchar	Solo el mes formato corto	Contracción del mes en tres caracteres

Ilustración 20: dimensión fecha común a las dos tablas de hechos

Dimensión PROYECTO			
campo	tipo	descripción	Observaciones
idpny	(FK) INT	Clave foránea de los proyectos	
Codpny	Varchar	Referencia del proyecto	

Ilustración 21: dimensión Proyectos común con a las dos tablas de hechos.

Dimensión UNIDAD DE NEGOCIO			
campo	tipo	descripción	Observaciones
idun	(FK) INT	FK de las Unidades de Negocio	
codun	Varchar	Codificación UN	

Ilustración 22: dimensión Unidad de Negocio común a las dos tablas de hechos.

Dimensión TIPO NO CONFORMIDAD			
campo	tipo	descripción	Observaciones
idtnc	(FK) INT	FK de los tipos de No Conformidades	
codtnc	Varchar	Descripción de los tipos de NC	

Ilustración 23: dimensión Tipo NC

Dimensión TIPO COSTE EVITABLE			
campo	tipo	descripción	Observaciones
idtce	(FK) INT	FK de los tipos de Costes Evitables.	
destce	Varchar	Descripción de los tipos de CE	

Ilustración 24: dimensión Tipo CE

Dimensión PROVEEDOR			
campo	tipo	descripción	Observaciones
idproveedor	(FK) INT	FK de los tipos de Costes Evitables.	
codproveedor	Varchar	Descripción de los tipos de CE	

Ilustración 25: dimensión Proveedor común a las dos tablas de hechos.

**SCD (Slowly changing dimesions).** Respecto al factor que mide la velocidad de cambio del contenido en las tablas de dimensiones -cambio de domicilio de un proveedor o el nombre de algún producto, por ejemplo- comentar que no se tendrá en cuenta pese a que este puede ser un factor interesante a mantener registrado. El motivo principal de la decisión apuntada antes es la naturaleza académica del trabajo, pero no se descarta que pueda ser incluida como mejora del producto en ampliaciones futuras.



**Granularidad.** Se ha previsto la incorporación de elementos que faciliten granularidad según determinadas dimensiones transversales y propias a cada tabla de hechos.

Como anticipo sobre cómo llevar a la práctica la capacidad de navegabilidad de la solución, señalar que se incluirán tablas o columnas que faciliten la agrupación valores de manera que resulte más práctico el empleo de los contenidos desde las soluciones de visualización. Pese a que puede parecer una forma poco optimizada de gestión de los datos, en este caso prima por encima de ello la practicidad y, por otro lado, el número de registros de estas dimensiones no es muy elevado.

Por ejemplo, se establecerá una dimensión FECHA en la que se incorporará columnas con año, número de mes, nombre corto de mes y otra columna conteniendo solo el día, con esta separación de la fecha en columnas se facilitan los diferentes niveles de navegación en el cuadro de mandos.

Las dimensiones que seguro permitirán granularidad serán FECHA, UNIDAD DE NEGOCIO y en la medida de lo posible, la tipificación tanto de los Costes Evitables y de las No Conformidades.

## 2.5 Elección de software y pruebas

En lo que respecta a la elección de las aplicaciones que permitan alcanzar el cuadro de mando ideado, se debe indicar que no se considera fundamental ninguna particularidad que condicione la elección de una u otra solución. Dicho de otra manera; la gestión de datos, la disponibilidad de las soluciones de visualización o el alcance de las funciones de análisis pueden ser cubiertos por la funcionalidad de cualquiera de los productos bajo licencia OpenSource o en calidad de demostración.

Con este fin también se ha empleado un referente como elemento de apoyo en la toma de decisiones, el cuadro mágico de Gartner.

La versión de este año, al igual que los últimos, destaca a los productos de Microsoft como tendencia. Dicho sea de paso, los expertos en BI señalan, sobre todo, al auto servicio como característica deseable en un producto BI, es decir, que pueda ser el propio usuario consumidor de *dashboard* quien pueda desarrollar a voluntad las pantallas de visualización. Incorporar dicha característica en los productos implica la anulación de la curva de aprendizaje en el uso de la solución y este es un factor relevante en el proyecto que se está llevando a cabo.



Ilustración 26: Cuadro Mágico de Gartner, Inc.

En sí, la elección se ha basado en un criterio práctico, principalmente por la facilidad de adaptación a la aplicación -curva de aprendizaje nula o mínima- y acceso a su documentación, dado que los conocimientos sobre la materia BI son limitados.

### **Criterios aplicados en la elección.**

El criterio de elección de los candidatos focaliza cuatro factores imprescindibles para el alcance del proyecto y, por otro lado, frente a la posibilidad de conflicto, el peso de cada uno de ellos viene definido por el orden de aparición en la lista.

- OpenSource o sin coste por el uso: la calidad de los productos actuales de este tipo de solución es elevado y anula la necesidad de inversión.
- Disponer de un grupo de apoyo o comunidad amplia y en activo.
- Ofrecer la capacidad de intercambio de información con otras fuentes de datos, así como el racionamiento de las otras partes. Este factor debería reducir el número de errores la simplificar el trabajo.
- Qué su explotación pueda ser, principalmente, local y con capacidad de exportación de los sub productos desarrollados por cada aplicación, además de que se facilite el acceso al dashboard para cualquier interesado.

También se valorará en positivo toda aquella característica que se considere desde la perspectiva del alumno, subjetivamente, una ayuda para el desarrollo del alguno de los sub productos.

En otra medida, se han dividido en tres grupos las soluciones candidatas según su desempeño y funcionalidad: herramientas *back end* para la implantación ETL, capa almacenamiento, herramientas *front end* y capa visualización. Los contrastes que se aportan son un compendio de diferentes fuentes de comparación:

### **Soluciones y propiedades**

#### **Herramientas *back end* ETL.**

	<b>Pentaho Data Integration v8.2 (KETTLE)</b>	<b>Talend Open Studio v7.1.1</b>
Lenguaje de desarrollo	JAVA con sus virtudes e inconvenientes.	JAVA con sus virtudes e inconvenientes.
Comunidad	Muy activa, en varios idiomas y con gran apoyo por parte del propietario.	Muy activa, en varios idiomas y con gran apoyo por parte del propietario.
Curva de aprendizaje	Gran facilidad de uso y orientado a profesionales BI.	Experto o desarrollador de soluciones BI.

Interoperabilidad	Alto nivel de interoperabilidad por la variedad de conectores.	Alto nivel de interoperabilidad por la variedad de conectores.
Gestión de trabajos	ETL (extracción, transformación y carga).	ETL (extracción, transformación y carga) y ELT (extraer, cargar, transformar). Generador de código para lenguajes Perl y Java.
Conjunto de acciones	Amplio y en constante revisión.	Amplio y en constante revisión.
Posicionamiento	Reconocido como uno de los productos líder BI <i>open source</i> .	Reconocido como el primer producto <i>open source</i> de integración de datos.
Gestión de datos	Gran flexibilidad en el desarrollo de transformaciones.	Muy potente, Optimiza su rendimiento empleando los motores de las DDBB relacionales (RDBMS).
Entorno de trabajo	Desarrollo del entorno gráfico basado en JBoss.	desarrollado bajo un entorno gráfico basado en Eclipse orientado a expertos.
Modalidad licencia	Community Edition bajo licencia GPL.	Version Community bajo licencia GPL v2.
Observaciones	Compuesto por las aplicaciones: Spoon, Pan, Chef y Kitchen.	Pertenece al paquete Talend Integration Suite. A su vez se compone de tres aplicaciones: modelador de negocios, Job Designer y Metadata Manager dentro.

### Capa de almacenamiento<sup>13</sup>.

	<b>PostgreSQL v10.7</b>	<b>MySQL v8.0:</b>
API	Entorno limitado y poco intuitivo.	Entorno multiplataforma para administración de los esquemas, MySQL Workbench, aceptable y suficiente para tal propósito.
Requisitos	Alto consumo de recursos hardware	Uso local sin necesidad de grandes recursos.

<sup>13</sup> <https://www.2ndquadrant.com/es/postgresql/postgresql-vs-mysql/>

Rendimiento	Equilibrio cuando se trabaja con grandes volúmenes de datos	Destaca por su velocidad
Interoperabilidad	Menos posibilidades en el intercambio de datos con soluciones de terceros	Capacidad de intercambio de información muy elevada.
escalabilidad	Especializado para el almacenamiento escalable.	Trabajo en <i>clusters</i> , pero limitada.
Enfoque	Funcionalidad avanzada y especializada en la gestión de grandes cantidades de datos y transacciones.	Enfoque transversal y general de almacenamiento de datos
comunidad	Comunidad muy potente y activa ( <i>PostgreSQL Global Development Group</i> )	Comunidad muy potente y activa.
Licencia	PostgreSQL License	Licenciamiento GPL

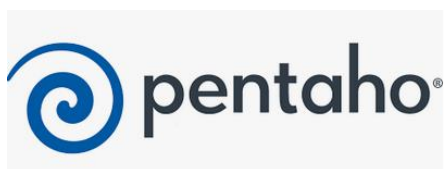
### Capa de visualización y herramientas *front end*<sup>14</sup>.

	Pentaho v8.2	Microsoft Power BI (desktop)	Qlik Sense (desktop)
entorno	No es completamente intuitivo, interface compleja.	Muy intuitivo.	Muy intuitivo
productividad	Ágil para el desarrollo de informes	Ágil para el desarrollo de informes	Ágil para el desarrollo de informes
Curva de aprendizaje	Aprendizaje lento, orientado a expertos	Curva de aprendizaje mínima, orientado al auto servicio	Curva de aprendizaje mínima, orientado al auto servicio
Posicionamiento	Reconocido como uno de los productos líder BI <i>open source</i> .	Según Gartner, Microsoft es líder en cuanto a soluciones BI propietarias	Reconocimiento como líder y competencia directa con Power BI.
Lenguaje de desarrollo	JAVA	Varias tecnologías	Varias tecnologías
Tools ETL	Incorpora potentes herramientas ETL	Incorpora conjunto de	No incorpora herramientas ETL.

<sup>14</sup> <https://es.slideshare.net/zanorte/comparativa-herramientas-business-intelligence>

		herramientas suficientes ETL	
comunidad	Amplia comunidad y muy activa.	Amplia comunidad y muy activa.	Amplia comunidad y muy activa.
Interoperabilidad	Gran variedad de conectores propios y de terceros.	Gran variedad de conectores propios y de terceros.	Gran variedad de conectores propios y de terceros.
Servicio Cloud	Dispone de servicio <i>cloud</i> , pero en modalidad de pago.	Dispone de servicio <i>cloud</i> .	Dispone de servicio <i>cloud</i> gratuito o mediante suscripción.
Tipo de licencia	Licenciamiento GLP	Propietario	Propietario
Observaciones		Admite programación mediante DAX <sup>15</sup> .	Versión desktop no incorpora el conjunto de funciones que ofrece la on-line.

**Decisiones y conclusiones.** Los experimentos y pruebas que se han llevado a cabo con intención de identificar las soluciones elegidas se han centrado en la usabilidad y en la velocidad de aprendizaje y por encima de ello que el acceso a las aplicaciones a coste 0. En resumen, las soluciones y las justificaciones son:



En primer lugar, para los trabajos relacionados con la extracción, transformación y carga de datos se ha seleccionado Pentaho, en concreto la solución Spoon. Algunos de los motivos de peso que ha decantado la balanza a su favor han sido posibilidades de intercambio de información y el equilibrio que ofrece entre su potencia transformadora de datos vs curva de aprendizaje. En el primer contacto se ha podido hacer una conexión de forma relativamente rápida. También se ha encontrado abundante información de uso y con ejemplos de varios escenarios y una comunidad muy activa.

<sup>15</sup> DAX: [ver significado en el glosario.](#)



En segundo lugar, la solución que facilitará la construcción del Data Warehouse y facilitar el acceso a los datos contenidos será MySQL. Un motivo importante es que se ha trabajado en alguna ocasión en dicho software y es por ello que se cree la integración de su empleo será muy rápido. Por otro lado, comentar que es una solución muy extendida y el acceso a los datos por parte de soluciones de terceros está garantizada.



En tercer y último lugar, la solución Microsoft Power Bi es la candidata escogida sobre la que se desarrollarán las ventanas de visualización de datos. El motivo decisivo ha sido las posibilidades que incorpora para la gestión de los datos a modo de herramienta *front end*. Esta integración simplifica la necesidad de manipulación íntima entre la carga de datos y la configuración de la visualización de estos, bien sea mediante la implementación de cubos como el acceso directo a tablas o adecuación de la información. Es muy probable que es esta funcionalidad simplifique la carga de datos y su refinamiento, así como la aplicación de cálculos mediante el lenguaje DAX. Otro factor inductor de la elección ha sido la facilidad de uso del entorno de desarrollo de las ventanas de visualización y la configuración de los controles.

En conclusión, colocando las aplicaciones que han sido seleccionadas en la gráfica de arquitectura de R. Kimball resulta de la siguiente manera:

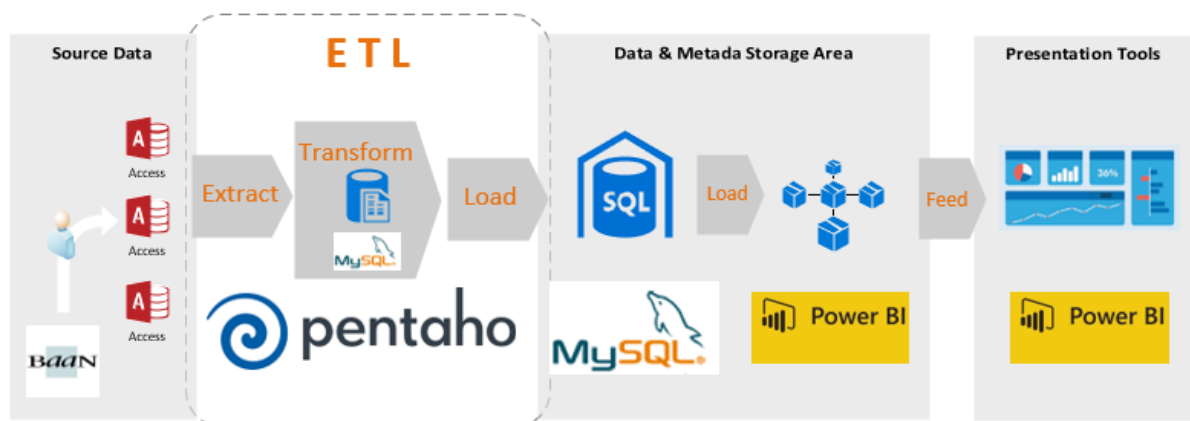


Ilustración 27: localización de las aplicaciones en el modelo de R. Kimball

Hay que destacar de esta gráfica dos puntos: se ha creado un espacio de almacenamiento para la transformación, o *staging*, después de considerar las capacidades de la solución Spoon a la hora de transformar los datos y, el segundo punto, se ha le ha asignado la gestión de los cubos a la aplicación M. Power BI por entender que esto facilitará el acceso a los controles de visualización.

## 3. Desarrollo del producto

### 3.1 Construcción del espacio de desarrollo.

El escenario del prototipo se construirá sobre un equipo con los siguientes recursos:

- CPU: Intel® Core™ i5-4570 CPU @ 3.20GHz
- RAM: 24 GB
- DD: Kingstom 480 GB SSD

Las aplicaciones que facilitarán la recreación del escenario y su motivo son se describen a continuación:

- SO del equipo anfitrión: Windows 10 Pro x64.
- Gestión Máquinas virtuales: VMware Workstation 15.0.4 Player16.
- SO equipo virtualizado: Windows 10 x6417, versión de evaluación de 90 días (2 CPU; 8 Gb RAM; 50 G de espacio de disco).

