

Universidad Nacional Autónoma de Honduras

Facultad de Ingeniería

Departamento de Ingeniería en Sistemas

Proyecto final: Data Warehouse

**Grupo:**

|  |  |
| --- | --- |
| Christhoper Daniel Portillo Reyes | 20161002154 |
| Hesler Fernando Gamez Perez | 20161001057 |

**Catedrático:**

Ing. Emilson Acosta

**Materia:**

IS-601 Base de Datos II

**Sección:**

0800

Tegucigalpa M.D.C. 17 de septiembre del 2019

# Introducción

El presente informe detalla los conceptos y la construcción de un Datawarehouse mediante los conocimientos adquiridos en la clase Base de Datos II enfocado en una empresa de aerolínea en este caso ficticia haciendo uso de una base de datos OLTP y mediante un ETL que se tuvo que programar totalmente en nuestro caso se programó en PHP y que cumpliese los requisitos establecidos en los lineamientos del proyecto y en este se cargarán los datos en una base de datos OLAP, para poder generar reportes usando la herramienta Tableau. Además, se incluye la configuración de del método de replicación en el motor de bases de datos SQL Server para simular la alta disponibilidad.

# Data warehouse

**¿Que es data warehouse?**

Un data warehouse es un repositorio unificado para todos los datos que recogen los diversos sistemas de una empresa. El repositorio puede ser físico o lógico y hace hincapié en la captura de datos de diversas fuentes sobre todo para fines analíticos y de acceso.

Normalmente, un data warehouse se aloja en un servidor corporativo o cada vez más, en [la nube](https://www.powerdata.es/el-valor-de-la-gestion-de-datos/beneficios-y-soluciones-para-un-data-warehouse-en-la-nube). Los datos de diferentes aplicaciones de procesamiento de transacciones Online (OLTP) y otras fuentes se extraen selectivamente para su uso por aplicaciones analíticas y de consultas por usuarios.

