

# **Sistema de Inteligencia de negocio para el análisis de los factores del donante que influyen en la supervivencia de los trasplantes de órganos sólidos, y de sus complicaciones**

**Fuencisla Martín Martín**

Máster en Ingeniería Informática

Trabajo final de Máster - M1.321-Business Intelligence

**David Amorós Alcaraz**

**Ferran Prados Carrasco**

07-enero-2018



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/)

## FICHA DEL TRABAJO FINAL

<b>Título del trabajo:</b>	Sistema de Inteligencia de negocio para el análisis de los factores del donante que influyen en la supervivencia de los trasplantes de órganos sólidos, y de sus complicaciones.
<b>Nombre del autor:</b>	Fuencisla Martín Martín
<b>Nombre del consultor:</b>	David Amorós Alcalá
<b>Nombre del PRA:</b>	Ferran Prados Carrasco
<b>Fecha de entrega:</b>	01/2018
<b>Titulación:</b>	Máster en Ingeniería Informática
<b>Área del Trabajo Final</b>	M1.321 - TFM-Business Intelligence
<b>Idioma del trabajo:</b>	Español
<b>Palabras clave:</b>	Inteligencia de negocio, trasplantes, toma de decisiones

### Resumen del trabajo:

España es líder mundial en donación de órganos. Sin embargo, la lista de espera para TOS (trasplante de órgano sólido) está aumentando con el aumento de la edad de la población. Conocer los factores derivados de la edad o de las enfermedades asociadas en los donantes de órganos que pueden influir en los resultados del trasplante y en la morbilidad y supervivencia de los receptores, son de gran importancia a la hora de decidir qué órgano puede ser de mayor utilidad y para qué receptor en lista de espera.

El objetivo principal de este proyecto es diseñar e implementar un sistema de Inteligencia de negocio (BI) que facilite la adquisición, el almacenamiento y la explotación de los datos asociados a los donantes y receptores de TOS. Para ello, siguiendo las premisas del ciclo de vida de Kimball para el desarrollo de sistemas de Inteligencia Empresarial y haciendo uso de la metodología ágil Kanban, se ha diseñado e implementado un almacén de datos que permite almacenar la información adquirida de los donantes y receptores. Este Data Warehouse se alimenta de datos transformados previamente por una serie de procesos ETL (extracción, transformación y carga) que también se han implementado. Finalmente, se ha desarrollado una aplicación BI que ofrece varios cuadros de mando que responden a los requerimientos analíticos definidos con el usuario final teniendo en cuenta su usabilidad para explotar y analizar la información.

Como conclusión se ha obtenido un un sistema BI que permite analizar la influencia de los factores del donante en la supervivencia y calidad de vida del receptor para ayudar a la toma de decisiones.

**Abstract:**

Spain is a world leader in organ donation. However, the waiting list for SOT (solid organ transplant) is increasing along with the age of the population. Knowing the factors derived from age or associated diseases in organ donors that can influence the results of transplantation and the morbidity and survival of recipients is of great importance when deciding which organ may be of greater utility and for what receiver on the waiting list.

The main objective of this project is to design and implement a Business Intelligence (BI) system that facilitates the acquisition, storage and exploitation of data associated with donors and recipients of SOT. To do this, following the premises of Kimball's life cycle for the development of Business Intelligence systems and making use of the Kanban agile methodology, a data warehouse has been designed and implemented to store the information acquired from donors and recipients. This Data Warehouse is powered by data previously transformed by a series of ETL processes (extraction, transformation and loading) that have also been implemented. Finally, a BI application has been developed that offers several control panels that respond to the analytical requirements defined with the end user taking into account their usability to exploit and analyze the information.

In conclusion we have obtained a BI system that allows us to analyze the influence of donor factors on the survival and quality of life of the recipient to help decision-making.

## ÍNDICE

1.	Introducción .....	1
1.1	Contexto y justificación del Trabajo.....	1
1.2	Objetivos del trabajo.....	2
1.2.1	Preguntas analíticas .....	2
1.2.2	Conocimientos previos.....	3
1.3	Enfoque y método seguido .....	3
1.3.1	Definición de tareas .....	4
1.4	Planificación del Trabajo .....	5
1.4.1	Recursos .....	5
1.4.2	Tareas .....	6
1.4.3	Diagrama de Gantt .....	6
1.5	Productos obtenidos .....	7
1.6	Contenido de los capítulos.....	8
2.	Análisis de Requisitos.....	8
3.	Arquitectura .....	9
3.1	Fuentes de datos .....	10
3.2	Procesos ETL.....	10
3.3	Almacén de datos.....	11
3.4	Explotación de datos .....	11
4.	Selección del entorno tecnológico .....	11
4.1	Selección de la base de datos para el Datawarehouse .....	12
4.1.1	MySQL Workbench.....	12
4.2	Selección de la herramienta de integración de datos.....	13
4.2.1	Pentaho data integration .....	14
4.3	Selección de la herramienta de explotación de datos .....	14
4.3.1	Microsoft Power BI.....	14
4.4	Entorno tecnológico seleccionado .....	15
5.	Modelo de datos .....	16
5.1	Modelado dimensional .....	16
5.1.1	Ámbito del proyecto.....	16
5.1.2	Granularidad de la información .....	17
5.1.3	Dimensiones.....	17

5.1.4	Hechos.....	17
5.1.5	Modelo conceptual .....	18
5.1.6	Modelo lógico.....	21
5.2	Modelo físico.....	22
6.	Implementación del Data warehouse .....	23
7.	Diseño y desarrollo de los procesos ETL .....	26
7.1	ETL para la carga de Hospitales.....	26
7.2	ETL para la carga de Donantes .....	28
7.3	ETL para la carga de Receptores .....	31
7.4	ETL para la carga de Injertos .....	33
7.5	ETL para la carga de Estado del receptor .....	34
7.6	ETL para la carga de Estado del injerto .....	37
8.	Diseño de la aplicación BI para la explotación de los datos.....	38
9.	Implementación y análisis de resultados .....	40
9.1	Informe Análisis por Órgano Sólido Trasplantado .....	46
9.2	Informe Análisis Riesgo .....	49
9.3	Informe Factores Donante .....	50
9.4	Informe Análisis Hospitales .....	51
9.5	Explotación de los datos publicados .....	53
10.	Conclusiones.....	55
	Bibliografía .....	57

## LISTA DE ILUSTRACIONES

Ilustración 1. Ciclo de vida definido por Kimball (The Data Warehouse Toolkit) .....	4
Ilustración 2. Diagrama de Gantt .....	7
Ilustración 3. Detalle Gantt 1 .....	7
Ilustración 4. Detalle Gantt 2 .....	7
Ilustración 5. Arquitectura BI definida por Kimball (The Data Warehouse Toolkit) .....	10
Ilustración 6. Pentaho Open BI Suite ( <a href="https://www.spec-india.com/blog/a-single-key-to-unlock-business-analytics-pentaho-bi/">https://www.spec-india.com/blog/a-single-key-to-unlock-business-analytics-pentaho-bi/</a> ).....	13
Ilustración 7. Flujo de trabajo Power BI .....	15
Ilustración 8. Ciclo de trabajo Power BI y productos resultantes .....	15
Ilustración 9. Entorno tecnológico seleccionado .....	16
Ilustración 10. Entradas para el modelado dimensional en 4 pasos (The Data Warehouse Toolkit) .....	16
Ilustración 11. Modelo conceptual 1.....	18
Ilustración 12.. Modelo conceptual 2 .....	19
Ilustración 13. Modelo conceptual 3.....	19
Ilustración 14. Modelo conceptual 4.....	20
Ilustración 15. Modelo conceptual 5.....	20
Ilustración 16. Modelo lógico.....	22
Ilustración 17. Modelo físico .....	23
Ilustración 18. Diseño del modelo físico en MySQL .....	23
Ilustración 19. Relaciones en el diseño físico en MySQL.....	24
Ilustración 20. Sincronización del modelo .....	24
Ilustración 21. Estructura de tablas en MySQL .....	25
Ilustración 22. Despliegue de columnas de cada tabla .....	25
Ilustración 23. Fuentes de datos .....	26
Ilustración 24. Fuente CSV para Hospitales .....	27
Ilustración 25. Proceso de transformación para Hospitales .....	27
Ilustración 26. Salida a Tabla para Hospitales.....	27
Ilustración 27. Ejecución de transformación para Hospitales.....	28
Ilustración 28. Comprobación en MySQL para Hospitales.....	28
Ilustración 29. Fuente de datos CSV para Donantes.....	29
Ilustración 30. Previsualización del input para Donantes .....	29
Ilustración 31. Proceso ETL para Donantes .....	29
Ilustración 32. Configuración de transformación para Donantes.....	30
Ilustración 33. Configuración de salida para Donantes.....	30
Ilustración 34. Ejecución de ETL para Donantes .....	31
Ilustración 35. Comprobación de carga de Donantes .....	31
Ilustración 36. Fuente de datos CSV para Receptores .....	32
Ilustración 37. Transformación para Receptores .....	32
Ilustración 38. Ejecución ETL de Receptores.....	32
Ilustración 39. Comprobación carga de Receptores .....	33
Ilustración 40. Fuente de datos CSV para Injertos .....	33
Ilustración 41. Ejecución de ETL para Injertos .....	34
Ilustración 42. Comprobación carga de Injertos .....	34

Ilustración 43. Fuente de datos CSV para Estado del Receptor .....	35
Ilustración 44. Input para Estado del Receptor.....	35
Ilustración 45. Configuración de salida para Estado del Receptor.....	36
Ilustración 46. Ejecución ETL de Estado del Receptor.....	36
Ilustración 47. Comprobación de carga de datos para Estado del Receptor .....	37
Ilustración 48. Ejecución ETL de Estado del Injerto .....	37
Ilustración 49. Comprobación de carga de datos de Estado del Injerto .....	38
Ilustración 50. Conexiones disponibles Power Bi.....	40
Ilustración 51. Conexión Power Bi con DW .....	41
Ilustración 52. Selección de tablas de trabajo en Power BI .....	41
Ilustración 53. Entorno de trabajo Power Bi .....	42
Ilustración 54. Vista de datos de Power BI.....	42
Ilustración 55. Vista de relaciones de Power BI .....	43
Ilustración 56. Configuración de relación en Power BI .....	43
Ilustración 57. Modelo relacional en Power BI .....	44
Ilustración 58. Editor de Power Query de Power BI.....	44
Ilustración 59. Estilo de los informes en archivo .json.....	45
Ilustración 60. Selección de temas en Power BI .....	45
Ilustración 61. Componentes de diseño de Power BI ( <a href="http://radacad.com/bookmarks-and-buttons-making-power-bi-charts-even-more-interactive">http://radacad.com/bookmarks-and-buttons-making-power-bi-charts-even-more-interactive</a> ) .....	46
Ilustración 62. Informe Análisis TOS 1.....	47
Ilustración 63. Informe Análisis TOS 2.....	47
Ilustración 64. Medida calculada en DAX 1.....	48
Ilustración 65. Medida calculada en DAX 2.....	48
Ilustración 66. Ejemplo para Informe Análisis TOS 1 .....	48
Ilustración 67. Ejemplo para Informe Análisis TOS 2 .....	49
Ilustración 68. Informe Análisis Riesgo .....	49
Ilustración 69. Ejemplo para Informe Análisis Riesgo .....	50
Ilustración 70. Informe Factores Donante .....	50
Ilustración 71. Ejemplo para Informe Factores Donante .....	51
Ilustración 72. Informe Análisis Hospitales 1 .....	51
Ilustración 73. Informe Análisis Hospitales 2 .....	52
Ilustración 74. Ejemplo para Informe de Análisis de Hospitales.....	52
Ilustración 75. Posibilidades de análisis 1 .....	53
Ilustración 76. Posibilidades de análisis 2 .....	53
Ilustración 77. Posibilidades de análisis 3 .....	53
Ilustración 78. Modo enfoque 1.....	54
Ilustración 79. Modo enfoque 2.....	54
Ilustración 80. Mostrar datos de gráfico 1 .....	54
Ilustración 81. Mostrar datos de gráfico 2 .....	55



# 1. Introducció

## 1.1 Contexto y justificación del Trabajo

El trasplante de órganos sólidos (TOS), entendiéndose como tal los de riñón, hígado, corazón, pulmones, páncreas o intestino son el tratamiento establecido de diversas enfermedades que llevan al fracaso o insuficiencia terminal de estos órganos. El trasplante permite salvar la vida en el caso fundamentalmente de insuficiencia hepática, cardíaca o respiratoria terminal, y mejora sustancialmente la calidad de vida y la supervivencia a medio y largo plazo de los pacientes con insuficiencia renal crónica terminal, de los diabéticos con afectación renal y de aquellos pacientes con insuficiencia absortiva grave del intestino. Además, el trasplante es un tratamiento costo-eficiente y, en el caso del trasplante renal, el más frecuentemente realizado, en España cada paciente trasplantado permite ahorrar más de 160.000 euros en los primeros cinco años de tratamiento al Sistema Nacional de Salud, que es la diferencia entre el coste del trasplante junto con el tratamiento antirrechazo de mantenimiento y el coste de la alternativa terapéutica que es el tratamiento con diálisis [4]

España es líder mundial en donación de órganos de forma mantenida desde el año 1992 y desde 2016 es el único país del mundo con más de 100 pacientes trasplantados por cada millón de habitantes cada año. Sin embargo, la lista de espera para TOS sigue manteniéndose y aumentando al aumentar las indicaciones con el aumento de la edad de la población, de manera que todavía entre un 5 y un 6 % de los pacientes fallecen estando en lista de espera para recibir un trasplante, ya que las donaciones de órganos de personas fallecidas no llegan a cubrir las necesidades crecientes de los mismos para trasplante.

Esta brecha entre el número de donaciones de órganos para trasplante y la creciente lista de espera ha llevado a explorar en nuestro país y en el resto países desarrollados la donación de órganos de personas fallecidas que hace años no se tenían en consideración debido a que eran de edad avanzada, tenían enfermedades crónicas, como la diabetes o la hipertensión arterial, infecciones o habían padecido un cáncer, o se les detectaba un tumor en el momento de la donación. Estos donantes, conocidos como donantes con criterios expandidos y donantes con riesgo no estándar (DRNE), son muchas veces la única solución para pacientes que pueden morir en cuestión de días si no reciben el trasplante y por lo tanto se utilizan estos órganos [1][3].

Se han descrito resultados peores en la supervivencia del injerto trasplantado o de los receptores de trasplante cuando se emplean los órganos donados por este tipo de donantes, pero como se ha referido anteriormente son la única solución para la supervivencia de muchos pacientes en lista de espera. [2] Órganos donados que, a priori, se supone pueden tener una calidad subóptima, y por lo tanto una menor duración de su función, pueden ser idóneos para pacientes en lista de espera de edad avanzada en los que su supervivencia y calidad de vida se va a ver muy mejorada por el trasplante y la duración prevista de la buena función de este excede la expectativa media de vida de la población de la misma edad que el receptor.

Conocer los factores derivados de la edad o de las enfermedades asociadas en los donantes de órganos que pueden influir en los resultados del trasplante y en la morbilidad y supervivencia de los receptores son de gran importancia para la toma de decisiones a la hora de decidir qué órgano puede ser de mayor utilidad y para qué receptor en lista de espera [5][6].

## 1.2 Objetivos del trabajo

El objetivo principal de este proyecto es diseñar e implementar un sistema de Business Intelligence que facilite la adquisición, el almacenamiento y la explotación de los datos asociados a los donantes de órganos y a los receptores de TOS y las relaciones entre estos, en especial los resultados a medio y largo plazo del TOS, en cuanto a supervivencia y calidad de vida.

Los objetivos son los siguientes:

1. Diseñar un almacén de datos (Data Warehouse) que permita almacenar la información adquirida de los diferentes orígenes de datos. Teniendo en cuenta las diferentes fuentes y las bases de datos ya existentes.
2. Implementar este almacén de datos y programar los procesos ETL (extracción, transformación y carga) que permitan alimentar el Data Warehouse a partir de ficheros base.
3. Analizar las diferentes plataformas BI disponibles al mercado que permitan explotar la información almacenada. Seleccionar e implantar una de estas herramientas de tal forma que se disponga de una capa de software para análisis de la información.
4. Obtener una herramienta capaz de ofrecer un análisis de los resultados teniendo en cuenta las preguntas analíticas e indicadores descritos mediante cuadros de mando para la presentación de los resultados.

### 1.2.1 Preguntas analíticas

- Para cada uno de los tipos de TOS (riñón, hígado, corazón, pulmones, páncreas e intestino) cuales son los factores asociados del donante que pueden influir negativamente en los resultados de supervivencia o de morbilidad en los receptores, en qué medida y en comparación con la población general de individuos sanos.
- Cuál es el riesgo de trasplantar órganos de donantes con infecciones o con tumores activos o en los antecedentes del donante, y hasta qué punto es asumible en el balance riesgo-beneficio.
- Que factores del donante son los que en mayor medida influyen en los resultados.
- Diferencias en los resultados para mismo tipo de donantes y receptores en función del hospital en el que se produce la donación o el trasplante.

## 1.2.2 Conocimientos previos

Para poder modelar adecuadamente el problema, aquí se presentan las principales reglas de negocio, así como los principales actores y conceptos.

**DONANTE:** Existirá un registro único con todas las variables significativas por cada **donante**, con un identificador único que sirva para indexarlo y relacionarlo con los **receptores** de los órganos, teniendo en cuenta que una variable importante es la **fecha de la donación**.

**RECEPTOR:** Existirá un registro basal para cada trasplante con las variables significativas demográficas y clínicas de inicio del **receptor**, con un identificador único que permita indexarlo, una de las variables será el identificador único del **donante**, del **órgano donado**, y otra variable fundamental será la **fecha del trasplante**, que nunca puede ser anterior a la de la donación.

A tener en cuenta:

- Un donante puede donar a más de un receptor.
- Un mismo receptor puede recibir más de un trasplante de más de un donante, sea del órgano primeramente trasplantado (retrasplante) o de otro.

**TRASPLANTE:** Para cada **trasplante**, identificado por su identificador único, habrá tantos registros de **seguimiento** como revisiones clínicas se realizan, en las que se recogerá el estado de salud del receptor y la función buena, deficiente o nula del injerto trasplantado. Siendo el último registro de seguimiento el que recoja el fallecimiento del receptor, que puede ser con buena función del trasplante o con fracaso del trasplante.

## 1.3 Enfoque y método seguido

Teniendo en cuenta la complejidad de un proyecto de Business Intelligence, donde se necesitará flexibilidad especialmente en la entrada de tareas, así como una validación periódica junto con expertos en los datos a analizar, se enfocará la gestión personal del proyecto con una metodología ágil. En concreto, Kanban, cuyo objetivo es gestionar de manera general cómo se van completando las tareas.

Kanban es una palabra japonesa que significa “tarjetas visuales”, donde Kan es “visual”, y Ban corresponde a “tarjeta”. Las principales ventajas que ofrece esta metodología es que es muy fácil de utilizar, actualizar y asumir por parte del equipo. Se trata de técnica de gestión de las tareas muy visual, que permite ver a golpe de vista el estado del proyecto, así como también pautar el desarrollo del trabajo de manera efectiva.

Kanban una metodología ágil muy similar a la popular Scrum, pero tiene algunas peculiaridades. Al igual que Scrum, Kanban se basa en el principio de desarrollo incremental, dividiendo el trabajo en distintas partes. Esto significa que se habla de la tarea en sí, sino que se divide en distintos pasos para agilizar el proceso de producción.

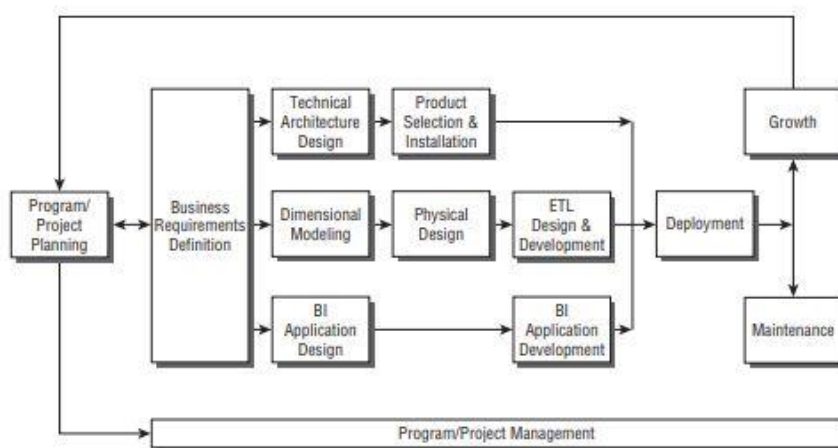
Kanban prioriza el trabajo que está en curso en vez de empezar nuevas tareas. Una de las principales diferencias y aportaciones de Kanban es que el trabajo en curso debe estar limitado y, por tanto, existe un número máximo de tareas a realizar en cada fase. En concreto se define el número de tareas máximo que se pueden tener en cada una de las fases. Esto restringe el trabajo en curso. En Kanban no se puede abrir una nueva tarea sin finalizar otra [7].

De esta manera, se pretende dar respuesta al problema habitual de tener muchas tareas abiertas, pero con un ratio de finalización muy bajo. Para la metodología Kanban lo importante es que las tareas que se abran se cierren antes de empezar con la siguiente.

### 1.3.1 Definición de tareas

Según diversos autores la inteligencia de negocio se puede definirse como “El conjunto de habilidades, tecnologías, aplicaciones y buenas prácticas utilizadas para ayudar a una organización determinada a obtener mayor conocimiento de su contexto comercial, imprescindible para tomar mejores decisiones encaminadas a la adquisición de ventajas competitivas”. Para este proyecto, enmarcado en el ámbito sanitario, se pretende ayudar a obtener mayor conocimiento de los resultados de los pacientes que reciben un trasplante de órgano sólido, en cuanto a supervivencia y calidad de vida, e identificar los factores clave que ayuden a la toma de decisiones a la hora de asignar un determinado órgano donado al receptor más idóneo.

Aunque existen patrones de diseño concretos para los datos de sanidad, que se tendrán en cuenta, la metodología para construir un sistema de BI es genérica, y ayudará a simplificar la complejidad que un proyecto de estas características supone. En la siguiente imagen se representan 13 áreas de tareas fundamentales involucradas en el desarrollo exitoso de una solución de inteligencia de negocio [8].



**Figure 17-1:** Kimball Lifecycle diagram.

*Ilustración 1. Ciclo de vida definido por Kimball (The Data Warehouse Toolkit)*

Como se puede observar, la tarea Definición de Requerimientos del Negocio interviene como base para los tres caminos que le siguen. Estos caminos que aparecen en el medio del ciclo de vida definido por Kimball (The Data Warehouse Toolkit), se concentran en tres áreas:

- **Tecnología**

Tareas relacionadas con la selección, instalación y configuración de las tecnologías que van a ser empleadas en el desarrollo de la solución.

- **Datos**

Tareas relacionadas con el diseño e instanciación del modelo dimensional y el desarrollo de los procesos de carga, extracción y transformación de datos (ETL) para alimentar el Data Warehouse (DWH).

- **Aplicaciones de inteligencia de negocio**

Tareas están relacionadas con el diseño y desarrollo de las aplicaciones BI para los usuarios del negocio.

Desarrollar cada una de las tareas propuestas expuestas en el diagrama del ciclo de vida de Kimball es de vital importancia para obtener un sistema de inteligencia de negocio bien pensado, que cumpla con las necesidades del proyecto y permita apoyar la toma de decisiones de forma exitosa [8].

## 1.4 Planificación del Trabajo

A continuación, se definen los recursos y cada una de las tareas que se pretenden llevar a cabo, así como las diferentes entregas a lo largo del ciclo de vida del proyecto.

### 1.4.1 Recursos

El equipo de trabajo estará formado por:

- David Amorós como director experto en BI, que llevará a cabo el seguimiento y control de este proyecto.
- Fuencisla Martín, como autora de este proyecto, que llevará a cabo la planificación, el análisis y la implementación.
- Se contará con la colaboración de médicos expertos de la ONT (Organización Nacional de Trasplantes) para obtener toda la información necesaria y las validaciones pertinentes de los resultados obtenidos en este proyecto.

