SISTEMA DE ANALISIS DE ECONOMÍA Y DOCENCIA DE LA ETSISI

Autor: Diego Cruz Cuervo

Tutor: Borja Bordel Sánchez

# 

# **RESUMEN**

En el contexto actual, la gestión y análisis de datos se ha convertido en una herramienta crucial para la toma de decisiones informadas y estratégicas. Las universidades públicas, enfrentan el desafío también de optimizar sus recursos y mejorar así la calidad de la docencia.

Este Trabajo de Fin de Grado se centra en la creación de un sistema de herramientas capaz de automatizar y facilitar el análisis económico y de profesorado al que se enfrenta frecuentemente la Escuela Técnica Superior de Ingenieros y Sistemas Informáticos.

Para ello, el sistema proporcionará un proceso ETL que extraerá de un documento Excel la información que la escuela necesite para sus procesos de negocios.

Una vez hecho esto, se creará un Data Warehouse que nos servirá de base de datos analítica donde se guardarán los datos procesados para facilitar el análisis de estos.

Por último, se creará un cuadro de mando que se podrá proporcionar a los usuarios finales que les servirá de gran utilidad para facilitar sus tareas de toma de decisiones económicas y de docentes.

# **ABSTRACT**

In the current context, data management and analysis have become crucial tools for making informed and strategic decisions. Public universities also face the challenge of optimizing their resources and thus improving the quality of teaching.

This Final Degree Project focuses on the creation of a system of tools capable of automating and facilitating the economic and faculty analysis that the Higher Technical School of Engineers and Computer Systems frequently faces.

To achieve this, the system will provide an ETL process that will extract from an Excel document the information the school needs for its business processes.

Once this is done, a Data Warehouse will be created, which will serve as an analytical database where the processed data will be stored to facilitate its analysis.

Finally, a dashboard will be created that can be provided to end users, which will be of great use in facilitating their economic and teaching decision-making tasks.

**Índice**

[**RESUMEN** 2](#_Toc168418498)

[**ABSTRACT** 3](#_Toc168418499)

[**INTRODUCCIÓN** 5](#_Toc168418500)

[**Contexto** 5](#_Toc168418501)

[**Objetivo** 7](#_Toc168418502)

[**Estructura del documento** 9](#_Toc168418503)

[**PASOS PREVIOS** 10](#_Toc168418504)

[**Conceptos** 10](#_Toc168418505)

[**Herramientas** 12](#_Toc168418506)

[**Pentaho Data Integration:** 12](#_Toc168418507)

[**Mysql:** 16](#_Toc168418508)

[**Power Bi:** 19](#_Toc168418509)

[**Versiones de las herramientas utilizadas** 20](#_Toc168418510)

[**Planificación** 21](#_Toc168418511)

[**PROYECTO** 22](#_Toc168418512)

[**Análisis** 22](#_Toc168418513)

[**Origen de los datos** 22](#_Toc168418514)

[**Estructura de los datos** 23](#_Toc168418515)

[**Diseño** 25](#_Toc168418516)

[**Data Warehouse** 25](#_Toc168418517)

[**Apartado Económico** 29](#_Toc168418518)

[**Apartado Profesorado** 32](#_Toc168418519)

[**Implementación** 36](#_Toc168418520)

[**ETL** 37](#_Toc168418521)

[**BBDD** 54](#_Toc168418522)

[**Cuadro de mando** 55](#_Toc168418523)

[**RESULTADOS** 60](#_Toc168418524)

[**CONCLUSIONES** 63](#_Toc168418525)

[**BIBLIOGRAFÍA** 65](#_Toc168418526)

[**ARCHIVOS** 65](#_Toc168418527)

# **INTRODUCCIÓN**

## **Contexto**

Sin lugar a duda, saber la importancia que los datos tienen hoy en día y su forma de utilizarlos a nuestro favor es algo esencial en la mentalidad de un gran número de empresas en la actualidad.

A raíz del auge de ciertas tecnologías como el Business Inteligence y el uso de las IAs, muchas empresas han optado por mejorar su relación con el apartado analítico para así poder sacar el máximo rendimiento a todos los datos que de una forma u otra han obtenido a lo largo del tiempo.

Según un estudio de la Universidad ISEP, las empresas que optan por utilizar tecnologías Bi obtienen un 5% más de retorno de inversión, aumentan un 53% su eficiencia operativa y mejoran un 48% su toma de decisiones respecto de aquellas empresas que no la utilizan.

Posiblemente uno de los factores que han producido el aumento del Business Inteligence ha sido el gran aumento de datos a lo largo de los últimos años, debido al continuo uso de los dispositivos tecnológicos. Los seres humanos pasamos alrededor de 7 horas navegando en internet y, normalmente, compartiendo los datos por los que se pasa, así como sus búsquedas y por ello sus gustos.

No solo empresas privadas hacen uso del BI, instituciones públicas como con la que nos encontramos en este Trabajo de Fin de Grado también sacan un gran partido a los datos que les rodean para así poder mejorar su rendimiento general.

En este caso hablamos de la Escuela Técnica Superior de Sistemas Informáticos, campus perteneciente a la Universidad Politécnica de Madrid, la cual maneja una gran cantidad de datos internos relacionados con todo tipo de aspectos propios de una universidad. En este trabajo se utilizaron sus datos económicos y de profesorado para poder realizar una herramienta que les permita poder analizar de una forma más sencilla y duradera los datos que extraigan a lo largo de los años.