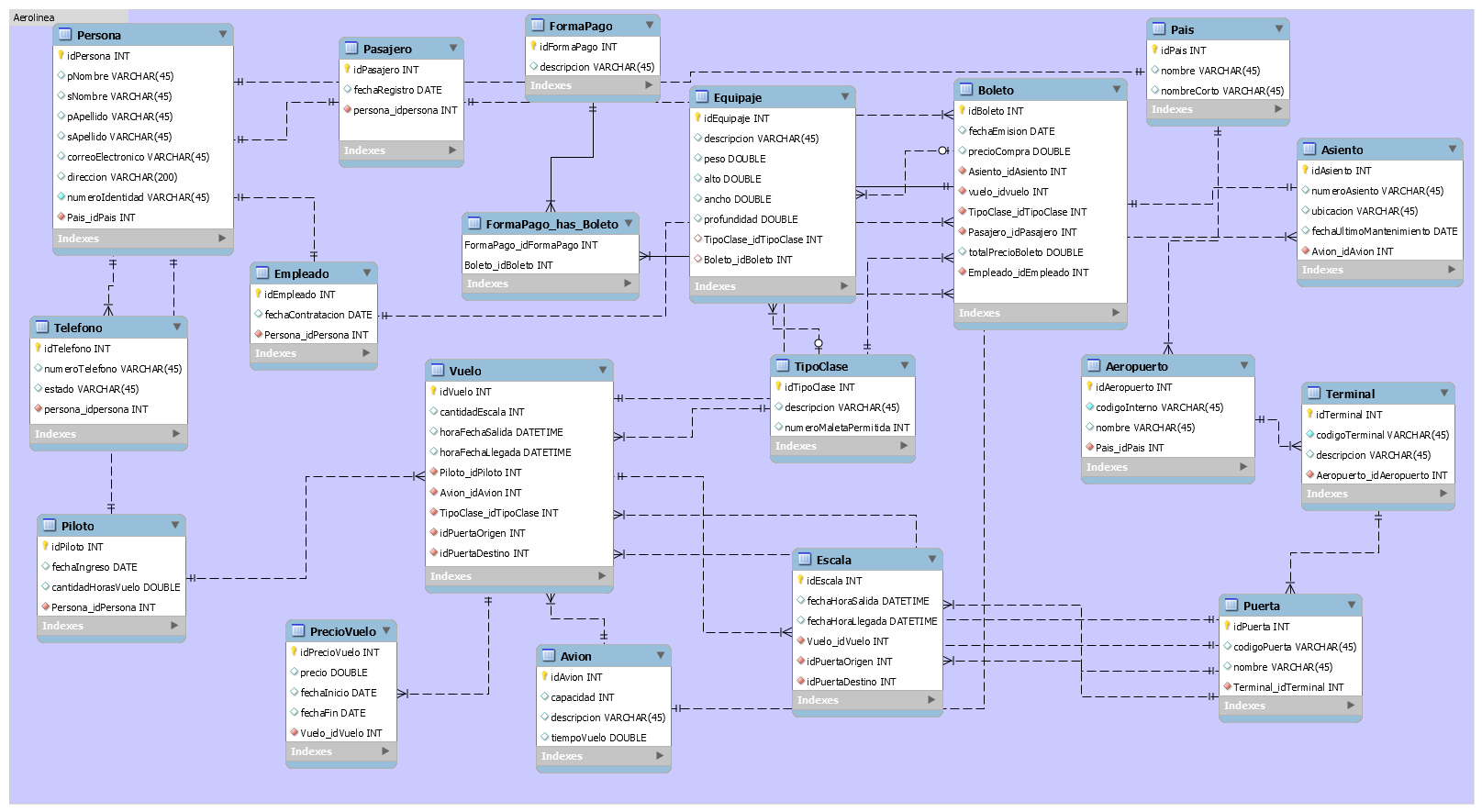
**Data mart**

En el data mart del proyecto, haciendo uso de un modelo en estrella, tomamos como métrica los boletos vendidos por la aerolínea con sus respectivas dimensiones

## Base de datos OLTP

Los sistemas OLTP (On-Line Transactional Processing) son bases de datos orientadas al procesamiento de transacciones. Una transacción genera un proceso atómico (que debe ser validado con un *commit*, o invalidado con un *rollback*), y que puede involucrar operaciones de inserción, modificación y borrado de datos. El proceso transaccional es típico de las bases de datos operacionales.

A continuación, el modelo relacional de la base OLTP del proyecto.

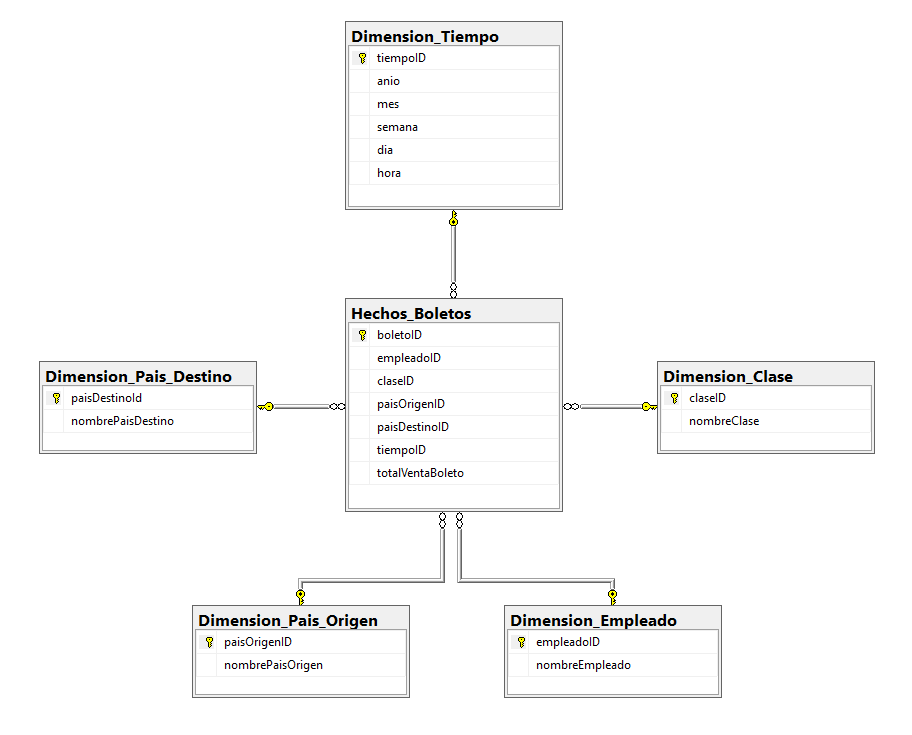


## Base de datos OLAP

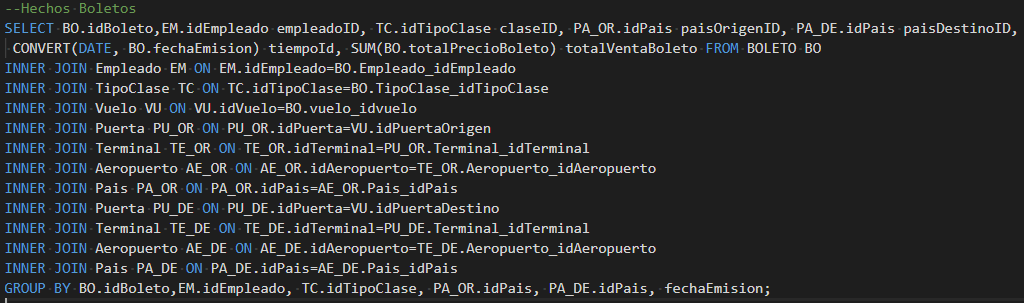
Los sistemas OLAP son bases de datos orientadas al procesamiento analítico. Este análisis suele implicar, generalmente, la lectura de grandes cantidades de datos para llegar a extraer algún tipo de información útil: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejos… etc. Este sistema es típico de los datamarts.