Recursos tecnológicos

- PC empleado para el desarrollo del proyecto.
- Herramientas software seleccionadas para el desarrollo del proyecto.

## 1.4.2 Tareas

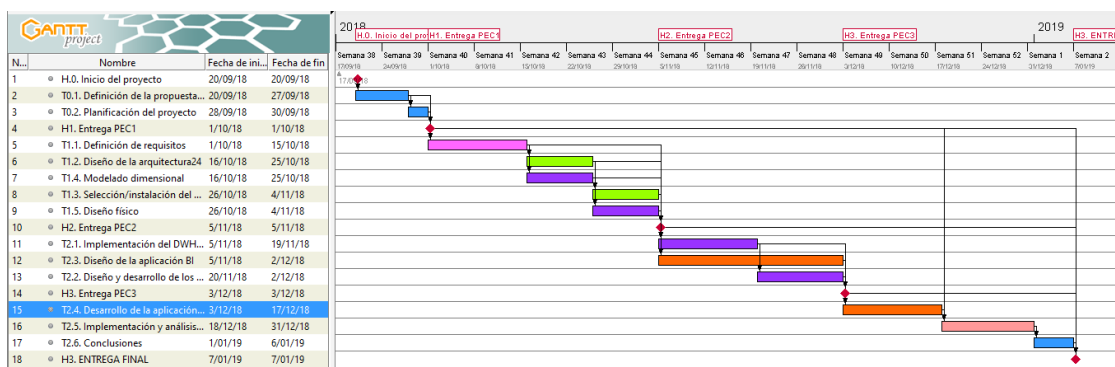
Seguindo los hitos o entregas definidos para el TMF (PECs) y el ciclo de vida propuesto por Kimball, se van a llevar a cabo las siguientes tareas:

Tareas	Dependencias	Fecha inicio	Fecha fin
<b>GESTIÓN</b>			
➤ H.0. Inicio del proyecto			20/09/2018
• T0.1. Definición de la propuesta de proyecto	H.0	20/09/2018	27/01/2010
• T0.2. Planificación del proyecto	T0.1	28/01/2010	30/09/2018
➤ H1. Entrega PEC1			01/10/2018
<b>ANÁLISIS</b>			
• T1.1. Definición de requisitos	H1	01/10/2018	15/10/2018
<b>TECNOLOGÍA</b>			
• T1.2. Diseño de la arquitectura	T1.1	16/10/2018	25/10/2018
• T1.3. Selección/instalación del entorno tecnológico	T1.2	26/10/2018	04/11/2018
<b>DATOS</b>			
• T1.4. Modelado dimensional	T1.1	16/10/2018	25/10/2018
• T1.5. Diseño físico	T1.4	26/10/2018	04/11/2018
➤ H2. Entrega PEC2			05/11/2018
• T2.1. Implementación del DWH	H2	05/11/2018	19/11/2018
• T2.2. Diseño y desarrollo de los procesos ETL	T2.1	20/11/2018	02/12/2018
<b>APLICACIÓN BI</b>			
• T2.3. Diseño de la aplicación BI	T1.1	05/11/2018	02/12/2018
➤ H3. Entrega PEC3			03/12/2018
• T2.4. Desarrollo de la aplicación BI	H3	03/12/2018	17/12/2018
<b>IMPLEMENTACIÓN</b>			
• T2.5. Implementación y análisis de resultados	H1,H2,T2.4	18/12/2018	31/12/2018
<b>GESTIÓN</b>			
• T2.6. Conclusiones	T2.5	01/01/2019	06/01/2019
➤ H3. ENTREGA FINAL			07/01/2019
▪ Gestión de proyecto - Documentación - Control y seguimiento		20/09/2018	07/01/2019

Se puede observar que como máximo se trabajará en dos tareas en paralelo, a parte de la gestión del proyecto, seguimiento, control y documentación que será una tarea continua, a lo largo de toda la vida del proyecto. Estas tareas se descompondrán en pequeñas subtareas que no se podrán solapar. Hasta que una subtarea no haya finalizado, no se llevará a cabo la siguiente, como se ha explicado en la metodología de trabajo Kanban, que se va a aplicar.

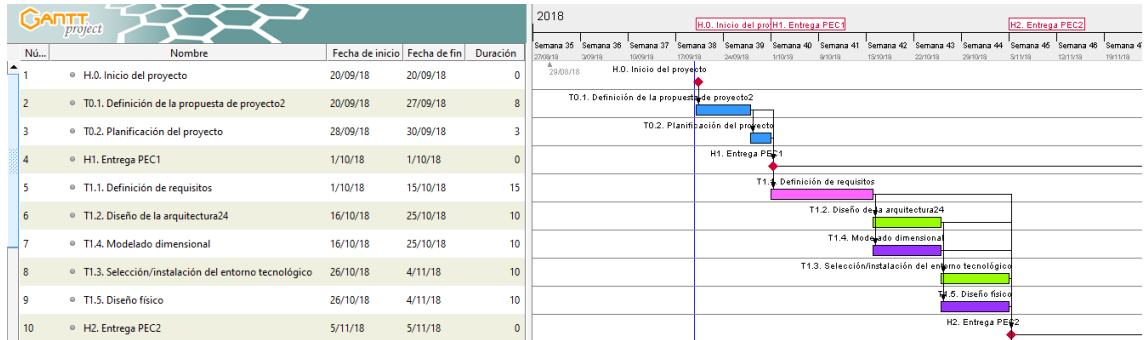
## 1.4.3 Diagrama de Gantt

A continuación, se muestra el diagrama de Gantt que se utilizará para la gestión y seguimiento del desarrollo del proyecto, con todas las tareas e hitos definidos y sus dependencias temporales:

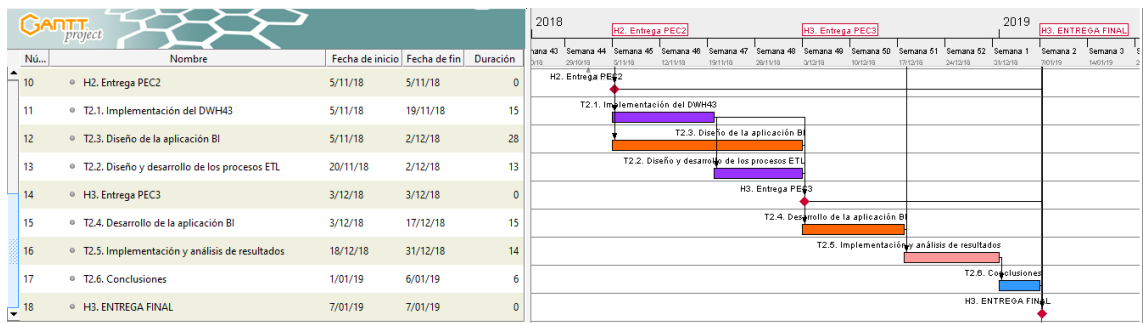


Il·lustració 2. Diagrama de Gantt

Detalle diagrama temporal, tareas, fechas y duración:



Il·lustració 3. Detalle Gantt 1



Il·lustració 4. Detalle Gantt 2

## 1.5 Productos obtenidos

Los productos obtenidos en el presente proyecto son los siguientes:

- Un Data Warehouse: dwh\_trasplante como almacén de datos.
- Conjunto de procesos ETL para la extracción y transformación de las fuentes de datos, y su carga en el Data Warehouse.
- Una aplicación BI para la explotación de datos con cuadros de mandos para el análisis de la información.

## 1.6 Contenido de los capítulos

A continuación, se hace un breve resumen de los capítulos que vienen a continuación y que describen todo el ciclo de desarrollo de este proyecto:

- **Análisis de Requisitos:** Este capítulo trata de los requisitos a alcanzar para lograr los objetivos de la herramienta de inteligencia de negocio que se desea desarrollar.
- **Arquitectura:** En este apartado se explica la estructura técnica del sistema, así como cada uno de los elementos a implementar
- **Selección del entorno tecnológico:** Se explica el por qué de las tecnologías seleccionadas para la realización de este proyecto.
- **Modelo de datos:** El contenido de este capítulo consiste en los pasos dados para alcanzar el modelo relacional del almacén de datos.
- **Implementación del Data Warehouse:** Contiene los pasos seguidos para la creación del modelo físico de la base de datos para la creación de las tablas y las relaciones en el Data Warehouse.
- **Diseño y desarrollo de los procesos ETL:** Explica las características de los procesos creados para la extracción, transformación y carga de datos en el almacén, así como su ejecución y comprobación en la base de datos.
- **Diseño de la aplicación BI para la explotación de los datos:** Este apartado contiene la realización de la conexión de la herramienta de explotación con los datos y los requisitos de diseño para los informes analíticos.
- **Implementación y análisis de los resultados:** En este capítulo se muestran los cuadros de mando creados, así como las diferentes posibilidades analíticas que permite la aplicación BI.

## 2. Análisis de Requisitos

El objetivo principal de este proyecto es diseñar e implementar un sistema de Business Intelligence que facilite la adquisición, el almacenamiento y la explotación de los datos asociados a los donantes de órganos y a los receptores de TOS y las relaciones entre estos, en especial los resultados a medio y largo plazo del TOS, en cuanto a supervivencia y calidad de vida.

Los requisitos genéricos de este proyecto son comunes a los requisitos de cualquier sistema de Inteligencia empresarial, por eso se van a tener en cuenta los requisitos definidos por Ralph Kimball y Margy Ross en su libro "The Data Warehouse Toolkit, 3rd Edition" [8][9].

- **Información accesible de forma simple y rápida**

El contenido del sistema BI debe ser comprensible, los datos deben ser intuitivos y evidentes para el usuario y no solo para el desarrollador. Las estructuras y etiquetas de los datos deben imitar los procesos de pensamiento y el vocabulario de los usuarios finales.

Las herramientas y aplicaciones de inteligencia empresarial que acceden a los datos deben ser simples y fáciles de usar. También deben devolver los resultados de la consulta al usuario con tiempos de espera mínimos.

- **Información presentada de manera consistente**

Los datos deben ser recolectados cuidadosamente de una variedad de fuentes, limpiados, con calidad asegurada y liberados solo cuando sean aptos para el consumo del usuario. La coherencia también implica que las etiquetas y definiciones comunes para los contenidos del



sistema BI se utilicen en todas las fuentes de datos. Si dos medidas de estado del trasplante tienen el mismo nombre, deben significar lo mismo. Por el contrario, si dos medidas no significan lo mismo, se deben etiquetar de manera diferente.

- **Adaptabilidad al cambio**

Las necesidades de los usuarios, las condiciones del estudio del éxito del trasplante, los datos y la tecnología están sujetos a cambios. El sistema BI debe estar diseñado para manejar este inevitable cambio sin invalidar los datos o aplicaciones existentes. Los datos y aplicaciones existentes no se deben cambiar ni interrumpir cuando se necesiten analizar nuevos parámetros o se agregan nuevos datos al almacén. Finalmente, si los datos descriptivos en el sistema BI deben modificarse, debe considerar adecuadamente los cambios y hacer que estos cambios sean transparentes para los usuarios.

- **Presentación la información de manera oportuna**

Como el sistema BI se usa más intensamente para las decisiones operativas, los datos sin procesar pueden necesitar convertirse en información procesable en horas, minutos o incluso segundos. El equipo de BI y los usuarios deben tener expectativas realistas de lo que significa entregar datos cuando hay poco tiempo para limpiarlos o validarlos

- **Protección de la información confidencial**

El sistema BI debe controlar de manera efectiva el acceso a la información confidencial. En el contexto de este proyecto, al tratarse de datos médicos, la delicadeza de estos datos es extrema, por lo que este requisito es especialmente importante.

- **Sistema de soporte de decisiones**

El almacén de datos debe tener los datos correctos para respaldar la toma de decisiones. Los resultados más importantes de un sistema BI son las decisiones que se toman en base a la evidencia analítica presentada; estas decisiones proporcionan el impacto y el valor empresarial atribuibles al sistema.

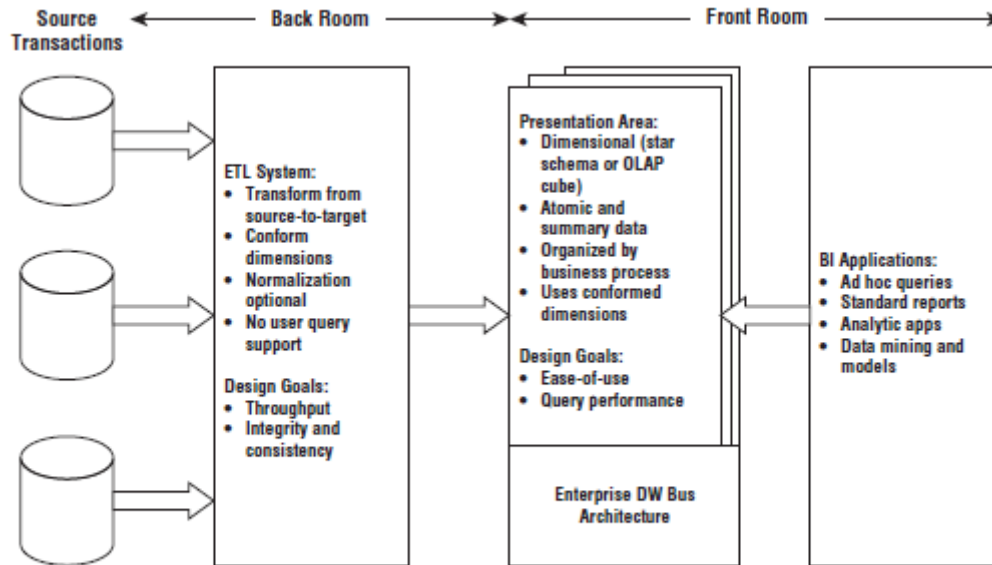
- **Aceptación del sistema BI**

A diferencia de una implementación de sistema operacional donde los usuarios no tienen más remedio que usar el nuevo sistema, el uso BI a veces es opcional. Los usuarios adoptarán el sistema DW / BI si es la fuente “simple y rápida” de información procesable.

### 3. Arquitectura

El modelo de arquitectura empleado para el diseño del software es, con diferencia, uno de los rasgos más críticos en una solución de BI.

Una arquitectura de BI es un marco donde organizar los datos, gestionar la información, los componentes tecnológicos que son usados para construir sistema de BI y sistemas de reporte y análisis de de datos. La base de la arquitectura de BI juega un papel importante en los proyectos de inteligencia de negocio porque afecta al desarrollo y a las decisiones de implementación. Dada la importancia de la arquitectura, nos vamos a basar en el modelo Kimball, representado en la siguiente figura:



**Figure 1-7:** Core elements of the Kimball DW/BI architecture.

*Il·lustració 5. Arquitectura BI definida por Kimball (The Data Warehouse Toolkit)*

Los componentes de la arquitectura tecnológica describen el proceso que se va a seguir durante el proyecto, desde la recolección de los datos hasta la explotación de los datos por el usuario final. Los elementos de la arquitectura son [8]:

- Fuentes de datos
- Procesos ETL
- Almacén de datos
- Explotación de datos

A continuación, se describe en detalle cada uno de estos elementos.

### 3.1 Fuentes de datos

El primer nivel de de la arquitectura son los datos. Estos se encuentran almacenados en los diferentes sistemas de información transaccionales/operacionales internos y externos a la organización.

En nuestro contexto, enmarcado en el ámbito sanitario, la fuente de datos se encuentra en las consultas médicas de seguimiento de los pacientes trasplantados, su historia clínica. La Organización Nacional de Trasplantes dispone un registro con los resultados de los trasplantes para cada paciente con datos de evolución.

### 3.2 Procesos ETL

El sistema de extracción, transformación y carga (ETL) de la arquitectura BI consta de un área de trabajo, estructuras de datos instanciados y un conjunto de procesos.

Un sistema ETL realiza funciones de extracción de la fuente de datos, transformación de los datos y la carga de los mismos en un Data Warehouse.

La **extracción** es el primer paso en el proceso de obtener datos en el entorno del almacén de datos. Extraer significa leer y comprender los datos de origen y copiar los datos necesarios en el sistema ETL para su posterior manipulación. En este punto, los datos pertenecen al almacén de datos.

Una vez que se extraen los datos al sistema ETL, hay numerosas **transformaciones** potenciales, como la limpieza de los datos, combinar datos de múltiples fuentes y analizar la duplicidad.

El último paso del proceso de ETL es la estructuración física y la **carga** de datos en los modelos dimensionales objetivo del área de presentación [8].

### 3.3 Almacén de datos

El Almacén de datos (Data Warehouse) o área de presentación es donde los datos se organizan, almacenan y ponen a disposición de los usuarios, los informes y otras aplicaciones analíticas de BI para realizar consultas directas.

El Data Warehouse es una gran base de datos corporativa que almacena los datos extraídos a partir de diferentes fuentes, la información almacenada se utilizará posteriormente en las consultas e informes. La información se almacena en el Data Warehouse en diferentes periodos de tiempo, y no se modifica (como en los sistemas transaccionales) sino que se incrementa [8].

### 3.4 Explotación de datos

El componente principal final de la arquitectura Kimball es la aplicación de inteligencia empresarial (BI). El término aplicación de BI se refiere a la gama de capacidades proporcionadas a los usuarios para aprovechar el área de presentación para la toma de decisiones analíticas.

La consulta, es el punto central del uso de datos para mejorar la toma de decisiones [8].

## 4. Selección del entorno tecnológico

Una vez definidos los requisitos y arquitectura del entorno BI que se desea conseguir, se va a realizar un análisis de las tecnologías existentes, priorizando las de **software libre**, para seleccionar las más adecuadas para este proyecto.

En general los programas que utilizan arquitectura y framework de open source ofrecen un mayor nivel de integración con otras plataformas. Además, no suponen las habituales restricciones ligadas a la utilización de herramientas propietarias, ni supeditan al departamento de IT a un único suministrador, es por esto por lo que analizará este tipo de componentes

tecnológicos para satisfacer cada elemento de la arquitectura descrita teniendo en cuenta los requerimientos que se desean alcanzar. Por otro lado, existen en el mercado herramientas comerciales muy consolidadas y muy orientadas a usuario que se tendrán también en consideración a la hora de seleccionar el entorno de explotación de datos, ya que, se requiere un entorno amigable y de fácil y rápido aprendizaje para el usuario final.

## 4.1 Selección de la base de datos para el Datawarehouse

El motor de base de datos para el Datawarehouse debe ser capaz de trabajar con grandes volúmenes de datos, garantizar la velocidad de realización de consultas, el particionado y la replicación de datos. Actualmente, entre los motores de base de datos Open Source que cumplen estos requisitos, PostgreSQL y MySQL son las bases de datos más conocidas y utilizadas.

PostgreSQL es un sistema de gestión de bases de datos relacional orientado a objetos y de código abierto, publicado bajo la licencia PostgreSQL, similar a la BSD o la MIT.

El proyecto de desarrollo de MySQL ha hecho disponible su código fuente bajo los términos de la Licencia Pública General de GNU, así como una variedad de acuerdos privativos. Ahora es propiedad de Oracle Corporation y ofrece varias ediciones pagadas para uso privativo.

Por la facilidad de instalación y la rapidez de aprendizaje para su uso, se ha optado finalmente por MySQL y en concreto por su herramienta visual para diseño de bases de datos: MySQL Workbench [14].

### 4.1.1 MySQL Workbench

MySQL Workbench es una herramienta visual de diseño de bases de datos que integra desarrollo de software, administración de bases de datos, diseño de bases de datos, creación y mantenimiento para el sistema de base de datos MySQL.

Este software permite entre otras funciones, el modelado de datos, desarrollo de SQL y herramientas de administración completas para la configuración de servidores, administración de usuarios y copias de seguridad. MySQL Workbench está disponible en Windows, Gnu/Linux y Mac OS X.

Uno de los principales motivos por los que se ha escogido esta herramienta es porque su facilidad para crear modelos de esquema de base de datos de manera gráfica. También permite invertir y enviar ingeniería entre un esquema y una base de datos en directo al tiempo que permite editar todos los aspectos de la base de datos utilizando el Editor de tablas. El Editor de tablas proporciona facilidades para editar tablas, columnas, índices, desencadenadores, particiones, opciones, inserciones, etc.

Con esta herramienta se creará el Datawarehouse y se dispondrá de su capacidad de ejecutar consultas SQL en las conexiones de la base de datos mediante el Editor de SQL que viene incorporado.

## 4.2 Selección de la herramienta de integración de datos

Para la sección de la herramienta de integración de datos, donde se van a crear los procesos ETL, se ha decidido comparar entre soluciones BI integradas, para valorar el aprovechar también sus herramientas de informes y análisis de la información. Actualmente existen varias soluciones **integradas de BI Open Source** cada vez más competitivas. Algunas de las soluciones de BI Open Source más populares existentes en el mercado son: **Pentaho y JasperSoft**

Tanto Pentaho como Jaspersoft ofrecen la ventaja de eliminar el coste de la licencia de software y se ofrecen dos **suits de herramientas open source** para cubrir los procesos de BI bastante completas y similares, pero se pueden encontrar algunas diferencias en cuanto a sus características. Si bien la herramienta de Jaspersoft para diseñar informes es comparativamente mejor que la que ofrece Pentaho, las capacidades del panel de Pentaho en términos de funcionalidad son mejores. Esto se debe a que la funcionalidad de Dashboard solo está presente en la edición Enterprise de Jaspersoft, mientras que en Pentaho está disponible en la edición básica. La capa intermedia de Pentaho también proporciona más flexibilidad en términos de diseño de complementos e integración con aplicaciones [10][11].

En cuanto a herramientas ETL, el integrador de datos Pentaho es comparativamente mejor. Cuando se trata del análisis OLAP, el motor Pentaho Mondrian también es más potente en comparación con Jaspersoft [13].

El enfoque de Jaspersoft es más sobre informes y análisis, y el enfoque de Pentaho es la integración de datos, ETL y la automatización del flujo de trabajo.

Aunque también se requiere la construcción ágil de informes para el análisis de la información, finalmente, se ha escogido la herramienta Pentaho porque, además de su potencia para la integración de datos, se encuentra bastante documentado y extendido su uso [12].

En la siguiente ilustración se muestra la arquitectura funcional de Pentaho:

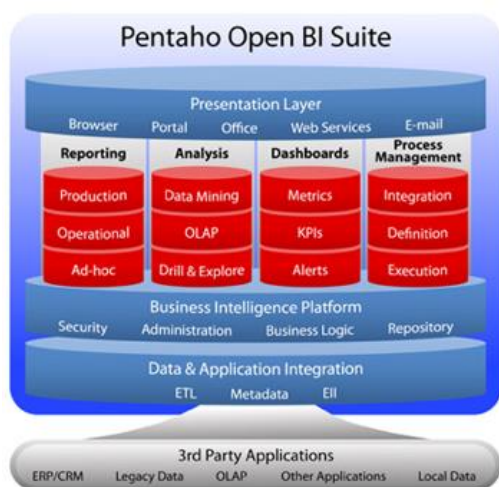


Ilustración 6. Pentaho Open BI Suite (<https://www.spec-india.com/blog/a-single-key-to-unlock-business-analytics-pentaho-bi/>)

### 4.2.1 Pentaho data integration

Pentaho ha sido desarrollado bajo la filosofía del software libre y proporciona potentes herramientas de explotación y visualización de datos para la gestión de procesos ETL.

**Pentaho Data Integration**, solución tecnológica conocida como **Kettle**, es una de las herramientas ETL opensource más potentes y versátiles, con el objetivo de construir y explotar un Data Warehouse, entre otras utilidades [12].

Uno de sus aspectos más ventajosos consiste en su capacidad de adaptación, lo que permite desarrollar fácilmente los procesos de integración de datos en el Data Warehouse.

Como herramienta ETL, Kettle es visual y consigue una mayor productividad del equipo de desarrollo y mantenimiento, permitiendo conectarse a diversas fuentes y arquitecturas para finalmente cargar los datos en un repositorio, si bien también puede utilizarse para otras muchas funciones.

