

Lab-Book

Data Warehouse

Beginning 9-02-2022

Jesús Baltasar Fernández

Carlos Gómez Fernández

Domenico Fasciano

Bases de Datos Avanzadas

Contents

Miércoles, 9 Febrero 2022	7
1 Introducción	7
1.1 Hito 1	7
1.2 Hito 5	7
Jueves, 10 Febrero 2022	9
1 Introducción	9
1.1 Hito 2	9
Martes, 15 Febrero 2022	11
1 Introducción	11
1.1 Hito 3	11
Jueves, 17 Febrero 2022	13
1 Introducción	13
1.1 Hito 4	13
1.2 Hito 2	13
Lunes, 21 Febrero 2022	15
1 Introducción	15
1.1 Hito 2	15
Martes, 22 Febrero 2022	17
1 Introducción	17
1.1 Hito 3	17
Miércoles, 23 Febrero 2022	19
1 Introducción	19
1.1 Hito 3	19
Jueves, 24 Febrero 2022	21
1 Introducción	21
1.1 Hito 3	21
Viernes, 25 Febrero 2022	23
1 Introducción	23
1.1 Hito 4	23

Sábado, 26 Febrero 2022	25
1 Introducción	25
1.1 Hito 4	25
Domingo, 27 Febrero 2022	27
1 Introducción	27
1.1 Hito 4	27
Lunes, 28 Febrero 2022	29
1 Introducción	29
1.1 Hito 6	29
Dificultades	31

Miércoles, 9 Febrero 2022

1 Introducción

Comenzará el desarrollo del trabajo sobre el diseño e implementación de un *Data Warehouse*, consensuando el *Hito 1*, correspondiente con la selección de tecnologías a emplear, e iniciando el *Hito 5*, correspondiente con la elaboración del cuaderno de laboratorio.

1.1 Hito 1

Tras discutir las tecnologías necesarias para el desarrollo del trabajo, se ha optado por emplear inicialmente *Python*, para la extracción, transformación y carga de datos; y *Microsoft Access*, como motor de base de datos relacional.

1.2 Hito 5

Comenzará la elaboración del cuaderno de laboratorio, en el cual se plasmarán los hitos más importantes durante el desarrollo del trabajo, haciendo uso de la plantilla *Latex* proporcionada por el profesor.

Jueves, 10 Febrero 2022

1 Introducción

Consensuando el *Hito 2*, correspondiente con la selección de la fuente de datos a emplear en el *Data Warehouse*.

1.1 Hito 2

Tras consultar diversas *webs* que mostraban *Open Data*, se seleccionó un *dataset*, en formato *Excel*, con el precio medio del metro cuadrado de suelo urbano en municipios de menos de 1.000 habitantes, obtenido de la *web* del *Ministerio de Transporte, Movilidad y Agenda Urbana*.

Martes, 15 Febrero 2022

1 Introducción

Consensuando el *Hito 3*, correspondiente con la selección y comienzo del diseño del *Data Warehouse*, pudiendo ser mediante un diseño en estrella o en copo de nieve.

1.1 Hito 3

Tras consultar diversas fuentes de información, muy probablemente, se seleccionará un diseño en estrella, ya que presenta una estructura más sencilla; al incluir muchas menos tablas, pero más grandes y con más información, a diferencia del modelo en copo de nieve, en el cual se tienen muchas más tablas, lo que obligaría a hacer muchos más *Join* y cada uno de estos ralentizaría la consulta; aún así se consultará con el profesor.

Jueves, 17 Febrero 2022

1 Introducción

Iniciando el *Hito 4*, correspondiente con el comienzo de la implementación del *Data Warehouse*, y cambio en el *Hito 2*, correspondiente con la selección de la fuente de datos a emplear en el *Data Warehouse*.

1.1 Hito 4

Inicio de la implementación del *Data Warehouse*, comenzando con la creación de la base de datos en *Microsoft Access*, previo tratamiento del *dataset* seleccionado, y las primeras pruebas de extracción, transformación y carga de datos desde un programa en *Python*.

1.2 Hito 2

Al iniciar el *Hito 4*, se analizaron en profundidad los datos seleccionados en el *Hito 2*, el pasado día 10 de febrero de 2022, los cuales eran insuficientes para la construcción del *Data Warehouse*, por lo que se optó por modificar la fuente de datos por una con una cantidad mayor tanto de atributos como de registros, la cual corresponde con el índice de precios de propiedades por ciudad desde el año 2009 hasta el año 2021.

Lunes, 21 Febrero 2022

1 Introducción

Cambio en el *Hito 2*, correspondiente con la selección de la fuente de datos a emplear en el *Data Warehouse*.

1.1 Hito 2

Al iniciar el *Hito 4*, se analizaron en profundidad los datos seleccionados en el *Hito 2*, el pasado día 17 de febrero de 2022, los cuales eran insuficientes para la construcción del *Data Warehouse*, por lo que se optó por modificar la fuente de datos por una con una cantidad mayor tanto de atributos como de registros, la cual corresponde con los condenados mayores de edad según sexo, edad, nacionalidad y número de delitos en España desde el año 2013 hasta el año 2020.