Las dimensiones incluidas para la base de datos OLAP en el proyecto son las siguientes:

* Clase
* Empleados
* Pais Origen
* Pais Destino
* Tiempo



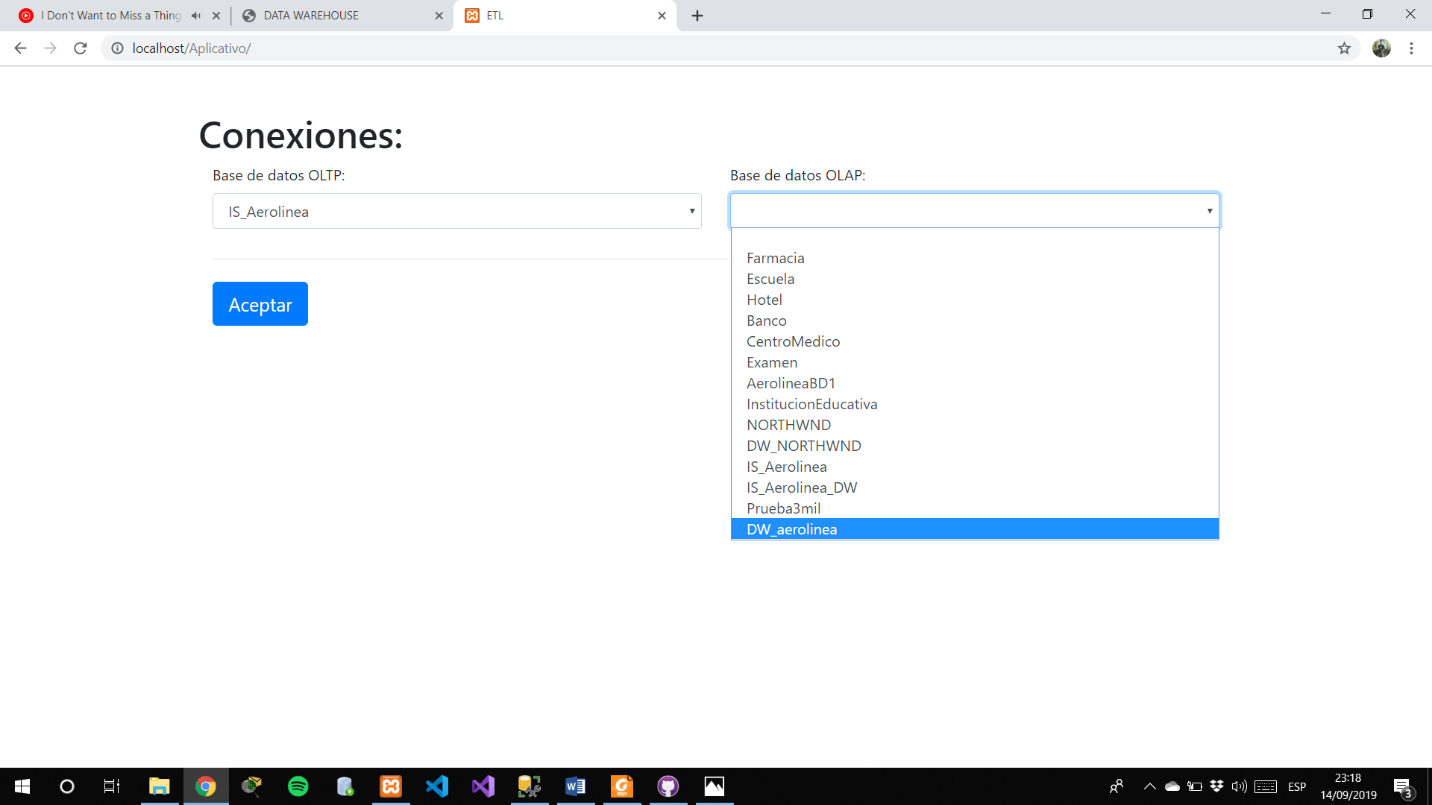
La siguiente imagen muestra la consulta final que obtiene los datos para la tabla hechos boletos de la base OLAP para la obtención necesaria de datos de cada una de las dimensiones.