Aplicaciones que facilitarán el desarrollo del producto:

- Pentaho v8.2 (pdi-ce-8.2.0.0-342.zip)
- MySQL Server v8.0 (mysql-installer-community-8.0.15.0.msi)
- MySQL Workbench v8.0 (mysql-workbench-community-8.0.15-winx64.msi)
- Microsoft Power BI Desktop 2.67.5404.981 64-bit (PBIDesktop\_x64.msi)
- Otros:
  - mysql-connector-odbc-8.0.15-winx64/win32.msi: para el acceso a las DDBB MDB desde el motor de base de datos MySQL.
  - mysql-connector-net-8.0.15: acceso a los datos de MySQL desde Power BI.
  - vc\_redist.x86.exe.
  - jdk-8u201-windows-x64.exe: máquina virtual JAVA.
  - python-3.7.2.exe: requerido por Kettle en algunos de las transformaciones y MySQL.

Respecto a la descripción de cómo se ha construido el escenario sólo se apuntarán los pasos que caracterizan el producto o configuraciones específicas, se da por entendido que el resto son trámites de aceptación o sin relevancia para el trabajo.

---

<sup>16</sup> [https://my.vmware.com/en/web/vmware/free#desktop\\_end\\_user\\_computing/vmware\\_workstation\\_player/15\\_0](https://my.vmware.com/en/web/vmware/free#desktop_end_user_computing/vmware_workstation_player/15_0)

<sup>17</sup> <https://developer.microsoft.com/en-us/microsoft-edge/tools/vms/>

## Plataforma de ensayo.

La plataforma virtualizada se ha descargado del propio fabricante Microsoft, la cual se ofrece para como herramienta de desarrollo y permite un margen de tiempo de uso de 90 días. Descargar la plataforma virtualizada no exige registro o proporcionar información de ningún tipo sobre cuál será su empleo final.

Una vez descomprimido el archivo que contiene la máquina virtual es posible su lanzamiento desde la consola gestora. Un paso requerido es el registro del SO siguiendo los pasos especificados en el escritorio [ver figura]. Por último, sí que es necesario configurar los parámetros de “Región”, si no realizan estos cambios se pueden producir problemas durante el intercambio de datos entre soluciones o se deben hacer más transformaciones de estos.

## Herramientas ETL (Back end)

La instalación de Kettle exige, únicamente, la descompresión del archivo comprimido y volcar los el directorio en su ubicación desde donde se explotará, esto es posible gracias a su implementación en lenguaje JAVA.

Por otro lado, si se debe tener en cuenta la necesidad de instalar los conectores a otras fuentes – MySQL ofrece los conectores son desarrollados por la comunidad y nativos de JAVA-. Cabe decir que esta tarea tampoco conlleva conocimientos específicos de la tecnología ni aplicar configuraciones propias del proyecto.

Para el uso de los conectores desde Kettle se debe tener en cuenta la parametrización exigida por el motor de DDBB. En el caso que se está desarrollando los parámetros exigidos se deriban de la combinación de la versión del SO Windows 10, su idioma y la configuración del motor de DDBB.

En el caso de no parametrizar la conexión se produce el siguiente error durante la prueba de conexión

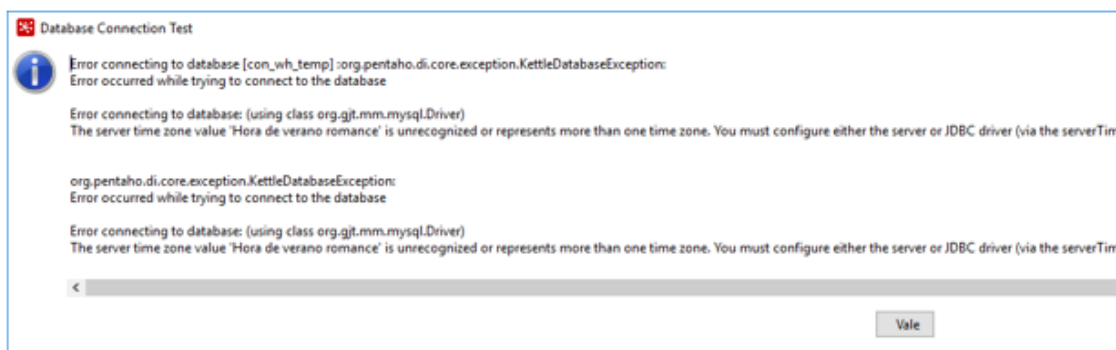


Ilustración 28: captura del error producido durante la conexión desde Kettle a MySQL

Resolver dicho problema requiere incorporar a la conexión los siguientes detalles.



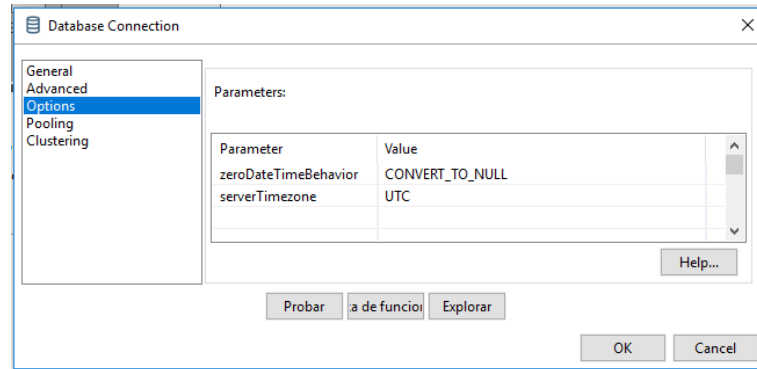


Ilustración 29: parametrización de la conexión en Kettle para acceder a MySQL

## Base de datos

El motor de BBDD empleado en el proyecto sí que requiere de alguna configuración que deberá ser utilizada desde otras soluciones.

Antes de iniciar la instalación se deben atender los requisitos que después MySQL empleará en diferentes tareas como la conectividad con otras aplicaciones o la confección de scripts que automaticen tareas. Un complemento requerido en el marco de ejecución según el lenguaje de programación en el que se ha desarrollado la solución. En el caso de MySQL el marco de ejecución exige las de Microsoft Visual C++ 2015 Resdistributables.

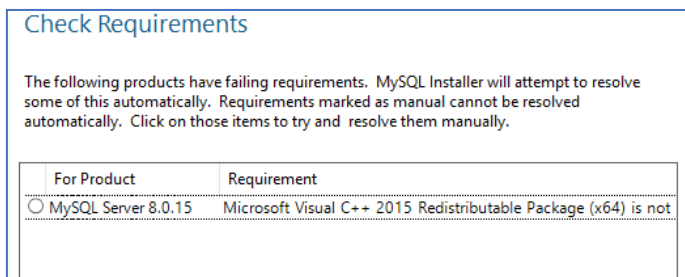


Ilustración 30: requisito fundamental para la instalación de MySQL.

La parametrización que se debe recordar para acceder al contenido de la DDBB se recoge en la captura siguiente. Se puede apreciar en ella que canales de acceso permite: sockets, Pipe y memoria compartida. Pese mantener las tres configuraciones ofrecidas por defecto, únicamente emplearemos la configuración TCP/IP en los conectores.

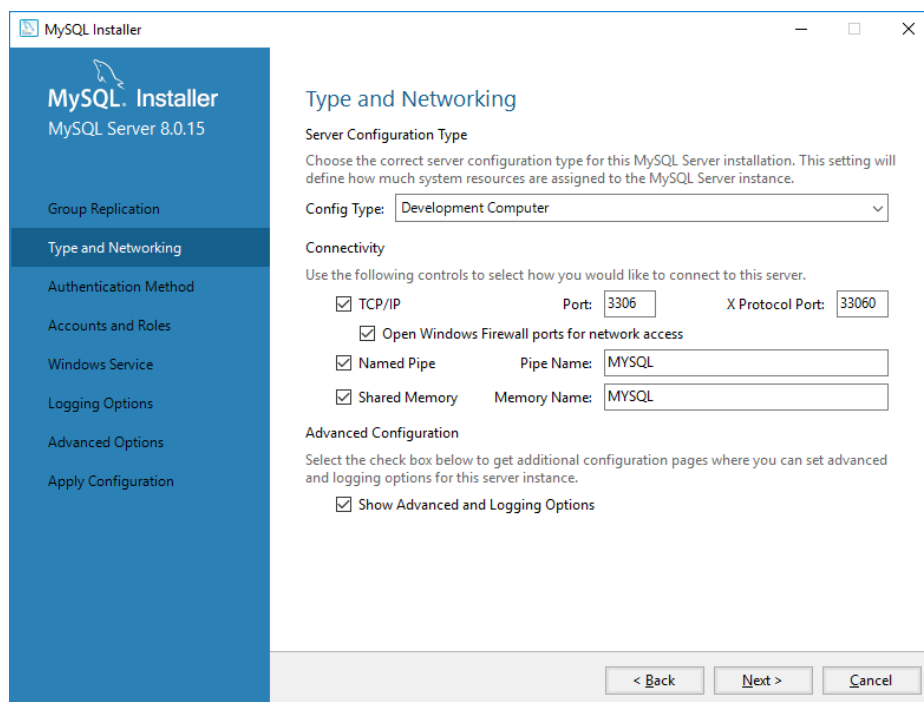


Ilustración 31: configuración de los canales de acceso a la BBDD.

Como último detalle a tener en cuenta de la instalación, se ha decidido configurar el acceso de los usuarios mediante credenciales compatibles con versiones de MySQL 5.x por temas de simplificación, obviando el uso de protocolos de seguridad SSL.

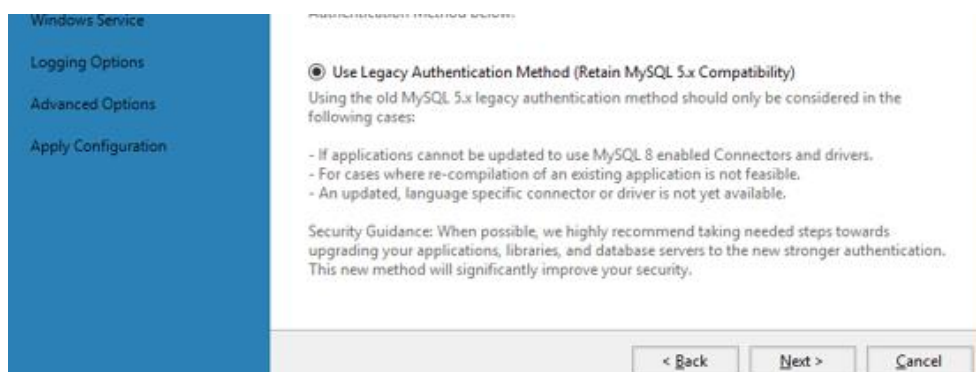


Ilustración 32: selección del mecanismo de autenticación (sin SSL).

## Dashboard

Tal y como se indicó en apartados anteriores, la herramienta escogida para la confección del *dashboard* es Microsoft Power BI y como es costumbre, la instalación de las aplicaciones de escritorio del fabricante Microsoft es un proceso muy sencillo y, por consiguiente, no requiere de apunte alguno.

Respecto a la configuración de la sesión local de trabajo tampoco requiere aclaración alguna, únicamente recordar que se debe tener instalado el conector al servidor MySQL.

Comentar un detalle de configuración, a la hora de realizar la conexión para la carga de datos la primera vez puede ser un poco confuso si no se presta atención sobre en qué ventana se está introduciendo el usuario y la contraseña puesto que ofrece la posibilidad de uso de las credenciales de sistema o las propias de la solución MySQL. Para el caso en desarrollo se debe emplear la opción “Base de datos”.

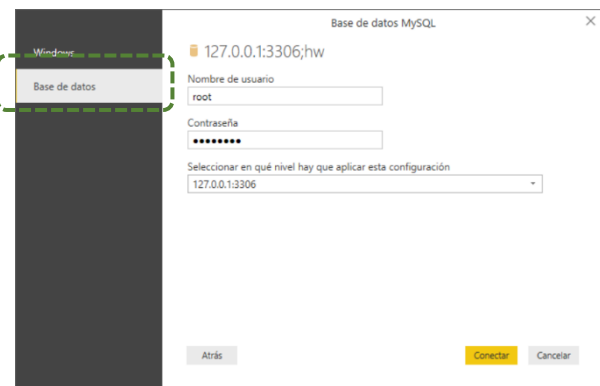


Ilustración 33: Ingreso de credenciales de MySQL desde Power BI.

### 3.2 Construcción del Data Warehouse.

Continuando con la confección del Data Warehouse y en base al diseño lógico ideado, se ha construido la estructura que albergará los datos una vez extraídos, transformados y cargados -zona enmarcada en naranja- bajo el esquema “wh”. A propósito de la transformación, tal y como se predijo, ha sido necesaria la incorporación de un repositorio intermedio para la transformación de los datos (Data Staging Area) bajo el nombre “wh-temp” -se ha remarcado en verde el laboratorio de datos-.

La captura hecha de la modelización, y mostrada más abajo, muestra el tipo de relación entre tablas, así como los tipos de datos empleados y *Primary Keys* de cada tabla.

Tal y como también se adelantó, se ha conseguido una estructura en estrella con la intención de mejorar el rendimiento del *dashboard*, como aporte de valor.

Acompañando a la memoria se entregará el código en SQL que generará los esquemas del laboratorio y del Data Warehouse.

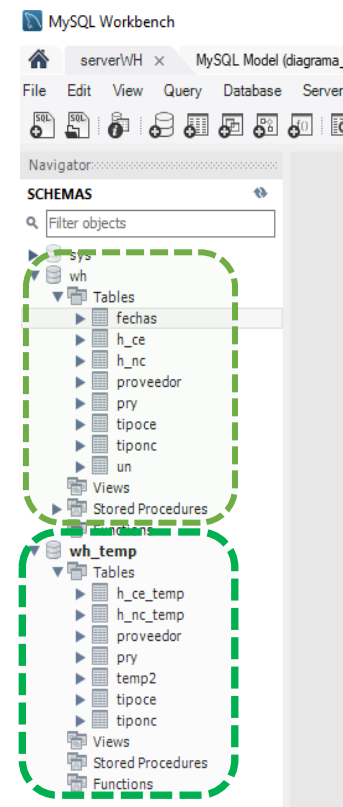


Ilustración 34: vista de los Schemas para el Data Warehouse y el laboratorio

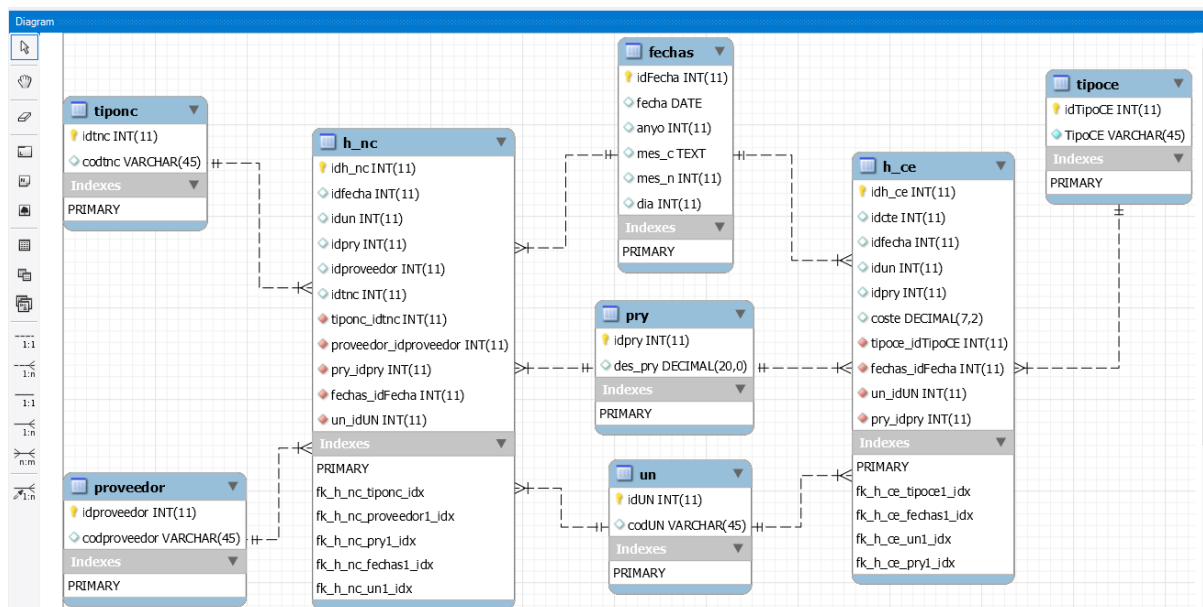


Ilustración 35: campos, relaciones y "Primary Keys" según la modelización del DWH.