## **Objetivo**

El objetivo principal de este Trabajo de Fin de Grado es diseñar y desarrollar un cuadro de mando para el análisis de datos económicos y de profesorado de la Escuela Técnica Superior de Ingenieros y Sistemas Informáticos de la Universidad Politécnica de Madrid. Este cuadro de mando tiene como propósito proporcionar una herramienta intuitiva y eficaz que permita a la institución optimizar la gestión de sus recursos, mejorar la calidad de la docencia y facilitar la toma de decisiones informadas.

El cuadro de mando se desarrollará utilizando Power BI debido a sus capacidades avanzadas de visualización de datos y su facilidad de uso. Power BI permitirá integrar diversas fuentes de datos, proporcionando una plataforma única para el análisis y la toma de decisiones. A través de este proyecto, se buscará diseñar visualizaciones claras y significativas que permitan a los usuarios comprender fácilmente la información presentada. Estas visualizaciones incluirán gráficos, tablas y otros elementos interactivos que faciliten la explotación y el análisis de los datos.

La creación del cuadro de mando implicará el desarrollo de un proceso ETL (Extracción, Transformación y Carga) que permitirá la recolección y preparación de los datos necesarios. Este proceso garantizará que los datos sean limpiados, transformados y cargados de manera eficiente en un Data Warehouse, donde serán almacenados para su posterior análisis. El Data Warehouse servirá como una base de datos analítica que proporcionará un acceso rápido y confiable a los datos procesados.

La implementación del cuadro de mando también considerará la usabilidad y la accesibilidad para los usuarios finales. El objetivo es que los usuarios finales puedan aprovechar al máximo las capacidades del cuadro de mando para mejorar la toma de decisiones y la gestión de los recursos en la escuela.

En resumen, este Trabajo de Fin de Grado no solo tiene como objetivo mejorar la eficiencia operativa de la Escuela Técnica Superior de Ingenieros y Sistemas Informáticos, sino también fortalecer su capacidad para tomar decisiones basadas en datos. Al proporcionar una herramienta avanzada de análisis y visualización de datos, se espera contribuir significativamente a la misión educativa y administrativa de la institución, promoviendo una gestión más informada y estratégica.

## **Estructura del documento**

Para su mejor comprensión, estructuraré este documento de una forma lineal de proyecto.

Empezaremos con los aspectos generales del proyecto donde se describirán una serie de conceptos necesarios para su claro entendimiento, las herramientas que han sido utilizadas para producir el resultado esperado y la planificación que se siguió a la hora de desarrollarlo.

Una vez vistos los aspectos generales y necesarios del proyecto, nos internaremos en el desarrollo de este. Por ello el documento continuará con el análisis, diseño e implementación del proyecto.

Una vez terminada su explicación se relatarán los resultados obtenidos, así como las conclusiones que se pueden sacar del desarrollo del TFG.

Como final del documento se encuentra las referencias usadas para el proyecto y una explicación de los diferentes archivos generados durante el desarrollo de este.

# **PASOS PREVIOS**

## **Conceptos**

* **Business Inteligence:** Se define como el conjunto de herramientas y buenas prácticas que utilizan las empresas para realizar un análisis eficaz de los datos que poseen y que así sirvan de ayuda para la toma de decisiones de la empresa.
* **ETL:** En inglés, son las siglas de “Extract, Transform and Load”. Se trata de un conjunto de procesos que se encargan de recoger los datos necesarios de la fuente de origen, realizar una serie de operaciones/transformaciones de esos datos para adaptarlos y facilitar así su análisis y, por último, los cargan en el sistema de almacenamiento de destino del que después se puede realizar el análisis directo o, en nuestro caso, montar un panel de control que nos facilitará su visionado y manejo.
* **Data Warehouse:** Es el nombre que se le dan a las bases de datos encargadas de almacenar los datos que van a ser usados para la toma de decisiones de la empresa. Normalmente, son las bases de datos de destino de los procesos ETl, anteriormente comentados. En este caso, esta base de datos está conformada por dos tipos de tablas: dimensión y hechos. Las tablas de dimensión están formadas por los datos que proporcionan información característica del proceso de negocio que se está analizando, mientras que las tablas de hechos contienen los valores que nos proporcionan la información analítica.
* **Control de versiones:** El control de versiones es una práctica fundamental en la gestión de proyectos de desarrollo de software, que permite registrar y supervisar los cambios realizados a lo largo del tiempo. Este proceso es crucial para el trabajo colaborativo, ya que facilita la integración de contribuciones de múltiples autores sin perder el historial de modificaciones. Mediante el uso de repositorios, commits, ramas y fusiones, se puede mantener un registro detallado de todas las versiones, facilitando la recuperación de datos, la colaboración eficiente y el control de calidad del proyecto. Herramientas como Git, Subversion (SVN) y Mercurial son comúnmente utilizadas para implementar el control de versiones, asegurando un manejo ordenado y seguro de los avances y contribuciones en el desarrollo de un proyecto.

## **Herramientas**

### **Pentaho Data Integration:**

figura


Ilustración 1: Logo Pentaho

Se trata de una herramienta de código abierto perteneciente a la familia de Pentaho. Esta herramienta permite al usuario la creación de procesos ETL.