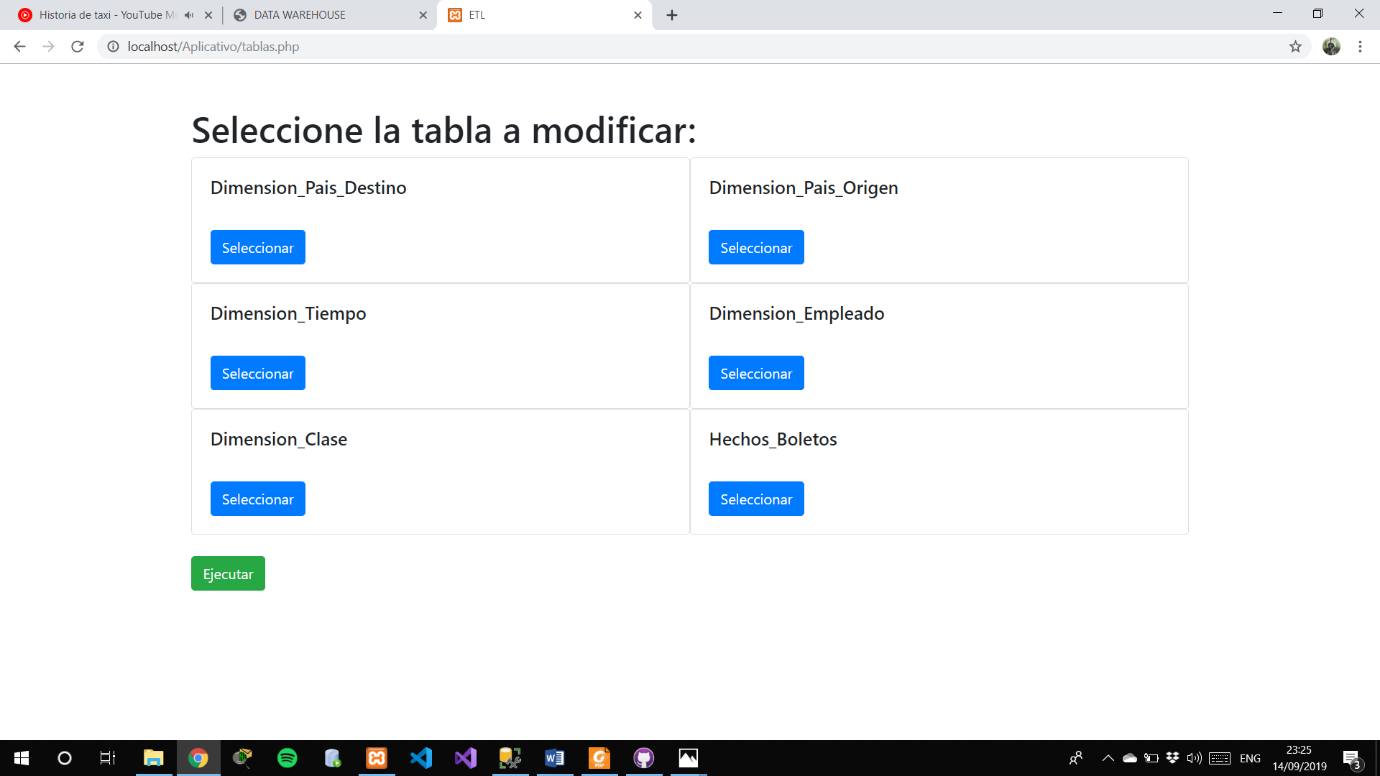
## ETL

Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y así cargarlos en otra base de datos, data mart, o data warehouse para analizar, o en otro sistema operacional para apoyar proceso de negocio.