Martes, 22 Febrero 2022

1 Introducción

Consensuando el *Hito 3*, correspondiente con la selección y comienzo del diseño del *Data Warehouse*, pudiendo ser mediante un diseño en estrella o en copo de nieve.

1.1 Hito 3

Tras consultar diversas fuentes de información, muy probablemente, se seleccionará un diseño en estrella, ya que presenta una estructura más sencilla; al incluir muchas menos tablas, pero más grandes y con más información, a diferencia del modelo en copo de nieve, en el cual se tienen muchas más tablas, lo que obligaría a hacer muchos más *Join* y cada uno de estos ralentizaría la consulta; aún así se consultará con el profesor.

Miércoles, 23 Febrero 2022

1 Introducción

Consensuando el *Hito 3*, correspondiente con la selección y comienzo del diseño del *Data Warehouse*, pudiendo ser mediante un diseño en estrella o en copo de nieve.

1.1 Hito 3

Inicio del diseño del *Data Warehouse*, el cual, al ser un diseño en estrella, se comenzó con la creación de las tablas de dimensiones del mismo entre las cuales se encuentra la tabla con el género del delincuente, la nacionalidad del delincuente, la edad del delincuente, la fecha de comisión del delito y la cantidad de delitos cometidos por el delincuente.

Jueves, 24 Febrero 2022

1 Introducción

Consensuando el *Hito 3*, correspondiente con la selección y comienzo del diseño del *Data Warehouse*, pudiendo ser mediante un diseño en estrella o en copo de nieve.

1.1 Hito 3

Avance en el diseño del *Data Warehouse*, creando la tabla de hechos del mismo, la cual incluyo como atributos el género del delincuente, la nacionalidad del delincuente, la edad del delincuente, la fecha de comisión del delito, la cantidad de delitos cometidos por el delincuente y la cantidad de delincuentes que tienen todas las características.

Viernes, 25 Febrero 2022

1 Introducción

Iniciando el *Hito 4*, correspondiente con el comienzo de la implementación del *Data Warehouse*.

1.1 Hito 4

Se cargaron los archivos *Excel* en *Microsoft Access* generando la base de datos preliminar del *Data Warehouse*, además de las relaciones entre las tablas que la componen, tras lo cual se comenzó con el programa en *Python* para acceder a la misma y realizar las consultas correspondientes sobre ella. Además se dio soporte a nuestro compañero *Erasmus*, en conocimientos de la asignatura, así como en las herramientas utilizadas (*Github*, *Microsoft Access*, *Python*) y en el desarrollo de la práctica.

Sábado, 26 Febrero 2022

1 Introducción

Avanzando en el *Hito 4*, correspondiente con la implementación del *Data Warehouse*.

1.1 Hito 4

Tras obtener la versión definitiva del *Data Warehouse* se comenzó con la creación de posibles sentencias *SQL*, para analizar los datos recogidos en el mismo, las cuales se han implementado en un *script* en *Python*, el cual establece conexión con la base de datos desarrollada en *Microsoft Access* y realiza consultas sobre la misma, mostrando los resultados de las mismas en la terminal.

Domingo, 27 Febrero 2022

1 Introducción

Finalización del *Hito 4*, correspondiente con la implementación del *Data Warehouse*.

1.1 Hito 4

Finalmente, se ha optado por representar el análisis de los datos recogidos en el *Data Warehouse*, haciendo uso de una interfaz gráfica e interactiva que facilite la comprensión de los mismos, creada en el mismo *script* en *Python* que contenía las consultas *SQL*.

Lunes, 28 Febrero 2022

1 Introducción

Desarrollo del *Hito 6*, correspondiente con la elaboración de la presentación.

1.1 Hito 6

Realización de la presentación sobre el trabajo desarrollado, así como del análisis de los datos almacenados en el *Data Warehouse*, haciendo uso de la herramienta *PowerPoint*.

Dificultades

Entre las principales dificultades encontradas durante el desarrollo del trabajo ha sido la selección del *dataset* con el que trabajar.

Bibliography

- [1] Instituto Nacional de Estadística (INE), *L^AT_EX: Condenados según sexo, edad, nacionalidad y número de delitos (CONA)*. España, 2021.
- [2] RevistaDigital, *L^AT_EX: Guía para construir un DataWarehouse. Caso práctico*. Granada, España, 2017.
- [3] Manejando datos, *L^AT_EX: Conexión con MS Access desde Python*. España, 2014.
- [4] Datos para pescar, *L^AT_EX: Cómo conectar Python a la base de datos de MS Access usando Pyodbc*.
- [5] Plotly, *L^AT_EX: Gráficos básicos de la biblioteca de gráficos de código abierto de Python Plotly*. 2022.
- [6] Datacamp, *L^AT_EX: Aprenda a crear tableros en Python usando Dash..* 2018.