Entre sus características más importantes se pueden destacar: conectividad, capacidad de entrega de datos, de metadatos y modelados de datos, de diseño y entorno de desarrollo, de gestión de datos de administración, además de capacidades SOA y un cierto grado de compactación, consistencia e interoperabilidad.

## 4.3 Selección de la herramienta de explotación de datos

Entre las herramientas que se han valorado para la explotación de los datos y desarrollo de informes se encuentra Pentaho Reporting y Microsoft Power BI.

Pentaho Reporting es el generador de informes incluido en la suite Pentaho. Se pueden generar informes tanto a nivel de cliente utilizando el asistente o el diseñador de informes o también se pueden generar los informes a través de la web.

Se han estudiado estas dos alternativas porque ya se está usando Pentaho para los procesos de integración ETL pero por su parte, Power BI, aunque se trate de una herramienta que no es Opensource, se encuentra ampliamente extendido su uso y se considera mucho menor el tiempo a invertir para su aprendizaje.

### 4.3.1 Microsoft Power BI

Microsoft Power BI es una colección de servicios de software, aplicaciones y conectores que funcionan conjuntamente para convertir orígenes de datos sin relación entre sí en información coherente, interactiva y atractiva visualmente.

Power BI consta del cliente *desktop*, donde se van a generar los informes; la parte **online** donde se consumen esos informes y se puede visualizar los gráficos, y la parte **mobile**, que cuenta con su propia app tanto para móvil como para tablet, en la que también se pueden ver los informes.

Un flujo de trabajo habitual de Power BI comienza en **Power BI Desktop**, donde se integran los datos y se crean informes. Estos informes se **publican** en el **servicio Power BI** y se **comparten** para que los usuarios puedan usar la información.

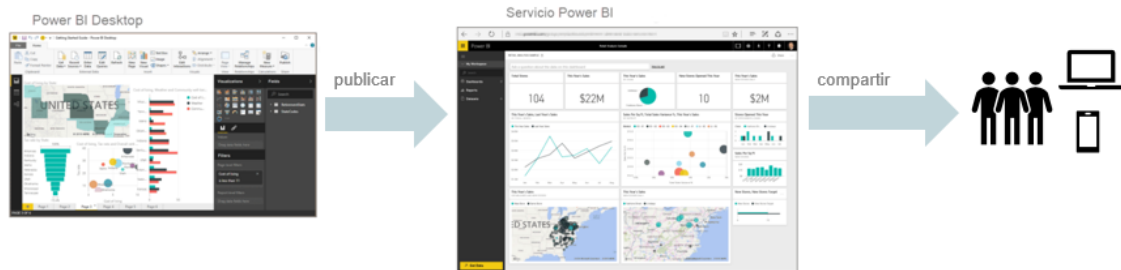


Ilustración 7. Flujo de trabajo Power BI

En este caso se va a utilizar como fuente de datos el Dataware House alimentado por el proceso ETL que vamos a construir a partir de las fuentes de datos facilitadas, pero cabe destacar que con Power BI podríamos añadir diferentes orígenes de datos. Lo cual puede resultar útil si queremos analizar nuevas fuentes de datos antes de incluirlas en nuestros procesos ETLs y Dataware House.

El resultado de los informes que desarrollemos en la aplicación de escritorio, Power BI Desktop se almacenarán en un archivo de tipo .pbix que podremos guardar y compartir con usuarios que tengan instalado esta aplicación. Además, estos informes se podrán publicar y compartir a través del servicio Power BI para poder visualizarlos on-line [15].

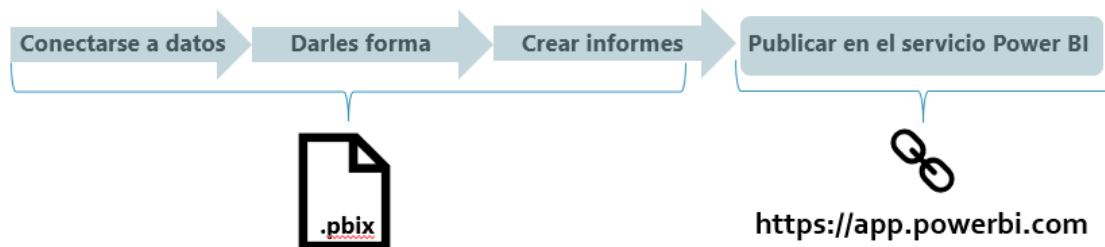


Ilustración 8. Ciclo de trabajo Power BI y productos resultantes

#### 4.4 Entorno tecnológico seleccionado

A modo de resumen se muestra una imagen con los componentes tecnológicos escogidos para el desarrollo del sistema de inteligencia de negocio:

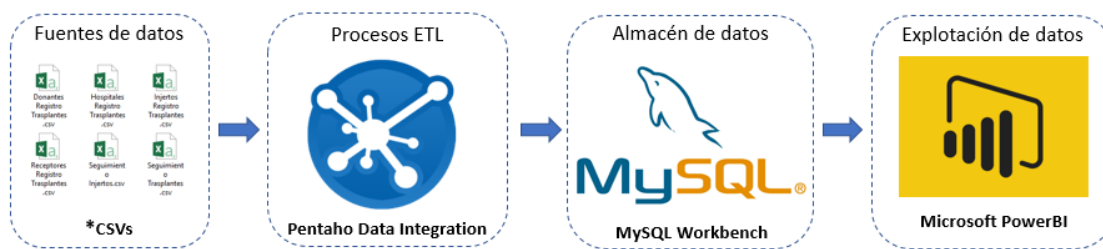


Ilustración 9. Entorno tecnológico seleccionado

## 5. Modelo de datos

En los siguientes apartados se describen los pasos para el diseño y construcción del almacén de datos.

### 5.1 Modelado dimensional

Siguiendo el método de Diseño Dimensional en 4 Etapas que se describió por primera vez en el año 1996 por Ralph Kimball como propuesta para el diseño de un Data Warehouse, se van a definir: El ámbito del proyecto, la granularidad, las dimensiones y los hechos [8].

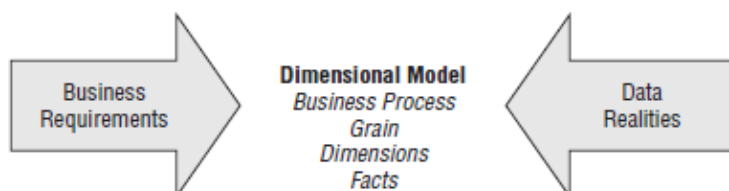


Figure 3-1: Key input to the four-step dimensional design process.

Ilustración 10. Entradas para el modelado dimensional en 4 pasos (The Data Warehouse Toolkit)

#### 5.1.1 Ámbito del proyecto

El ámbito del proyecto es, como ya se ha definido anteriormente, los resultados de los pacientes que reciben un trasplante de órgano sólido. El objetivo es tener un mejor conocimiento de los factores clave que ayuden a la toma de decisiones a la hora de asignar un determinado órgano donado al receptor más idóneo.



## 5.1.2 Granularidad de la información

A partir de los requerimientos de la información nos centraremos en el detalle del **estado del receptor del trasplante**, y analizaremos el estado del injerto/s trasplantado/s, teniendo en cuenta el tiempo transcurrido, las características del donante, el tipo de injerto y el hospital en el que se realiza la donación/trasplante.

Por otro lado, analizaremos el **estado del injerto trasplantado**, teniendo en cuenta también esos parámetros.

## 5.1.3 Dimensiones

Las perspectivas por las que se quiere analizar la información almacenada en los registros de la ONT son las siguientes:

- **Dimensión Tiempo:** Incluirá la fecha de la donación
- **Dimensión Donante:** Incluirá su edad, localización, hipertensión arterial (HTA), diabetes mellitus (DM), infección activa, antecedentes de cáncer.
- **Dimensión Receptor:** Incluirá su edad, localización, veces trasplantado (puede ser su primer trasplante o haber tenido trasplantes previos)
- **Dimensión Injerto:** Incluirá el tipo de órgano trasplantado, tiempo de isquemia (tiempo desde que cesa la circulación sanguínea en el donante hasta que se restablece en el receptor)
- **Dimensión Hospital Donación:** Localización, dispone o no de programa de trasplante
- **Dimensión Hospital Trasplante:** Localización, dispone o no de programa de trasplante

## 5.1.4 Hechos

Los hechos serán el **estado del receptor** y el **estado del injerto**. Las métricas de estos hechos son los indicadores del estado de salud del paciente y del estado del injerto trasplantado, es decir, los conceptos cuantificables para medir el **éxito del trasplante**.

Indicadores para el **estado del receptor** del injerto trasplantado son los siguientes:

- **Estado Vital de los receptores:**
  - Fallecido (0)
  - Vivo (1)
- **Estado de salud:**
  - Bueno (4)
  - Con mínima morbilidad no relacionada con el trasplante (3)
  - Con mínima morbilidad relacionada con el trasplante (2)
  - Morbilidad importante no relacionada con el trasplante (1)
  - Morbilidad importante relacionada con el trasplante (0)

(Morbilidad: proporción de seres vivos que se enferman en un sitio y tiempo determinado)

El principal indicador de la función del **injerto trasplantado** varía según el tipo de trasplante, pero para cada uno de ellos se puede asignar una variable de laboratorio o de prueba funcional o se puede clasificar la función del injerto en grados cualitativos (normalmente de 1 a 5), evaluados por el médico que realiza el seguimiento:

- **Estado funcional del injerto trasplantado:**
  - Función normal (4)
  - Mínimo deterioro funcional (3)
  - Deterioro funcional moderado (2)
  - Importante deterioro funcional (1)
  - Nula función del órgano trasplantado de forma irreversible (0)

De este análisis se construye el modelo conceptual preliminar del almacén de datos.

### 5.1.5 Modelo conceptual

Una vez identificados, los hechos, sus indicadores y sus dimensiones, se ha definido su esquema dimensional. En la siguiente imagen, se puede observar que el diagrama resultante es en estrella, ya que, para cada hecho, (estado del receptor y estado del injerto), existen un grupo de dimensiones, que para este contexto son comunes a ambos hechos.

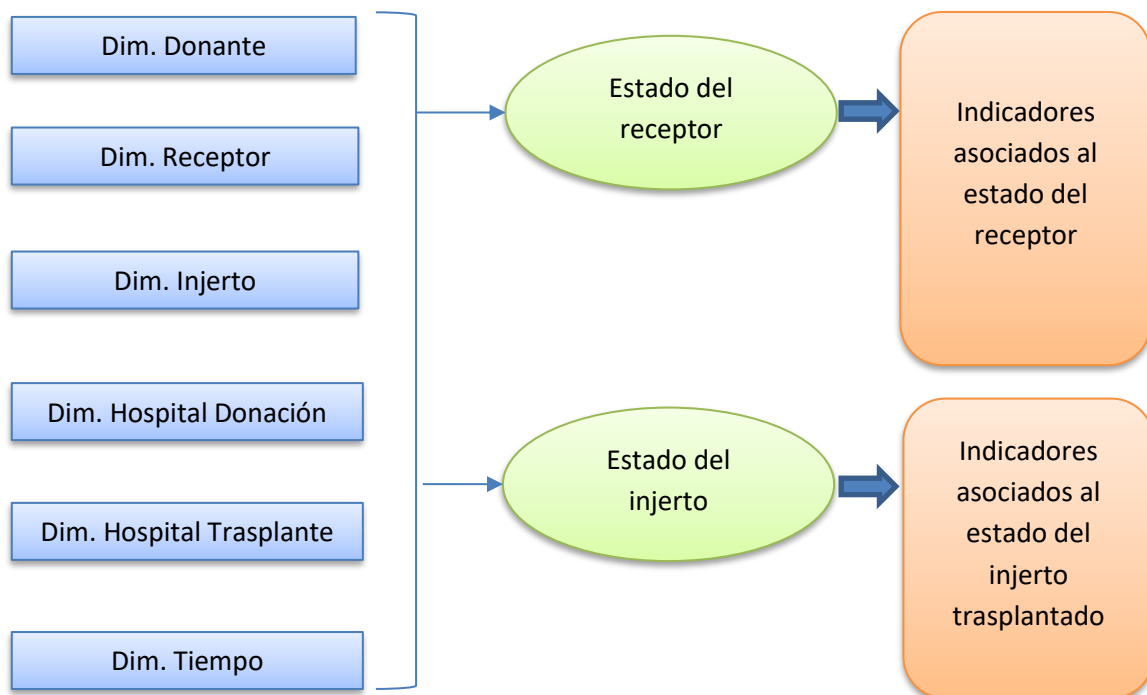
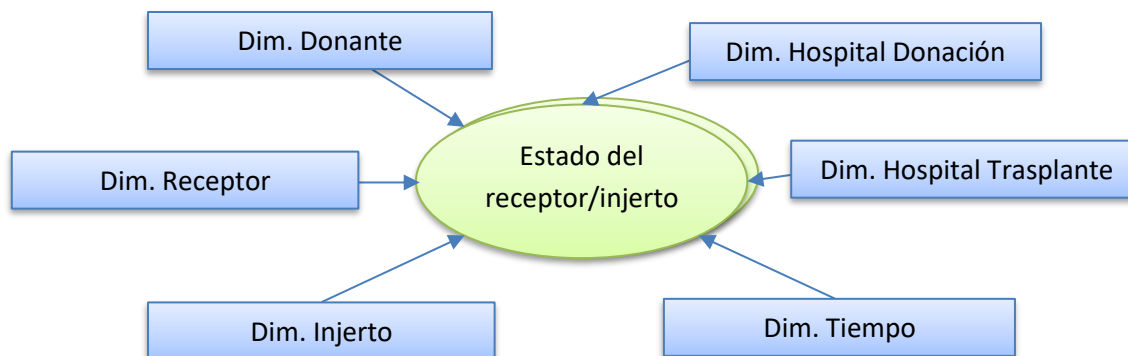


Ilustración 11. Modelo conceptual 1

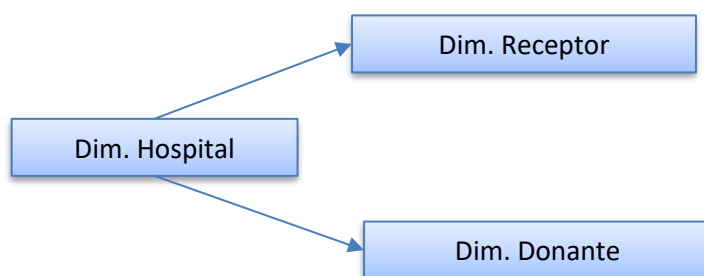
Aunque el esquema anterior, no sugiera una estructura en estrella, por simplificar la representación gráfica, en la siguiente imagen se observa como en la parte central tenemos la cada tabla de hechos y rodeándola, las tablas de dimensiones.



*Ilustración 12.. Modelo conceptual 2*

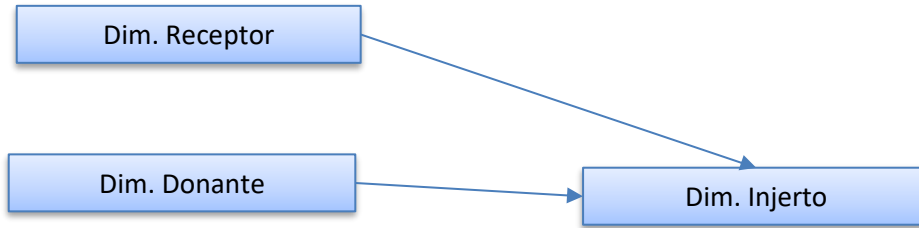
Al normalizar estas tablas, eliminando renunciarias y creando jerarquías, obtendremos un esquema en copo de nieve.

Las dos tablas de hospitales, **hospital trasplante** y **hospital donación**, se pueden simplificar en una sola tabla de hospitales, que, al relacionarse con **receptor**, se tratará del hospital donde se realiza el trasplante, y al relacionarse con **donante**, se tratará del hospital donde se prepara el órgano para la donación. Es decir, esta información se encontrará almacenada en las tablas de Receptor y de Donante.



*Ilustración 13. Modelo conceptual 3*

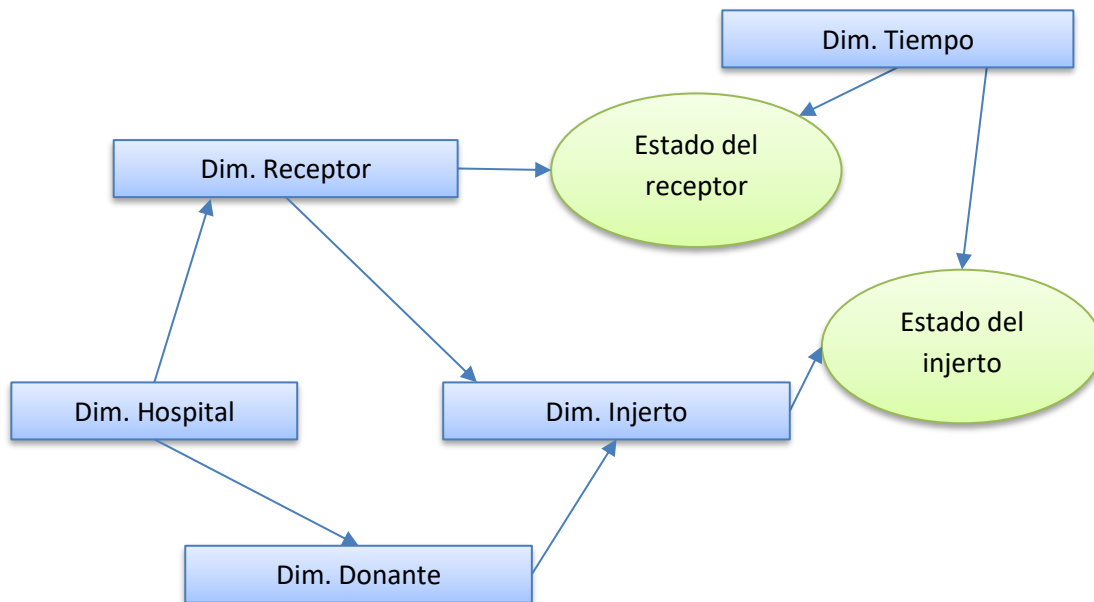
En la tabla de **injertos**, tendremos por cada injerto u órgano trasplantado, la información del **donante** y del **receptor**, por lo que esta tabla servirá de nexo para extraer esos datos. Para el estado del receptor, conoceremos el donante (o donantes) por su relación con injerto.



Il·lustració 14. Modelo conceptual 4

De esta forma, el receptor se relaciona con sus donantes por la tabla injerto.

El resultado de esta revisión del modelo inicial es el siguiente:



Il·lustració 15. Modelo conceptual 5

Como se puede observar, los datos necesarios para la tabla de hechos de **estado del receptor** se extraerán a partir de la tabla de la dimensión receptor y los datos necesarios para la tabla de hechos de **estado del injerto**, se extraerán a partir de la tabla de la dimensión injerto.

Con el modelo conceptual de los datos ya definido se diseña el modelo lógico de los datos, donde se indican las métricas de las tablas de hechos, sus demás atributos y dimensiones, así como los atributos de estas dimensiones.

### 5.1.6 Modelo lógico

En este apartado, se define el diseño lógico de datos, indicando las métricas de las tablas de hechos, sus demás atributos y dimensiones, así como los atributos de estas dimensiones. Las tablas de hechos, como se muestra a continuación, tienen una clave primaria, para identificar de forma única cada uno de los registros. Su relación con las tablas de dimensión se identifica con las claves externas y sus métricas, son los indicadores definidos anteriormente.

Tabla de hecho	Clave primaria	Claves externas	Métricas
<b>hch_estado_receptor</b>	id_estado_receptor	id_receptor	estado_vital
		id_tiempo	estado_salud

Tabla de hecho	Clave primaria	Claves externas	Métricas
<b>hch_estado_injerto</b>	id_estado_injerto	id_injerto	estado_funcional
		id_tiempo	

A continuación, se muestra el diseño lógico de cada una de las tablas de dimensión:

Tabla de dimensión	Clave primaria	Claves externas	Atributos
<b>dim_receptor</b>	id_receptor	id_hospital	fecha_nacimiento
			sexo
			diagnostico
			grupo_sanguineo
			num_trasplantes
			fecha_transplante
			CCAA

Tabla de dimensión	Clave primaria	Claves externas	Atributos
<b>dim_injerto</b>	id_injerto	id_receptor	tiempo_isquemia
		id_donante	
		id_tipo	

Tabla de dimensión	Clave primaria	Claves externas	Atributos
<b>dim_donante</b>	id_donante	id_hospital	fecha_nacimiento
			sexo
			grupo_sanguineo
			talla
			peso
			tipo_donante
			hipertension
			diabetes
			hipotension
			VIH
			VHC
			alcoholismo
			tabaquismo
			causa_muerte
			infeccion_activa
			ant_cancer
			fecha_donacion
			CCAA

Tabla de dimensión	Clave primaria	Atributos
<b>dim_tiempo</b>	id_tiempo	dd
		mm
		aa
		quarter

Tabla de dimensión	Clave primaria	Atributos
<b>dim_hospital</b>	id_hospital	CCAA

A partir de las tablas de hechos y las tablas de dimensiones, se obtiene el siguiente modelo lógico de datos:

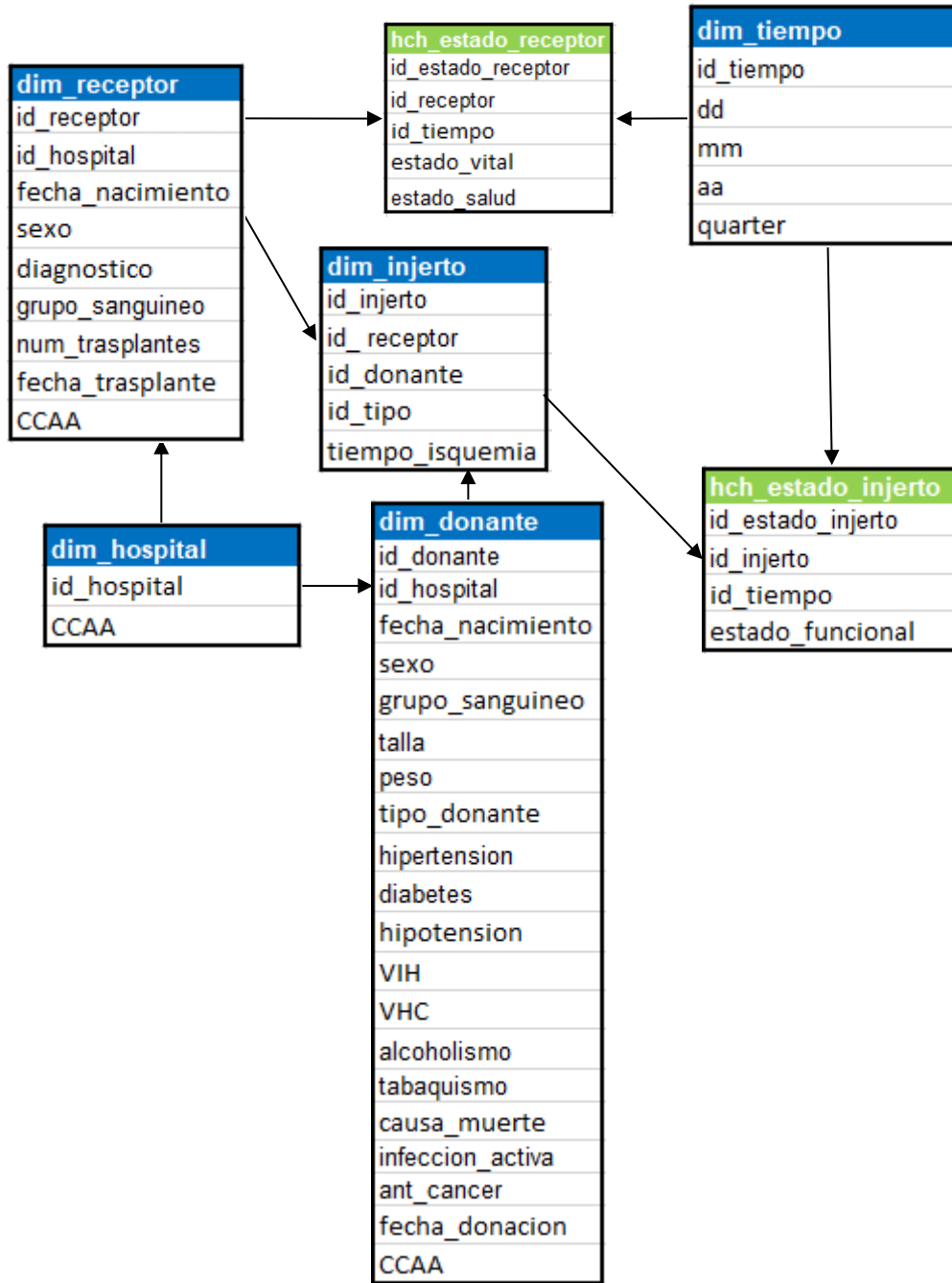


Ilustración 16. Modelo lógico

## 5.2 Modelo físico

Basándonos en el modelo lógico, se define el modelo físico de los datos con la representación de los objetos datos relacionales (tablas, columnas, claves principales y claves externas) y sus relaciones.