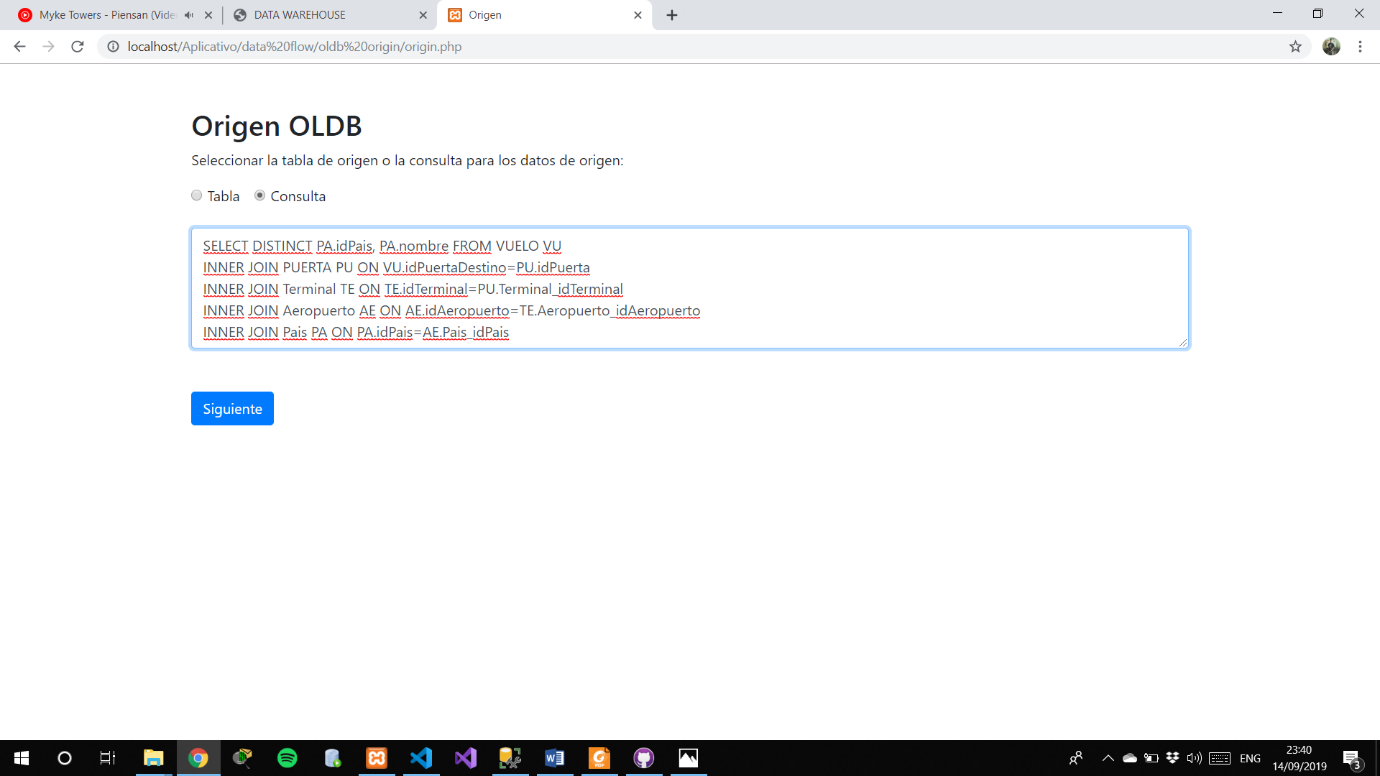
Para el ETL del data warehouse del proyecto se programó en este caso en PHP , JAVASCRIPT :

****

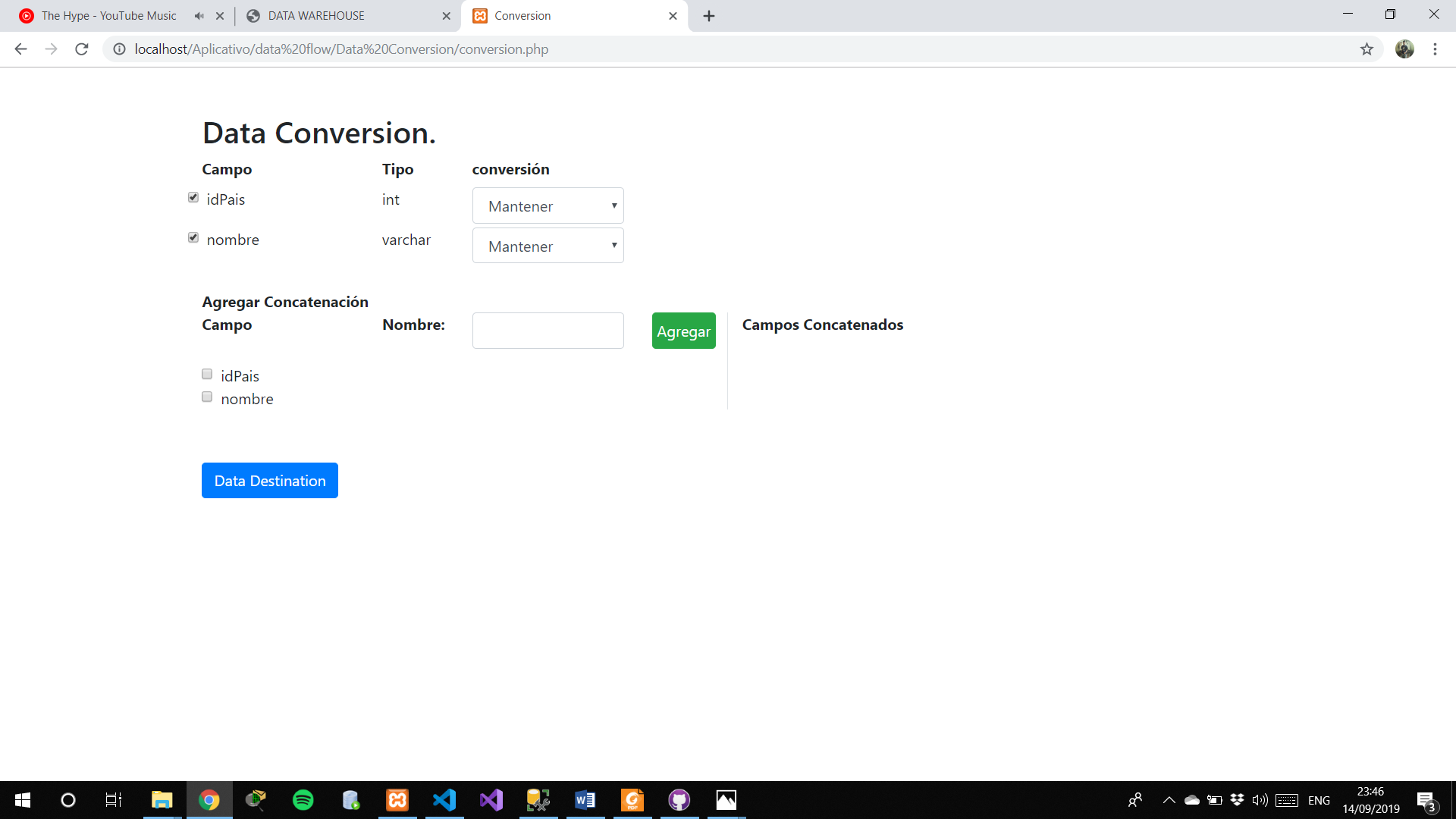
1. Se muestra la pantalla principal del aplicativo ETL donde se eligen las bases, primero la OLTP de donde se extraerán los datos para las tablas dimensiones de la base OLAP y se da click en aceptar