Su interfaz es bastante amigable, ya que solamente se basa en dos tipos de procesos, transformaciones y Jobs.

Interfaz de usuario gráfica, Texto, Aplicación, Sitio web

Descripción generada automáticamente

Estos primeros son los que se encargarán de realizar todo el proceso ETL, haciendo uso de su gran número de herramientas que permiten realizar gran cantidad de operaciones, además de las herramientas que permiten realizar la extracción y la carga de los datos de forma sencilla y amable.

Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamente

Por otro lado, los Jobs son los procesos que permiten orquestar varias transformaciones.

Texto

Descripción generada automáticamente con confianza baja

En este caso existen otras herramientas de creación de procesos ETL como puede ser Talent Open Studio o Microsoft SQL Server Integration Services, entre otras. En el caso de mi proyecto he optado por la alternativa de Pentaho debido sus ventajas:

* **Gratuidad:** Su uso no supone un gasto en licencias o mantenimiento de recursos como algunas otras competidoras. Si que existe una edición de pago, pero en este caso la parte gratuita sirve con creces para la realización de este proyecto.
* **Gran comunidad:** Gracias posiblemente a la existencia de la versión gratuita y gracias a que se trata de una herramienta de código libre, se ha creado una gran comunidad alrededor de Pentaho lo que origina un gran abanico de blogs y páginas donde consultar ciertos problemas que uno se puede encontrar al utilizar la herramienta.
* **Conocimiento previo:** En este caso, esta ventaja es propia de mí ya que gracias a experiencias previas en las que he podido participar, partía ya con cierto conocimiento de la herramienta que me ha hecho balancearme sobre utilizarla como herramienta de procesos ETL.

### **Mysql:**



Para la gestión de los datos analíticos he usado como base de datos de destino una base de datos MySQL.

MySQL es un sistema de gestión de bases de datos relacionales de código abierto desarrollado por Oracle. Esta herramienta nos va a permitir administrar la base de datos de destino de los procesos ETL, en esencia, lugar en el que se van a almacenar los datos una vez procesados por Pentaho Data Integration.

Esta herramienta funciona con un modelo de base de datos relacional, esto significa que los datos se organizan en tablas definidas con una estructura. Existen otras bases de datos que no son relacionales, pero en el caso de realizar una base de datos analítica como la que se debe utilizar en este proyecto

Como su propio nombre indica, utiliza el lenguaje SQL para tratar los datos.

En este caso también existen otras opciones como SQL Server o SQLite, pero al igual que con Pentaho, decidí optar por la opción de MySQL debido a sus características:

* **Gratuidad:** Al igual que con Pentaho, se trata de un software de código abierto haciéndolo bastante accesible.
* **Compatibilidad:** Una gran ventaja a la hora de tomar la decisión de elegir MySQL como gestor de bases de datos es claramente su compatibilidad con Pentaho el cual volcará los datos transformados sin problema en la base de datos.
* **Escalabilidad:** Puede escalar para adaptarse a las necesidades de diferentes aplicaciones. Idóneo en el caso de que se quiera perfeccionar el proyecto y ampliarlo a las distintas áreas que posee la escuela.
* **Conocimiento:** Claramente MySQL se estudia a lo largo de toda la carrera que he estudiado (Ingeniería del Software) es por ello por lo que poseo un conocimiento previo de la herramienta que me ha ayudado bastante a tomar la decisión.
* **Facilidad**: A parte de este gestor, a lo largo de la carrera he tenido la oportunidad de probar otros de ellos, y no he visto ninguno que sea tan secillo a la hora de realizar casi todas las operaciones necesarias en una base de datos. Además, la instalación de MySQL también es de lo más sencilla necesitando únicamente el instalador de la herramienta el cual te creará el servido, el promt y el propio gestor.

La segunda opción que se tuvo en mente a la hora de elegir una base de datos de destino para los datos analíticos fue SQL Server ya que es bastante parecida a MySQL, pero al final opté por la sencillez que nos proporciona esta última.

### **Power Bi:**



Con todo ello llegamos a la herramienta sobre la que recae el final del proyecto. Power Bi se trata de un conjunto de servicios que permiten al usuario analizar diferentes orígenes de datos y montar en la herramienta un panel de mandos dinámico con el que se puede realizar una visual analítica sobre esos datos anteriormente mencionados.

Esta herramienta consta de una aplicación de escritorio sobre la que se genera el panel de mandos como si de un PowerPoint se tratase. Consta también de un servicio de software en el que puedes publicar los cuadros de mando que crees para así poder compartirlos con otras personas sin que ellos puedan modificar la aplicación. Por último, también tiene una aplicación móvil con la que puedes modificar y visualizar los paneles de mando de tu cuenta.

En este caso la opción fue bastante sencilla ya que Power Bi es de las mejores herramientas para la creación de cuadros de mando y también poseía un conocimiento previo de la herramienta, aunque en este caso en menor medida lo que me ha permitido mejorar con ella a lo largo del proyecto.