### 3.3 ETL (Extracción, Transformación y Carga)

La creación de los trabajos y secuencias de transformación se han ideado de manera que se maximice a la automatización de visualización del *dashboard*. Dado el escenario de trabajo el cual se ha diseñado el producto, el único requisito que el usuario debe tener en cuenta para que todo funcione se limita a localizar en una ubicación específica el archivo o archivos CSV contenedor de los datos relacionados con los Costes Evitables. Respecto a las extracciones de datos almacenados en fuentes "MDB" apuntar que no es necesario que el usuario tenga en cuenta nada puesto que dicha fuente de datos no sufre cambio alguno.

Volviendo a las tareas ETL, cabe decir que se han agrupado en dos trabajos, dos archivos independientes entre sí, cada uno centrado en una fuente diferente: CSV y MDB. Un motivo por el que se ha tomado esta decisión es porque cada fuente contiene datos específicos de análisis y los hechos no están relacionados. Los hechos No Conformidades se almacena en los MDB y los hechos de los Costes Evitables en los archivos CSV. Otro motivo de la división de los trabajos se debe a un tema práctico puesto que, de este modo, resulta más fácil depurar los errores y entender la operativa de cada salto.

Para finalizar, comentar que se ha empleado un *schema* a modo de laboratorio, tal y como se previó durante la definición del producto. Según se indicó, dicha necesidad viene dada por la falta de un conocimiento profundo entorno la capacidad de la solución Kettle y sus mecanismos de cache o trabajo en temporal en memoria. Es verdad que hay documentación [4][5] al respecto y ayudas en la comunidad [6][7], pero la necesidad de entender cómo evoluciona la transformación de los datos durante este primer contacto con la aplicación, ha obligado esta forma de trabajar.

## Trabajos ETL de archivos CSV.

### Trabajo Main.

El primero de los trabajos tratados es el encargado de gestionar las transacciones que incorporan los Costes Evitables.

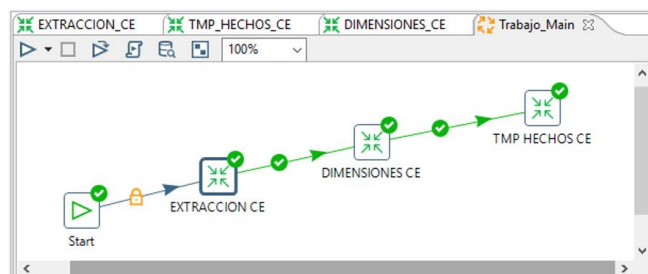


Ilustración 36: JOB principal ETL para ficheros CSV

Tal y como se explicará más adelante, se empleará un script capaz de revisar todos los archivos CSV, de forma recursiva, localizados en un directorio. Esta operativa es necesaria, principalmente, por la necesidad de inicializar las bases de datos gestionando un conjunto de 108 archivos. Dicho sea de paso, estos archivos contienen información registrada desde febrero del 2017 hasta febrero del 2019, sumando un total que ronda las 1.200.000 transacciones.

El conjunto de transformaciones se divide en tres archivos: "extraccion\_ce.ktr", "dimensiones\_ce.ktr" y "tmp\_hecho\_ce.ktr".

### EXTRACCION CE

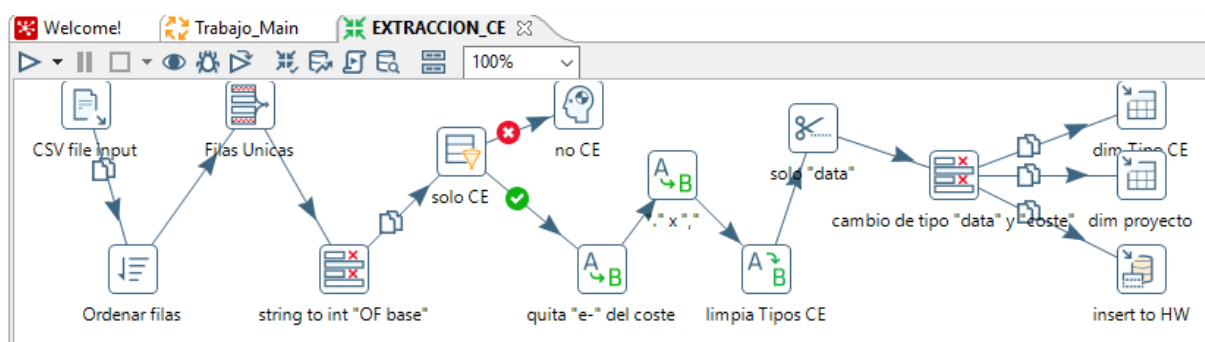


Ilustración 37: saltos para la extracción de los archivos CSV

Las acciones principales que recoge este conjunto de saltos se centran en extraer las transacciones validas de los archivos CSV, aplicar transformación sobre los costes y fecha y conseguir su adaptación al formato del esquema del Data Warehouse.

También se han configurado dos pasos para la extracción de dimensiones "Proyectos" y "Tipos de CE". Estos dos últimos pasos también podrían dirigirse sobre la estructura final "wh", pero se ha decidido mantener este flujo de datos

por necesidades de revisión del contenido y porque el coste del proceso es despreciable en cuanto al tiempo de carga.

### TMP HECHOS CE

El conjunto de saltos incluidos en este conjunto ha sido diseñado para identificar dimensiones y almacenar temporalmente los datos en el laboratorio “wh\_temp”, además se hace un filtrado y transformación previa que termina de refinar la columna “fechas” de los hechos.

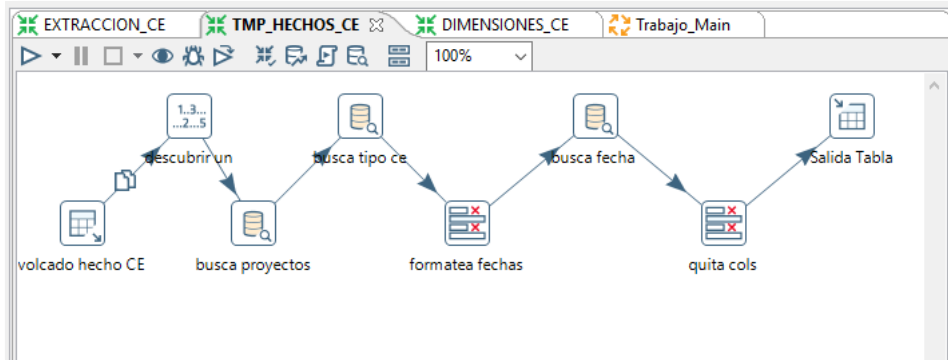


Ilustración 38: saltos de transformación de ficheros CSV

### DIMENSIONES CE

En este grupo se ejecutan dos saltos en paralelo volcando los proyectos afectados por los Costes Evitables y los tipos de CE identificados. Esta última dimensión actualiza la tabla en busca de nuevos tipos de CE y lo mismo sucederá con los proyectos identificados.

Los pasos realizados son únicamente de extracción de datos del laboratorio, estructura “wh\_temp”, filtrando las transacciones no afectadas por Costes Evitables.

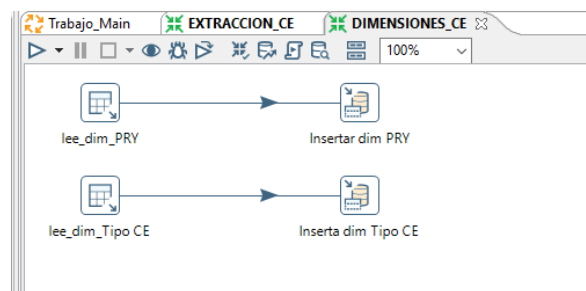


Ilustración 39: extracción de dimensiones de los archivos CSV

### **ETL para las No Conformidades.**

#### Trabajo Main MDB.

El segundo trabajo, especializado por el acceso a las bases de datos en Microsoft Access, gestionará los trabajos dedicados a las No Conformidades.

Su estructura secuencía un conjunto de llamadas a archivos de transformaciones:

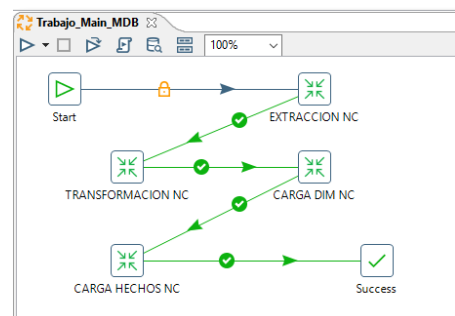
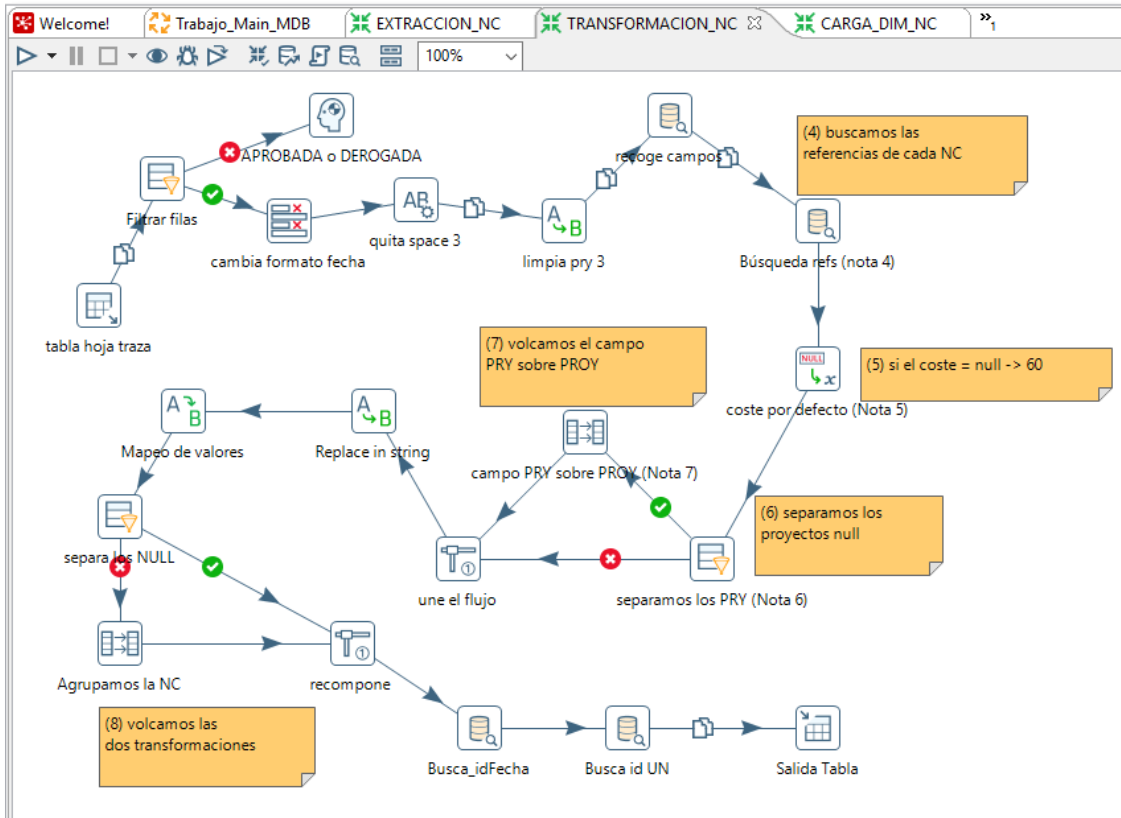


Ilustración 40: JOB ETL para las fuentes MDB



“extraccion\_NC.ktr”, “transformacion\_NC.ktr”, “carga\_dim\_nc.ktr” y “carga\_hechos\_nc.ktr”.

**EXTRACCION NC**

Las transformaciones configuradas tienen como única función extraer de la tabla con las transacciones de No Conformidad, de la tabla de los proveedores involucrados y de la tabla con la trazabilidad de cada transacción de No Conformidad. De cada tabla se recogerán las columnas con información útil.

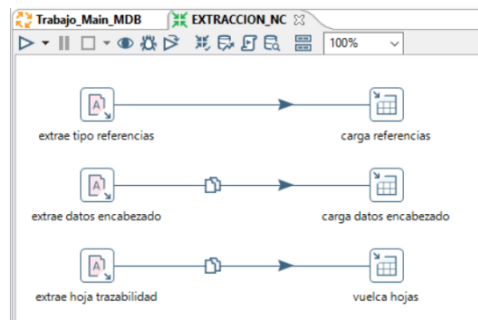


Ilustración 42: extracción de las dimensiones NC

**TRANSFORMACION NC**

En general, este conjunto de saltos da como resultado la tabla de hechos para las No Conformidades. Como tabla incluyen todas las columnas que definen cada hecho y, además, incluyen los campos que relacionan los hechos con las dimensiones. El conjunto de datos resultantes aún se mantiene en el laboratorio porque se debe extraer dimensiones y aplicar la selección de columnas final.

Como se puede apreciar han hecho falta aplicar un conjunto importante y variado de saltos hasta conseguir coherencia y calidad en los datos extraídos.

Ilustración 41: saltos de transformación de los hechos NC de la fuente MDB

## CARGA DIM NC

Acto seguido a la creación de la tabla de hechos, que incluye los datos organizados y estandarizados, se procede a la extracción de las dimensiones “tipo de NC”, “proyectos” y “proveedores” relacionados con las No Conformidades. La fuente sobre la que se trabaja es la propia tabla temporal de hechos y cada valor tratado será añadido a la dimensión solo si no existe previamente en la tabla, de esta manera nos aseguramos de que los componentes de las dimensiones representarán con fiabilidad al hecho que corresponde.

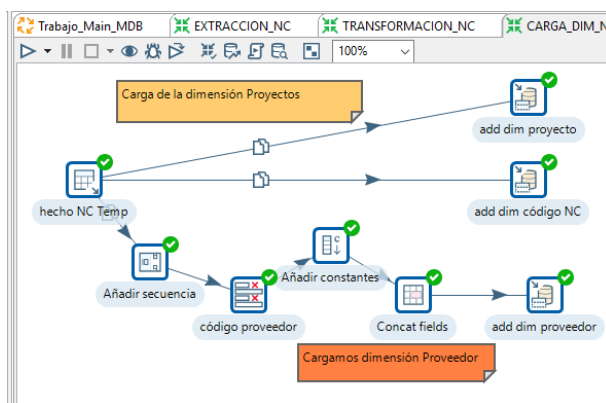


Ilustración 43: saltos para la carga de las dimensiones en Data Warehouse

## CARGA HECHOS NC

El último grupo de saltos se encarga de componer la tabla de hechos No conformidad definitiva en el DataWarehouse. Las acciones consisten en recoger los índices de las dimensiones que sustituirán los valores propios de las dimensiones. Con esta acción se conforma la estructura en estrella de la bases de datos para los hechos NC.

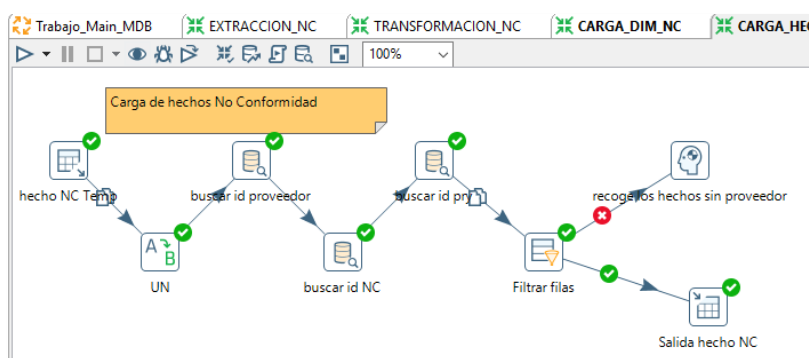


Ilustración 44: saltos para la carga de los hechos en el Data Warehouse

### 3.3.1 Pruebas con datos

Ahora se llevará a cabo una batería de experimentos que pondrán a prueba la ejecución de todos los trabajos y transformaciones que agrupan y organizan.