****

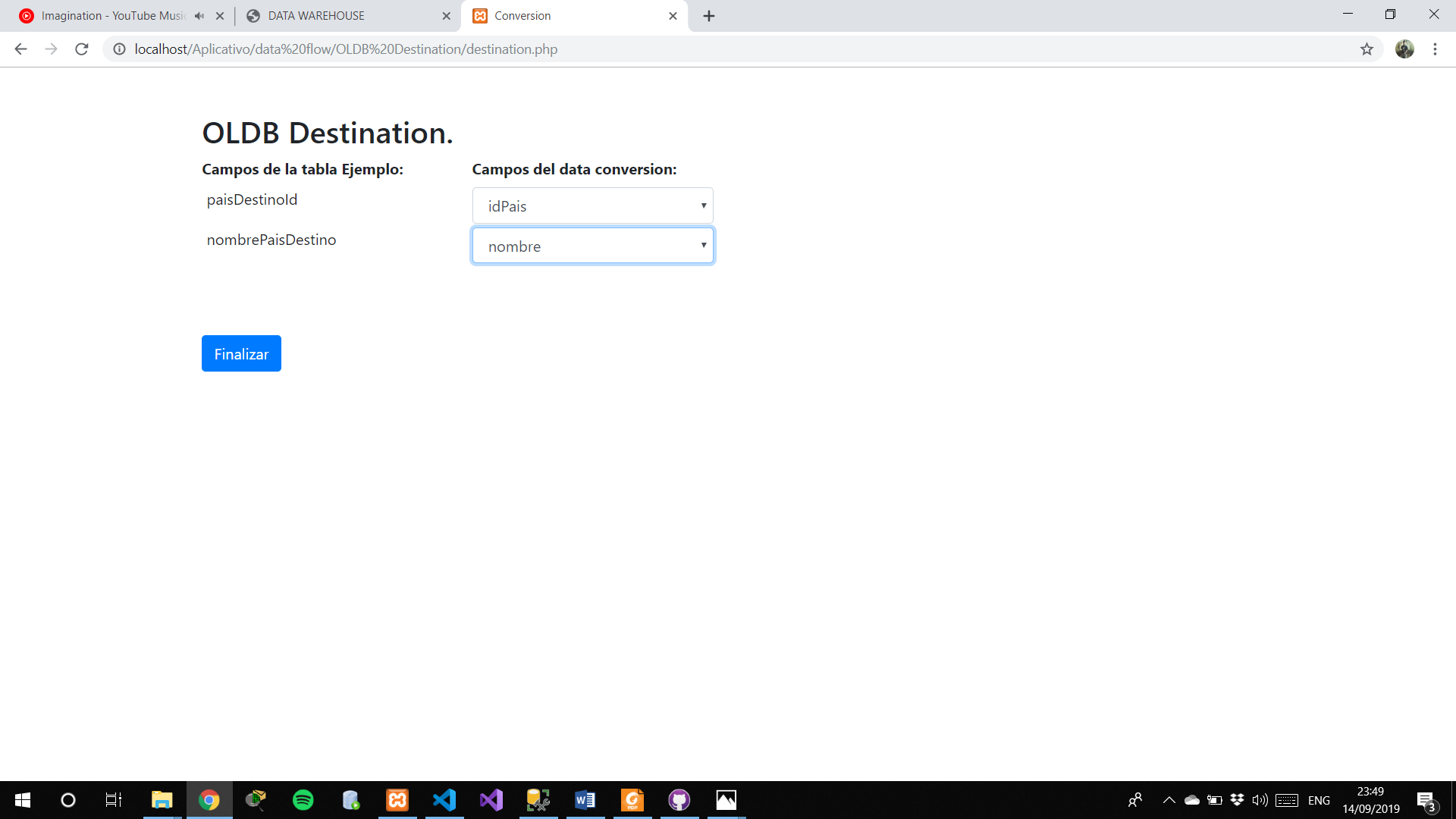
1. Se muestran las tablas de la base OLAP que vamos a llenar con la información de la base OLTP y seleccionamos una por una las dimensiones