### **Versiones de las herramientas utilizadas**

* Pentaho: 9.4
* MySQL:
  + Server: 8.0.36
  + Workbench: 8.0.36
* Power Bi: 2.129.1229.0 (Mayo 2024)

## **Planificación**

Lo primero que se tiene que hacer a la hora de afrontar un proyecto es realizar una planificación de lo que se debe de hacer, en que orden y de qué forma. Así, podremos tener una visual del alcance que tiene el proyecto, poder hacer una estimación de la duración que tendrá e incluso observar de una forma clara los pasos a seguir para que todo vaya acorde al plan.

Es por ello, que antes de comenzar a trabajar directamente sobre el proyecto, realicé un estudio para tener claro todo lo que se iba a llevar a cabo en el TFG, hablar con mi tutor para aclarar bien el objetivo de este y poder, como se ha mencionado en el apartado anterior, tomar una decisión de las herramientas que se han utilizado.

Como punto de comienzo encontraríamos la primera reunión que tuvimos mi tutor y yo en el que él, me enseño el tema del TFG, el cual me pareció bastante interesante ya que abarca un tema que, personalmente, me parece muy atractivo, además que podría generar una herramienta que aporte gran valor a los futuros análisis que pueda realizar la escuela en sus datos.

Tras ello, analicé las posibles herramientas que se podrían utilizar en el proyecto y elegí las anteriormente mencionadas. Una vez hecho todo lo anterior, era momento de comenzar el análisis de los documentos que mi tutor me proporcionó tras esa primera reunión.

# **PROYECTO**

## **Análisis**

A la hora de afrontar un proyecto con una metodología tradicional como con el que nos enfrentamos en este caso, lo primero que se debe realizar (tras la planificación y pasos previos) sería un análisis del objetivo que tiene el proyecto, así como de los datos que tenemos para poder realizarlo.

### **Origen de los datos**

Los datos que se van a utilizar en este TFG provienen de los documentos generados por la propia escuela proporcionados por mi tutor.

Primeramente, la mayoría de los datos son extraídos de forma manual de un PDF que contiene una gran cantidad de datos de los años comprendidos entre 2009 y 2013. Sobre ellos se recogieron los datos económicos y de docencia.

Tabla

Descripción generada automáticamente

Tabla

Descripción generada automáticamente

Tabla

Descripción generada automáticamente

Por otra parte, se completaron los años entre 2014 y 2023 para el apartado económico con los PDF económicos que se generaron esos años. Así, se pudo realizar un mejor análisis de este apartado de datos.

### **Estructura de los datos**

Como se ha comentado en el apartado anterior, los datos son extraídos de documentos PDF de forma manual, algo que claramente es muy ineficiente para el futuro plan que pueda tener este proyecto.

Observando los datos generé una plantilla Excel, tanto del apartado económico como del de docencia, en el que los encargados de realizar el análisis de los datos puedan incorporar allí directamente los datos y así facilitar luego los procesos que van a actuar encima de ellos.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Tabla

Descripción generada automáticamente con confianza media

## **Diseño**

Una vez hecho el análisis de los datos que me proporcionó mi tutor para poder realizar el proyecto y sabiendo como están estructurados, comenzó una fase de diseño de la base de datos de destino.

Recuerdo que estos datos, que en este punto encontramos en unos archivos Excel estructurados, van a sufrir una serie de transformaciones y van a ser depositados en una base de datos especial llamada Data Warehouse.

Por eso en esta fase del proyecto, estructuré esta base de datos.

### **Data Warehouse**

Sabiendo el concepto de Data Warehouse y analizando la estructura que había generado en la sección anterior, se generó el Data Warehouse correspondiente:

**NOTA:** Se utilizarán dos prefijos para referirnos a los dos tipos de tablas que encontramos en un Data Warehouse.

**Dim – Dimensión**

**Fact – Hechos**

Cabe mencionar también que este tipo de bases de datos se pueden estructurar de diferentes formas dependiendo del objetivo final que se quiera lograr con ella:

**-Estrella:** Este esquema está compuesto por una tabla de hechos, que será la tabla principal de la tabla, la cual contará con los datos necesarios para realizar cualquier tipo de cálculo de medidas para realizar el análisis y también tendrá una referencia a cada tabla de dimensión que haya en el modelo, las cuales tendrán información adicional que nos dará el contexto de la situación. Las tablas de dimensión no se conectan a nada que no sea la tabla de hechos.

Este esquema es el óptimo para herramientas de BI, además de ser el más sencillo y óptimo para el disco.

Diagrama

Descripción generada automáticamente

**-Copo de nieve:** Este esquema se parece bastante al esquema anterior, con la única diferencia que existen en él dimensiones que contienen referencias a otras dimensiones las cuales no se conectan directamente con la tabla de hechos.

Los datos en este esquema se encuentran normalizados, es por ello por lo que ocupa menos espacio en disco y se reduce así el tiempo de consulta. La única desventaja es que el trabajo de mantenimiento es mayor.

Diagrama

Descripción generada automáticamente

**-Constelación:** En ella hay dos tablas de hechos que comparten alguna tabla de dimensión. Esto nos permite en la herramienta BI poder filtrar por esa dimensión y que automáticamente ambas tablas estén filtradas. Tiene un mayor espacio en disco y puede llegar a dificultar las consultas y por ende el rendimiento.

Diagrama

Descripción generada automáticamente

Este esquema es el que se terminó siguiendo ya que encontraremos una dimensión común entre las dos tablas de hechos que representan los sectores en los que se divide el cuadro de mando. En un principio iban a ser dos esquemas de estrella totalmente independiente uno del otro.