El resultado obtenido es positivo, las pruebas realizadas, tal y como se muestra en la captura del log (ver Anexo II), confirman que se llevan a cabo todos los saltos de forma correcta. Es más, de la ejecución del trabajo cabe decir que ha

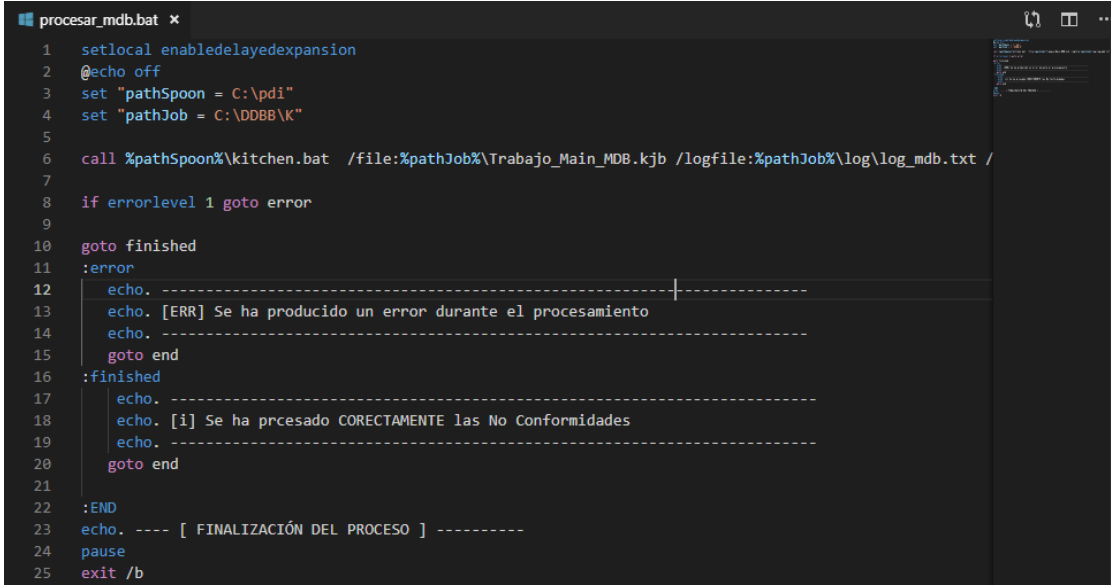


sorprendido tanto la estabilidad de la aplicación como la velocidad en la que se lleva a cabo todas la tareas -una de las pruebas consistía en gestionar 109.766 transacciones de las que se extrajo 1.797 relacionadas con los CE y dicha prueba se ejecutó en 8 segundos-.

Pese a la correcta ejecución de la programación de saltos, aún falta por confirmar la calidad de los datos obtenidos por la herramienta ETL y que solo será validado una vez se puedan contrastar los resultados visualizados por el *dashboard*. Dicho trabajo de valoración de calidad se realizará durante el experimento una vez se integre dentro del entorno de la empresa y se lleve la a práctica por algún integrante del departamento de calidad.

### Scripts para la automatización de extracción, transformación y carga de datos.

Además de las pruebas de ejecución las herramientas ETL, se ha implementado dos scripts utilizables desde el sistema del usuario con la intención de automatizar la importación de los datos. De esta manera el usuario únicamente entrar en la aplicación Kettle en el caso de reconfigurar la secuencia de trabajos ETL puesto que la ejecución puede ser programada según las necesidades de actualización.



```

procesar_mdb.bat x
1  setlocal enabledelayedexpansion
2  @echo off
3  set "pathSpoon = C:\pdi"
4  set "pathJob = C:\DDBB\K"
5
6  call %pathSpoon%\kitchen.bat /file:%pathJob%\Trabajo_Main_MDB.kjb /logfile:%pathJob%\log\log_mdb.txt /
7
8  if errorlevel 1 goto error
9
10 goto finished
11 :error
12 echo. -----
13 echo. [ERR] Se ha producido un error durante el procesamiento
14 echo. -----
15 goto end
16 :finished
17 echo. -----
18 echo. [i] Se ha procesado CORECTAMENTE las No Conformidades
19 echo. -----
20 goto end
21
22 :END
23 echo. ---- [ FINALIZACIÓN DEL PROCESO ] -----
24 pause
25 exit /b

```

Ilustración 45: script para la automatización de las acciones ETL sobre las fuentes MDB

El uso de los scripts tiene una exigencia clave y es que el origen de las fuentes de datos no puede variar en ninguno de los dos casos.

```

procesar_mdb.bat  procesar_CSVs.bat x
1  setlocal enabledelayedexpansion
2  @echo off
3  set pathCSV=C:\DDBB\CSV
4  set pathSpoon=C:\pdI
5  set pathJob=C:\DDBB\K
6
7  for %%f in (*.csv) do (
8      call :procesar "%%f"
9  )
10 echo. ---- [ FINALIZACIÓN DEL PROCESO ] -----
11 pause
12 goto END
13
14 :procesar %1
15 echo. -----
16 echo. [i] Se procesa el archivo csv: %1
17 echo. -----
18 move /y "%pathCSV%\todos\%~1" %pathCSV%\tmp\in.csv
19
20 echo. -----
21 echo. [i] Se ejecuta el trabajo sobre el archivo csv: %1
22 echo. -----
23 call %pathSpoon%\kitchen.bat /file:%pathJob%\Trabajo_Main.kjb /logfile:%pathJob%\log\log.txt /level:basic
24 if errorlevel 1 goto error
25 goto ok
26
27 :error
28 echo. -----
29 echo. [E] Se ha producido un error durante el procesamiento del archivo: %1
30 echo. [E] Se para el proceso
31 echo. -----
32 pause
33 goto END
34
35 :ok
36 echo. -----
37 echo. [i] Se ha procesado CORECTAMENTE el archivo: %1
38 echo. -----
39 echo. [i] Se almacena el archivo: %1
40 echo. -----
41 move /y %pathCSV%\tmp\in.csv %pathCSV%\_salida%\random%\random%\procesado.csv
42 GOTO END
43
44 :END
45 exit /b

```

Ilustración 46: script para la automatización de las tareas ETL sobre los archivos CSV

En conclusión, respecto las pruebas con datos en los procesos ETL, el resultado es satisfactorio, la velocidad de ejecución de la herramienta ETL aceptable: para el caso de los archivos MDB la gestión es muy rápida y para el caso de los archivos CSV es comprensible la inversión de 17 minutos puesto que se han gestionado 108 archivos, pero se debe recordar que este volumen de datos solo será tratado una sola vez, el resto de meses solo se cargará un único archivo CSV.

### 3.4 Confección de OLAP

Una vez concluida la extracción, transformación y carga de datos en el Data Warehouse es el momento de confeccionar los diferentes cubos base para la representación de los análisis.

Se debe decir antes, de continuar con el análisis de los datos, que la construcción del Data Warehouse se ha estructurado conforme la idea de cómo debían ser los cubos. Esta decisión ha simplificado enormemente el tratamiento específico

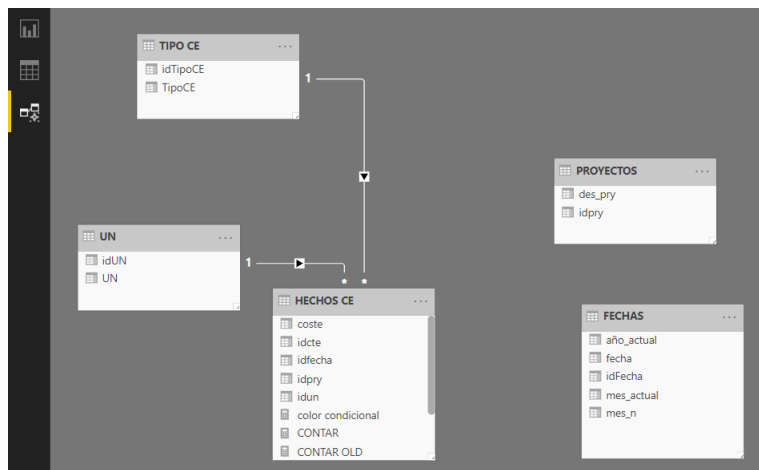
de cada estudio de los KPIs. Ahora resta la definición final que configura cada uno los casos de estudio y que serán visualizados en desde Power BI.

Dicho sea de paso, tal y como se ha mencionado en el apartado de elección de aplicaciones, una de los razones por las que se ha escogido Microsoft Power BI es la posibilidad que ofrece en cuanto el refinamiento de datos y configuración de las estructuras de datos adecuándolas para su visualización.

### Carga de datos.

El primer paso del proceso de carga de datos desde la aplicación Power BI es la configuración de la conexión al motor de DDBB. Este proceso de configuración es sencillo y claro: selección del conector específico del motor MySQL y aplicación de las credenciales configuradas en dicho motor.

Acto seguido, establecida la conexión, se debe seleccionar las tablas de interés y solo resta aprobar la conexión.



Una opción interesante que ofrece Power BI es la gestión de la estructura entre tablas importadas (ver ilustración XX, con el detalle de la estructura concerniente a los Costes Evitables).

*Ilustración 47: resultado de la carga de datos desde Power BI.*

Tal y como se puede apreciar en la captura el resultado de la carga, esta difiere respecto la estructura establecida en el Data Warehouse, concretamente, no se han importado todas las relaciones.

Pese a que puede parecer un problema no se ha profundizado en busca de los motivos de dicho defecto, en su lugar se ha reconfigurado la estructura desde la propia vista del modelo. Pese a resolver dicho contratiempo se ha buscado algún motivo y se cree (puesto que no se ha confirmado) que es consecuencia de alguna incongruencia detectada durante el análisis de los metadatos que realiza la solución durante el proceso de importación.

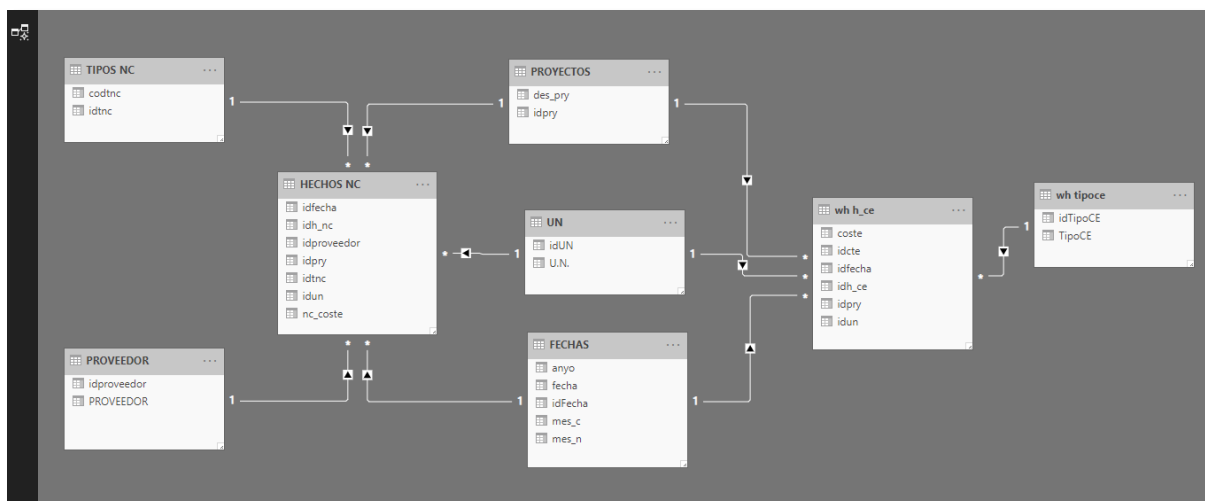


Ilustración 48: rectificación de las relaciones defectuosas después de la carga de datos

En lo que respecta al tipo de sincronización de los datos se ha dejado la opción propuesta de “Importar” que se ofrece por defecto. Esta configuración no ha sido modificada por considerar que el volumen de datos es fácilmente tratable por la aplicación y no es necesario un estado de sincronización permanente con el Data Warehouse. Por otro lado, puesto que dicha configuración implica la recogida del conjunto de datos, esto nos obliga a ejecutar el proceso de sincronización desde Power BI cada vez que se extraigan, transformen y carguen datos sobre el Data Warehouse.

### Refinado de la tabla de hechos.

Es muy probable que los datos cargados sobre el Data WH mediante la herramienta ETL y su posterior importación desde Power Bi puedan generar alguna inconsistencia en por los metadatos de tipos de datos *string*, *date* o respecto al contenido vacío de los campos.

Dichos defectos en los datos son consecuencia de la falta de estandarización en la configuración del idioma o interpretación de la información transferida. Por ejemplo: en alguno de los pasos de importación de datos, los acentos no se han interpretado correctamente, y fueron sustituidos por símbolos de otras tablas de caracteres.

En el caso de las tablas de hechos, puesto que contiene las referencias a los PK de las dimensiones, no es tan probable que se den casos de falta de estandarización (o por lo menos según lo ideado en desde el proyecto en desarrollo), pero si se pueden dar casos de campos vacíos.

Es cierto que el control de estos defectos se puede incorporar en la capa ELT, pero se ha decidido las acciones de depuración durante la carga desde Power BI por varios motivos: por trabajar en modo “Importado”, por asegurar que aplicará según las necesidades de la solución y por resultar extremadamente cómo aplicar acciones de este tipo en la configuración de sincronización.

Es más, aun teniendo la certeza que la calidad de los es aceptable, siempre se recomienda realizar tareas de refinamiento desde el editor de consultas que incorpora Power BI por ser este quien debe trabajar con ellos. Estas acciones ayudarán a prevenir valores falsos y desviaciones en los análisis.

Como ejemplo de la facilidad que ofrece Power BI en la gestión de los modelos, han incorporado una función de análisis automático sobre los datos brindando la calidad de estos y acciones correctoras.

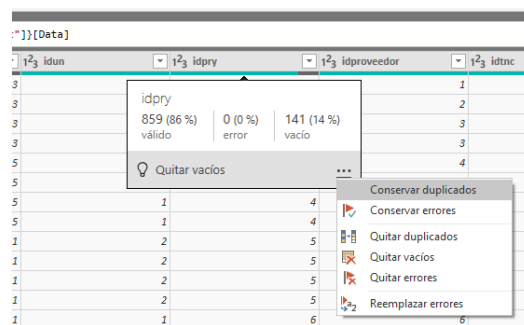


Ilustración 49: funcionalidad de análisis automático y corrección rápida de errores

### Refinamiento de dimensiones.

En el caso de las dimensiones la acciones que se debe efectuar apuntan más hacia el refinamiento de los campos con de tipo *string*.

En la captura se pueden apreciar algunas de las acciones que se han programado en Power BI para ser ejecutadas durante el proceso de actualización de datos, principalmente las acciones a realizar se centran en la rectificación de caracteres particulares y eliminación de valores similares con la correspondiente reconfiguración de su *Primary Key*.

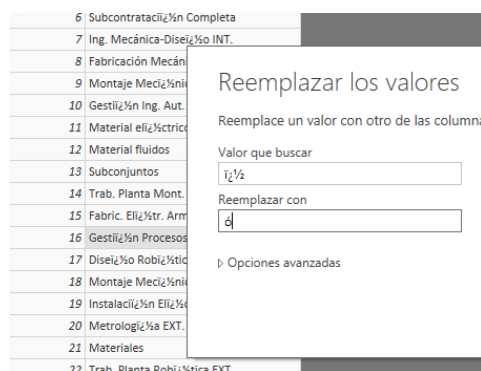


Ilustración 50: acciones rápidas sobre los campos desde Power BI

Otra acción que se debe programar está relacionada con las características que se deben otorgar a la solución en cuanto a una “Inteligencia de Tiempo” eficaz. La parametrización de estos campos debe darnos como resultado las siguientes propiedades:

- Granularidad de la fecha que comprende desde la mínima porción que es el mes hasta la máxima porción que es un año,
- los saltos se valoran por meses, no se emplearán otro tipo de división del año como pueden ser los trimestres, cuatrimestres o semanas.
- las comparaciones entre periodos de tiempo a sobre un año natural.
- Dicha parametrización puede hacerse empleando campos de “año” y “mes” o trabajando con un campo que declarado con el tipo fecha y formateado “dd-mm-aaaa” desde la modelización de las tablas. En el caso presentado se incorporan las dos fórmulas con intención de cubrir todas las posibilidades, pero es muy probable, según las pruebas realizadas durante la elección de software, que se emplee el campo con formato fecha debido a que la parametrización de las fechas en los informes es muy potente y ágil.