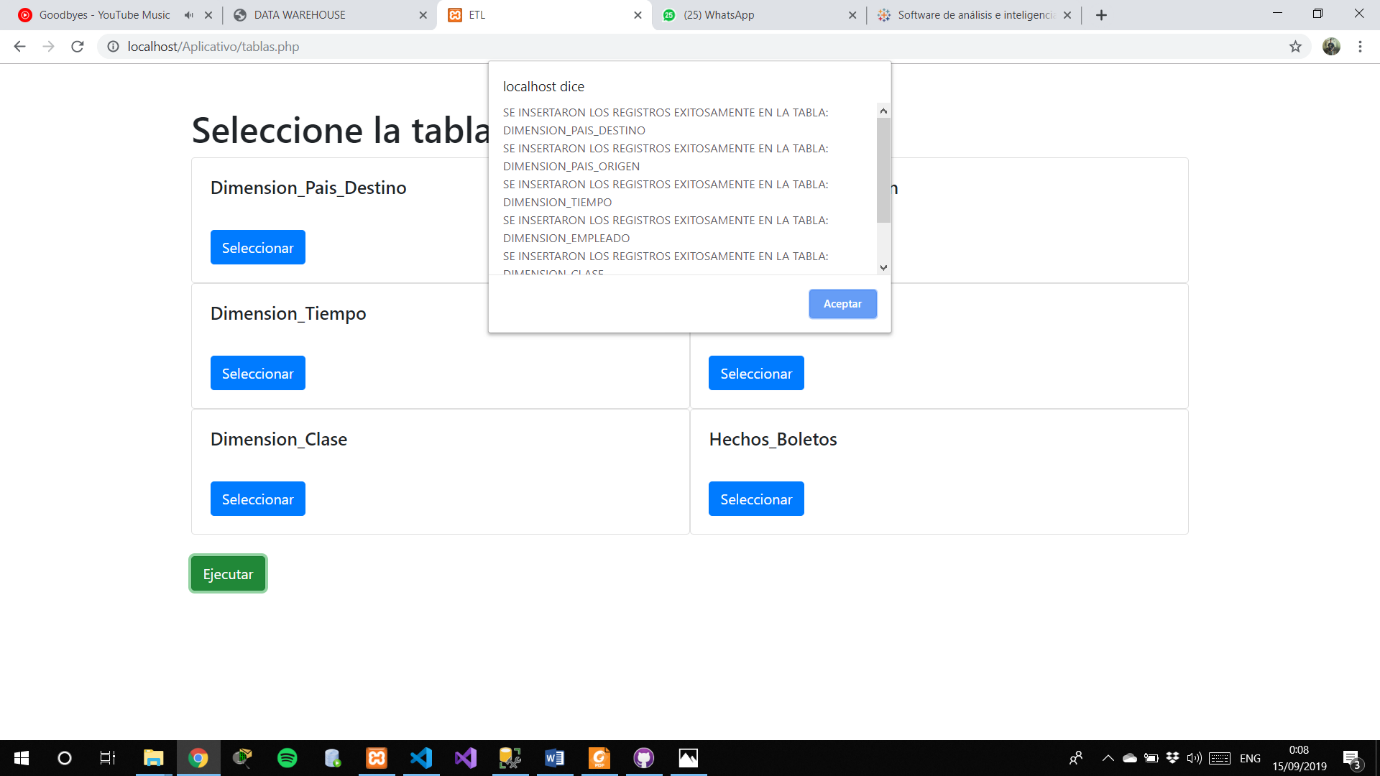
1. Se selecciona la forma de extraer los datos de la Base OLTP, en este caso lo haremos por medio de una consulta



1. Luego de elegir los campos, llegamos al Data conversion donde se pueden realizar las operaciones solicitadas por los lineamientos del proyecto



1. Luego se selecciona los campos en el data destination correspondiente a cada campo proveniente del data conversion y se repiten los demas pasos para las otras tablas de Dimension

****

1. Luego de llenar todas las tablas de la base OLAP, Se ejecuta y se hacen los respectivos inserts.

## TABLEAU

Los sistemas de reportes tienen como objetivo principal mostrar una visión general de la situación de la empresa. Consecuentemente, estos muestran la situación de las operaciones regulares de la empresa para que los directivos puedan controlar, organizar, planear y dirigir. Los reportes se pueden visualizar, exportar a otros formatos como PDF, HTML, XML, etc. y también se pueden imprimir en papel.