### **Apartado Económico**

**Dim\_concepto\_economico:**

Esta tabla contiene los datos de los distintos conceptos/artículos en los que la escuela ha gastado o ingresado dinero. En caso de que en algún año posterior se realice un gasto o ingreso sobre un concepto nuevo, solamente será necesario incorporar toda su información necesaria en esta tabla y luego referenciarlo en la tabla de hechos.

Los atributos de esta tabla son los siguientes:

Imagen que contiene Texto

Descripción generada automáticamente

**Dim\_tipo\_economico;**

Esta tabla contiene los dos tipos de datos que puede haber en la tabla económica: Gasto o Ingreso.

Los atributos de esta tabla son los siguientes:

Texto

Descripción generada automáticamente con confianza media

**Fact\_economico:**

Esta tabla posee los datos analíticos necesarios para posteriormente realizar cualquier estudio que se quiera realizar. Aparte de estos datos analíticos, también posee referencias a las tablas de dimensión que nos proporcionarán información extra sobre los datos para poder agruparlos o incluso filtrarlos.

Los atributos de esta tabla son los siguientes:

Texto

Descripción generada automáticamente con confianza media

### **Apartado Profesorado**

**Dim\_categoria\_profesorado:**

Esta tabla contiene los diferentes tipos de profesores que ha tenido la escuela a lo largo de los años 2009 y 2013. A través de esta podremos en la tabla de hechos hacer un análisis de cuantos profesores de un tipo hubo o incluso poder agruparlos por uno de estos tipos.

Los atributos de esta tabla son los siguientes:

Texto

Descripción generada automáticamente

**Dim\_departamento\_profesorado:**

Esta tabla contiene los diferentes departamentos que tiene la escuela, que nos servirán para posteriormente en la tabla de hechos poder analizar el profesorado respecto a cada departamento haciendo referencia desde esa tabla a esta.

Los atributos de esta tabla son los siguiente:

Texto, Carta

Descripción generada automáticamente

**Fact\_profesorado:**

Esta tabla es la tabla principal de la sección de profesorado. En ella al igual que en la tabla de hechos de economía, encontramos los datos analíticos de la sección. Además, también tiene las referencias a las distintas tablas de dimensión para poder filtrar y agrupar por ellas.

Los atributos de esta tabla son los siguientes:

Texto

Descripción generada automáticamente

A parte de las dimensiones anteriormente mencionadas, en el modelo final de Power Bi encontramos una última dimensión común a ambas tablas de hechos: Dim\_ejercicio.

Esta dimensión servirá para poder filtrar ambas tablas de hechos por un año en concreto y así poder ver y analizar los datos de una forma más sencilla. Esta tabla, únicamente cuenta con una columna de sencilla. Esta tabla, únicamente cuenta con una columna de ejercicio en el que tiene los años entre 2000 y 2050 para poder incorporar sin problema los datos de esos años.

En caso de continuar con el proyecto, y en caso de que se quiera incorporar la información de algún año posterior a 2050, tendremos que añadir en esta tabla esos años.

## **Implementación**

Una vez puesto todo lo anterior en contexto comienza la parte de implementación, que como su propio nombre indica se trata de la parte de desarrollo del proyecto en la que se realiza todo lo anteriormente mencionado.

Para ello el orden de realización de las diferentes partes será el orden lógico que se tiene que seguir para poder obtener los resultados queridos. Es decir, nuestro objetivo es realizar un panel de mando sobre unos datos que nos llega en formato PDF.

Estos datos fueron pasados a Excel para así poder incorporarlos en Pentaho y que nos resultara más fácil su manejo.

Una vez hecho eso pasarán por los procesos de transformación para adaptar los datos a nuestro modelo y para poder acomodarlos se depositarán en una base de datos propia.

Con ello en la base de datos se terminará por el desarrollo del cuadro de mando con una última herramienta.

### **ETL**

Como se acaba de comentar, tras la actualización de los datos a formato Excel, se realizaron 6 Transformaciones y 3 Jobs en la herramienta de Pentaho para así poder acomodar los datos y conformar el Data Warehouse.

Texto

Descripción generada automáticamente

Comenzaremos por la categoría de docencia y dentro de ella con las transformaciones de las tablas de dimensión.

Hay que destacar que lo más habitual es comenzar por las tablas de dimensión ya que la tabla de hechos está formada por referencias a las tablas de dimensión con las que tiene relación, es por ello por lo que, para poder coger esas referencias, se tienen que tener las tablas ya creadas y formadas. Dicho de otro modo y como veréis posteriormente, se necesitan las tablas de dimensión para crear la tabla de hechos.

**Dim\_categoria\_profesorado:**

Diagrama

Descripción generada automáticamente

Esta transformación consta de 4 pasos:

* **Tab\_profesorado:** Se trata del paso de obtención de los datos del Excel para poder introducirlos en la herramienta. Aquí lo más importante es coger bien los datos de la hoja en la que se encuentran del Excel. En mi caso adicionalmente parametricé la ruta desde donde se coge el Excel para que en caso de que se cambie la ruta solamente sea necesario acceder al archivo de la parametrización y ahí cambiarlo, sin tener que tocar nada de la transformación.

