

Universidad San Carlos
Facultad de Ingeniería
Escuela de Ciencias y Sistemas
Seminario de Sistemas 2
Ing. Luis Alberto Vettorazzi Espana
Ing. Fernando José Paz Gonzáles
Tutor José Alejandro Lorenty Herrera
Tutor Sergio Enrique Cubur Chalí



Proceso ETL (Análisis de Vuelos)

1. Resumen Ejecutivo

La práctica consiste en desarrollar el proceso ETL (extraer, transformar y cargar) de fuentes de datos proporcionadas que se les brindaran. Cada parte del proceso debe ser realizado con precisión para poder adaptar los datos a un modelo de inteligencia de negocios (BI) que el estudiante determinará que es el más adecuado posteriormente se podrá realizar consultas a dicho modelo para obtener información de los datos en el modelo.

2. Objetivos de Aprendizaje

2.1. Objetivo General

- Desarrollar una solución de software para poder obtener información de datos de una fuente utilizando los pasos del proceso ETL.

2.2. Objetivos Específicos

- Extraer de manera efectiva los datos de una fuente específica.
- Transformar los datos obtenidos con el fin de limpiar, homologar y estandarizar un modelo.
- Cargar los datos transformados a un destino para su futura utilización.
- Determinar en base a los datos un modelo de inteligencia de negocios adecuado para un caso específico.
- Realizar consultas a el modelo de inteligencia de negocios para obtener información.

3. Enunciado del proyecto

3.1. Descripción del problema a resolver

Los datos pueden estar en cualquier fuente, sin embargo, al momento de centralizar y tener la necesidad de almacenar todos los datos en un solo lugar buscando la estandarización y homologación de dichos datos. En los aeropuertos se desarrollan cantidades inmensas de vuelos, dichos registros de vuelos históricos se encuentran en un archivo fuente el cual debe ser procesado con ETL, en una aplicación en consola, para poder transformar los datos y cargar en un modelo para realizar su análisis mediante consultas.

3.2. Alcance del proyecto

- Implementación completa del proceso ETL.
- Diseño de un modelo de inteligencia de negocios que se adapta a las fuentes de datos solicitadas.
- Obtención de información mediante consultas en un modelo de inteligencia de negocios.

3.3. Requerimientos técnicos

- Implementación del lenguaje Python.
- Implementación de la base de datos SQL Server.
- Conocimiento de los pasos del proceso ETL.
- Conocimiento de los modelos de inteligencia de negocios.

3.4. Entregables

- **Aplicación en consola en lenguaje Python con las siguientes funcionalidades:**
 - Borrar el modelo existente
 - Crear nuevo modelo de datos
 - Extraer información de los archivos fuentes
 - Capacidad de procesar archivos de entrada específicos
 - Implementación de consultas analíticas predefinidas
- **Scripts SQL para la creación del modelo de inteligencia de negocios (modelo de datos)**
- **Documentación técnica:**
 - Diagrama del modelo realizado
 - Manual técnico de la aplicación
- **Resultados de consultas solicitadas**
 - **SELECT COUNT (*)** de todas las tablas para ver que si realizo la carga en las tablas del modelo.
 - Porcentaje de pasajeros por género.
 - Nacionalidades con su mes año de mayor fecha de salida.

Nacionalidad	01-2021	02-2021	03-2021
Guatemala	23	12	32

- **COUNT** de vuelos por país.
- Top 5 aeropuertos con mayor número de pasajeros.
- **COUNT** dividido por estado de vuelo.
- Top 5 de los países más visitados.
- Top 5 de los continentes más visitados.
- Top 5 de edades dividido por género que más viajan.
- **COUNT** de vuelos por **MM-YYYY**.

4. Restricciones

- La aplicación debe ser desarrollada con el lenguaje de programación Python.
- La aplicación debe ser una implementación en consola
- El modelo de inteligencia de negocios será creado en SQL Server
- Se deberá crear un repositorio para el desarrollo de prácticas y proyectos con el nombre (**SS2_1S2025_#carné**) crear una carpeta con relación al número de práctica o proyecto realizándose y agregar al tutor a dicho repositorio. Usuario: **BLorenty**
- Copias detectadas obtendrán una nota de 0 puntos y se reportarán a la Escuela de Ciencias y Sistemas.

5. Metodología

1. Desarrollo de la aplicación:

- Implementación de la interfaz de consola
- Desarrollo de funcionalidades ETL
- Creación de scripts SQL

2. Modelado de datos:

- Diseño del modelo dimensional
- Implementación en SQL Server
- Validación del modelo

3. Implementación de consultas:

- Desarrollo de consultas analíticas
- Validación de resultados
- Generación de reportes

6. Cronograma

- Asignación del Proyecto: 10 de febrero de 2025
 - Diseño de aplicación en consola: 2 días
 - Diseño de modelo de datos: 5 día
 - Implementación de proceso ETL: 5 días
 - Implementación de consultas SQL: 3 días
- Entrega del proyecto: 28 de febrero de 2025
- Calificación del proyecto: 01 de marzo de 2025