### **Configuración de métricas.**

En general el filtrado y segmentación<sup>18</sup> de datos, que conforman y delimitan los cubos, no se ha incluido en las métricas, esta parametrización se implementará mediante la funcionalidad ofrecida desde la ventana de informe, por consiguiente, la configuración de estas propiedades pertenece a la fase de desarrollo de las visualizaciones de las gráficas de los datos.

Las métricas consideradas durante el diseño de los KPIs son varias y algunas son dependientes de otras. Es cierto que la programación mediante DAX[8] permite alcanzar complejas configuraciones concentrando todas las métricas un una sola, pero el grado de conocimiento de este lenguaje no ha permitido explotar todas sus posibilidades. Durante la descripción del informe se apuntarán las más relevantes.

### **3.5 Configuración de elementos gráficos.**

Componente importante de la filosofía de Power BI es la automatización en cuanto al *self service*. Bajo esta perspectiva esta solución incorpora diversas funcionalidades automatizadas que permiten al usuario despreocuparse de la necesidad programar la tipificación los datos visualizados. Los principales conceptos que permiten estas acciones sobre la navegación son los niveles de contexto <sup>19</sup>y los controles de segmentación.

### **Segmentación y Contextos**

Sin profundizar en los detalles de estos conceptos, comentar que son factores que van de la mano: la segmentación aplica, por ejemplo, la granularidad o variación del valor de la parametrización permitida sobre los elementos de visualización, por su parte, el nivel de contexto se refiere a que alcance que tendrá la segmentación o las medidas predefinidas. El alcance que define cada contexto (nivel de informe, nivel de página, elemento específico de visualización, nivel de tabla, fila, etc.) también puede ser alterado mediante comandos DAX [9].

El efecto que la segmentación y los niveles de contexto son una potente herramienta de parametrización a la hora de diseñar el *dashboard*, pero su efecto debe ser medido constantemente puesto que puede producir resultados no deseado y de difícil detección y, como consecuencia, llevando a la confusión a quién interprete los resultados.

---

<sup>18</sup> Segmentación: [ver significado en glosario](#).

<sup>19</sup> Contexto: [ver significado en el glosario](#).

El principal elemento empleado en la segmentación es el aportado por Power BI incluido en el espacio de visualizaciones, (núm. 1 de la figura 51)

Por otro lado, para el caso de las fechas también se integrará un aporte de la comunidad de desarrolladores, “Time Line 2.0.2”, que es descargable desde Marketplace de Power BI. Este complemento aporta una parametrización específica para la gestión de la “Inteligencia de Tiempo<sup>20</sup>”[8] (núm. 2 de la figura 51).

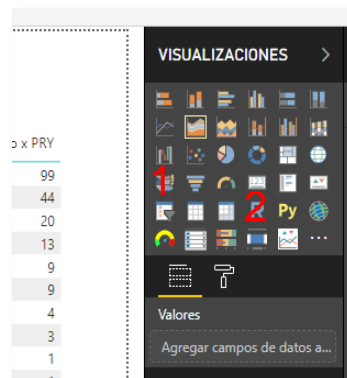


Ilustración 51: elementos de segmentación de Power BI

Un buen ejemplo de parametrización de los niveles de contexto es el aplicado mediante el filtro que permite limitar la segmentación del tiempo. Hay que recordar que estamos trabajando con dos fuentes de datos diferentes y se da el caso que el intervalo de fechas es diferente: los hechos No Conformidad (01-01-2013 hasta 28-02-2019) es mucho mayor que los hechos registrados de los Costes Evitables (01-01-2017 hasta 28-02-2019). Gracias a la posibilidad de aplicar filtro a nivel de página, podemos limitar el espacio de fechas para simplificar que datos con fechas nos interesa visualizar independientemente del resto de páginas del documento.

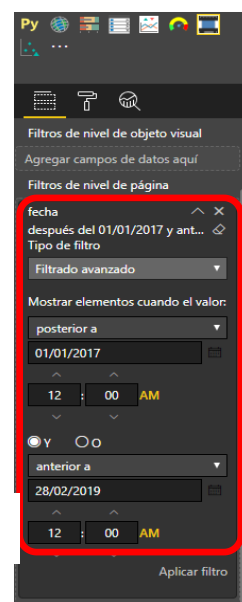


Ilustración 52; parametrización del elemento de visualización de las fechas

## Visualización las métricas y KPIs

En su conjunto, la visualización de los análisis se ha ideado para ser divididos en dos grupos. Tal y como se comentó, la intención del producto se centra en la visualización de dos perspectivas de la actividad de la Cadena de Valor y la Cadena de Suministro. Es por dicho motivo que se cree acertada y conveniente la decisión de separar los análisis en diferentes páginas que albergarán las métricas y KPIs dependiendo de la intención de estudio de los Costes Evitables y No Conformidad.

También se ha optado por la fórmula de combinar diferentes KPIs en una misma página. Esta combinación permite complementar una visión más general de la información con una perspectiva más concreta de los datos. Los casos de combinación son las páginas NC 1 y NC 2 en donde se ha incluido el índice de desempeño NC 3 consiguiendo ampliar las posibilidades de perspectiva del estudio.

<sup>20</sup> Inteligencia de Tiempo: [ver significado en glosario.](#)

Un aspecto que se ha tenido en cuenta es la distribución de los elementos de visualización agrupándolos por colores según su intención. Los elementos de segmentación, comunes en todas las páginas, se han enmarcado de color morado, el resto de elementos han sido coloreados según su pertenencia a cada uno de los análisis, de esta forma son fácilmente identificables aun compartan página.

**Navegabilidad**

Las funcionalidades que ofrece Power BI realmente son sorprendentes. Las posibilidades interactivas ofrecidas permiten alcanzar diferentes niveles de zoom de todas las perspectivas de los datos analizados.

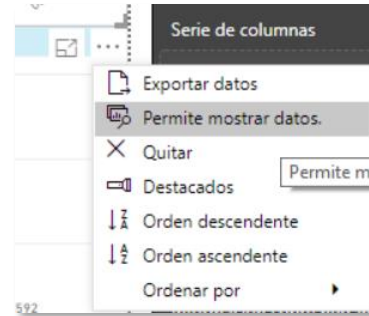


Ilustración 53: funcionalidad para la visualización de los datos de una gráfica

En general, en todo el informe, se han mantenido las relaciones entre los elementos de visualización y sus posibilidades de segmentación. Con esto se quiere decir que además de seleccionar el intervalo de tiempo, o la unidad de negocio, si se pulsa la tecla “control” y clicamos sobre cualquier valor de cualquier gráfica se profundiza en el nivel de filtrado. Como es lógico se ha limitado su acción al contexto de página para que la navegación resulte ágil.

Dentro de las posibilidades de la navegabilidad Power BI ofrece la posibilidad de visualizar los datos a la par que el elemento de visualización de estos datos. Esta es una funcionalidad y que tampoco se ha limitado por su potencia para la focalización un análisis.

**Indicador de Desempeño CE-1**

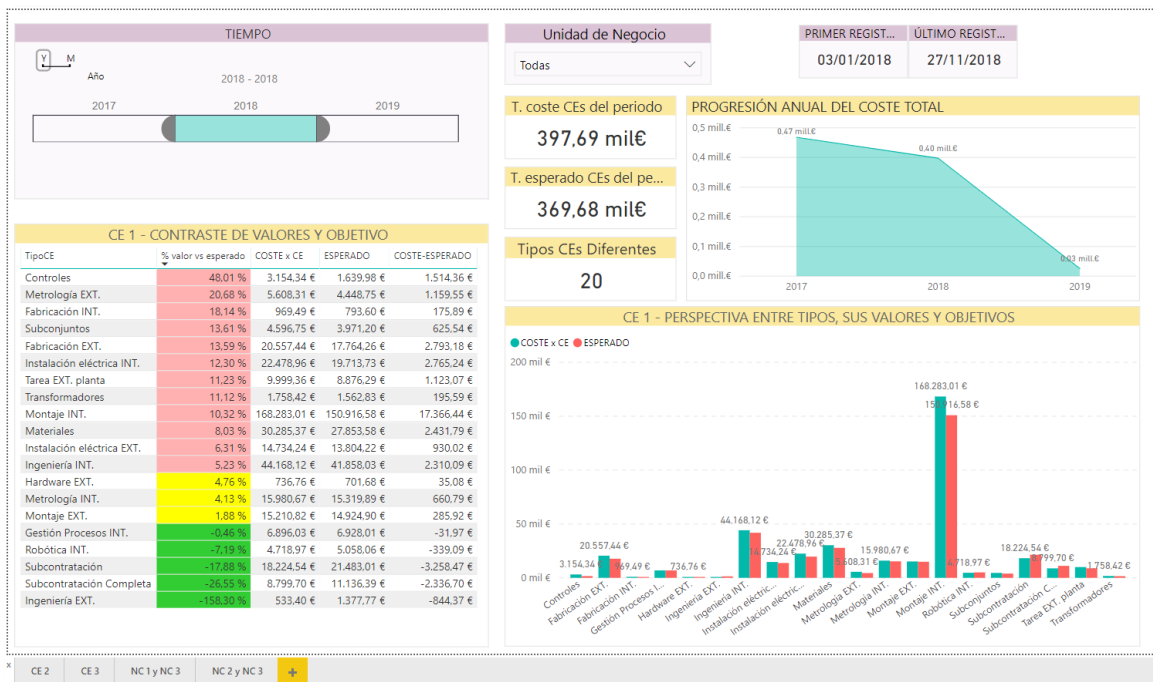


Ilustración 54: composición de visualizaciones para la página CE 1



El objeto de la visualización de este KPI está relacionado con las consecuencias que suponen los Costes Evitables según su tipo. En la “gráfica de barras agrupadas” (CE 1- PRESPECTICA ENTRE TIPOS, SUS VALORES Y OBJETIVOS) se han dispuesto dos valores en contraste: por un lado, los costes del mes estudiado en color magenta y, por otro lado, la estimación u objetivo esperado del KPI en color rojo.

Además, a la izquierda de la gráfica se ha incluido una matriz complementaria en donde se concreta los valores de la gráfica anterior. Esta representación tabulada de los datos desgrana los tipos de CE: diferencia entre el cote y el objetivo del KPI, la diferencia entre ambos valores y el desvió en porcentajes según el objetivo predefinido.

Respecto al resto de elementos de visualización comentar que ponen de relieve una serie de totales que no salen reflejados en los elementos principales.

### *Cálculos y parametrización*

Los elementos de segmentación y gestión de granularidad son el control de tiempo y el de la unidad de negocio.

La manipulación de los datos visualizados se llevará a cabo por las “medidas”:

```

COSTE X CE.dax ✕
1  COSTE X CE = CALCULATE(
2    SUM('HECHOS CE'[coste]);USERRELATIONSHIP('HECHOS CE'[idcte];'TIPO CE'[idTipoCE])
3  )

```

*Ilustración 55: métrica para el cálculo del coste por cada tipo CE*

```

ESPERADO.dax ✕
1  ESPERADO =
2  var cantidad = CALCULATE(COUNT('HECHOS CE'[idcte])/10;USERRELATIONSHIP('HECHOS CE'[idcte];'TIPO CE'[idTipoCE]))
3  var coste_t_x_CE =
4    CALCULATE(
5      SUM('HECHOS CE'[coste]);
6      ALL(FECHAS[fecha]);
7      USERRELATIONSHIP('HECHOS CE'[idcte];'TIPO CE'[idTipoCE])/100
8  var cantidad_t_x_CE=
9    CALCULATE(
10     COUNT('HECHOS CE'[idcte]);
11     ALL(FECHAS[fecha]);
12     USERRELATIONSHIP('HECHOS CE'[idcte];'TIPO CE'[idTipoCE]))
13  return
14  (cantidad * (coste_t_x_CE/cantidad_t_x_CE))/1,05

```

*Ilustración 56: cálculo del objetivo según el tipo CE*

En cuanto a la configuración gráfica de los elementos incluidos en la página se puede destacar la técnica de cambio dinámico de formato del campo “% COSTE vs OBJETIVO” de la matriz “CE 1 – CONTRASTE DE VALORES Y OBJETIVOS”.

A grandes rasgos, para colorear el fondo de las celdas se ha aplicado una condición mediante una medida en código DAX. El código que condiciona el color [1] es el siguiente.



Ilustración 57: formato condicional mediante código DAX

Dicha medida se ha incluido en el formato condicional del valor, de esta forma se evaluará el contenido de cada celda y configurando su aspecto de forma individual.

Dicho sea de paso, comentar que esta técnica es la empleada en el resto de páginas del informe. Se pueden emplear otros dos métodos de configuración dinámica a nivel de objeto, pero no son exportables, por el contrario, esta técnica permite ser reutilizada y es igual de eficaz.

### Métrica CE-2

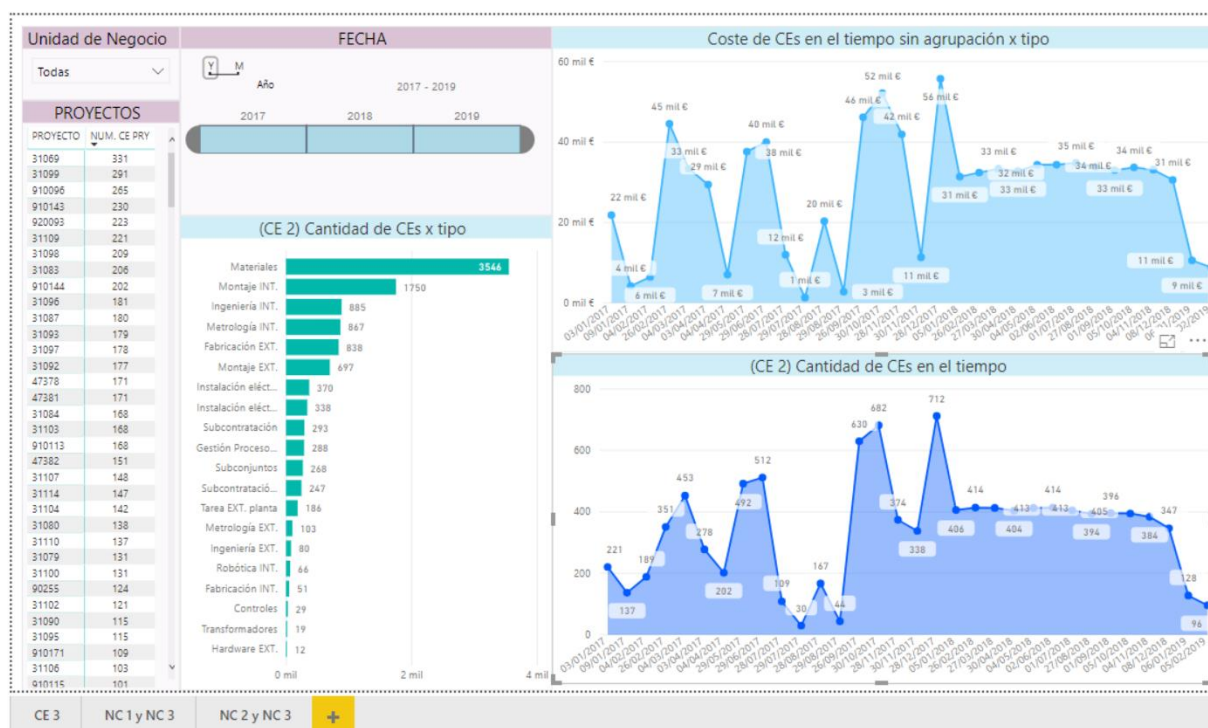


Ilustración 58: composición de visualizaciones para la página CE 2

Esta página recoge una serie de análisis sobre los Costes Evitables con la intención de conocer el número de veces que aparece un tipo de CE a lo largo de la vida de un proyecto. El KPI consistía en contrastar un valor marcado desde la dirección técnica de ingeniería sobre cada tipo de cada proyecto. Dichos valores han resultado ser inalcanzables. En un principio, durante la fase de prospección, se mal interpretó la intención de unos valores localizados en las fuentes de información MDB relacionadas con los Costes Evitables, lo cual ha provocado que no se puedan contrastar los datos respecto los objetivos del KPI. Este contratiempo ha obligado a tomar la decisión de generar la representación de las métricas sin contraste con objetivos, pese a ello, la información que

despunden de las gráficas también resulta útil como herramienta de seguimiento de progresión del CE 2.