Con Tableau tambien podemos conectarnos a diferentes tipos de BD y con Tableau Server compartir internamente los reportes.

Ahora, en algunas ocasiones no basta tener la información online y es necesario exportar el reporte en pdf o en formato de imagen.

Esto es una tarea fácil si solo es un reporte o si no usamos filtros en la visualización, pero que pasa si tenemos 100, 1000 o 5000 reportes que debemos exportar, hacerlo manualmente definitivamente no es una opción.

Para hacer uso de la herramienta Tableau se creó primero una conexión a SQL Server mediante la base de datos del modelo OLAP, luego se creó una nueva conexión y a partir de eso se procedió a la generación de reportes.

## Alta disponibilidad - configuración

La disponibilidad es una de las características de las arquitecturas empresariales que mide el grado con el que los recursos del sistema están disponibles para su uso por el usuario final a lo largo de un tiempo dado. Ésta no sólo se relaciona con la prevención de caídas del sistema (también llamadas tiempos fuera de línea, downtime u offline), sino incluso con la percepción de “caída” desde el punto de vista del usuario: cualquier circunstancia que nos impida trabajar productivamente con el sistema – desde tiempos de respuesta prolongados, escasa asistencia técnica o falta de estaciones de trabajo disponibles – es considerada como un factor de baja disponibilidad.

Para la configuración de Alta disponibilidad:

**Preparación del servidor para la Replicación**

Primero se crearon las cuentas de Windows para replicación. Las cuales son:

* repl\_snapshots (publicador)
* repl\_logreader (publicador)
* repl\_distribuition (publicador y suscriptor)
* repl\_merge (publicador y suscriptor)

Estos usuarios de Windows se utilizaron como Agente de la replicación de la base de datos. Seguidamente se crearon las cuentas locales de Windows en el agente publicador y después se crearon en el agente suscriptor. Luego se creó la carpeta de instantáneas que es la que se utiliza para crear y almacenar la instantánea de publicación. Después se hizo la configuración de la distribución en el publicador y establecer los permisos requeridos en las bases de datos de publicación y distribución.

**Configuración de replicación transaccional**

Primero se configuró el publicador para la replicación, luego se creó una suscripción a la publicación transaccional agregando un suscriptor a la publicación que fue creada anteriormente.

**Configuración de replicación de mezcla.**

Primero se configuró un publicador para la replicación de mezcla mediante SQL Server Management Studio para publicar un subconjunto de las tablas. Luego se creó una suscripción a la publicación de mezcla que se creó anteriormente. Seguidamente se establecen los permisos en la base de datos de suscripciones y se genera manualmente la instantánea de datos filtrados para la nueva suscripción. Y por último se sincroniza la suscripción con la publicación de mezcla.

# 

# Conclusiones

1. Para diseñar una buena arquitectura de data warehouse es necesario como primer paso conocer bien los requerimientos del negocio y hacer un estudio profundo de las fuentes externas que nos van a suministrar los datos.
2. El Data warehouse ha adquirido gran aplicación en lo que respecta al manejo de datos para el buen desarrollo de una empresa.
3. Gracias al data warehouse se da una reducción de costes de almacenamiento y una mayor velocidad de respuesta frente a las consultas de los usuarios.
4. Además, hacer un buen diseño del área de transformación de datos, cuáles son las transformaciones que se van a realizar y cómo se va a implementar el modelo dimensional con sus tablas de hechos y de dimensiones es el segundo paso a seguir.

# 

# Recomendaciones

* En el ETL, para conectar con la base de datos, fue necesaria la utilización de Xampp en lugar de WampServer64 debido a incompatibilidad con los drivers y la versión de PHP.
* No tuvimos éxito al conectar la base de datos con Pentaho, sin embargo, con Tableau no hubo inconvenientes.
* Para la configuración de la replicación de la base de datos, fue necesaria la utilización de windows 10 pro y que la versión de SQL Server no sea express o Compact.

# Bibliografía

*Inteligencia de Negocios.* (2018). Obtenido de https://inteligenciadenegocios1107.wordpress.com/sistemas-de-reportes/

Josep Curto, J. C. (2018). *El conocimiento imprescindible .* Obtenido de http://reader.digitalbooks.pro/content/preview/books/43005/book/OEBPS/chapter02.xhtml

Microsoft. (2018). *Microsoft Docs.* Obtenido de https://docs.microsoft.com/es-es/sql/relational-databases/replication/tutorial-preparing-the-server-for-replication?view=sql-server-2017

<https://www.datapoetry.org/2015/07/generacion-masiva-de-reportes-en-pdf/>

*Wikipedia.* (4 de Abril de 2018). Obtenido de https://es.wikipedia.org/wiki/Cubo\_OLAP#Un\_ejemplo