Imagen que contiene Texto

Descripción generada automáticamente

* **Seleccionar categoría:** En este caso, por el Excel, nos vienen todos los datos del Excel, algo que no necesitamos puesto que solo nos interesa el apartado de categoría, es por ello que este paso se encarga de seleccionar únicamente la columna de categoría que nos viene del Excel.

Interfaz de usuario gráfica, Tabla

Descripción generada automáticamente

* **Filas únicas:** Tras el paso anterior, nos hemos quedado únicamente con la columna de categoría, pero como os podéis imaginar existen valores repetidos dentro de los datos que nos quedan. Es por ello por lo que este paso nos elimina este inconveniente quedándose una única referencia por categoría, por lo que ya tenemos las categorías que van a conformar la dimensión.

Interfaz de usuario gráfica, Texto, Aplicación, Tabla

Descripción generada automáticamente

* **Dim\_categoria\_profesorado:** Y, con todo esto, llegamos al último paso de la transformación. Este último paso, como os podéis imaginar, se trata de cargar la tabla de dimensión.

Para ello, Pentaho trae consigo un paso que nos permite crearla y mantenerla actualizada en caso de que se quiera realizar más cargas. En ella puedes poner el nombre que quieras a la tabla, seleccionar los campos que tendrá y también puedes realizar un control de versiones con su rango de fechas y, lo más importante, puedes crear una clave que se vaya incrementando de forma automática para cada valor de la tabla.

De esta manera, podrás luego desde la tabla de hechos a la tabla de dimensión.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Cosa importante también que mencionar es la conexión a la base de datos ya que en este último paso es necesario ya tener creada la base de datos, aunque no las tablas ya que estas también se pueden crear a medida pulsando en el botón inferior que pone SQL.

Con esto ya no solo se explica la dimensión de categoría sino que las otras 4 dimensiones que tiene el Data Warehouse también se realizan de la misma forma, añadiendo alguna de ellas un paso de ordenación que es opcional y cambiando claramente el paso de la selección o el paso del Excel en el que se recogen los datos económicos en vez de los de docencia.

Texto, Tabla

Descripción generada automáticamente

**Dim\_departamento\_profesorado:**

Diagrama

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

**Dim\_tipo\_economico:**

Diagrama

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

**Dim\_concepto\_economico:**

Diagrama

Descripción generada automáticamente

Tabla

Descripción generada automáticamente

Una vez creadas las tablas de dimensión como se ha comentado anteriormente se realizaron las tablas de hechos que contienen referencias a las tablas de dimensión con las que están relacionadas, así como los datos que van a servir para realizar los distintos cálculos para hacer el análisis querido.

Aquí encontramos 2 tablas de hechos que en su construcción se parecen bastante:

**Fact\_economico:**

Interfaz de usuario gráfica, Diagrama

Descripción generada automáticamente

**Fact\_profesorado:**

Gráfico de cajas y bigotes

Descripción generada automáticamente con confianza baja

El flujo de esta transformación es sencillo:

Comienza con el paso de extracción de datos de los excels en los que se encuentran la información para cada una de las categorías. Como se ha comentado anteriormente se encuentran parametrizadas para así poder modificarlo más fácilmente en caso de querer cambiar la zona en la que se quiera depositar.





Una vez tenemos los datos en el flujo, vienen los pasos necesarios para poder hacer referencias desde la tabla de hechos a las tablas de dimensión con las que están relacionadas cada una.

Para ello hay que tener en cuenta que existe un campo común entre la tabla de hechos y las tablas de dimensión. Por ese campo se realizará una unión entre las tablas teniendo como resultado, que en cada fila existirá tanto la descripción de la dimensión como su referencia:

Interfaz de usuario gráfica, Aplicación, Tabla, Excel

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

(Las capturas anteriores son un ejemplo con la dimensión del tipo económico, esto se repite para las demás dimensiones).

Como podéis observar, se nos quedan ciertos datos los cuales son repetidos o innecesarios. Estamos consiguiendo la referencia a la tabla de dimensión para no sobrecargar la tabla de hechos con datos los cuales son simplemente de contexto.

Es por ello por lo que el siguiente paso a realizar es una selección de campos de la tabla para poder eliminar las descripciones que están guardadas ya en las tablas de dimensión.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamenteInterfaz de usuario gráfica, Aplicación, Tabla

Descripción generada automáticamente

Con ello ya tendríamos formadas las tablas de hechos asique lo último que nos quedaría para completar la tabla de hechos es el paso de carga en el Data Warehouse.

Para ello, al igual que en las tablas de dimensión, uniendo el flujo con el paso final y dándole internamente en ese paso al botón de SQL, se nos creará la tabla con la estructura que se necesita para así solo tener que terminar de configurar el paso y poder ejecutarlo.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Con ello ejecutado, ya tendríamos el Data Warehouse creado y con los datos ya transferidos desde el archivo Excel.

**Fact\_economico:**

Imagen que contiene Tabla

Descripción generada automáticamente

**Fact\_profesorado:**

Tabla

Descripción generada automáticamente

**Jobs**

Una vez realizada la implementación de las transformaciones de las diferentes tablas que van a formar el Data Warehouse, se realizó la creación de los Jobs, que son por así decirlo procesos a mayor nivel de abstracción los cuales sirven normalmente para coordinar las transformaciones, así como poder realizar alguna actividad de más alto nivel como mandar un mail de confirmación o incluso hacer un manejo de archivos para transportarlos.