La interpretación de la página apunta a cómo evolucionan los costes que han generado los CE a lo largo de un proyecto con la posibilidad de seleccionar un momento en el tiempo, reflejado en la gráfica de líneas inferior [(CE 2) Cantidad de CEs en el tiempo]. En la gráfica de líneas superior [Coste de CEs en el tiempo sin agrupación x tipo] se refleja que coste total supuso el conjunto de CE filtrado. Por otro lado, el conjunto de CE estudiado se desgrana en la gráfica de barras horizontales, concretando cuantos CE de cada tipo se han dado según la configuración del momento en la página.

Por último, frente a la imposibilidad de configurar los elementos de visualización con los valores objetivo que definen el comportamiento KPI CE-2, se ha tomado nota para tener en cuenta este punto de mejora en futuras versiones del *dashboard* en desarrollo.

### Métrica CE-3

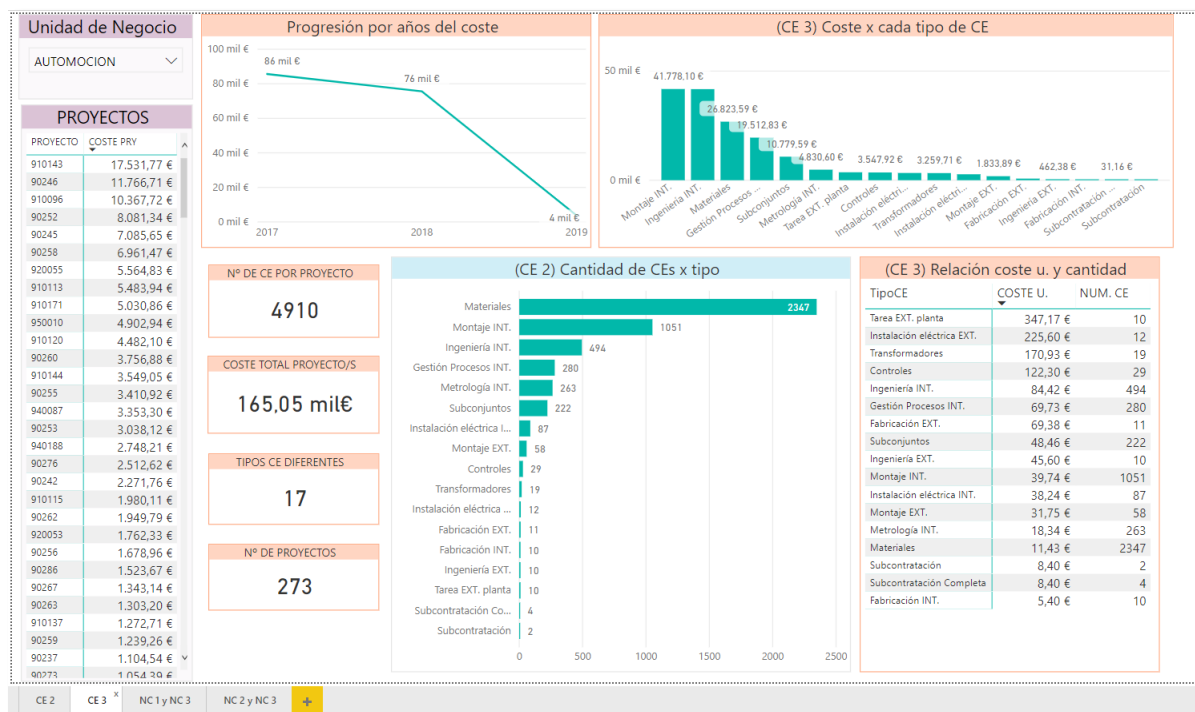


Ilustración 59: composición de visualizaciones para la página CE 3

Este caso de análisis sufre las mismas consecuencias que el caso anterior por el mismo motivo descrito en el punto anterior. Su configuración como KPI no es posible por las complejidades derivadas de alcanzar los datos de cada uno de los tipos de CE que se han dado en cada uno de los proyectos.

Al igual que le caso de la página anterior, es posible reutilizarlo como monitor del estado sobre el progreso de una métrica. En particular, este análisis pretende conocer las consecuencias según cada tipo de Coste Evitable aparecido durante el desarrollo de un proyecto.

**Indicador de Desempeño NC-1 y NC-3**

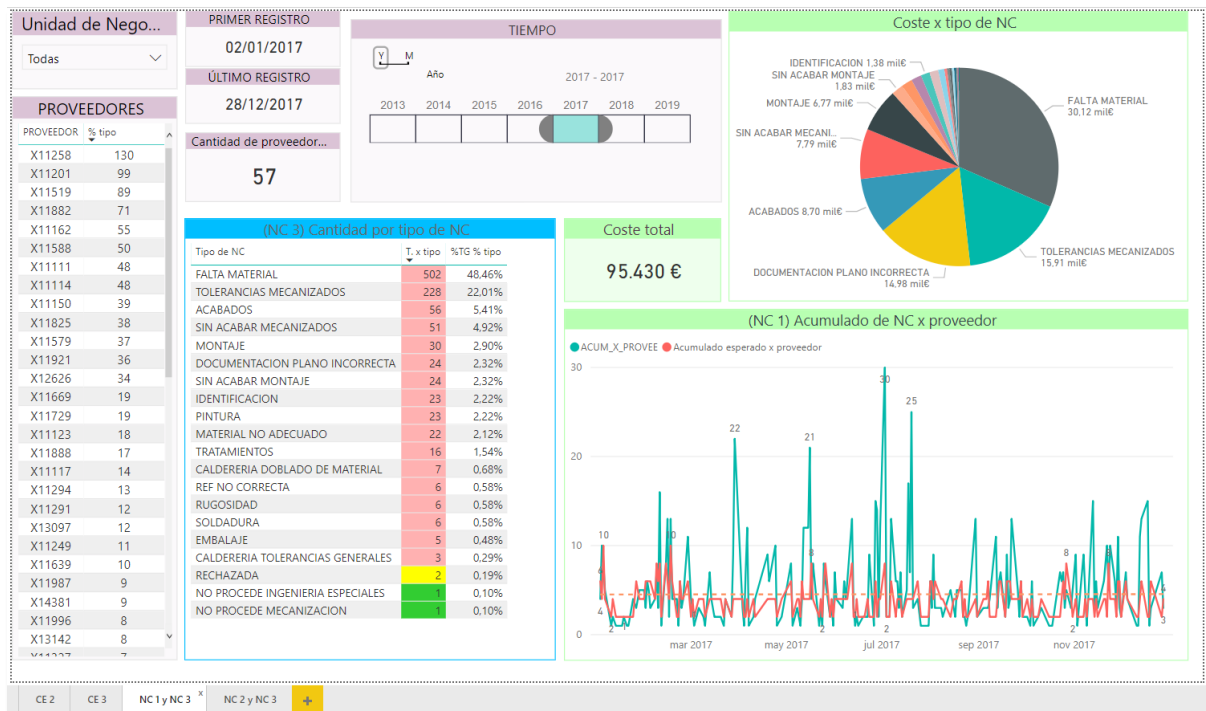


Ilustración 60: composición de visualizaciones para la página NC 1 y NC 3

En el estudio de este índice de desempeño NC -1 se ha pretendido dar relevancia al factor tiempo, es por este motivo que el núcleo del KPI sea una gráfica de líneas. En la gráfica se recogen dos valores que entran en contraste: el número de casos No Conformidad frente al objetivo estimado desde el departamento de Calidad y con protagonista los proveedores.

Esta cantidad es desgranada desde dos perspectivas, por un lado, una visión general del coste por cada tipo de No Conformidad, sobre una gráfica circular, y, por otro lado, una matriz que representa do aquel tipo que supera el objetivo, identificable según el color del fondo de la celda. Como se puede apreciar, por el color del título de la gráfica, la matriz corresponde al KPI NC-3. Como se ha dicho anteriormente, su función es secundaria y complementaria al índice de desempeño NC-1.

```

Acumulado esperado x proveedor =
VAR TIPOS_NC =
  CALCULATE(
    DISTINCTCOUNT('HECHOS NC'[idtn]);
    |ALL(FECHAS[fecha].[Año];FECHAS[fecha].[Mes])
  )
RETURN
TIPOS_NC * 2
    
```

Ilustración 61: métrica para el acumulado objetivo por proveedor

```

% tipo = CALCULATE(count('HECHOS NC'[idtn]);ALL(FECHAS[fecha].[Año];FECHAS[fecha].[Mes]);USERRELATIONSHIP('HECHOS NC'[idtn];'TIPO NC'[idtn]))
    
```

Ilustración 62: métrica para el recuento de las NC según su tipo

Cabe destacar que este estudio se ha ideado con la intención de evaluar cada proveedor de forma individual, pese a que también permite la selección múltiple de valores presionando la tecla “control”.

Ilustración 63: ejemplo de búsqueda



### Indicador de Desempeño NC-2 y NC-3

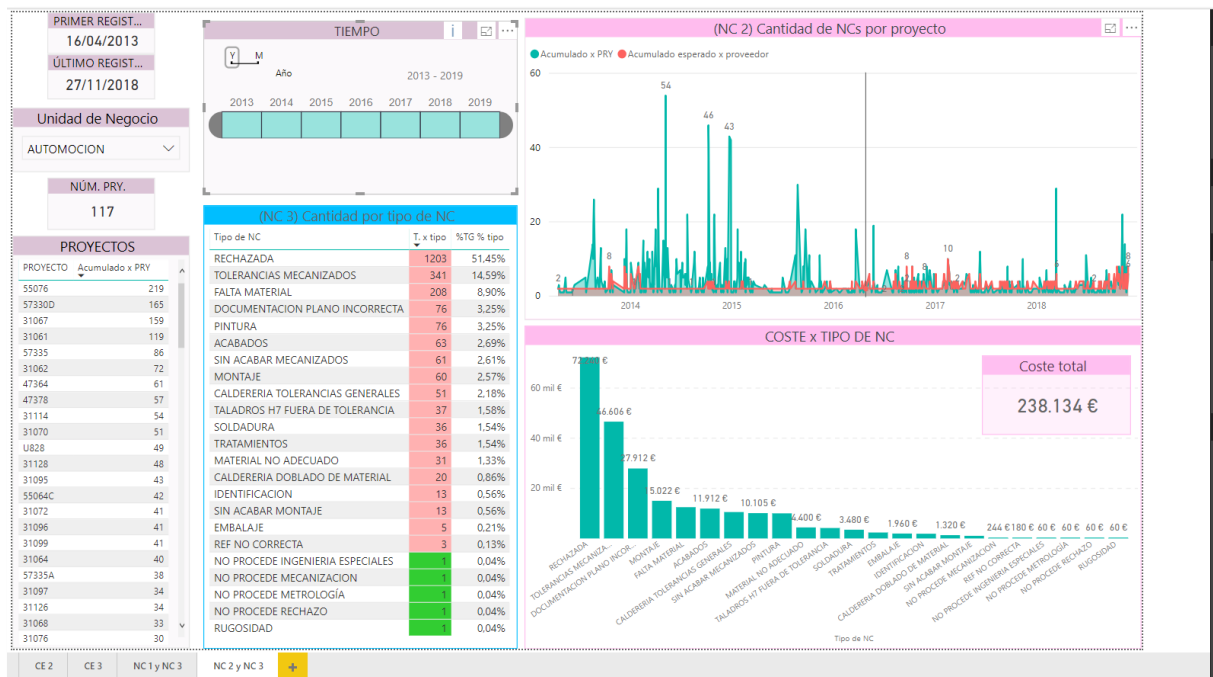


Ilustración 64: composición de visualizaciones para la página NC 2 y NC 3

En esta última página se estudia el KPI dedicado a analizar lo sucedido en los proyectos respecto las No Conformidades que se han dado durante su evolución. El elemento central de estudio es la gráfica de líneas representado lo sucedido a lo largo del tiempo. Su representación desvela el contraste entre el número de veces que se ha dado un NC, línea en color turquesa, frente el objetivo esperado, línea en color rojo.

Se debe tener en cuenta que la intención del estudio es la de evidenciar las particularidades cada proyecto por separado. Se ha mantenido la posibilidad de la selección múltiple por la posibilidad de experimentar y cruzar datos sin limitar la curiosidad del usuario.

En esta página también se puede observar la reutilización de la matriz del índice de desempeño NC 3, con título de color azul, que mide y estudia el contraste de las veces que se da un mismo NC durante la evolución de un proyecto.

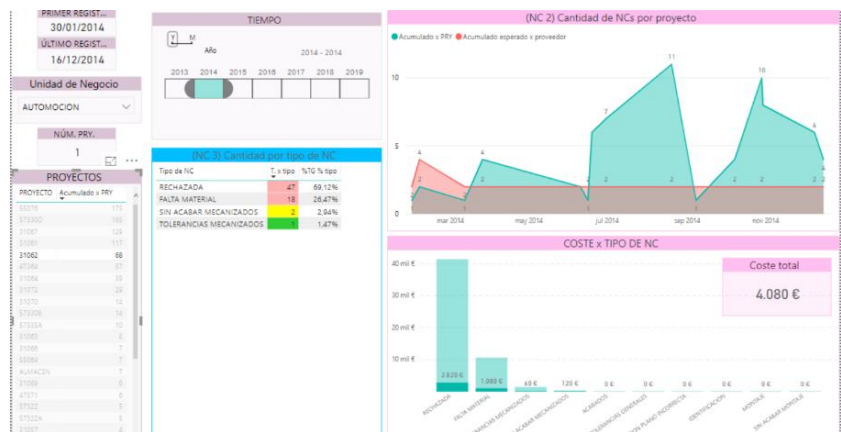


Ilustración 65: ejemplo de uso de la página NC 2 y NC 3

### 3.6 Pruebas de visualización de datos.

Respecto a las posibilidades de visualización que ofrece Power BI, aspectos comentados a lo largo del proyecto, se intenció centrar en la agilidad y rendimiento de la solución, condicionada por las decisiones entorno a la gestión de los datos, configuraciones de los elementos de visualización y el grado de relación entre ellos. También, se han hecho pruebas con el mismo propósito sobre una versión subida al servicio *cloud* de Power BI.

#### Versión desktop

Teniendo presente que el proyecto se ha desarrollado sobre una máquina virtual, se entiende que el rendimiento de la aplicación puede mejorar una vez se exporte a un equipo físico. También se ha de valorar que no se han restringido las relaciones entre elementos de visualización, hecho que obliga a la aplicación a hacer un mayor número de cálculos.

En otra medida, la transición entre páginas es aceptable, la reconstrucción de cada una sufre un *delay* de un par de segundos. Por otro lado, la navegación dentro de la propia página es muy ágil. Su velocidad va disminuyendo conforme se incrementa la selección (empleo de la multi selección mediante la tecla "control") de elementos en la segmentación, hasta llegar a los cinco segundos.

Es muy probable que el hecho de trabajar con todos los datos cargado en el propio archivo \*.PBIX de Power BI reduzca el rendimiento. También se cree que la dispersión de los cálculos en varias medidas no favorece al rendimiento de la visualización de los datos. Otra posible causa del *delay* puede ser dejar en manos de Power BI la segmentación en vez de ejecutar consultas directas sobre la base de datos.