A continuación, se muestra el diseño del modelo físico que se ha diseñado con la herramienta MySQL Workbench:

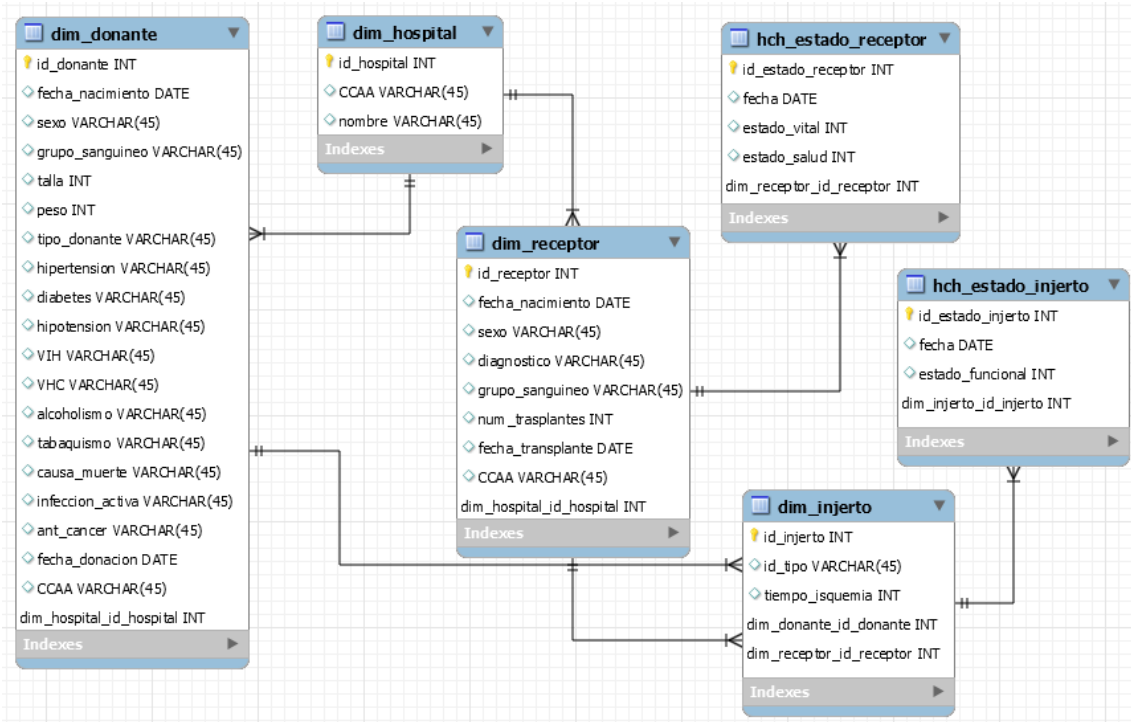


Ilustración 17. Modelo físico

## 6. Implementación del Data warehouse

Una vez creado el modelo físico en el diseñador de MySQL, se dispone de la estructura de datos para implementar el almacén de datos.

Se ha creado con la herramienta MySQL Workbench, tabla por tabla para luego relacionarlas entre sí:

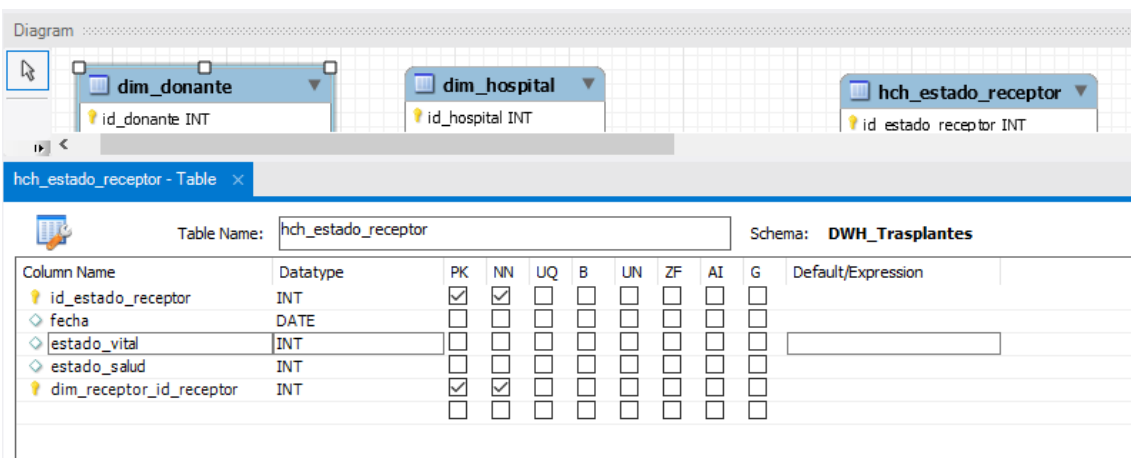


Ilustración 18. Diseño del modelo físico en MySQL

Las claves foráneas se crean automáticamente al relacionar las tablas.

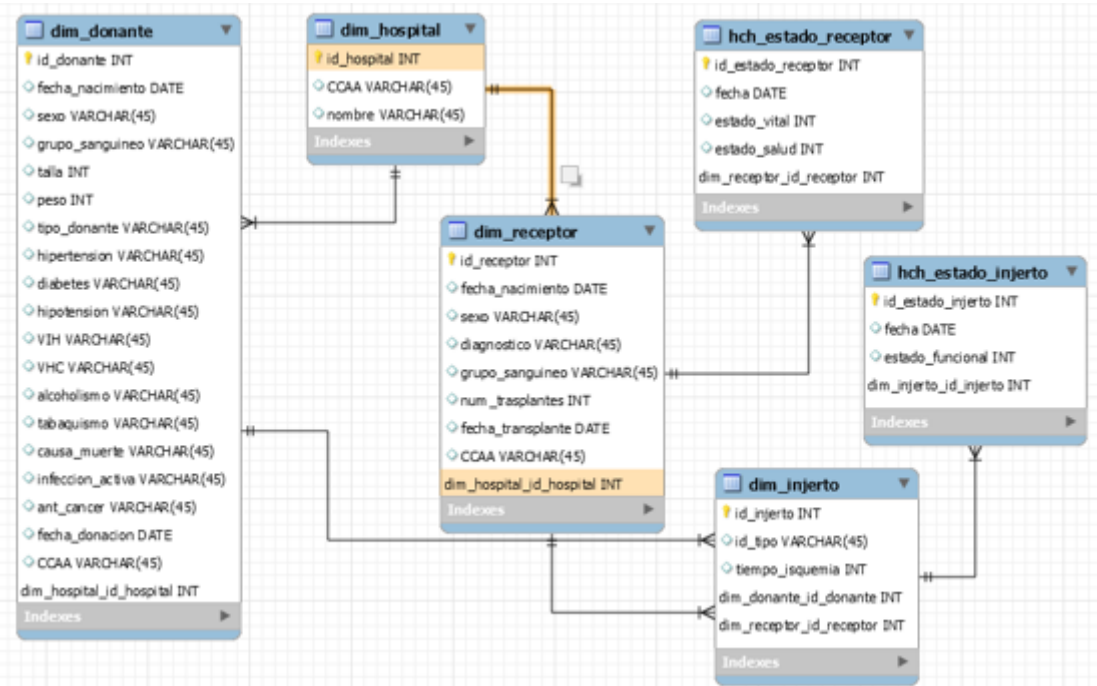


Ilustración 19. Relaciones en el diseño físico en MySQL

Esta estructura nos permite crear automáticamente la base de datos con sus tablas y relaciones, así como actualizar con modificaciones en caso de que sea necesario.

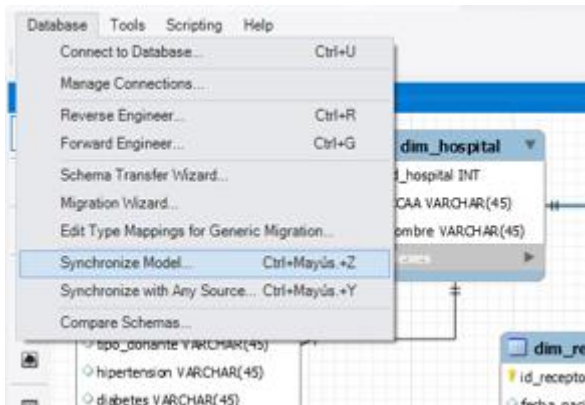


Ilustración 20. Sincronización del modelo

La base de datos que se ha nombrado como **dwh\_trasplantes** y tiene este aspecto dentro de la herramienta de MySQL:





Ilustración 21. Estructura de tablas en MySQL

A continuación se muestra imagen del despliegue de columnas de cada tabla:

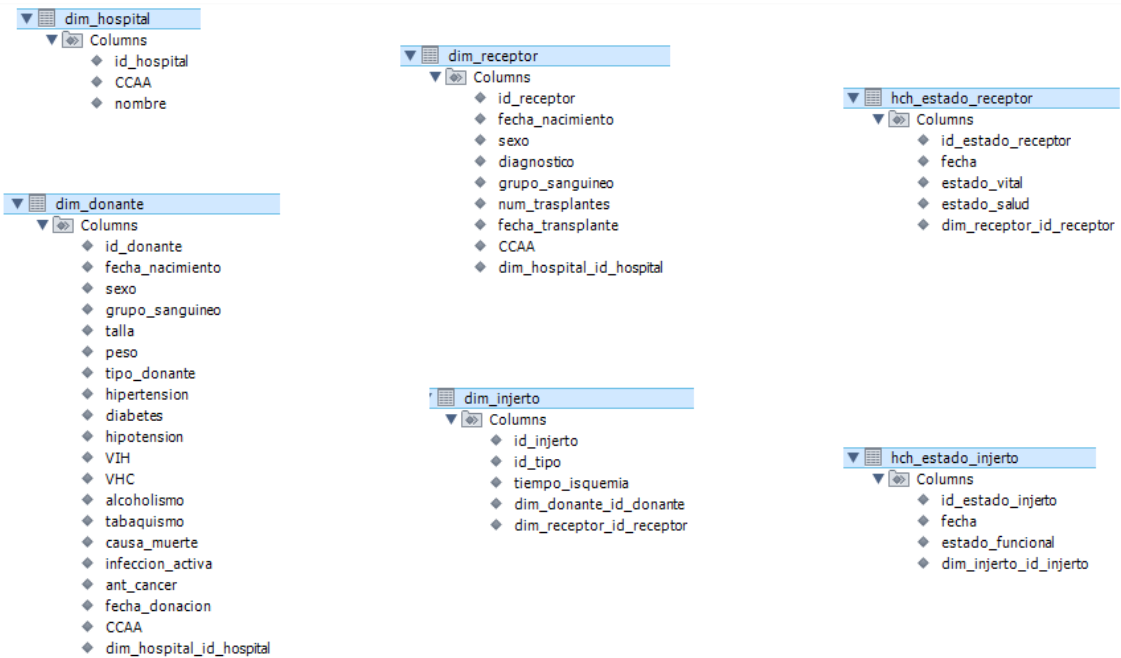


Ilustración 22. Despliegue de columnas de cada tabla

Estas tablas se crean en MySQL vacías. En la implementación de los procesos ETL se alimentarán a partir de las fuentes de datos proporcionadas.

## 7. Diseño y desarrollo de los procesos ETL

Una vez implementada la estructura del almacén de datos, el siguiente paso es la implementación de los procesos ETL para la extracción, transformación y carga de los datos en el Data Warehouse.

Cada proceso ETL toma los datos de diferentes fuentes de información, los procesa y, si es necesario los transforma para poder cargarlos en el almacén de datos.

Los datos facilitados para poder realizar este proceso se encontraban en diferentes hojas dentro de un Excel y lo primero que se ha realizado ha sido estudiar esa información y separarla en diferentes archivos .csv para su fácil proceso por el integrador de Pentaho.

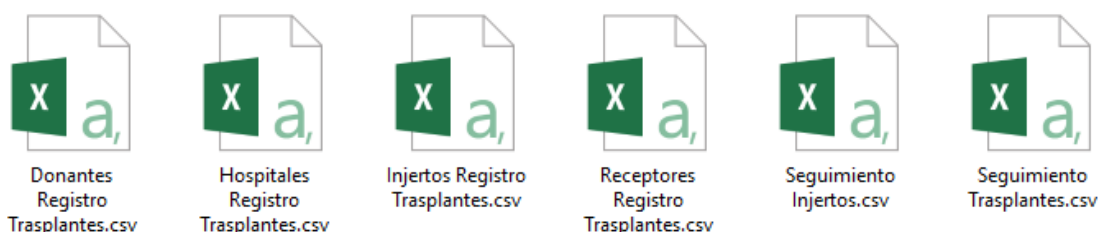


Ilustración 23. Fuentes de datos

A continuación, se va a explicar cada proceso de carga de los datos. Hay que tener en cuenta que este proceso se ha de realizar en un orden determinado para poder cargar las claves foráneas sin errores. Por lo tanto, en primer lugar, se cargará la tabla de **Hospitales**, a continuación, las tablas de **Donantes** y **Receptores** (en este caso el orden no es relevante), tras esto, la tabla de **Injertos** (ya que tiene relación con Donantes y Receptores) y por último las tablas de hechos: **Estado del receptor** y **Estado del injerto** sin importar en este caso el orden de carga ya que se dispondrá de los datos relacionados ya cargados previamente.

### 7.1 ETL para la carga de Hospitales

La fuente de datos e input de hospitales se encuentra en un fichero .csv y presenta las siguientes columnas:

HOSPITAL_DONACION_ID	HOSPITAL_DONACION	CCAA_HOSPITAL_DON
11111	H. SON DURETA	ILLES BALEARS
11112	H. VERGE DEL TORO	ILLES BALEARS
11113	H. MATERNO INFANTIL TERESA HERRERA	GALICIA
11114	C. H. DE OURENSE	GALICIA
11115	H. CLÍNIC BARCELONA	CATALUÑA
11116	CHUAC (H. JUAN CANALEJO)	GALICIA
11117	H. REINA SOFÍA	ANDALUCÍA
11118	H. UNIV. LA PAZ INFANTIL	MADRID

Ilustración 24. Fuente CSV para Hospitales

El proceso de transformación es sencillo, como se puede observar.



Ilustración 25. Proceso de transformación para Hospitales

Un input que es el fichero CSV y una salida a la tabla correspondiente de la base de datos. No han sido necesarias transformaciones o cálculos para los datos.

En la salida, se ha establecido la conexión con la base de datos y la tabla de destino. También se ha realizado el mapeo de los campos de entrada con los campos de salida como se puede comprobar en la siguiente imagen:

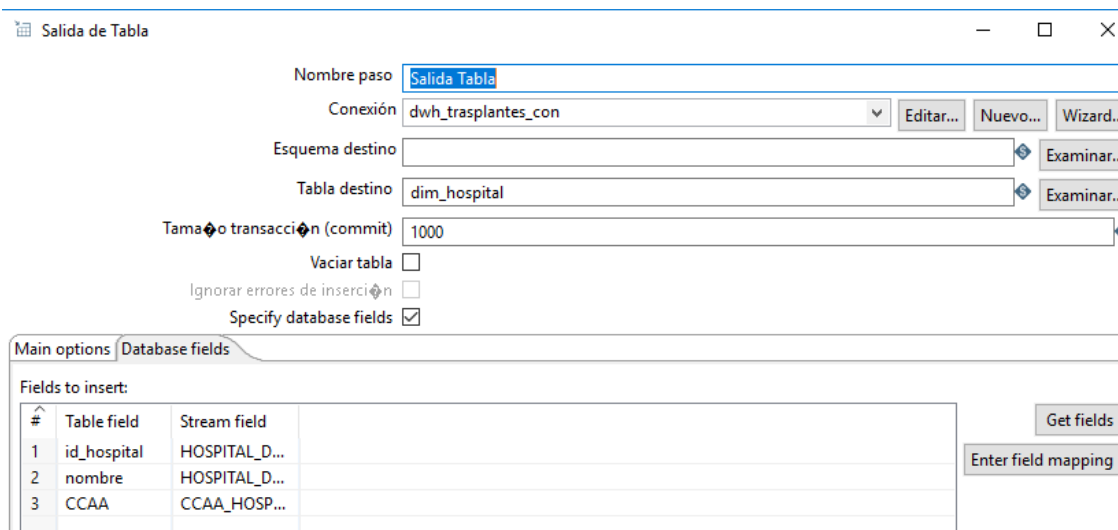


Ilustración 26. Salida a Tabla para Hospitales

Una vez que se establecieron estos parámetros se ejecutó la transformación:

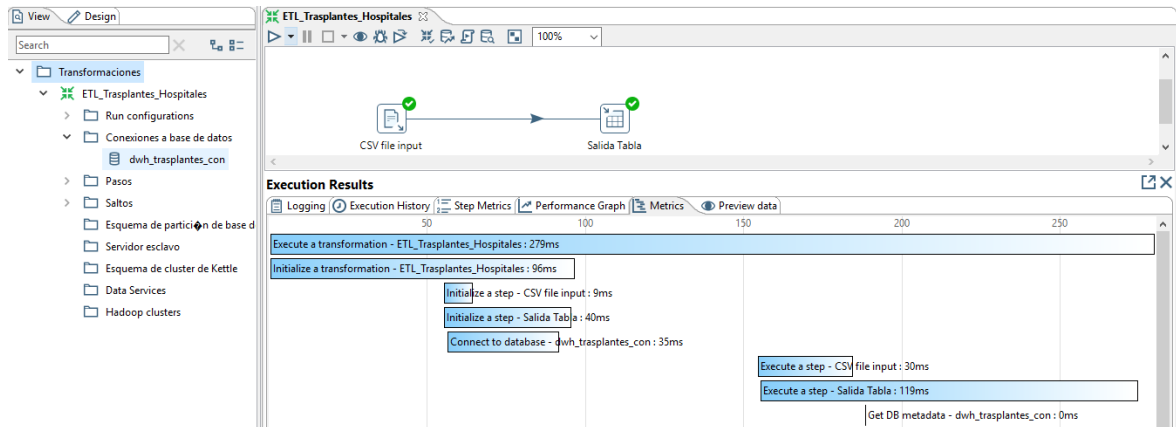


Ilustración 27. Ejecución de transformación para Hospitales

Y tras esto se pudo comprobar la carga de los datos en MySQL:

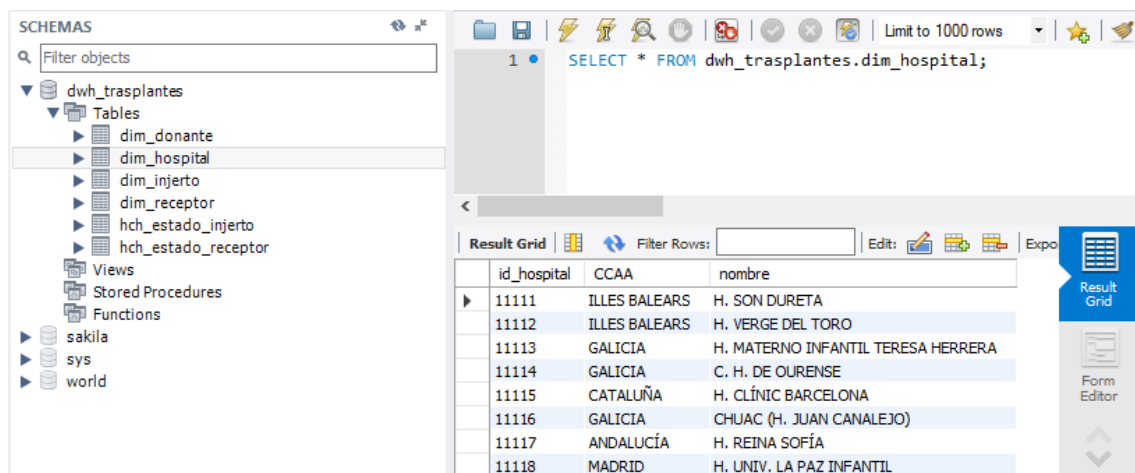


Ilustración 28. Comprobación en MySQL para Hospitales

## 7.2 ETL para la carga de Donantes

Para la carga de donantes tenemos un fichero .csv como fuente de datos:

FX_NACIMIENTO_DON	SEXO_DON	GS_DON	PESO_DON	TALLA_DON	TIPO_DONANTE	HIPERTENSIO	DIABETES	HIPOTENSIO	HIV	VHC	ALCOHOLIS	TABAQUISM	CAUSA_MUE	INFECT_ACT	AN
27/06/1971	Hombre	A		73	170 Muerte Cerebral	No	No	Si	Negativo	Negativo	No	Nunca	Tce Tráfico		Ne
13/03/1956	Hombre	A		75	180 Muerte Cerebral	No	No	No	Negativo	Negativo	No	Fumador act	Tce No Tráfico		
05/01/1992	Hombre	O		40	160 Muerte Cerebral	No	No	Si	Negativo	Negativo	No	Nunca	ACVA		
17/03/1982	Mujer	A		55	173 Muerte Cerebral	No	No	Si	Negativo	Negativo	No	Nunca	Encefalopatía Anóxica		
03/07/1987	Mujer	A		56	170 Muerte Cerebral	No	No	Si	Negativo	Negativo	No	Nunca	Tce Tráfico		Ne
03/03/1982	Hombre	A		66	162 Muerte Cerebral	No	No	Si	Negativo	Negativo	No	Nunca	Tce Tráfico		Po
20/03/1968	Mujer	AB		53	155 Muerte Cerebral	No	No	No	Negativo	Negativo	No	Fumador act	ACVA		
27/02/1974	Mujer	A		50	161 Muerte Cerebral	No	No	Si	Negativo	Negativo	No	Fumador act	ACVA		Ne
04/03/1992	Hombre	A		55	160 Muerte Cerebral	No	No	No	Negativo	Negativo	No	Nunca	Tce Tráfico		Ne
25/03/1968	Mujer	O		51	166 Muerte Cerebral	Si	No	Si	Negativo	Negativo	No	Nunca	ACVA		Ne
07/06/1977	Mujer	A		62	155 Muerte Cerebral	No	No	No	Negativo	Negativo	No	Nunca	Tce No Tráfico		
05/04/1978	Hombre	O		83	176 Muerte Cerebral	No	No	Si	Negativo	Negativo	No	Nunca	Tce No Tráfico	Negativo	Ne
26/03/1981	Hombre	O		86	184 Muerte Cerebral	No	No		Negativo	Negativo	No	Nunca	Tce Tráfico		
30/04/1967	Mujer	A		66	178 Muerte Cerebral	No	No	No	Negativo	Negativo	Moderado (F	Fumador act	Tce Tráfico		Ne

Ilustración 29. Fuente de datos CSV para Donantes

Como se puede observar en la fuente de datos y en la previsualización del input del .csv, existen valores a nulo que vamos a transformar para darles un valor en cada caso:

ION	DIABETES	HIPOTENSION	HIV	VHC	ALCOHOLISMO	TABAQUISMO	CAUSA_MUERTE	INFECT_ACT	ANT_CAN	FX
	No	Si	Negativo	Negativo	No	Nunca	Tce Tráfico	<null>	Negativo	27,
	No	No	Negativo	Negativo	No	Fumador actual	Tce No Tráfico	<null>	<null>	13,
	No	Si	Negativo	Negativo	No	Nunca	ACVA	<null>	<null>	05,
	No	Si	Negativo	Negativo	No	Nunca	Encefalopatía Anóxica	<null>	<null>	17,
	No	Si	Negativo	Negativo	No	Nunca	Tce Tráfico	<null>	Negativo	03,
	No	Si	Negativo	Negativo	No	Nunca	Tce Tráfico	<null>	Positivo	03,
	No	No	Negativo	Negativo	No	Fumador actual	ACVA	<null>	<null>	20,
	No	Si	Negativo	Negativo	No	Fumador actual	ACVA	<null>	Negativo	27,
	No	No	Negativo	Negativo	No	Nunca	Tce Tráfico	<null>	Negativo	04,
	No	Si	Negativo	Negativo	No	Nunca	ACVA	<null>	Negativo	25,
	No	No	Negativo	Negativo	No	Nunca	Tce No Tráfico	<null>	<null>	07,
	No	Si	Negativo	Negativo	No	Nunca	Tce No Tráfico	Negativo	Negativo	05,
	No	<null>	Negativo	Negativo	No	Nunca	Tce Tráfico	<null>	<null>	26,

Ilustración 30. Previsualización del input para Donantes

Para ello existe un paso intermedio en el ETL:

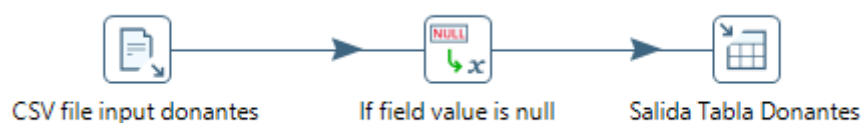


Ilustración 31. Proceso ETL para Donantes

Que se ha configurado como se muestra a continuación:

#	Field	Replace by value
1	HIPERTENSION	No
2	HIPOTENSION	No
3	DIABETES	No
4	HIV	Negativo
5	VHC	Negativo
6	TABAQUISMO	Nunca
7	ALCOHOLISMO	No
8	INFECC_ACT	Negativo
9	ANT_CAN	Negativo

Ilustración 32. Configuración de transformación para Donantes

Tras esto se puede configurar la salida con la tabla correspondiente de la base de datos y configurar el mapeo de los campos:

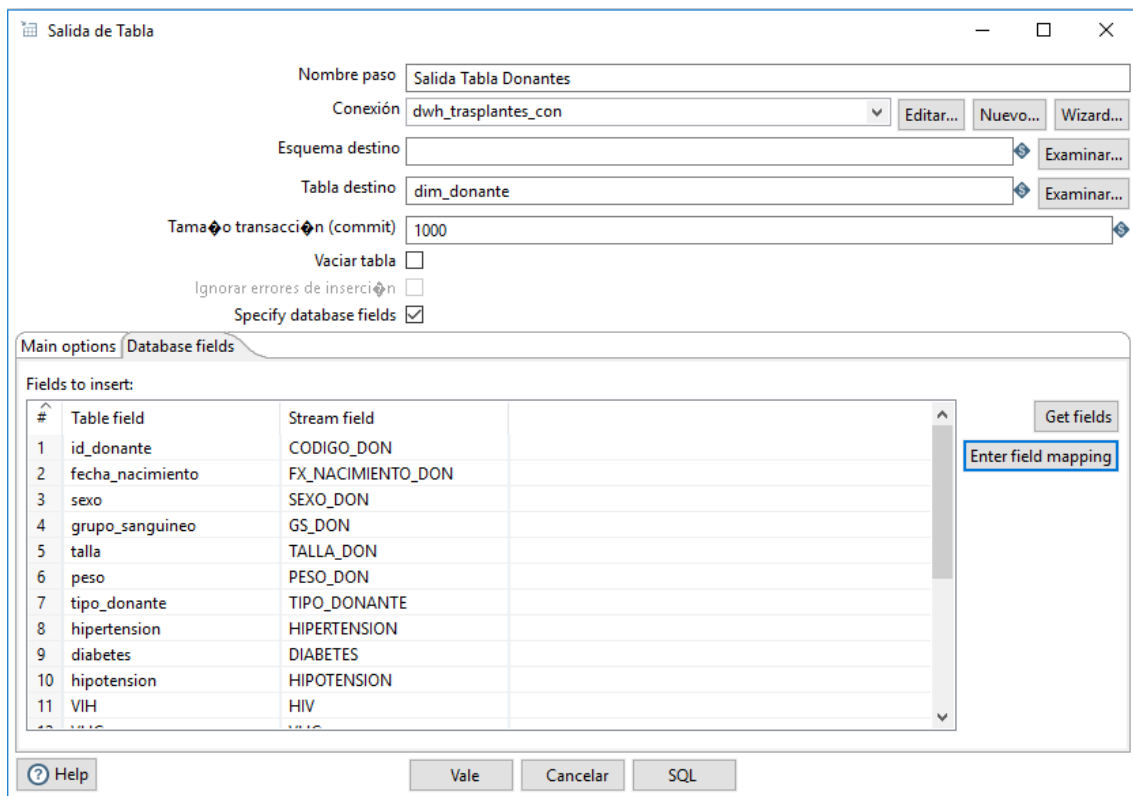


Ilustración 33. Configuración de salida para Donantes

Tras la ejecución del ETL, como se puede observar en la siguiente imagen:

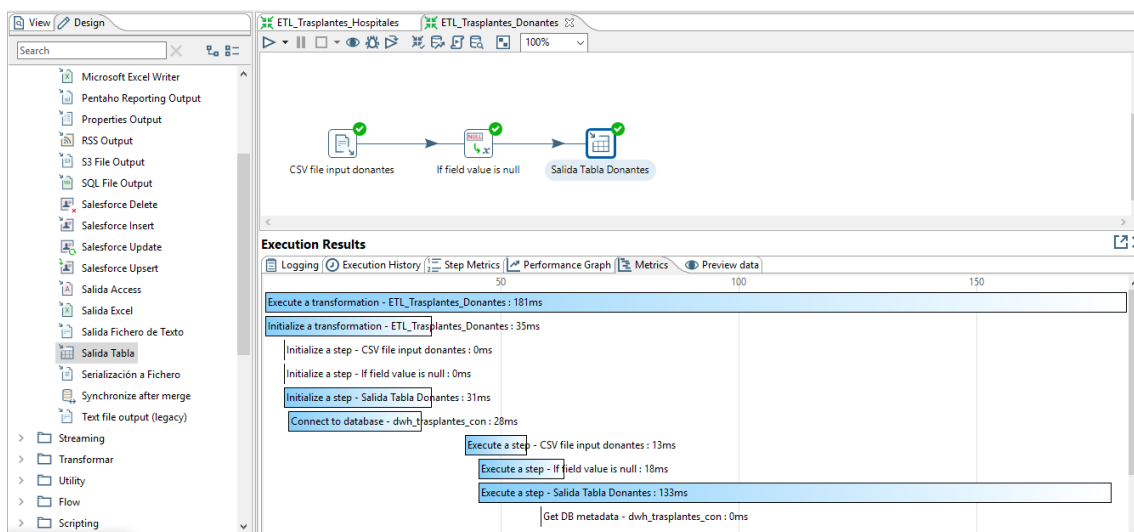


Ilustración 34. Ejecución de ETL para Donantes

Comprobamos que los datos se han cargado correctamente en el almacén de datos:

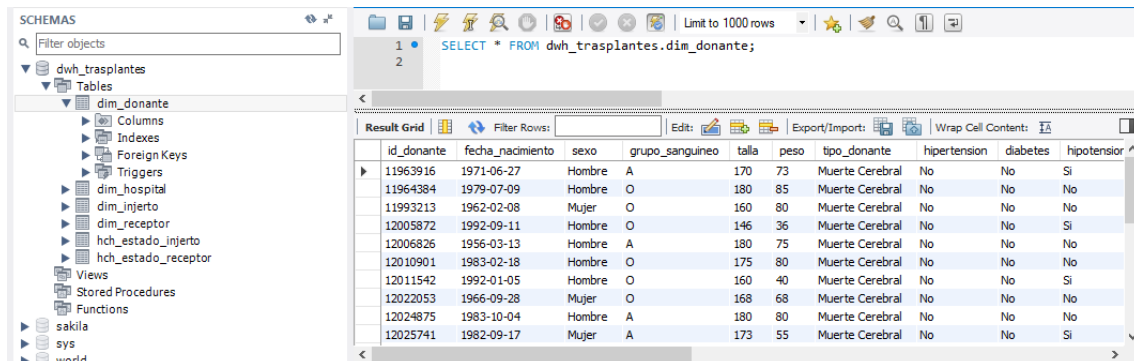


Ilustración 35. Comprobación de carga de Donantes

### 7.3 ETL para la carga de Receptores

El proceso de transformación y carga continua con los receptores de los trasplantes. En la siguiente imagen se puede observar la fuente de datos:

CODIGO_REC	CENTRO_TX	CENTRO_REC	EDAD_REC	SEXO_REC	DIAGNOSTICO_REC	GS_REC	NUM_TRAS	FX_TRASPLA	CCAA_RECEPTOR
12249391	11115	H. CLÍNIC BA	37	Hombre	Insuficiencia renal crónica + Diabetes	A	1	28/06/2000	ILLES BALEARS
12324486	11115	H. CLÍNIC BA	27	Mujer	Insuficiencia renal crónica + Diabetes	A	1	13/03/2002	COMUNIDAD VALENCIANA
12278920	11116	CHUAC (H. JU	56	Hombre	Insuficiencia renal crónica + Diabetes	O	1	06/03/2002	GALICIA
12425146	11115	H. CLÍNIC BA	37	Hombre	Insuficiencia renal crónica + Diabetes	A	1	18/09/2004	CATALUÑA
12287202	11115	H. CLÍNIC BA	14	Hombre	Insuficiencia renal crónica + Diabetes	A	1	04/07/2003	ILLES BALEARS
12288635	11115	H. CLÍNIC BA	36	Mujer	Insuficiencia renal crónica + Diabetes	A	1	04/02/2003	ARAGÓN
12375962	11115	H. CLÍNIC BA	33	Hombre	Retrasplante Crónico PA-R	AB	1	21/03/2003	CATALUÑA
12430945	11115	H. CLÍNIC BA	46	Mujer	Insuficiencia renal crónica + Diabetes	A	1	28/02/2006	COMUNIDAD VALENCIANA
12494656	11115	H. CLÍNIC BA	35	Mujer	Insuficiencia renal crónica + Diabetes	A	1	05/03/2006	ARAGÓN
12418286	11115	H. CLÍNIC BA	32	Mujer	ReTxPAN	O	1	26/03/2008	EXTREMADURA
12502673	11115	H. CLÍNIC BA	47	Mujer	ReTxPAN	A	1	08/06/2008	PAÍS VASCO
12559981	11115	H. CLÍNIC BA	40	Hombre	Insuficiencia renal crónica + Diabetes	O	1	06/04/2009	CATALUÑA
12530883	11117	H. REINA SOF	36	Hombre	Insuficiencia renal crónica + Diabetes	O	1	27/03/2008	ANDALUCÍA
12624230	11115	H. CLÍNIC BA	46	Mujer	Insuficiencia renal crónica + Diabetes	A	1	01/05/2009	CATALUÑA
12612341	11115	H. CLÍNIC BA	44	Hombre	Insuficiencia renal crónica + Diabetes	A	1	17/01/2009	CATALUÑA
12574773	11115	H. CLÍNIC BA	23	Hombre	Insuficiencia renal crónica + Diabetes	A	1	25/02/2009	CATALUÑA
12669354	11118	H. UNIV. LA F	0	Mujer	Sind. Intestino Corto	A	1	27/09/2010	ARAGÓN
12635290	11115	H. CLÍNIC BA	44	Mujer	ReTxPAN	A	1	09/09/2010	CATALUÑA
12267565	11117	H. REINA SOF	38	Mujer	Insuficiencia renal crónica + Diabetes	O	1	10/07/2000	ANDALUCÍA
12299651	11119	CHUAC (H. JU	43	Hombre	Insuficiencia renal crónica + Diabetes	O	1	09/02/2001	GALICIA
12321620	11119	CHUAC (H. JU	33	Mujer	Insuficiencia renal crónica + Diabetes	O	1	11/09/2001	GALICIA

Ilustración 36. Fuente de datos CSV para Receptores

Necesitamos la fecha de nacimiento, pero tenemos la edad con la que fue trasplantado, para ello se decide finalmente, realizar la siguiente transformación:

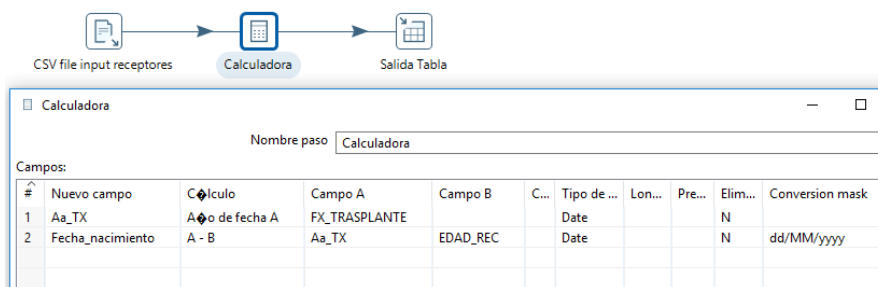


Ilustración 37. Transformación para Receptores

Tras el mapeo de campos con en pasos anteriores, se lleva a cabo la ejecución del ETL:

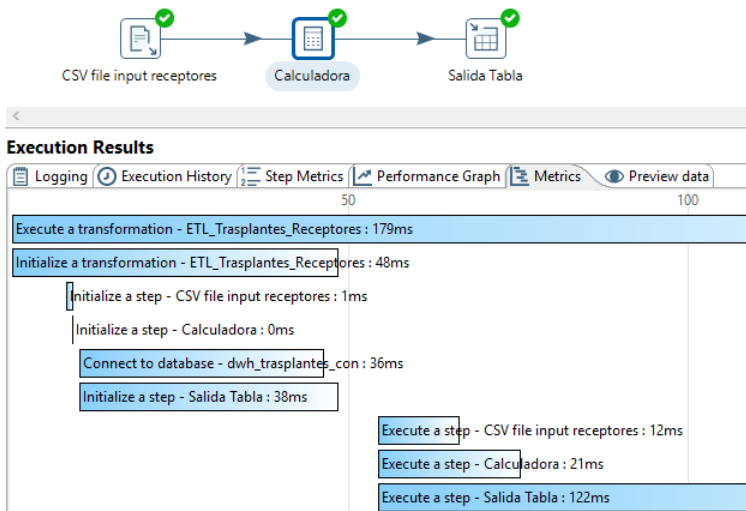


Ilustración 38. Ejecución ETL de Receptores



Y la comprobación de los datos en el data warehouse:

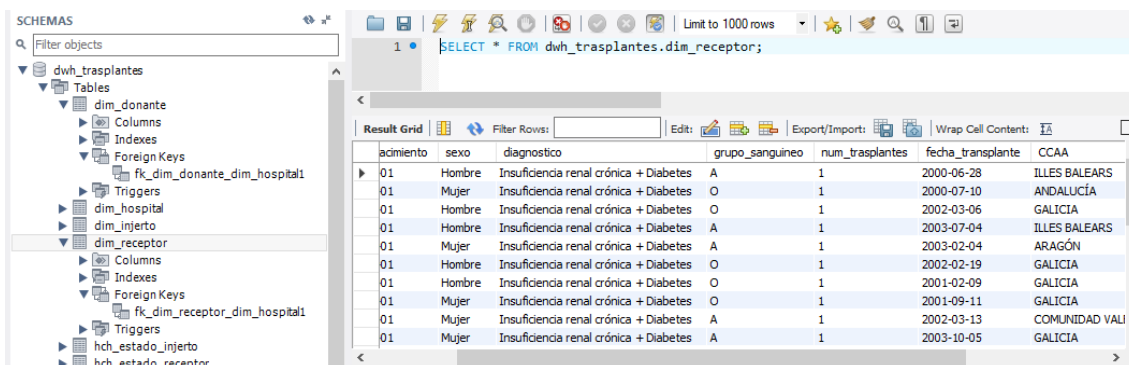


Ilustración 39. Comprobación carga de Receptores

## 7.4 ETL para la carga de Injertos

Como en los ETLs anteriores, se muestra a continuación la fuente de datos e input del proceso:

CODIGO_INJERTO	CODIGO_RECEP	CODIGO_DO	ORGANO	ISQUEMIA
1111111	12249391	11963916	Pancreas	60
1111112	12324486	12006826	Pancreas	120
1111113	12278920	12011542	Pancreas	90
1111114	12425146	12025741	Pancreas	60
1111115	12287202	12033872	Pancreas	120
1111116	12288635	12043591	Pancreas	90
1111117	12375962	12058195	Pancreas	60
1111118	12430945	12081100	Pancreas	120
1111119	12494656	12081144	Pancreas	90
1111120	12418286	12102402	Pancreas	60
1111121	12502673	12104933	Pancreas	120
1111122	12559981	12118546	Pancreas	90
1111123	12530883	12131905	Pancreas	60
1111124	12624230	12135346	Pancreas	120
1111125	12612341	12135545	Pancreas	90
1111126	12574773	12140725	Pancreas	60
1111127	12669354	12144730	Estómago	120
1111128	12635290	12156770	Pancreas	90
1111129	12267565	11964384	Estómago	60
1111130	12299651	11993213	Hígado	120
1111131	12321620	12005872	Intestino	90
1111132	12293200	12010901	Estómago	30
1111133	12353002	12022053	Pancreas	60
1111134	12348920	12024875	Pancreas	30
1111135	12443696	12049751	Pancreas	30
1111136	12249391	11963916	Riñón Izquie	60

Ilustración 40. Fuente de datos CSV para Injertos

En este caso no es necesaria ningún paso intermedio para la transformación de los datos, quedando la ejecución de la siguiente forma:

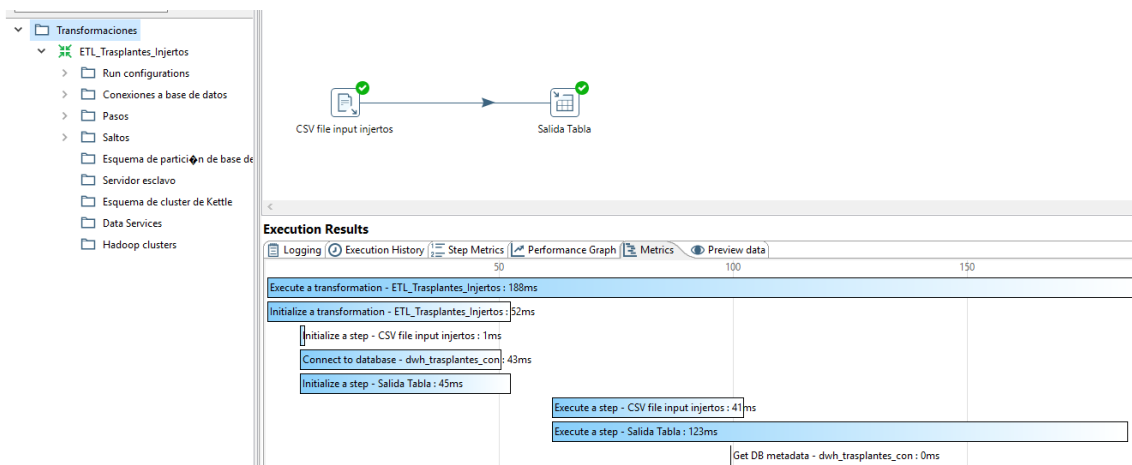


Ilustración 41. Ejecución de ETL para Injertos

Finalmente, se comprueba la carga de los datos en MySQL:

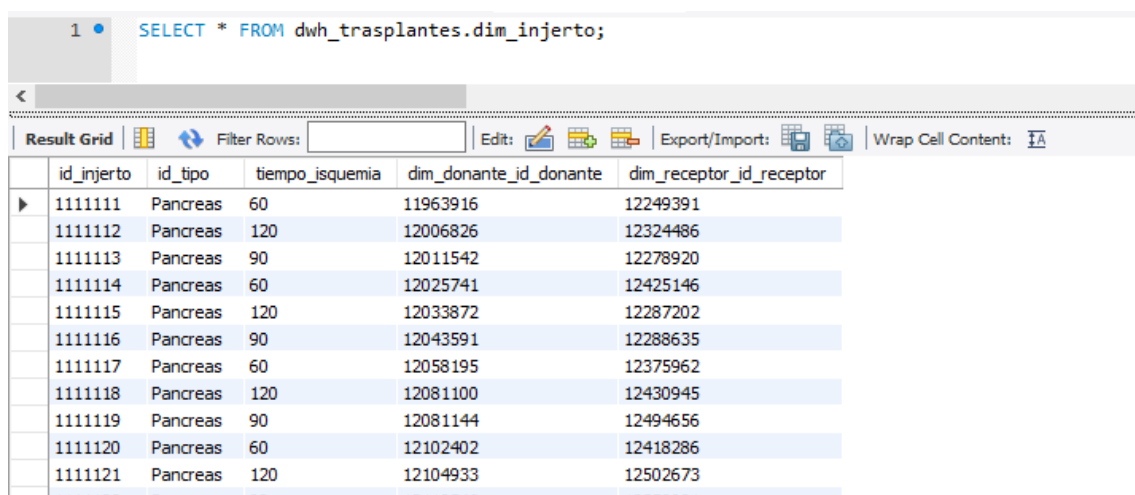


Ilustración 42. Comprobación carga de Injertos

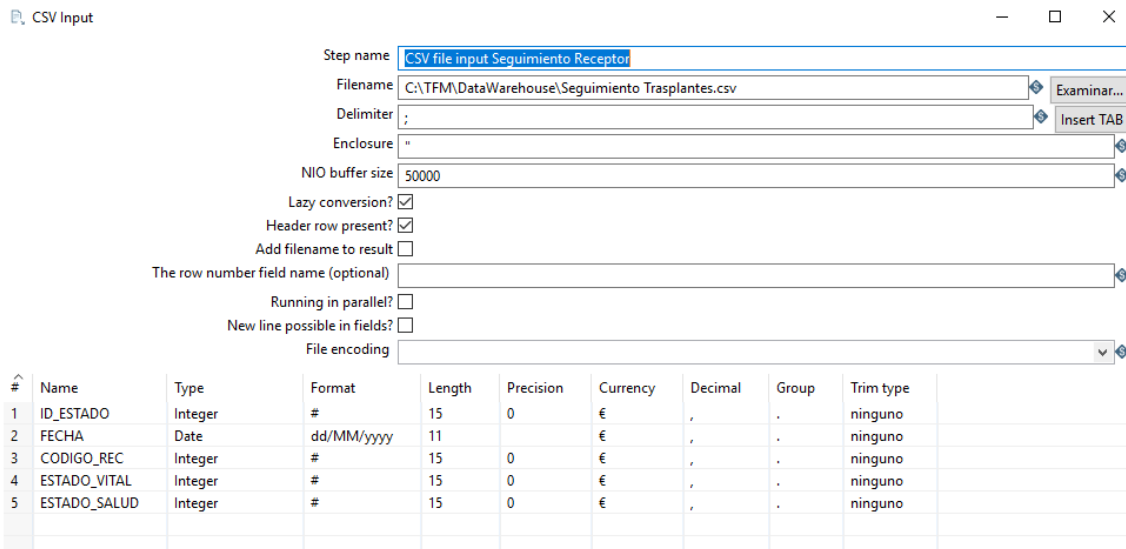
## 7.5 ETL para la carga de Estado del receptor

Una vez cargados los datos previos se puede llevar a cabo la carga de los datos de seguimiento del receptor

ID_ESTADO	FECHA	CODIGO_REC	ESTADO_VITAL	ESTADO_SALUD
1111111	28/07/2000	12249391	1	3
1111112	12/04/2002	12324486	1	3
1111113	05/04/2002	12278920	1	3
1111114	18/10/2004	12425146	1	3
1111115	03/08/2003	12287202	1	3
1111116	06/03/2003	12288635	1	2
1111117	20/04/2003	12375962	1	2
1111118	30/03/2006	12430945	1	2
1111119	04/04/2006	12494656	1	2
1111120	25/04/2008	12418286	1	2
1111121	08/07/2008	12502673	1	2
1111122	06/05/2009	12559981	1	2
1111123	26/04/2008	12530883	1	2
1111124	31/05/2009	12624230	1	2
1111125	16/02/2009	12612341	1	2
1111126	27/03/2009	12574773	1	2
1111127	27/10/2010	12669354	1	2
1111128	09/10/2010	12635290	1	2
1111129	09/08/2000	12267565	1	2
1111130	11/03/2001	12299651	1	2

Ilustración 43. Fuente de datos CSV para Estado del Receptor

El input, como en los pasos anteriores se trata de un archivo .CSV:



Step name: CSV file input Seguimiento Receptor

Filename: C:\TFM\DataWarehouse\Seguimiento Trasplantes.csv

Delimiter: ;

Enclosure: "

NIO buffer size: 50000

Lazy conversion?