Nuestro proyecto está formado por un Job por cada tabla de dimensión que se ha creado anteriormente, aunque serán algo sencillo únicamente para dejar creada su estructura en caso de que se quiera profundizar más en ellos.

Interfaz de usuario gráfica, Diagrama

Descripción generada automáticamente

Para las tablas de hechos, se ha realizado una Job para automatizar la ejecución de los Jobs de las dimensiones con las que tienen relación para que en caso de que haya algún campo nuevo se actualicen y tras ello llama a su respectiva transformación.

Diagrama

Descripción generada automáticamente con confianza baja

Así en caso de que solo quieras actualizar la dimensión se ejecutará el Job pertinente, pero en caso de que quieras crear la tabla de hechos de nuevo se ejecutará ese Job que primero actualizará las tablas de dimensión, en caso de que sea necesario y luego creará la tabla de hechos.

### **BBDD**

Con todo lo anterior realizado, los datos son depositados en nuestro Data Warehouse, base de datos creada para almacenar los datos que van a ser usados para el análisis.

Este Data Warehouse en principio contará con las tablas que se han creado en las transformaciones anteriores, pero se podría ir ampliando con las diferentes secciones que se quieran ir añadiendo a la herramienta final.

Ahora que tenemos los datos en el Data Warehouse, podemos validar que todo este acorde a lo planeado mediante una herramienta de administración de bases de datos como es, en este caso, MySQL Workbench.

Con esta aplicación se podrán realizar las consultas para poder observar los datos que han sido incorporados en las tablas en su totalidad:

SELECT \*

FROM [insertar nombre tabla]

### **Cuadro de mando**

**Definición**

Un cuadro de mando de Power BI es una herramienta que permite el diseño, la visualización y el análisis de datos clave en tiempo real para la toma de decisiones empresariales. Este cuadro de mando se compone de diversos gráficos, tablas, mapas y otros elementos visuales que enseñan la información más relevante y permiten a los usuarios explorar los datos mediante filtros y segmentaciones.

El objetivo principal de un cuadro de mando es proporcionar una visión clara y concisa de las métricas importantes de una organización, facilitando el monitoreo de los datos y la identificación de tendencias y anomalías.

Power BI, desarrollado por Microsoft, ofrece capacidades avanzadas de conectividad a diversas fuentes de datos, transformación y modelado de datos, y creación de visualizaciones personalizadas, lo que permite a los usuarios construir cuadros de mando altamente efectivos y adaptados a sus necesidades específicas.

**Diseño de un Cuadro de Mando en Power BI**

El diseño de un cuadro de mando en Power BI implica varios pasos clave que aseguran que la herramienta sea efectiva, intuitiva y visualmente atractiva. A continuación, se detallan los componentes y consideraciones principales para un diseño óptimo:

**1. Identificación de Objetivos y Usuarios**

**Objetivos:**

- Determinar los objetivos específicos del cuadro de mando, tales como monitoreo de las medidas, análisis financiero, control de docentes, etc.

- Definir las preguntas clave que el cuadro de mando debe responder.

**Usuarios:**

- Identificar a los usuarios finales para así comprender sus necesidades y preferencias.

- Asegurarse de que el nivel de detalle y la complejidad del cuadro de mando sea adecuado para su audiencia.

**2. Selección de Datos y Medidas**

**Fuentes de Datos:**

- Conectar Power BI al Data Warehouse empresarial.

- Asegurar la seguridad de accesos para prevenir futuros problemas.

**Medidas:**

- Seleccionar los datos más relevantes para los objetivos definidos por el usuario.

- Definir métricas claras y calculables que reflejen la evolución y progreso de los datos.

**3. Diseño Visual**

**Diseño y Navegación:**

- Organizar el cuadro de mando de manera lógica y coherente, con una estructura clara que guíe al usuario.

- Utilizar pestañas y enlaces para permitir una navegación fluida entre diferentes vistas o niveles de detalle.

**Elementos Visuales:**

- Elegir los tipos de visualización adecuados (gráficos de barras, líneas, mapas, tablas, etc.) según la naturaleza de los datos y la información que se desea resaltar.

- Asegurar que los gráficos sean fáciles de interpretar, evitando la sobrecarga de información y priorizando la claridad.

**Estilo:**

- Mantener una paleta de colores coherente y utilizar colores con propósito (por ejemplo, verde para la sección económica y rosado para la parte de profesorado).

- Aplicar formatos de número, fechas y texto de manera consistente.

**4. Segmentación y Detalle**

**Filtros y Segmentaciones:**

- Incluir opciones de filtrado y segmentación que permitan a los usuarios explorar los datos de diferentes maneras (por tipo de financiación, período de tiempo, partida, etc.).

**Vista de detalle:**

- Implementar una vista de detalle para que los usuarios puedan profundizar en los datos y analizar la procedencia de las métricas.

**5. Pruebas y Refinamiento**

**Feedback:**

- Realizar pruebas para recoger feedback sobre la usabilidad y efectividad del cuadro de mando con diferentes usuarios.

- Ajustar el diseño y funcionalidad en base a las observaciones y sugerencias recibidas.