Este puede ser otro aspecto del producto a incluir en el plan de mejora de futuros proyectos.

## Versión Cloud

En general, el rendimiento de la versión ejecutada en el servicio Power BI ofrece una experiencia de usuario mejor. Incorpora una serie de funciones orientadas al trabajo colaborativo e intercambio de la información muy potente. También se incluye la posibilidad de crear paneles resumen que recogen los elementos de visualización más representativos.

En lo referente al rendimiento y visualización de los informes cabe decir que la transición entre páginas es similar a la ofrecida por la versión *desktop*. Aun así, el *delay* en parte se puede achacar a la velocidad de la conexión a Internet.

De la misma manera sucede durante la navegación entre los elementos de visualización dentro de una misma página. El rendimiento es similar, se puede decir que ofrece una mejor experiencia de usuario.

El acceso a la versión publicada en el servicio *cloud* solo es posible mediante las credenciales de la organización UOC y sí se ha concedido permiso de acceso al informe por parte de alguno de los autorizados.

El documento estará publicado durante un periodo de 60 días máximo. Esta limitación se corresponde al periodo de favor por la promoción de la versión Power BI Pro.

<https://app.powerbi.com/groups/b1694280-0013-4386-b89f-22651734e702/reports/40705285-1644-4e3a-8a99-6e9c656f6823?ctid=aec762e4-3d54-495e-a8fe-4287dce6fe69>

## 4. Fase de ensayo y calidad

### 4.1 Disposición en el sistema real

La disposición de las aplicaciones en el sistema ha resultado ser rápida y certera según lo previsto. El proceso de instalación de las aplicaciones (Kettel, MySQL y Power BI) no han sufrido contratiempos, los sistemas de seguridad no han frenado el proceso de adaptación del escenario puesto que las aplicaciones cumplen requisitos de licencia, intrusión en el dominio y empleo.

La solución principal o de más interés para el usuario, Power BI, se ha considerado que sea la versión desktop por la previsión de posibles limitaciones de seguridad, velocidad de carga de datos y poder asegurar su puesta en marcha.

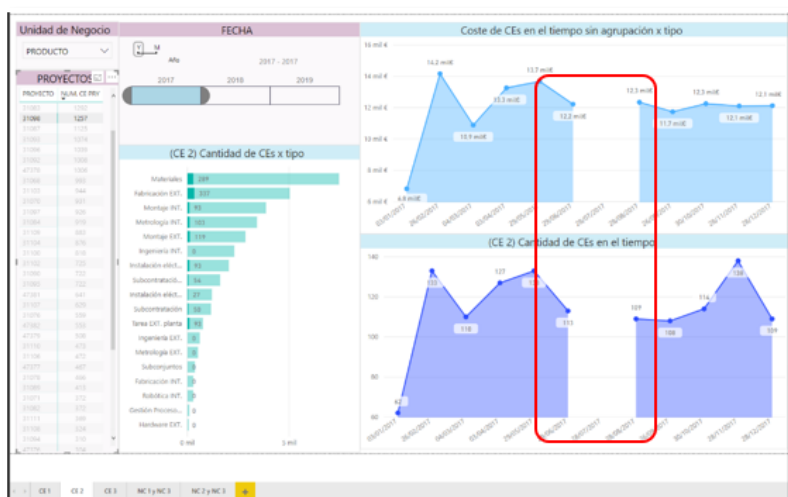
Los productos se han localizado en uno de los equipos del departamento de calidad. Respecto a las fuentes de información indicar que se han reconfigurado los saltos de extracción para las fuentes de datos CSV. Por otro lado, el acceso a las bases de datos MDB también se ha parametrizado y tampoco ha resultado un problema, pese a su localización en los servidores de archivos del dominio.

### 4.2 Pruebas de calidad

Finalizada la puesta en marcha del producto en el sistema, se ha presentado la solución frente al integrante del departamento de calidad valorador. Elogios al margen, indicar que se ha centrado en las posibilidades operativas y funcionales de la herramienta prescindiendo de la ejecución de análisis profundos de las visualizaciones y sin persiguiendo la veracidad de los datos expuesto.

Durante el ensayo se ha explicado los motivos de porque no se ha completado la confección de los objetivos para los KPI CE-2 y CE-3, los problemas encontrados a la hora de acceder a los datos que permitirían su implementación.

Por último, ha resultado sorprendente el resultado de una de las manipulaciones de la segmentación. Se ha detectado un *gap* en un intervalo de fechas que ha llamado la atención.



*Ilustración 66: identificación de una posible anomalía*



Después de una serie de combinaciones mediante los elementos de segmentación, se valora el caso como anómalo y con alta probabilidad de causa en el origen de la fuente de datos.



Ilustración 67: confirmación de la anomalía

A falta de confirmación del origen del problema, se ha considerado como revelador el hallazgo de lo representado por la solución.

La última captura indica que se importaron datos de ese intervalo de tiempo, pero que no están todos.

Una primera impresión apunta a que el conjunto de archivos CSV no está completo, posiblemente porque la exportación desde el ERP no se realizó completamente. Si este es el motivo del gap resolverlo requerirá un control sobre la duplicidad de las transacciones durante algún momento de la carga de los datos. Ante tal circunstancia se toma nota como posible mejora operativa a implementar en versiones futuras.

## 5. Conclusiones

Antes de tratar las conclusiones del TFG quisiera expresar que ha supuesto su ejecución desde la perspectiva personal:

- Destacar que gracias al trabajo se ha adquirido un punto de vista más preciso sobre el mundo Business Intelligence. Es cierto que no ha dado tiempo más que de arañar la superficie, pero, aun así, lo experimentado ha sido realmente enriquecedor y ha despertado la curiosidad.
- Por otro lado, el trabajo desarrollado me ha permitido experimentar y poner en práctica un conjunto de disciplinas y conocimientos necesarios para pilotar un proyecto con gran carga personal por mi relación con la empresa. Así mismo, también me ha supuesto un reto enriquecedor e intenso en cuanto los trabajos de investigación y desarrollo de la parte práctica.

### 5.1 Descripción de las conclusiones del trabajo.

Cada vez más las organizaciones están adquiriendo conciencia en cuanto la importancia de disponer de mejores herramientas capaces de explotar nuevos recursos que les haga ser competitivos. Un buen ejemplo de ello es el tándem formado por metodologías y tecnologías BI.

El trabajo desarrollado ha pretendido, en esencia, llevar a cabo la creación de un producto de naturaleza Inteligencia de Negocio hecho a medida para un sistema de información con un grado elemental de digitalización de la información. Como consecuencia de ello, el trabajo ha permitido contrastar y evidenciar que beneficios supone la integración de una solución de tal naturaleza.

Otro aspecto que ha inducido el trabajo es la visión que debe adquirir el propietario del proyecto respecto a la empresa y de cómo puede influir la calidad del producto en cuanto a su influencia sobre el criterio de quien toma decisiones dentro de una organización. Esta visión debe ser suficientemente sugestiva como para inducir prudencia y responsabilidad en cuanto las decisiones que definan el producto BI resultante.

Por su parte, el que escribe, ha podido experimentar en primera persona la casuística que supone poner en marcha un producto BI en un Sistema de Información sin adaptación para este. Realmente, la sensación final ha sido que se ha empezado la casa por el tejado debido a que la realidad del Sistema de Información no ha facilitado una gestión fácil de los datos.

### 5.2 Consecución de los objetivos planteados

El objetivo fundamental del Trabajo gira en torno a la construcción de una solución BI capaz de proveer de atributos innovadores y dinamización a los mecanismos de análisis sobre la casuística en determinados procesos. Estos atributos se han conseguido prácticamente por completo mediante la entrega del

producto en forma de cuadro de mando, sin embargo, la construcción del mismo no ha sido plena. En concreto, han sido dos KPIs los que no se han podido desarrollar por resultar inalcanzables los objetivos de contraste. El motivo del contratiempo no ha sido otro que el de dar excesivo crédito al interlocutor del departamento y no validar la información suministrada.

Un segundo objetivo general apunta a la promoción de las tecnologías BI dentro del ámbito organizacional. En este sentido la meta se ha alcanzado sobradamente, el ensayo por parte del integrante del departamento de Calidad se ha valorado de forma positiva.

### 5.3 Análisis crítico de la metodología del proyecto y su planificación

Desde un principio se ha planteado la planificación como un factor clave en la persecución de los objetivos planteados según la iniciativa ideada. Con esta misión en mente, se ha optado por estructurar los trabajos conforme el marco de referencia de R. Kimball y se cree que este ha sido un factor fundamental en éxito de la consecución del dashboard.

Estas son las conclusiones del plan de trabajo:

- Se ha seguido todos los pasos planteados en listado de tareas (EDT) sin dificultad. La secuencia de ítems y su desgrane en trabajos se ha seguido de forma natural.
- Es posible que otra secuencia de trabajos hubiese detectado antes el caso de los dos KPIs, pero se cree que dicho problema se ha dado no por la planificación sino por la experiencia del ejecutor de las prospecciones, es por ello que se considera acertada la hoja de ruta planteada desde un principio.
- El desconocimiento sobre el uso de las herramientas BI ha supuesto algo más de esfuerzo del esperado. Se debe reconocer que la fase de confección de la capa ETL ha resultado más intensa de lo esperado por el ejercicio de investigación sobre la solución PDI (Kettle).
- Se cree acertada la combinación de herramientas elegidas para la confección y desarrollo del proyecto y producto. Las funcionalidades colaborativas y trabajo sobre *cloud* han dado la posibilidad de desarrollo de la documentación sin dependencia del lugar.

En cuanto al marco de referencia seguido, como metodología de desarrollo del proyecto, destacar su eficacia. Aplicar una perspectiva *agile* a la confección del producto ha permitido lograr los diferentes ítems con un desvío en el tiempo sin efecto negativo sobre la entrega del producto, dicho de otra forma, no se ha acumulado un exceso de carga hasta el final del proyecto y esto ha permitido su finalización dentro de un tiempo aceptable.

También, se ha procurado, y conseguido, presentar entregas parciales de lo desarrollado además de todas las PEC. Esta estrategia a resultado fundamental puesto que ha sido gracias a las correcciones del consultor que se ha alcanzado una documentación del proyecto correcta tanto por su contenido como por su estructura.

## 5.4 Gestión del proyecto.

Tal y como era de esperar, las complejidades de un trabajo como el presentado ha exigido hacer uso de diferentes roles muy diferentes. Uno de ellos es el de gestor del proyecto y a tenor del resultado se puede concluir que el pilotaje se ha realizado sin sobre saltos ni grandes desviaciones respecto a lo planificado.

Cabe decir que, al margen de las habituales variaciones en los esfuerzos para resolver las tareas, la gestión que se ha hecho del caso sobre la modificación del alcance en dos de los seis KPIs ha conseguido redirigir el ritmo de las tareas planificadas. Las acciones correctoras se centraron en torno a la revisión del resto de objetivos para después confeccionar una alternativa a los KPIs afectados según la nueva realidad detectada.

En resumen, se cree que la correcta planificación, el seguimiento de la metodología escogida y aplicar las pautas propuestas por el consultor han conseguido un resultado más que aceptable del proyecto.

## 5.5 Líneas de trabajo futuro

A lo largo de las diferentes fases del proyecto se han identificado una serie de características extra o aspectos mejorables de alcance más allá de la intención académica del Trabajo, salvando una excepción.

Únicamente se ha dado un caso de no cumplimiento sobre lo planificado. La imposibilidad de implementar los KPIs CE 2 y CE 3. Esta puede ser una buena propuesta de mejora del producto incorporando un mecanismo que facilite la recogida de datos sobre los objetivos para la finalización de los KPIs

En otra medida, estas son las mejoras detectadas y no planteadas dentro del proyecto:

- Extracción de datos directamente del ERP; esta propuesta agilizaría la capa ETL y simplificaría la gestión de la solución (pg. 31)
- Gestión de la antigüedad de los datos extraídos: el producto gestiona datos de diferentes fuentes que son actualizadas en diferentes momentos y esto puede producir incoherencia (pg. 37).
- SCD: mejora que ayudaría en la coherencia de las dimensiones propensas al cambio en el tiempo. (pg 40).
- Mejorar el *delay* durante la navegación en el *dashboard* empleando otras técnicas para la creación de los cubos (pg 61)

Por último, pese a que no hay una declaración oficial de la integración del producto en el sistema este es operativo y está a la espera de su puesta en producción.

## 6. Glosario

- **Back-end tool:** este tipo de herramienta es una aplicación de software, normalmente residente tanto en el cliente como en el servidor, que ayuda en el proceso de extracción de datos de producción. (interface o puente entre la fuente de datos y el DW)
- **Contexto:** dentro del ámbito de trabajo de Power BI el contexto es el alcance de acción de la segmentación o la configuración de los filtros. Algunos alcances: a nivel de elemento de visualización, nivel del sub informe, nivel de informe o nivel de campo.
- **Coste Evitable:** registro de toda aquella decisión o acción o ejecución con algún efecto adverso y que si se hubiese realizado de otra forma habría permitido ahorro de dinero. Este es un factor que apunta a como se hacen las cosas dentro de la empresa y no tanto por el personal externo o proveedor.
- **Data Staging Area:** se puede traducir como área de *pruebas* o área de *ensayo*, también llamada **zona de landing** (zona de *aterrizaje*), es un área intermedia de almacenamiento de datos utilizada para el procesamiento de los mismos durante procesos de extracción, transformación y carga (ETL). Esta área se encuentra entre la fuente de los datos y su destino.
- **Dataset:** conjunto de datos encapsulados en un mismo repositorio y sin necesidad de arquitectura de datos.
- **Data Warehouse:** estructura de datos relacional específica por referirse sus datos a un ámbito o temática particular. Puede estar compuesta de sub estructuras más específicas y simples.
- **DAX (o Expresiones de análisis de datos):** basado en las fórmulas incluidas la aplicación Microsoft Excel. DAX es un conjunto de funciones, operadores y constantes orientados a la formulación y creación de expresiones para calcular y devolver uno o varios valores.
- **Front-end tool:** mediante esta herramienta de cliente se pueden obtener o manipular los datos almacenados en una base de datos relacional (interfaces o puentes entre el DW y la visualización de los datos).
- **Inteligencia de tiempo:** dentro del ámbito de Power BI, es conjunto de técnicas aplicadas en la gestión del tiempo como dimensión en los estudios de los hechos.
- **OLAP:** el Procesamiento Analítico en Línea es un conjunto de herramientas de software que proporciona un marco tridimensional para la toma de decisiones. (dependiendo de donde se localiza el software puede ser el Front-end tool, puede darse el caso de que este se localice dentro de la solución que visualiza los datos.)
- **No Conformidades (o NC):** registro que parecen como consecuencia del análisis de los productos adquiridos. Solo se generan en el momento que aparece una desviación a la hora de verificar que los productos cumplan una serie de parámetros y atributos mínimos requeridos/estipulados. La necesidad del registro de esta información es fundamental para que en un futuro se puedan tener criterio y recordar los detalles de lo sucedido.
- **Segmentación:** forma alternativa de filtro que limita la parte del conjunto de datos tratados y que se muestran en los elementos de visualización de datos.