Header row present?

Add filename to result?

The row number field name (optional):

Running in parallel?

New line possible in fields?

File encoding:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	ID_ESTADO	Integer	#	15	0	€	,	.	ninguno
2	FECHA	Date	dd/MM/yyyy	11		€	,	.	ninguno
3	CODIGO_REC	Integer	#	15	0	€	,	.	ninguno
4	ESTADO_VITAL	Integer	#	15	0	€	,	.	ninguno
5	ESTADO_SALUD	Integer	#	15	0	€	,	.	ninguno

Ilustración 44. Input para Estado del Receptor

La salida se configura con la conexión a la tabla del almacén de datos y se realiza el mapeo de los campos:

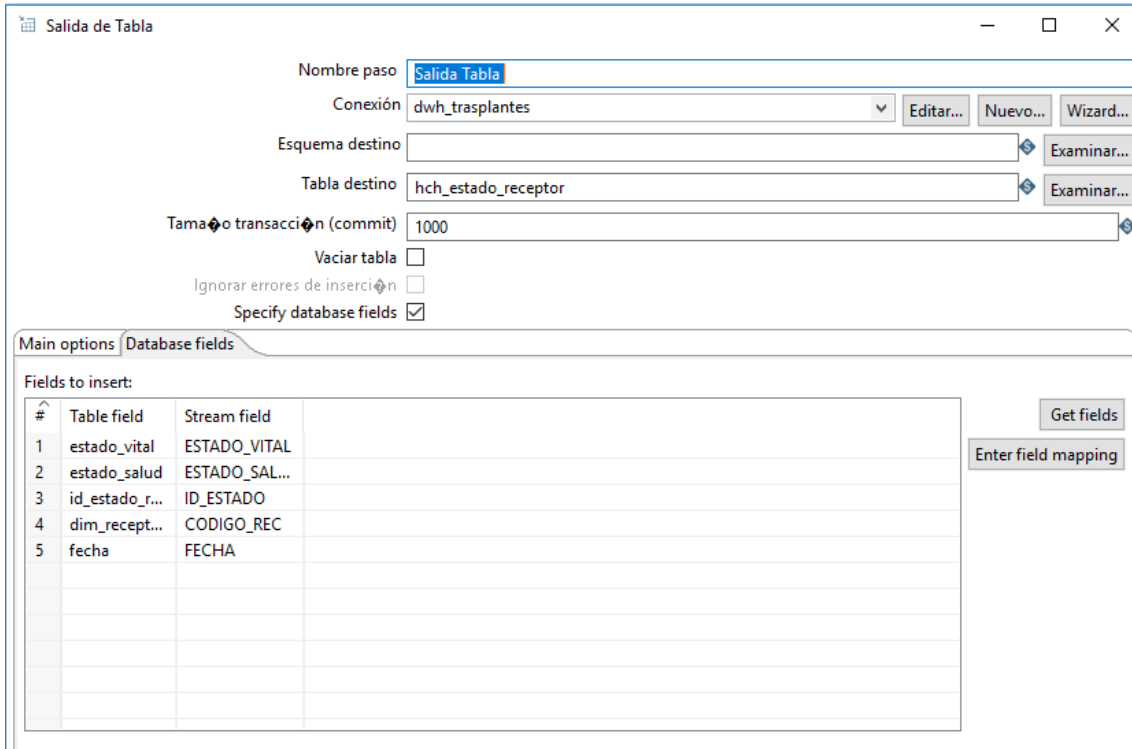
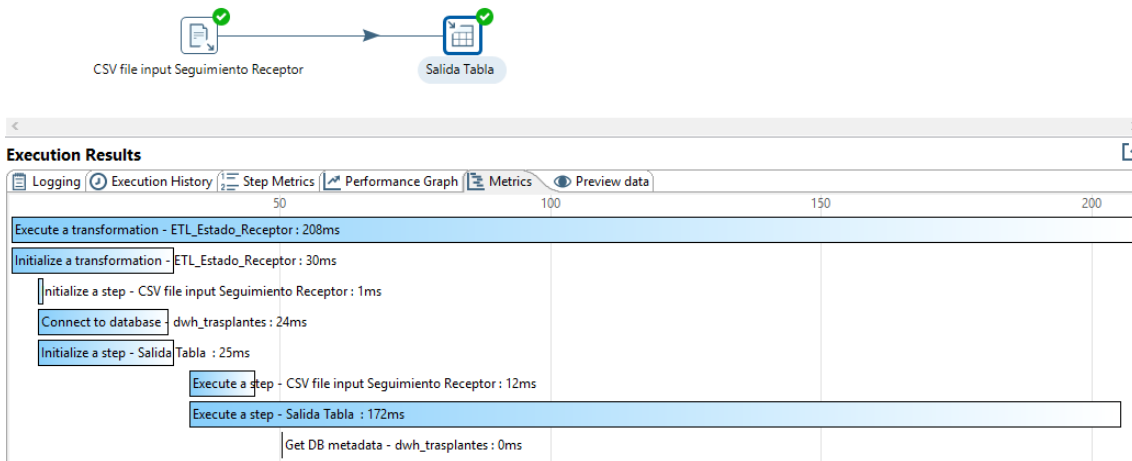


Ilustración 45. Configuración de salida para Estado del Receptor

Por último, se ejecuta comprobado que no se hayan producido errores en la transformación:



Y como se ha venido haciendo, se comprueba en la base de datos que los datos se hayan cargado según lo esperado:

```
1 • |SELECT * FROM dwh_trasplantes.hch_estado_receptor;
```

	id_estado_receptor	fecha	estado_vital	estado_salud	dim_receptor_id_receptor
▶	1111111	2000-07-28	1	3	12249391
	1111112	2002-04-12	1	3	12324486
	1111113	2002-04-05	1	3	12278920
	1111114	2004-10-18	1	3	12425146
	1111115	2003-08-03	1	3	12287202
	1111116	2003-03-06	1	2	12288635
	1111117	2003-04-20	1	2	12375962
	1111118	2006-03-30	1	2	12430945
	1111119	2006-04-04	1	2	12494656
	1111120	2008-04-25	1	2	12418286
	1111121	2008-07-08	1	2	12502673
	1111122	2009-05-06	1	2	12559981

Ilustración 47. Comprobación de carga de datos para Estado del Receptor

### 7.6 ETL para la carga de Estado del injerto

Al igual que la carga de datos para el seguimiento del receptor, ya es posible la carga del seguimiento de los injertos.

La fuente de datos, también se trata de un .csv que se configura en el primer paso del ETL. Este proceso tampoco requiere calculos ni transformación por lo que con una entrada y una salida, es suficiente para cargar los datos en MySQL:

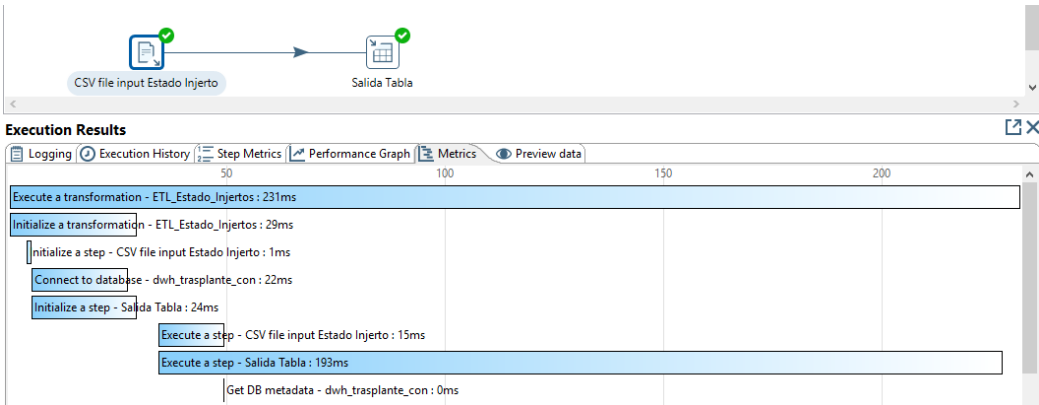
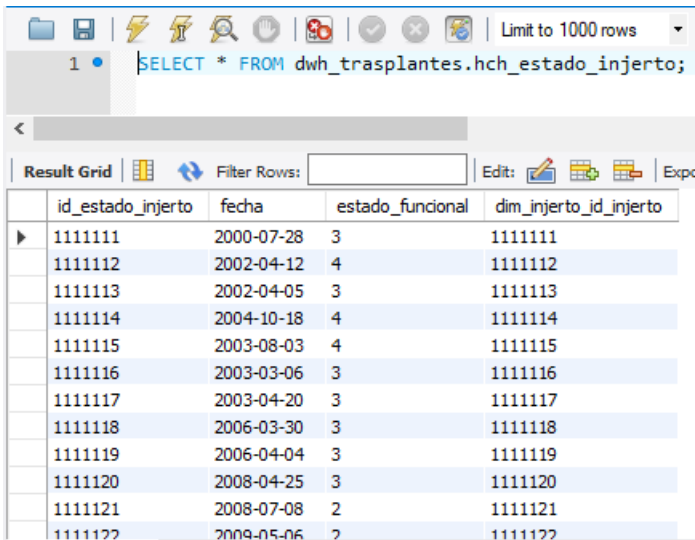


Ilustración 48. Ejecución ETL de Estado del Injerto

Una vez ejecutado el ETL, se pueden comprobar los datos cargados en la tabla de la base de datos:



id_estado_injerto	fecha	estado_funcional	dim_injerto_id_injerto
1111111	2000-07-28	3	1111111
1111112	2002-04-12	4	1111112
1111113	2002-04-05	3	1111113
1111114	2004-10-18	4	1111114
1111115	2003-08-03	4	1111115
1111116	2003-03-06	3	1111116
1111117	2003-04-20	3	1111117
1111118	2006-03-30	3	1111118
1111119	2006-04-04	3	1111119
1111120	2008-04-25	3	1111120
1111121	2008-07-08	2	1111121
1111122	2009-05-06	2	1111122

Il·lustració 49. Comprobación de carga de datos de Estado del Injerto

## 8. Diseño de la aplicación BI para la explotación de los datos

En este apartado se van a analizar la forma en la que vamos a explotar los datos cargados en los pasos anteriores.

Para ello, es necesario recordar las preguntas analíticas que tenemos como objetivo en este sistema de inteligencia de negocio que vamos a tener presentes para establecer los requisitos que se desea que cumplan los informes a desarrollar. También es importante tener en cuenta los requisitos que se establecieron para nuestro sistema BI, que en resumen, son los siguientes:

- Información accesible de forma simple y rápida
- Información presentada de manera consistente
- Adaptabilidad al cambio
- Presentación la información de manera oportuna
- Protección de la información confidencial
- Sistema de soporte de decisiones
- Aceptación del sistema BI

Se van a desarrollar un informe con 4 apartados diferenciados, basándonos en las preguntas analíticas:

1. Para cada uno de los tipos de TOS (riñón, hígado, corazón, pulmones, páncreas e intestino) cuales son los factores asociados del donante que pueden influir negativamente en los resultados de supervivencia o de morbilidad en los receptores, en qué medida y en comparación con la población general de individuos sanos.  
Este informe se va a llamar **Análisis por TOS** y permitirá seleccionar los diferentes tipos de órganos y diferentes casuísticas de factores asociados con el donante para poder analizar la información en varios gráficos que proporcionen la siguiente información:

- Estado vital por órgano: Mostrará el número de receptores vivos y de receptores fallecidos clasificados por órgano.
  - Estado vital: Mostrará el número de receptores vivos y de receptores fallecidos.
  - Estado vital según el año en el que se realizó el trasplante: Mostrará el número de receptores vivos y fallecidos clasificado por año en el que se llevó a cabo la intervención.
  - Tabla resumen por Órgano y Receptor con los datos de estado del receptor a fecha actual.
2. Cuál es el riesgo de trasplantar órganos de donantes con infecciones o con tumores activos en los antecedentes del donante, y hasta qué punto es asumible en el balance riesgo-beneficio. Este informe se llamará **Análisis Riesgo** y permitirá seleccionar tanto el rango de edad de Donante y Receptor como:
- Comparar para los diferentes órganos, la afectación tanto de infección activa como de antecedentes de cáncer
  - Según la edad de donante y receptor, analizar el número de receptores vivos y fallecidos
  - Analizar los diferentes factores (antecedente de cáncer, infección activa) en función del diagnóstico del receptor, por el cual se tuvo que someter a trasplante.
3. Que factores del donante son los que en mayor medida influyen en los resultados. Se creará un informe de **Análisis Factores Donante** que permitirá introducir como parámetros los factores asociados al donante y analizar:
- Estado de salud de los receptores
  - Estado de los injertos
- Además, para los injertos se añadirá como parámetro el tiempo de isquemia.
4. Diferencias en los resultados para mismo tipo de donantes y receptores en función del hospital en el que se produce la donación o el trasplante. Para analizar la influencia de los hospitales se creará un informe de **Análisis Hospitales**.  
El informe de de Análisis Hospitales permitirá seleccionar parámetros equivalentes de Donante-Receptor para poder analizar su influencia.
- Se estudiará en relación con los hospitales y comunidades autónomas en los que se hizo la intervención.
  - Además, se requiere comparar por fecha de trasplante la supervivencia de los receptores entre los diferentes hospitales.

En cuanto a usabilidad y diseño de la aplicación hay que tener en cuenta los siguientes requisitos de usuario:

- Los informes deberán adecuarse al diseño corporativo de la ONT.
- Deberán ser intuitivos para el usuario.
- Se requiere que sean fácilmente exportables a PDF, Power Point y Excel.

- Los tipos de gráficos se deben adecuar a la información a mostrar. Con colores diferenciadores y con posibilidad de cambiar el tema para adecuarlo a personas con problemas de visibilidad.
- Cada gráfico debe de tener la opción de hacer zoom para poder ver los datos de forma independiente y en detalle. También deben de poder mostrar en formato Excel si se desea.
- Los filtros aplicados en un informe deberán de permanecer activos en los subsiguientes informes.
- El usuario debe de poder conocer siempre los parámetros por los que se encuentra filtrado el informe.
- La navegabilidad debe de ser siempre clara y con posibilidad de regresar al punto anterior.
- Uso de gráficos de mapas para los parámetros con localización geográfica.

## 9. Implementación y análisis de resultados

Para la implementación de los informes, en primer lugar, es necesario crear la conexión con nuestro almacén de datos.

En el entorno de Power BI Desktop podemos observar que es posible selección un tipo de conexión con MySQL, como es el caso:

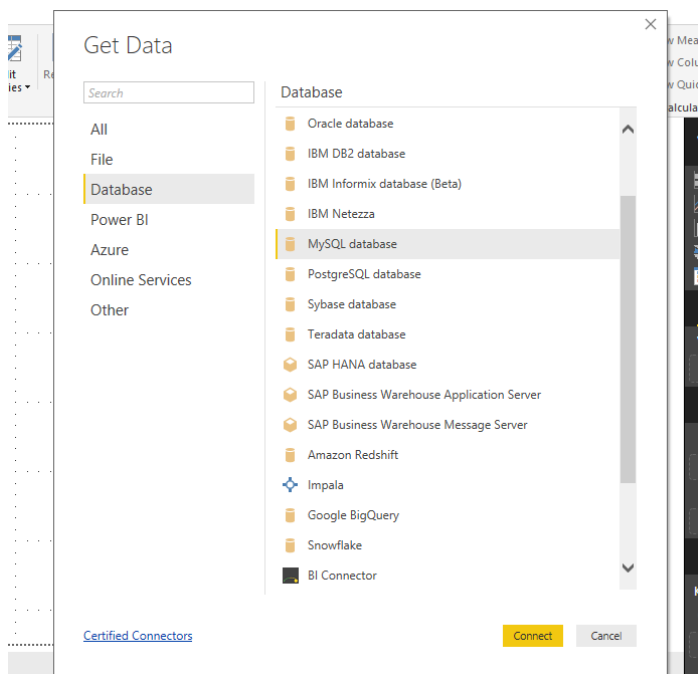


Ilustración 50. Conexiones disponibles Power Bi



Una vez configurado el acceso, este se mantiene permitiendo actualizar los datos.

MySQL database

localhost;dwh\_trasplantes

User name  
root

Password  
●●●●●●

Select which level to apply these settings to  
localhost;dwh\_trasplantes

Ilustración 51. Conexión Power Bi con DW

Establecida la conexión el sistema permite seleccionar las tablas con las que deseamos trabajar, también permite crear consultas.

Navigator

Display Options

- localhost: dwh\_trasplantes [6]
  - dwh\_trasplantes.dim\_donante
  - dwh\_trasplantes.dim\_hospital
  - dwh\_trasplantes.dim\_injerto
  - dwh\_trasplantes.dim\_receptor
  - dwh\_trasplantes.hch\_estado\_injerto
  - dwh\_trasplantes.hch\_estado\_receptor

dwh\_trasplantes.hch\_estado\_receptor

id_estado_receptor	fecha	estado_vital	estado_salud
1111111	28/07/2000	1	
1111112	12/04/2002	1	
1111113	05/04/2002	1	
1111114	18/10/2004	1	
1111115	03/08/2003	1	
1111116	06/03/2003	1	
1111117	20/04/2003	1	
1111118	30/03/2006	1	
1111119	04/04/2006	1	
1111120	25/04/2008	1	
1111121	08/07/2008	1	
1111122	06/05/2009	1	
1111123	26/04/2008	1	
1111124	31/05/2009	1	
1111125	16/07/2009	1	

Select Related Tables Load Edit Cancel

Ilustración 52. Selección de tablas de trabajo en Power BI

El entorno de trabajo de Power BI presenta el siguiente aspecto:

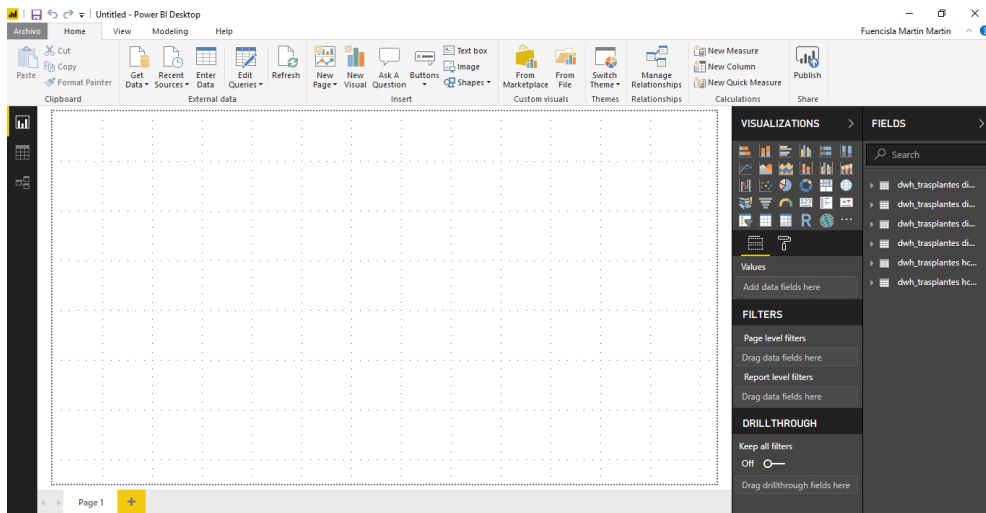


Ilustración 53. Entorno de trabajo Power Bi

En la barra de herramientas lateral izquierda podemos navegar en sus distintos modos de navegación.

Los tres iconos que aparecen en el borde izquierdo de la ventana se corresponden con las tres vistas disponibles en Power BI Desktop: Informe, Datos y Relaciones (de arriba abajo). En la imagen anterior se muestra la vista **Informe**.

En la vista de Datos podemos navegar por las tablas cargadas (en la parte derecha) y analizar los datos. En esta vista, y también en la de Informe, se puede valorar la creación de columnas o medidas:

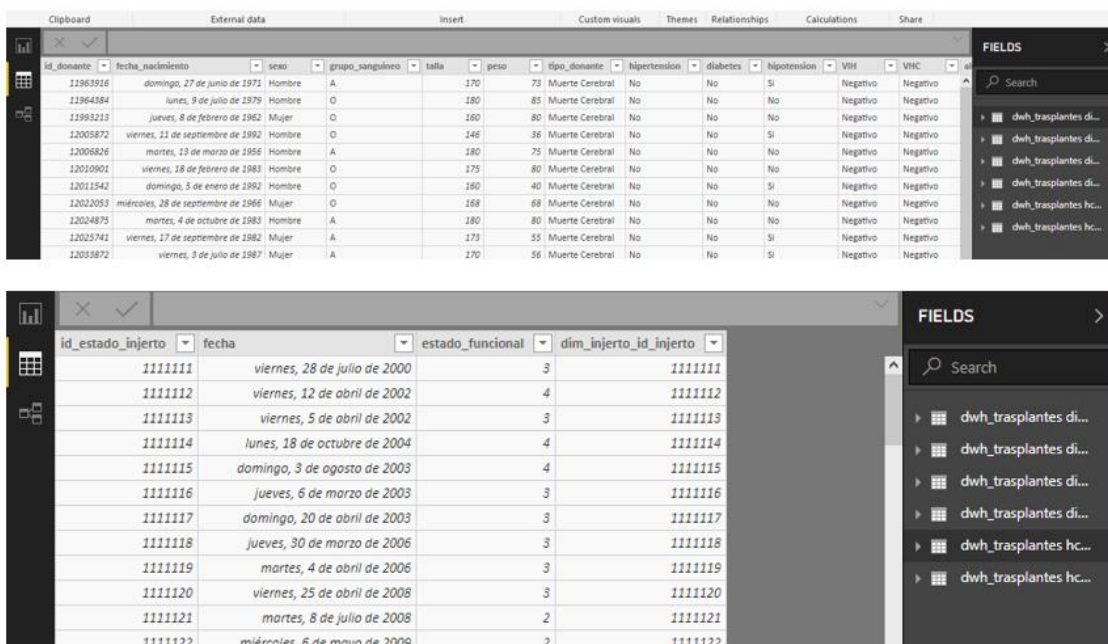


Ilustración 54. Vista de datos de Power Bi

Por último, la vista de Relaciones nos permite analizar, modificar o crear relaciones entre nuestras tablas.