**Rendimiento:**

- Optimizar las consultas y modelos de datos para asegurar un rendimiento eficiente y tiempos de carga rápidos.

- Revisar y depurar cualquier problema de rendimiento o errores en los datos visualizados.

**6. Implementación y Mantenimiento**

**Publicación:**

- Publicar el cuadro de mando en Power BI Service y configurar las opciones de seguridad y acceso según las necesidades de la organización.

- Asegurar que los usuarios tienen el acceso adecuado y que el cuadro de mando está actualizado regularmente.

**Mantenimiento:**

- Establecer un plan de mantenimiento para revisar y actualizar el cuadro de mando periódicamente, asegurando que los datos y medidas reflejen los datos correctamente.

- Incorporar mejoras y nuevas funcionalidades según sea necesario para mantener la relevancia y utilidad del cuadro de mando.

Un cuadro de mando bien diseñado en Power BI no solo proporciona una visión clara y concisa del desempeño organizacional, sino que también empodera a los usuarios para tomar decisiones informadas y basadas en datos.

Dicho todo esto, lo último que falta por mencionar es que al final Power Bi es una herramienta con gran parte de diseño la cual concede al usuario una gran libertad de decisiones para poder formar un cuadro de mando único.

# **RESULTADOS**

* **Inicio:** La primera vista es la portada del cuadro de mando. Además, sirve de página principal en el que seleccionar a que categoría de los datos ir en la zona superior.

Un letrero de color blanco

Descripción generada automáticamente con confianza media

* **Economía:** Esta vista es la página principal del apartado de economía del cuadro de mando. Contiene una serie de visualizaciones que muestra la variación de gasto e ingreso económico que ha tenido la ETSISI a lo largo de los años.

Interfaz de usuario gráfica, Aplicación, Tabla, Excel

Descripción generada automáticamente

* **Economía – Detalle:** Para poder observar de donde proceden los datos mostrados en la vista anterior, se ha implementado esta vista con la que puedes ir al detalle de la economía que ha tenido la escuela.

Tabla

Descripción generada automáticamente

* **Docencia:** La vista general de la categoría de docencia. En ella se puede observar la variación de profesorado a través del tiempo, así como su variación dentro de un departamento o la distribución de categoría de docencia por año.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

* **Docencia – Detalle:** Esta vista similar a la de “Economía – Detalle”, muestra los datos de los que parte la vista anterior para que, en caso de encontrar una anomalía o algo destacado, se pueda ver que datos influyen en ello.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

# **CONCLUSIONES**

La realización de este Trabajo de Fin de Grado, basado en la creación de un cuadro de mando utilizando Power BI, apoyado por procesos ETL desarrollados con Pentaho Data Integration y soportado por una base de datos SQL en MySQL, ha sido una experiencia altamente enriquecedora. A lo largo del proyecto, he tenido la oportunidad de aplicar y profundizar en diversas tecnologías y herramientas que son fundamentales en el ámbito de los analistas de datos.

El proyecto ha resultado ser entretenido y dinámico, permitiéndome explorar tanto el aspecto técnico como el creativo del desarrollo de soluciones de procesos de negocio. La integración de diferentes tecnologías para la gestión y visualización de datos ha sido un desafío que ha contribuido significativamente a mi formación.

En términos de utilidad, el cuadro de mando desarrollado posee un valor considerable para su implementación en contextos reales. La sección de economía proporciona una visión clara y detallada de los aspectos financieros, mientras que la sección dedicada al análisis del profesorado a lo largo del tiempo permite identificar tendencias y patrones relevantes para la toma de decisiones estratégicas. Estas funcionalidades demuestran la capacidad del cuadro de mando para ofrecer información crucial de manera accesible y comprensible.

Además, el uso de Pentaho Data Integration para la creación de procesos ETL ha permitido automatizar y optimizar la extracción, transformación y carga de datos, asegurando que la información visualizada en Power BI sea precisa y esté actualizada. El soporte de la base de datos MySQL ha sido fundamental para el almacenamiento y gestión eficiente de grandes volúmenes de datos, demostrando la robustez y fiabilidad de esta plataforma.

En conclusión, este proyecto no solo ha cumplido con los objetivos planteados, sino que ha superado mis expectativas en cuanto a aprendizaje y aplicabilidad. La combinación de tecnologías utilizadas y la implementación efectiva del cuadro de mando reflejan un trabajo integral y bien fundamentado, que estoy seguro será de gran utilidad para futuras aplicaciones en la Escuela Técnica Superior de Ingeniería y Sistemas Informáticos.

# **BIBLIOGRAFÍA**

Concepto de business intelligence : <https://www.tableau.com/es-es/learn/articles/business-intelligence>

Mysql: [https://www.arsys.es/blog/mysql#Que\_es\_MySQL](https://www.arsys.es/blog/mysql" \l "Que_es_MySQL)

PowerBi: <https://learn.microsoft.com/es-es/power-bi/fundamentals/power-bi-overview>

Esquemas Data Warehouse: <https://datamanagement.es/2020/04/03/esquemas-data-warehousing/>

# **ARCHIVOS**

Los archivos del proyecto se encuentran subidos en mi GitHub personal, esto permitirá al proyecto ser más accesible para poder ser evaluado o continuado si se da el caso.

[**Link**](https://github.com/DiegoCruzCuervo/TFG)