## 7. Bibliografía

- [1] Formato condicional en gráficos de columnas [URL: <https://www.artesaniadelsoftware.com/2018/01/07/formato-condicional-en-graficos-de-columnas/>] [fecha visita: 4/abril/2019] [fecha creación: 7 enero, 2018]
- [2] 5. Fases en la implantación de un sistema DW. Metodología para la construcción de un DW [URL: <https://churriwifi.wordpress.com/2009/12/05/5-fases-en-la-implantacion-de-un-sistema-dw-metodologia-para-la-construccion-de-un-dw/>] [fecha visita: 02/04/2019] [fecha creación: 5 diciembre 2009]
- [3] Principios para modelar exitosamente un Datawarehouse [URL: <https://www.youtube.com/watch?v=SHIAbB13djo>] [fecha visita: 02/04/2019] [fecha creación: 12/02/2015]
- [4] Manual del Usuario de Spoon [URL: <https://wiki.pentaho.com/display/EALes/Manual+del+Usuario+de+Spoon>] [fecha visita: 11/04/2019] [fecha creación: 01/04/2010]
- [5] Kettle no es una tetera, es la herramienta de ETL de Pentaho! [URL: <https://www.adictosaltrabajo.com/2014/03/04/kettle/#44-ejecutando-la-transformation>] [fecha visita: 11/04/2019] [fecha creación: 28/02/2014]
- [6] Hitachi Vantara Community [URL: <https://community.hitachivantara.com/community/products-and-solutions/pentaho>] [fecha visita: 11/04/2019] [fecha creación: sin definir]
- [7] Pentaho Community Forums [URL: <https://forums.pentaho.com/forums/66-International/>] [fecha visita: 11/04/2019] [fecha creación: sin definir]
- [8] Mejorar tablas de fechas para *Time Intelligence* [URL: <https://www.bitl.es/mejorar-tablas-de-fechas-para-time-intelligence/>] [fecha visita: 07/05/2019] [fecha creación: 16/11/2018]
- [9] Aspectos básicos de DAX en Power BI Desktop [URL: <https://docs.microsoft.com/es-es/power-bi/desktop-quickstart-learn-dax-basics>] [fecha visita: 13/05/2019] [fecha creación: 08/05/2019]

# 8. Anexos

## I. Log de la ejecución de los trabajos ETL

The screenshot shows an ETL job execution interface. At the top, there are tabs for 'Trabajo\_Main\_MDB', 'EXTRACCION\_NC', 'TRANSFORMACION\_NC...', 'CARGA\_DIM\_NC', and 'CARGA\_HECHOS\_NC'. Below the tabs is a flowchart showing the sequence of tasks: Start -> EXTRACCION\_NC -> TRANSFORMACION\_NC -> CARGA\_DIM\_NC -> CARGA\_HECHOS\_NC -> Success. Each task is represented by a green icon with a checkmark, indicating successful completion.

Below the flowchart is the 'Execution Results' section, which includes a table with the following columns: Trabajo / Entrada de Trabajo, Comentario, Resultado, Razón, Nombre Fichero, Núm, and Fecha registro.

Trabajo / Entrada de Trabajo	Comentario	Resultado	Razón	Nombre Fichero	Núm	Fecha registro
Trabajo_Main_MDB	Start of job execution		start			2019/05/29 20:17:20
Trabajo_Main_MDB	Start of job execution		start			2019/05/29 20:17:20
Start	Job execution finished	Exito		file:///C:/DDBB/K/EXTRACCION_N...	0	2019/05/29 20:17:20
EXTRACCION_NC	Start of job execution		Followed unconditional link	file:///C:/DDBB/K/EXTRACCION_N...		2019/05/29 20:17:20
EXTRACCION_NC	Job execution finished	Exito		file:///C:/DDBB/K/EXTRACCION_N...	1	2019/05/29 20:17:31
TRANSFORMACION_NC	Start of job execution		Followed link after success	file:///C:/DDBB/K/TRANSFORMAC...		2019/05/29 20:17:31
TRANSFORMACION_NC	Job execution finished	Exito		file:///C:/DDBB/K/TRANSFORMAC...	2	2019/05/29 20:17:38
CARGA_DIM_NC	Start of job execution		Followed link after success	file:///C:/DDBB/K/CARGA_DIM_N...		2019/05/29 20:17:38
CARGA_DIM_NC	Job execution finished	Exito		file:///C:/DDBB/K/CARGA_DIM_N...	3	2019/05/29 20:17:40
CARGA_HECHOS_NC	Start of job execution		Followed link after success	file:///C:/DDBB/K/CARGA_HECHO...		2019/05/29 20:17:40
CARGA_HECHOS_NC	Job execution finished	Exito		file:///C:/DDBB/K/CARGA_HECHO...	4	2019/05/29 20:17:41
Success	Start of job execution		Followed link after success			2019/05/29 20:17:41
Success	Job execution finished	Exito			4	2019/05/29 20:17:41
Trabajo_Main_MDB	Job execution finished	Exito	finished			2019/05/29 20:17:41

The screenshot shows the 'Execution Results' section with a detailed log for the 'CARGA\_DIM\_NC' and 'CARGA\_HECHOS\_NC' tasks. The log includes the following entries:

- 2019/05/29 20:17:38 - CARGA\_DIM\_NC - Using run configuration [Pentaho local]
- 2019/05/29 20:17:38 - CARGA\_DIM\_NC - Using legacy execution engine
- 2019/05/29 20:17:38 - CARGA\_DIM\_NC - Iniciado despacho de la transformación [CARGA\_DIM\_NC]
- 2019/05/29 20:17:38 - hecho NC Temp.0 - Finished reading query, closing connection.
- 2019/05/29 20:17:38 - hecho NC Temp.0 - Procesamiento finalizado (I=5560, O=0, R=0, W=16680, U=0, E=0)
- 2019/05/29 20:17:38 - Añadir secuencia.0 - Procesamiento finalizado (I=0, O=0, R=5560, W=5560, U=0, E=0)
- 2019/05/29 20:17:38 - código proveedor.0 - Procesamiento finalizado (I=0, O=0, R=5560, W=5560, U=0, E=0)
- 2019/05/29 20:17:38 - Añadir constantes.0 - Procesamiento finalizado (I=0, O=0, R=5560, W=5560, U=0, E=0)
- 2019/05/29 20:17:38 - Concat fields.0 - Procesamiento finalizado (I=0, O=0, R=5560, W=5560, U=0, E=0)
- 2019/05/29 20:17:38 - add dim código NC.0 - Procesamiento finalizado (I=5560, O=0, R=5560, W=5560, U=0, E=0)
- 2019/05/29 20:17:39 - add dim proveedor.0 - Procesamiento finalizado (I=5560, O=50, R=5560, W=5560, U=0, E=0)
- 2019/05/29 20:17:40 - add dim proyecto.0 - Procesamiento finalizado (I=5560, O=0, R=5560, W=5560, U=0, E=0)
- 2019/05/29 20:17:40 - Trabajo\_Main\_MDB - Starting entry [CARGA\_HECHOS\_NC]
- 2019/05/29 20:17:40 - CARGA\_HECHOS\_NC - Using run configuration [Pentaho local]
- 2019/05/29 20:17:40 - CARGA\_HECHOS\_NC - Using legacy execution engine
- 2019/05/29 20:17:40 - CARGA\_HECHOS\_NC - Iniciado despacho de la transformación [CARGA\_HECHOS\_NC]
- 2019/05/29 20:17:40 - Salida hecho NC.0 - Connected to database [wh] (commit=1000)
- 2019/05/29 20:17:40 - hecho NC Temp.0 - Finished reading query, closing connection.
- 2019/05/29 20:17:40 - hecho NC Temp.0 - Procesamiento finalizado (I=5560, O=0, R=0, W=5560, U=0, E=0)
- 2019/05/29 20:17:40 - UN.0 - Procesamiento finalizado (I=0, O=0, R=5560, W=5560, U=0, E=0)
- 2019/05/29 20:17:41 - buscar id proveedor.0 - Procesamiento finalizado (I=5510, O=0, R=5560, W=5560, U=0, E=0)
- 2019/05/29 20:17:41 - buscar id NC.0 - Procesamiento finalizado (I=23, O=0, R=5560, W=5560, U=0, E=0)
- 2019/05/29 20:17:41 - buscar id prj.0 - Procesamiento finalizado (I=425, O=0, R=5560, W=5560, U=0, E=0)
- 2019/05/29 20:17:41 - Filtrar filas.0 - Procesamiento finalizado (I=0, O=0, R=5560, W=5560, U=0, E=0)
- 2019/05/29 20:17:41 - recoge los hechos sin proveedor.0 - Procesamiento finalizado (I=0, O=0, R=50, W=50, U=0, E=0)
- 2019/05/29 20:17:41 - Salida hecho NC.0 - Procesamiento finalizado (I=0, O=5510, R=5510, W=5510, U=0, E=0)
- 2019/05/29 20:17:41 - Trabajo\_Main\_MDB - Starting entry [Success]
- 2019/05/29 20:17:41 - Trabajo\_Main\_MDB - Finished job entry [Success] (result=[true])
- 2019/05/29 20:17:41 - Trabajo\_Main\_MDB - Finished job entry [CARGA\_HECHOS\_NC] (result=[true])
- 2019/05/29 20:17:41 - Trabajo\_Main\_MDB - Finished job entry [CARGA\_DIM\_NC] (result=[true])
- 2019/05/29 20:17:41 - Trabajo\_Main\_MDB - Finished job entry [TRANSFORMACION\_NC] (result=[true])
- 2019/05/29 20:17:41 - Trabajo\_Main\_MDB - Finished job entry [EXTRACCION\_NC] (result=[true])
- 2019/05/29 20:17:41 - Trabajo\_Main\_MDB - Job execution finished
- 2019/05/29 20:17:41 - Spoon - Trabajo ha terminado.

**Execution Results**

Trabajo / Entrada de Trabajo	Comentario	Resultado	Razón	Nombre Fichero	Núm	Fecha registro
Trabajo: Trabajo_Main	Start of job execution		start			2019/05/29 20:25:52
Start	Start of job execution		start			2019/05/29 20:25:52
Start	Job execution finished	Exito			0	2019/05/29 20:25:52
EXTRACCION CE	Start of job execution		Followed unconditional link	file:///C:/DDBB/K/EXTRACCION_C...		2019/05/29 20:26:13
EXTRACCION CE	Job execution finished	Exito		file:///C:/DDBB/K/EXTRACCION_C...	1	2019/05/29 20:26:13
DIMENSIONES CE	Start of job execution		Followed link after success	file:///C:/DDBB/K/DIMENSIONES_...		2019/05/29 20:26:13
DIMENSIONES CE	Job execution finished	Exito		file:///C:/DDBB/K/DIMENSIONES_...	2	2019/05/29 20:26:13
TMP HECHOS CE	Start of job execution		Followed link after success	file:///C:/DDBB/K/TMP_HECHOS_...		2019/05/29 20:26:13
TMP HECHOS CE	Job execution finished	Exito		file:///C:/DDBB/K/TMP_HECHOS_...	3	2019/05/29 20:26:18
Trabajo: Trabajo_Main	Job execution finished	Exito	finished		3	2019/05/29 20:26:18

**Execution Results**

```

2019/05/29 20:25:53 - Filas Unicas.0 - Procesamiento finalizado (I=0, O=0, R=81760, W=11656, U=0, E=0)
2019/05/29 20:25:53 - string to int "OF base".0 - Procesamiento finalizado (I=0, O=0, R=11656, W=11656, U=0, E=0)
2019/05/29 20:25:53 - solo CE.0 - Procesamiento finalizado (I=0, O=0, R=11656, W=11656, U=0, E=0)
2019/05/29 20:25:53 - no CE.0 - Procesamiento finalizado (I=0, O=0, R=11215, W=11215, U=0, E=0)
2019/05/29 20:25:53 - quita "e-" del coste.0 - Procesamiento finalizado (I=0, O=0, R=441, W=441, U=0, E=0)
2019/05/29 20:25:53 - "x", "0" - Procesamiento finalizado (I=0, O=0, R=441, W=441, U=0, E=0)
2019/05/29 20:25:53 - limpia Tipos CE.0 - Procesamiento finalizado (I=0, O=0, R=441, W=441, U=0, E=0)
2019/05/29 20:25:53 - solo "data".0 - Procesamiento finalizado (I=0, O=0, R=441, W=441, U=0, E=0)
2019/05/29 20:25:53 - cambio de tipo "data" y "coste".0 - Procesamiento finalizado (I=0, O=0, R=441, W=1323, U=0, E=0)
2019/05/29 20:25:53 - dim Tipo CE.0 - Procesamiento finalizado (I=0, O=441, R=441, W=441, U=0, E=0)
2019/05/29 20:25:53 - dim proyecto.0 - Procesamiento finalizado (I=0, O=441, R=441, W=441, U=0, E=0)
2019/05/29 20:26:13 - insert to HW.0 - Procesamiento finalizado (I=441, O=191, R=441, W=441, U=250, E=0)
2019/05/29 20:26:13 - Trabajo_Main - Starting entry [DIMENSIONES CE]
2019/05/29 20:26:13 - DIMENSIONES CE - Using run configuration [Pentaho local]
2019/05/29 20:26:13 - DIMENSIONES CE - Using legacy execution engine
2019/05/29 20:26:13 - DIMENSIONES CE - Iniciado despacho de la transformaci3n [DIMENSIONES_CE]
2019/05/29 20:26:13 - lee_dim_Tipo CE.0 - Finished reading query, closing connection.
2019/05/29 20:26:13 - lee_dim_Tipo CE.0 - Procesamiento finalizado (I=20, O=0, R=0, W=20, U=0, E=0)
2019/05/29 20:26:13 - lee_dim_PRY.0 - Finished reading query, closing connection.
2019/05/29 20:26:13 - lee_dim_PRY.0 - Procesamiento finalizado (I=350, O=0, R=0, W=350, U=0, E=0)
2019/05/29 20:26:13 - Inserta dim Tipo CE.0 - Procesamiento finalizado (I=20, O=0, R=20, W=20, U=0, E=0)
2019/05/29 20:26:13 - Insertar dim PRY.0 - Procesamiento finalizado (I=350, O=0, R=350, W=350, U=0, E=0)
2019/05/29 20:26:13 - Trabajo_Main - Starting entry [TMP HECHOS CE]
2019/05/29 20:26:13 - TMP HECHOS CE - Using run configuration [Pentaho local]
2019/05/29 20:26:13 - TMP HECHOS CE - Using legacy execution engine
2019/05/29 20:26:13 - TMP HECHOS CE - Iniciado despacho de la transformaci3n [TMP_HECHOS_CE]
2019/05/29 20:26:13 - Salida Tabla.0 - Connected to database [con_wh] (commit=1000)
2019/05/29 20:26:14 - volcado hecho CE.0 - Finished reading query, closing connection.
2019/05/29 20:26:14 - volcado hecho CE.0 - Procesamiento finalizado (I=27668, O=0, R=0, W=27668, U=0, E=0)
2019/05/29 20:26:14 - descubrir un.0 - Procesamiento finalizado (I=0, O=0, R=27668, W=27668, U=0, E=0)
2019/05/29 20:26:14 - PRY pasa a INT.0 - Procesamiento finalizado (I=0, O=0, R=27668, W=27668, U=0, E=0)
2019/05/29 20:26:14 - PRY pasa a STRING.0 - Procesamiento finalizado (I=0, O=0, R=27668, W=27668, U=0, E=0)
2019/05/29 20:26:14 - busca proyectos.0 - Procesamiento finalizado (I=350, O=0, R=27668, W=27668, U=0, E=0)
2019/05/29 20:26:14 - busca tipo ce.0 - Procesamiento finalizado (I=20, O=0, R=27668, W=27668, U=0, E=0)
2019/05/29 20:26:14 - formatea fechas.0 - Procesamiento finalizado (I=0, O=0, R=27668, W=27668, U=0, E=0)
2019/05/29 20:26:15 - busca fecha .0 - Procesamiento finalizado (I=4027, O=0, R=27668, W=27668, U=0, E=0)
2019/05/29 20:26:16 - quita cols.0 - Procesamiento finalizado (I=0, O=0, R=27668, W=27668, U=0, E=0)
2019/05/29 20:26:18 - Salida Tabla.0 - Procesamiento finalizado (I=0, O=27668, R=27668, W=27668, U=0, E=0)
2019/05/29 20:26:18 - Trabajo_Main - Finished job entry [TMP HECHOS CE] (result=[true])
2019/05/29 20:26:18 - Trabajo_Main - Finished job entry [DIMENSIONES CE] (result=[true])
2019/05/29 20:26:18 - Trabajo_Main - Finished job entry [EXTRACCION CE] (result=[true])
2019/05/29 20:26:18 - Trabajo_Main - Job execution finished
2019/05/29 20:26:18 - Spoon - Trabajo ha terminado.
    
```