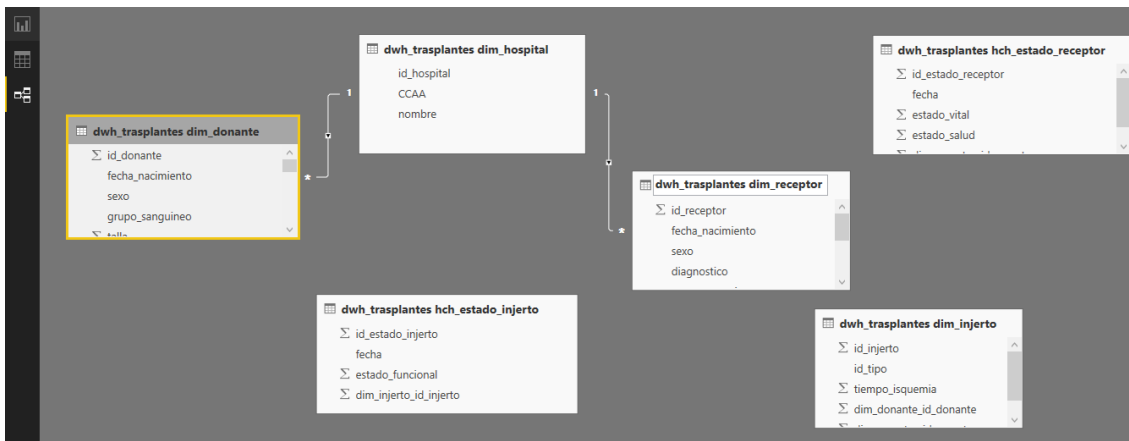


Ilustración 55. Vista de relaciones de Power BI

Como se puede observar en la imagen, es necesario completar este modelo ya que no se ha traído todas las relaciones al cargar las tablas en Power BI.

### Edit relationship

Select tables and columns that are related.

**dwh\_trasplantes dim\_injerto**

id_injerto	id_tipo	tiempo_isquemia	dim_donante_id_donante	dim_receptor_id_receptor
1111111	Pancreas	60	11963916	12249391
1111112	Pancreas	120	12006826	12324486
1111113	Pancreas	90	12011542	12278920

**dwh\_trasplantes dim\_receptor**

id_receptor	fecha_nacimiento	sexo	diagnostico	grupo_sanguineo	num_trasplai
12249391	jueves, 1 de enero de 1970	Hombre	Insuficiencia renal crónica + Diabetes	A	
12267565	jueves, 1 de enero de 1970	Mujer	Insuficiencia renal crónica + Diabetes	O	
12278920	jueves, 1 de enero de 1970	Hombre	Insuficiencia renal crónica + Diabetes	O	

Cardinality: Many to one (\*:1)

Cross filter direction: Single

Make this relationship active

Assume referential integrity

Apply security filter in both directions

OK
Cancel

Ilustración 56. Configuración de relación en Power BI

Una vez editadas las relaciones, el modelo quedaría configurado de la siguiente forma:

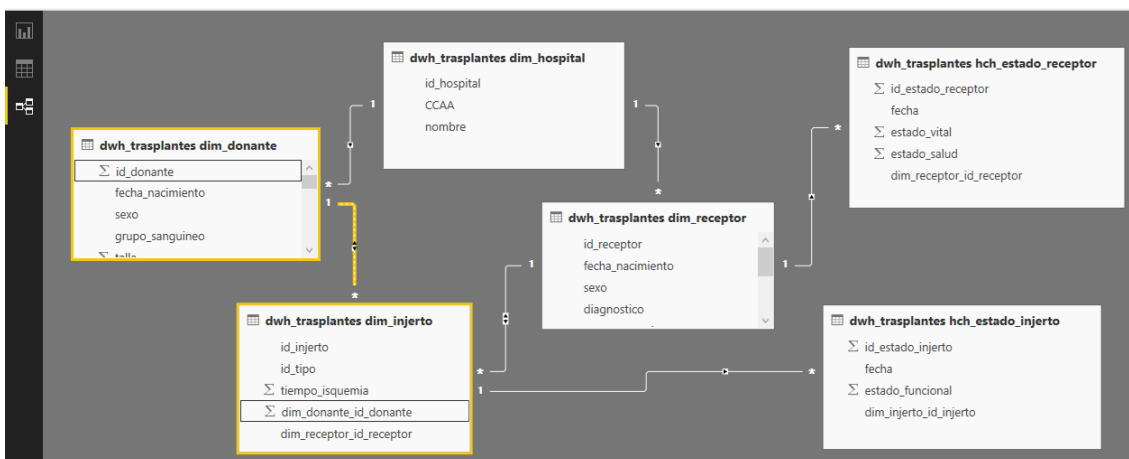


Ilustración 57. Modelo relacional en Power BI

Power BI Desktop también incluye el **Editor de Power Query**, que se abre en una ventana independiente.

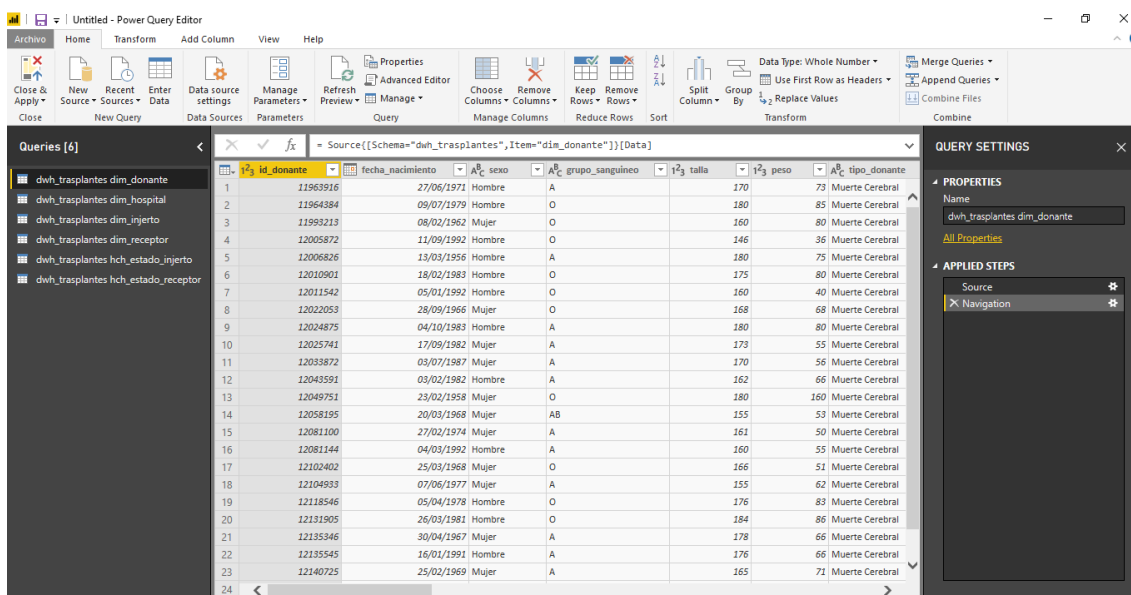


Ilustración 58. Editor de Power Query de Power BI

En el Editor de Power Query, se pueden crear consultas y transformar datos, a continuación, cargar dicho modelo refinado en Power BI Desktop y crear informes.

En nuestro caso, ya hemos realizado todas las transformaciones necesarias en los procesos ETL. Sin embargo, esta funcionalidad puede resultar útil para realizar joins “Merge Queries” y alcanzar datos que no son posibles mediante navegación entre las relaciones.

También se ha valorado el sustituir los valores de los indicadores de estado de injerto y de estado del receptor por literales, pero al tratarse de textos demasiado extensos se ha optado por dejarlos de forma numérica para después añadir una leyenda de ayuda en el cuadro de mando.

En cuanto al diseño del cuadro de mando a desarrollar, se ha creado un archivo .json con los colores corporativos:

```
{  
  "name": "ONTTheme",  
  "dataColors": ["#D34F6E", "#419C6D", "#EFA149", "#86B7D5"]  
}
```

Ilustración 59. Estilo de los informes en archivo .json

Al incluir el estilo de los informes de esta forma, tendremos la posibilidad de cambiar de tema y de volver al tema original.

Con esto cumpliríamos el requisito de poder modificar el tema para personas con baja visibilidad.

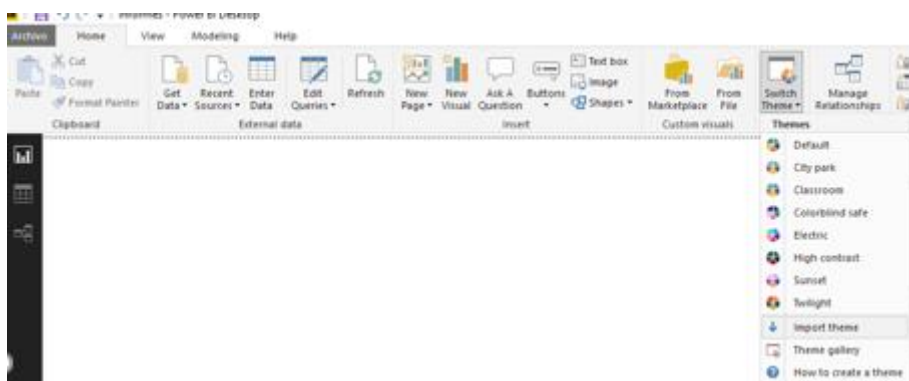


Ilustración 60. Selección de temas en Power BI

Ya que, como se puede observar, Power BI dispone un tema propio llamado “High contrast”.

Antes de realizar los informes, se ha realizado un esquema de lo que se desea incluir, para así poder diseñar los fondos de nuestro informe, otorgando al cuadro de mando un aspecto agradable e intuitivo para trabajar.

Para el informe de Análisis por Órgano Sólido Trasplantado, se han tomado la decisión de crear un área en la que se puede cambiar la vista por gráficos o por tabla y KPIs.

Para ello se va a hacer uso de una funcionalidad agregada hace relativamente poco a Power BI: Los Bookmarks.

Los **bookmarks** o marcadores son componentes que guardan el estado de la página, exactamente como está al momento de guardar ese marcador. De forma que luego se puede seleccionar el bookmark y ver la página con el estado en que se ha guardado. No se trata de una captura de pantalla, es una página dinámica e interactiva de Power Bi que guarda un estado del informe.

El uso de los **bookmarks** combinado con los **botones** y con el **panel de selección**, permite crear funciones adicionales y mejorar la experiencia de usuario.

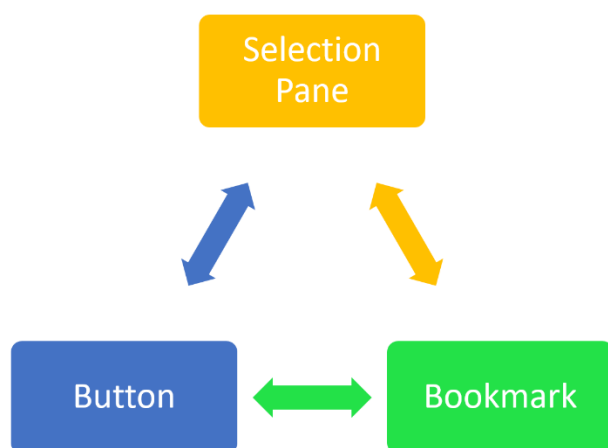


Ilustración 61. Componentes de diseño de Power BI (<http://radacad.com/bookmarks-and-buttons-making-power-bi-charts-even-more-interactive>)

A continuación, se explica el desarrollo de cada uno de los informes del cuadro de mandos para la explotación de la información transformada mediante los procesos ETL y almacenada en el Data Warehouse.

## 9.1 Informe Análisis por Órgano Sólido Trasplantado

Como se ha explicado anteriormente, para la realización de este informe, se ha hecho uso de la potencia de los marcadores de Power BI.

En la parte superior se muestran todos los tipos de órganos que disponemos en nuestro juego de datos. Estos injertos se pueden seleccionar, con la capacidad de multi selección para poder estudiar sus datos en los gráficos y tablas creadas para ello.

Además, como se puede observar en la figura siguiente, se podrán seleccionar diferentes parámetros que se mantendrán, aunque se navegue entre marcadores:

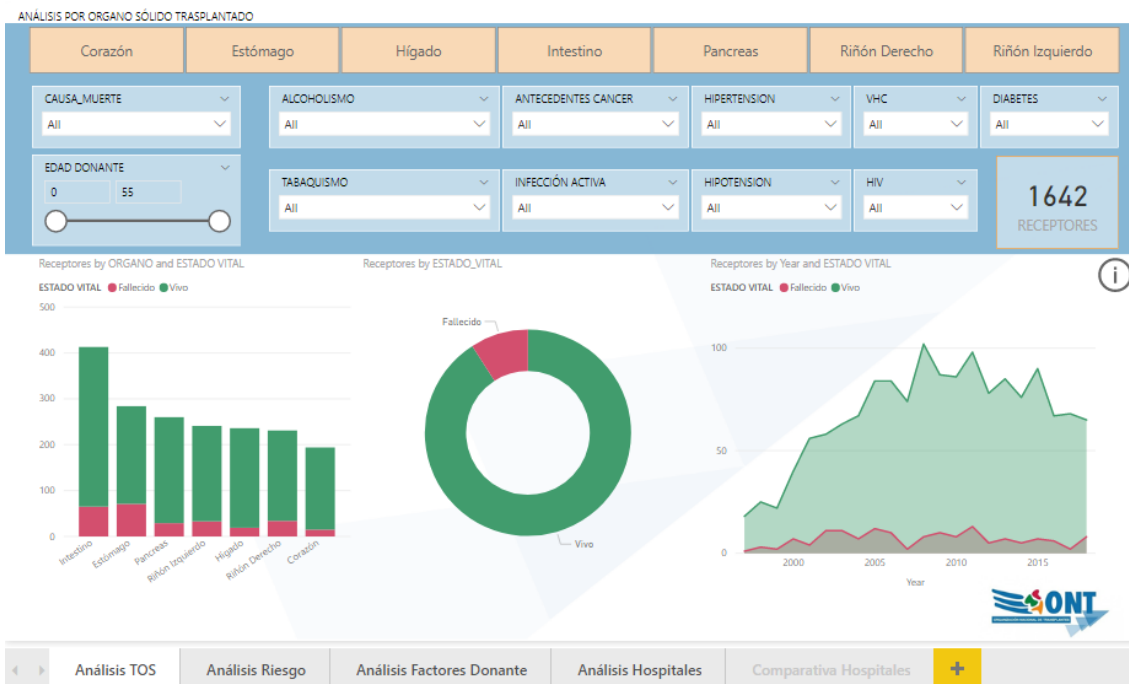


Ilustración 62. Informe Análisis TOS 1

El botón de información nos permitirá sustituir los gráficos por una tabla con los datos de los receptores y los KPIs en función del estado de salud del receptor:

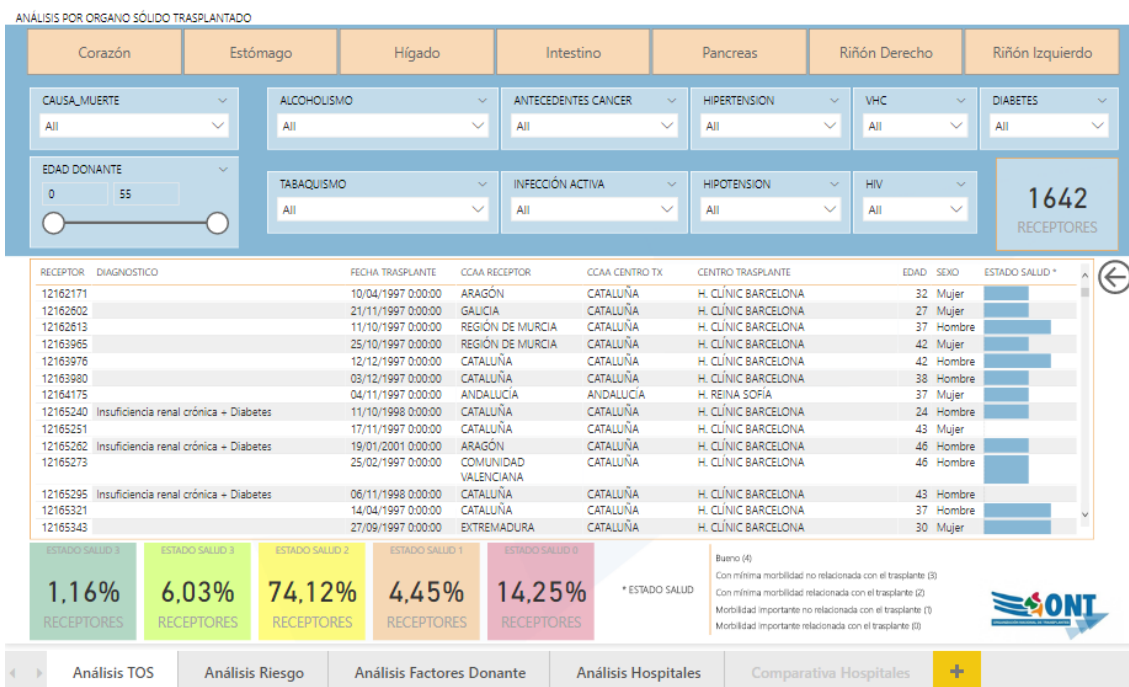


Ilustración 63. Informe Análisis TOS 2

Para la realización de este informe se han tenido que crear varias medidas adicionales, como, por ejemplo, el porcentaje de receptores con estado de salud 4:

**PORCENT 4 = [Count of CODIGO\_REC for 4]/COUNTA(Receptores[CODIGO\_REC])**

Ilustración 64. Medida calculada en DAX 1

Que a su vez hace uso de otra medida que también se ha tenido que crear:

```

1 Count of CODIGO_REC for 4 =
2 CALCULATE(
3     COUNTA('Receptores'[CODIGO_REC]);
4     'Receptores'[ESTADO_SALUD] IN { 4 }
5 )
    
```

Ilustración 65. Medida calculada en DAX 2

Ejemplo para receptores de Riñón derecho no Alcohólicos:

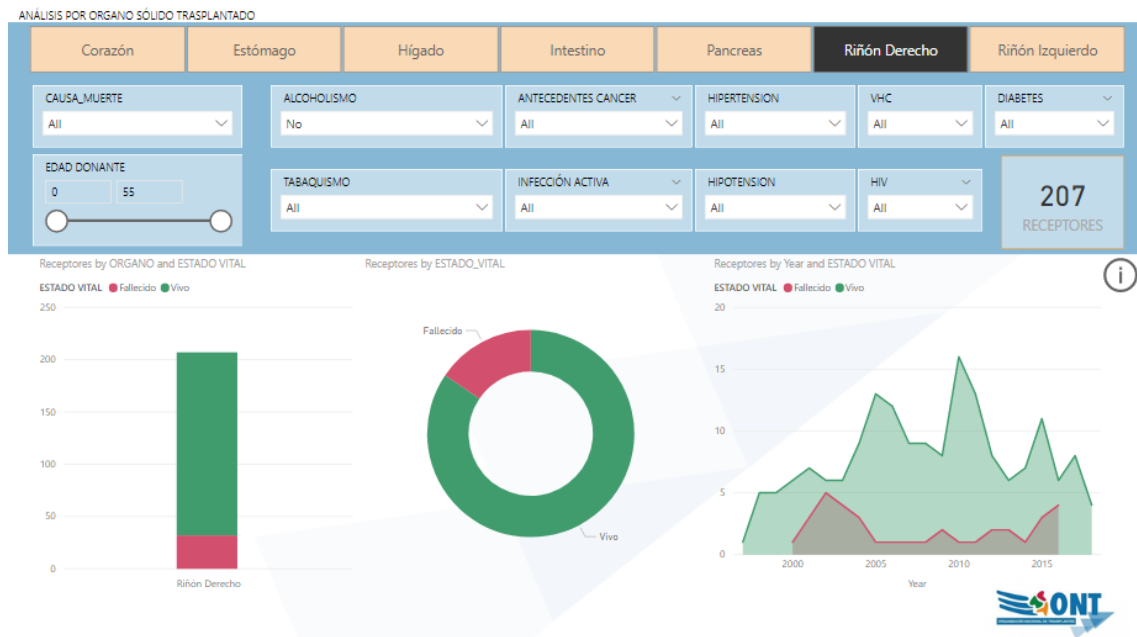


Ilustración 66. Ejemplo para Informe Análisis TOS 1



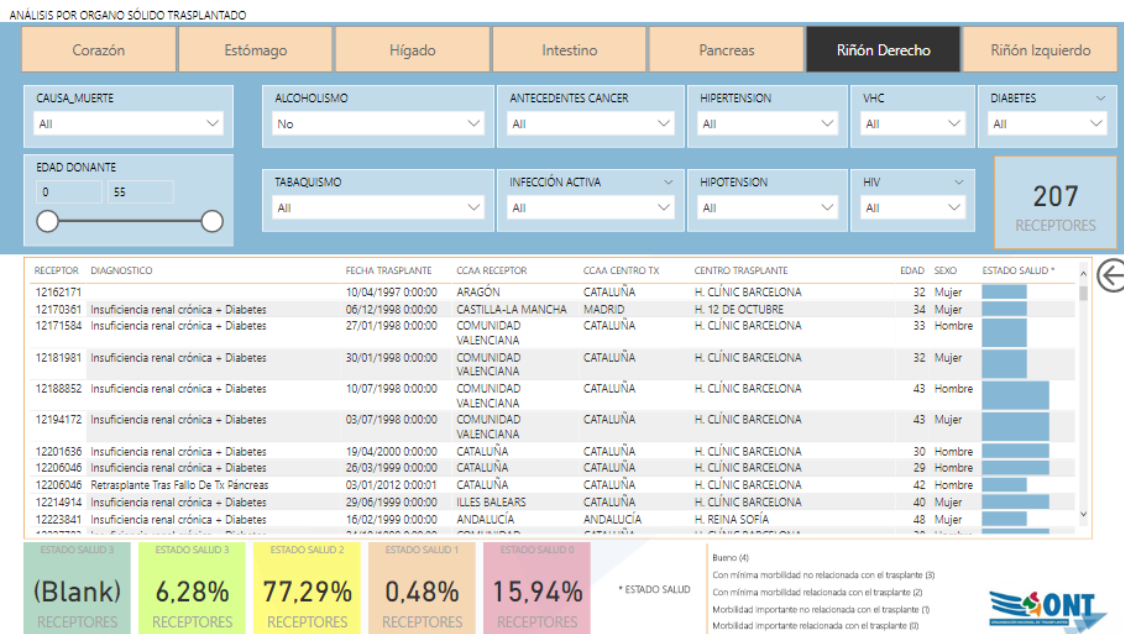


Ilustración 67. Ejemplo para Informe Análisis TOS 2

## 9.2 Informe Análisis Riesgo

En este informe se desea analizar cuál es el riesgo de trasplantar órganos de donantes con infecciones o con tumores activos en los antecedentes del donante, y hasta qué punto es asumible en el balance riesgo-beneficio. Para ello se ha desarrollado el siguiente informe en Power BI:

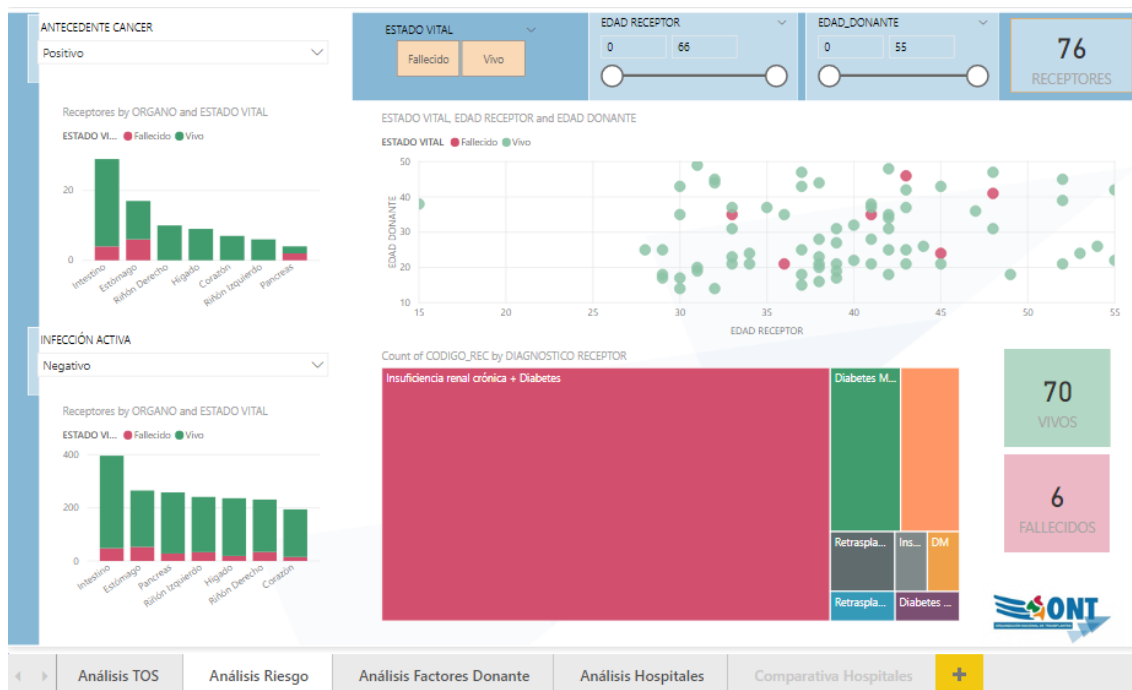


Ilustración 68. Informe Análisis Riesgo

Como se puede observar a la derecha, se han creado un par de gráficas idénticas. Estas gráficas dependerán solamente de su filtro correspondiente para poder analizar la información, como se puede observar en el siguiente ejemplo:

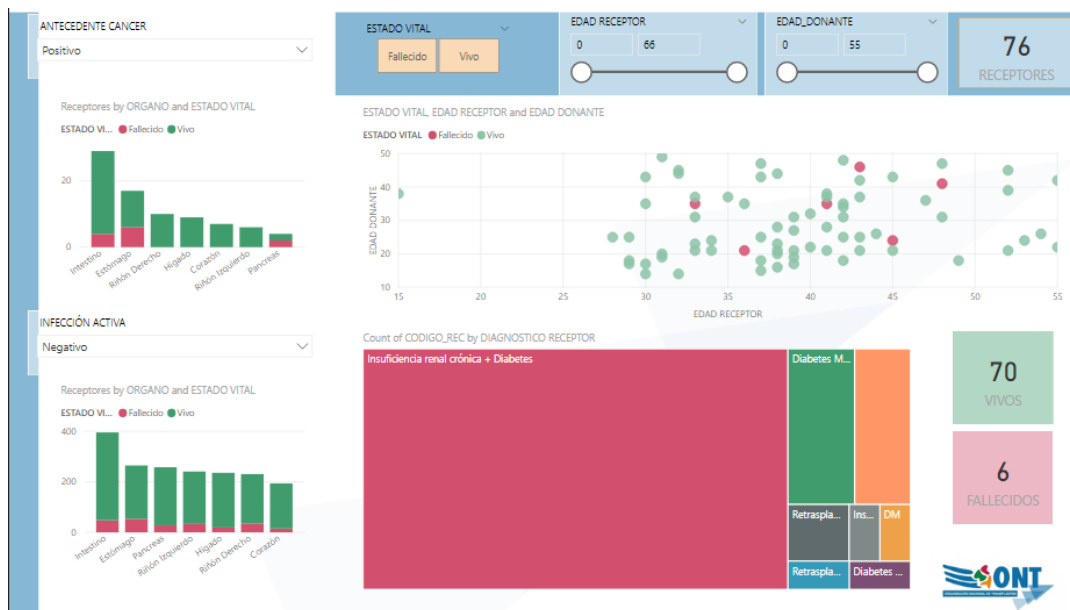


Ilustración 69. Ejemplo para Informe Análisis Riesgo

### 9.3 Informe Factores Donante

El informe de Factores Donante requiere poder analizar los factores del donante son los que en mayor medida influyen en los resultados. Se ha creado el siguiente informe:

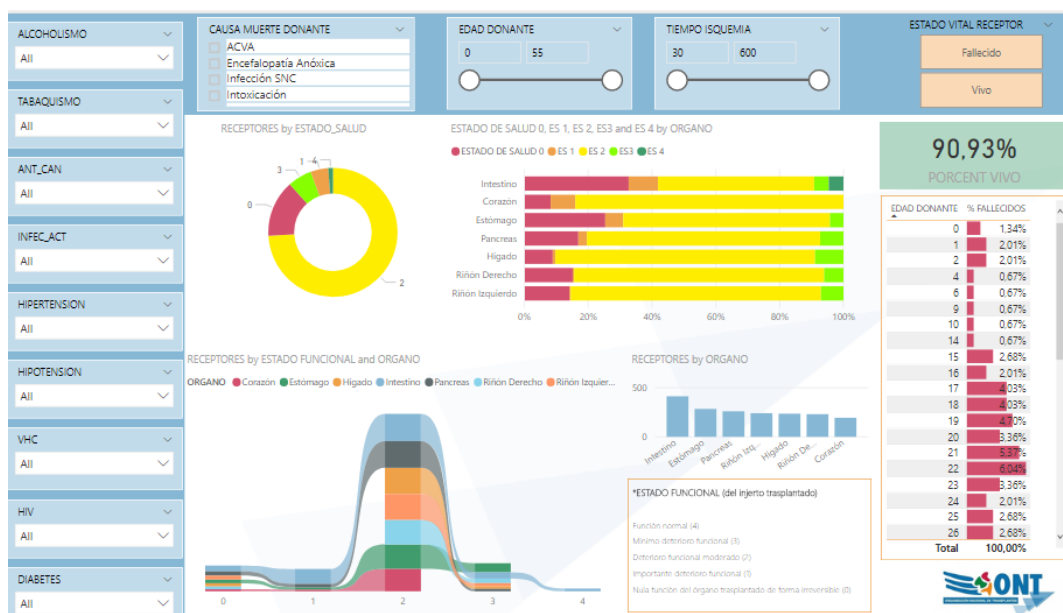


Ilustración 70. Informe Factores Donante

En este informe se analizan tanto los estados del receptor como los estados funcionales de los injertos. A continuación, se muestra un ejemplo, seleccionado la causa de muerte del donante, la edad del donante e hipertensión:

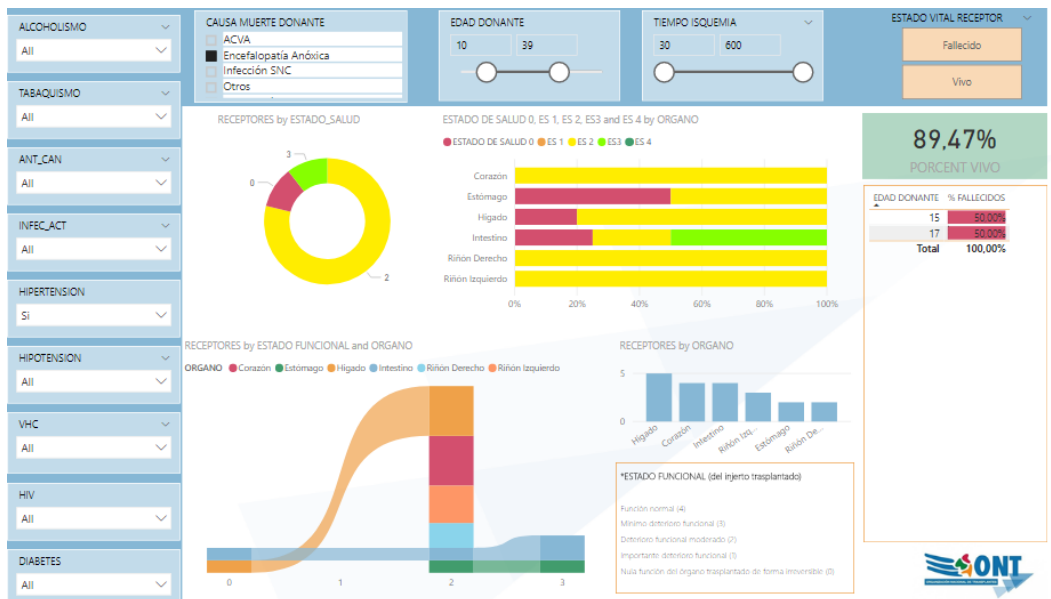


Ilustración 71. Ejemplo para Informe Factores Donante

## 9.4 Informe Análisis Hospitales

Para la creación del informe de Análisis de Hospitales dónde se realiza el transplante se ha realizado el siguiente informe:

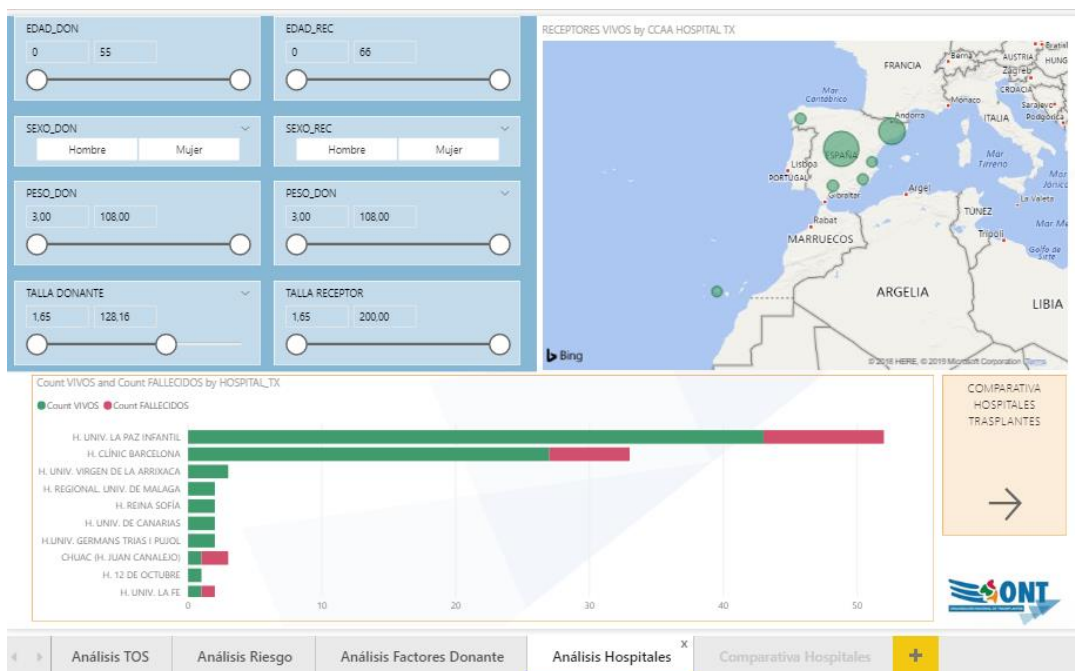


Ilustración 72. Informe Análisis Hospitales 1

Como se puede observar, se ha podido hacer uso de un gráfico de tipo mapa al tener disponibles las comunidades autónomas del hospital.

También se ha hecho uso de los bookmarks pero esta vez, para navegar entre una página y una página oculta de la siguiente forma:

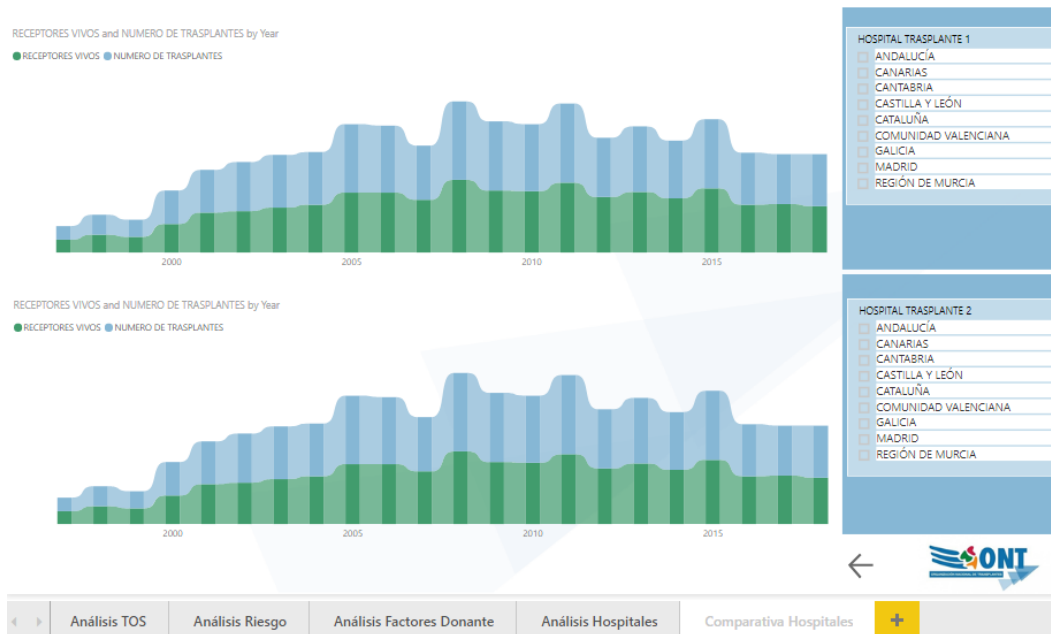


Ilustración 73. Informe Análisis Hospitales 2

En el segundo gráfico se han usado dos gráficos idénticos, pero se han editado las relaciones con sus filtros para poder ver de forma gráfica la comparativa de dos hospitales como se puede observar en el siguiente ejemplo:

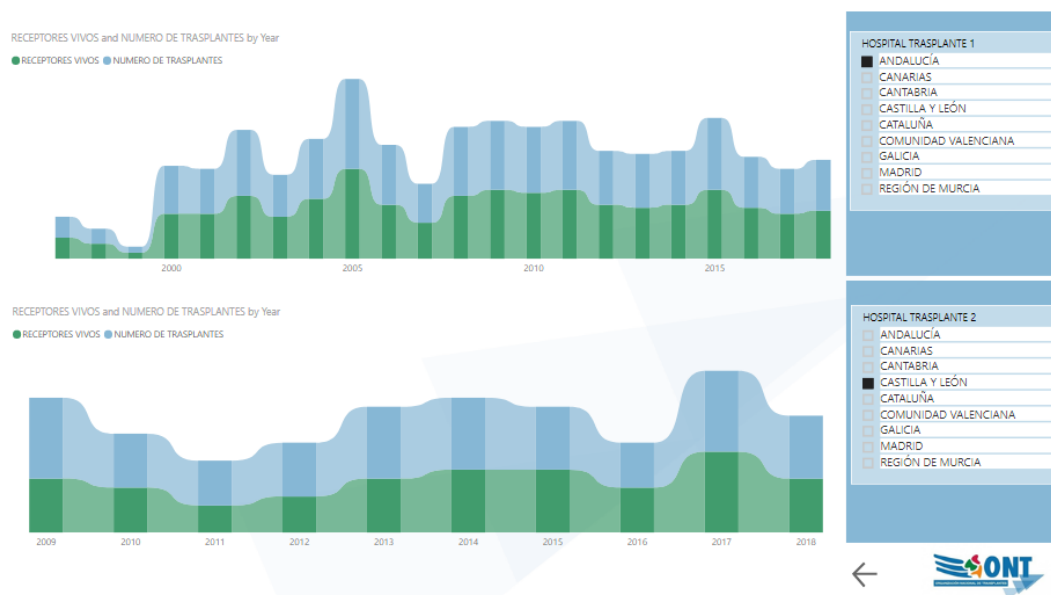


Ilustración 74. Ejemplo para Informe de Análisis de Hospitales

## 9.5 Explotación de los datos publicados

Una vez creado los informes en Power BI, los podemos publicar en su servicio y así poder analizar y exportar los datos con todas las funcionalidades que ofrece la herramienta. También, una vez publicados, es posible compartir con miembros de la misma organización. A continuación, se muestran algunas de las posibilidades que ofrece el servicio de Power BI:

Guardar, Imprimir, Insertar en SharePoint, Publicar en la Web, Exportar a PowerPoint y Descargar:

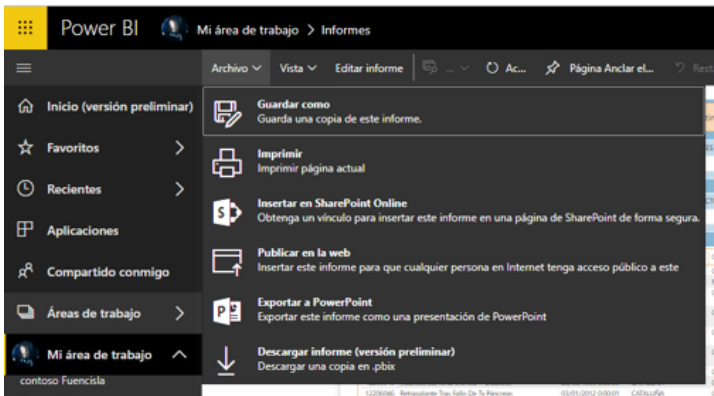


Ilustración 75. Posibilidades de análisis 1

Analizar métricas de uso:



Ilustración 76. Posibilidades de análisis 2

Actualizar, programar actualizaciones analizar en Excel, ejecutar conclusiones rápidas, administrar permisos y seguridad:



Ilustración 77. Posibilidades de análisis 3

Modo enfoque:



Ilustración 78. Modo enfoque 1

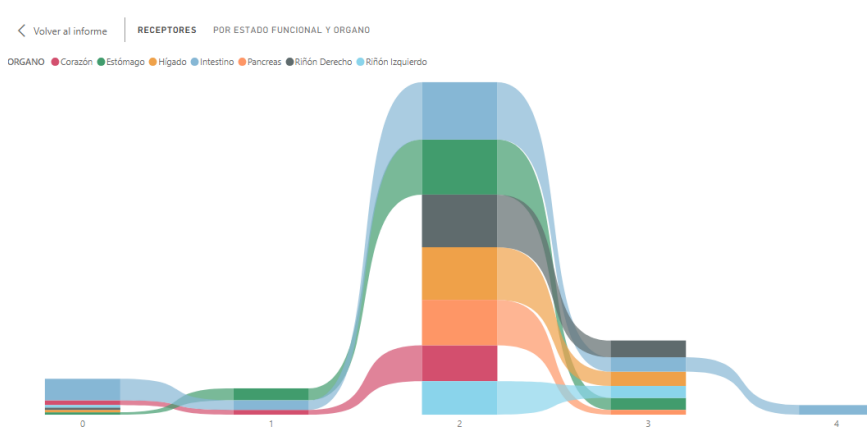


Ilustración 79. Modo enfoque 2

Mostrar datos:



Ilustración 80. Mostrar datos de gráfico 1

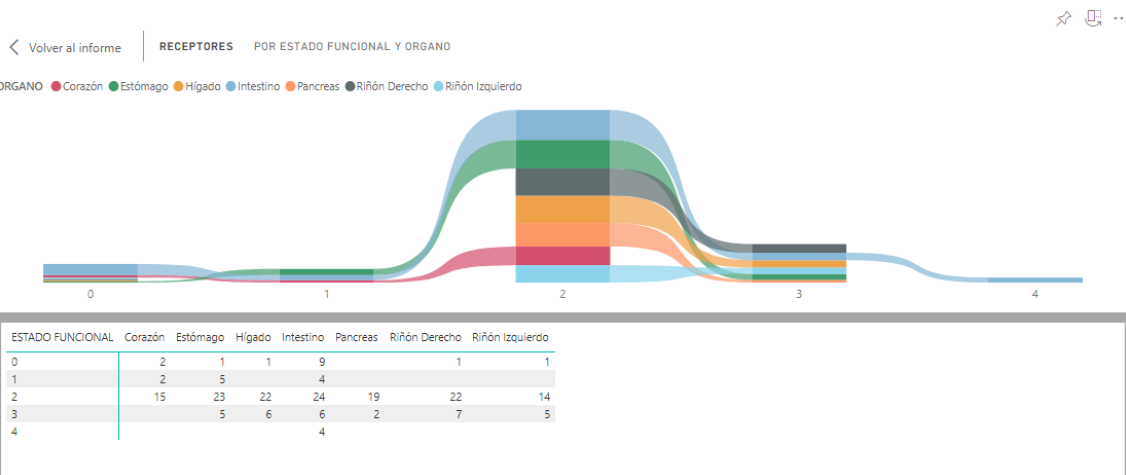


Ilustración 81. Mostrar datos de gráfico 2

Por lo tanto, con el desarrollo del cuadro de mandos en Power BI el usuario final tendrá muchísimas opciones de análisis a parte de las facilitadas en los informes creados a medida.

## 10. Conclusiones

Como conclusión de todo este proceso de elaboración del sistema de Sistema de Inteligencia de negocio para el análisis de los factores del donante que influyen en la supervivencia de los trasplantes de órganos sólidos, y de sus complicaciones, se han obtenido un conjunto de herramientas que permitirán analizar los indicadores de trasplante desde varias perspectivas para ayudar a la toma de decisiones de los médicos expertos de la ONT (Organización Nacional de trasplantes) y así mejorar el proceso actual de trasplantes.

Por lo tanto, los objetivos definidos al inicio del proyecto se han alcanzado. Cabe destacar que el usuario final de la herramienta ha manifestado su satisfacción al respecto de la explotación de los datos, haciendo especial mención a los cuadros de mando ofrecidos. Al conocer la potencia del desarrollo de informes para el análisis de datos con Power BI, incluso ha manifestado su deseo de añadir más cuadros de mando.

El proceso de análisis para el diseño y creación del almacén de datos y de procesos ETL para alimentarlo han resultado bastante duros debido en parte al desconocimiento de los datos y la dificultad para obtener un juego de datos de calidad. Pero lo que más ha podido impactar en el desarrollo normal del proyecto ha sido la conexión de los procesos ETL desde Pentaho con la base de datos MySQL. Se debería de haber previsto más tiempo para este punto, ya que, a parte de la dificultad de la configuración tecnológica, existía la falta de conocimiento de estas herramientas y procesos.

En contra partida, el desarrollo de la herramienta de explotación de datos con Power BI, ha sido mucho más ágil y agradable, básicamente porque ya se disponía de conocimiento previo.

Aunque, también, el desarrollo de este proyecto me ha permitido adquirir mayor conocimiento y posibilidades con la herramienta.

Como se ha comentado, con vistas a futuro, este proyecto podría ampliarse con la incorporación de más fuentes de datos y, sobre todo, como ha manifestado el usuario final, con más cuadros de mando para la explotación de la información.



## Bibliografia

- [1] Hernández Marrero D, González-Molina Alcaide M. Resultados Globales del Trasplante Renal. En: Lorenzo V, López Gómez JM (Eds) Nefrología al Día. <http://www.revistanefrologia.com/es-monografias-nefrologia-dia-articulo-resultados-globales-del-trasplante-renal-58>
- [2] VALJALO, Ricardo et al. Resultados a largo plazo en trasplante renal de donantes con criterios expandidos. Rev. méd. Chile [online]. 2016, vol.144, n.1 [citado 2018-09-29], pp.22-29.
- [3] Bergler, Tobias, Hutchinson, James A. 2017. Tools for Predicting Kidney Transplant Outcomes. Transplantation: September 2017 - Volume 101 - Issue 9 - p 1958–1959. doi: 10.1097/TP.0000000000001891.  
[https://journals.lww.com/transplantjournal/Fulltext/2017/09000/Tools\\_for\\_Predicting\\_Kidney\\_Transplant\\_Outcomes.4.aspx](https://journals.lww.com/transplantjournal/Fulltext/2017/09000/Tools_for_Predicting_Kidney_Transplant_Outcomes.4.aspx)
- [4] Durand F. How to improve long-term outcome after liver transplantation?. Liver Int. 2018;38(Suppl. 1):134–138. <https://doi.org/10.1111/liv.13651>
- [5] <http://www.ont.es/informacion/Paginas/Trasplante.aspx>
- [6] <https://samiuc.es/index.php/calculadores-medicos/calculadores-de-evaluadores-pronosticos/indice-de-comorbilidad-de-charlson-cci.html>
- [7] Kanban: <https://www.iebschool.com/blog/metodologia-kanban-agile-scrum/>
- [8] The Data Warehouse Toolkit, 3rd Edition. Ralph Kimball y Margy Ross
- [9] <https://gravitar.biz/bi/toma-de-requerimientos-en-solucion-business-intelligence/>
- [10] <http://www.innoventolutions.com/comparison-matrix.html>
- [11] <https://comparisons.financesonline.com/jaspersoft-vs-pentaho>
- [12] <http://www.stratebi.com/pentaho>
- [13] <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/ventajas-en-la-integracion-de-datos-con-la-herramienta-etl-pentaho>
- [14] <http://ftp.tcrc.edu.tw/MySQL/doc/refman/5.0/es/features.html>
- [15] <https://docs.microsoft.com/es-es/power-bi/guided-learning/>