*JAVIER CUARTERO CORREDOR & GONZALO DE LOS REYES SÁNCHEZ*

*BASES DE DATOS AVANZADAS | ESCUELA SUPERIOR DE INFORMATICA, UCLM*

*TRABAJO 1*

DATA WAREHOUSE | ATAQUES DE CIBERSEGURIDAD

Tabla de contenido

[1. Resumen 1](#_Toc191138407)

[2. Introducción 1](#_Toc191138408)

[3. Desarrollo 1](#_Toc191138409)

[3.1 Procesamiento de datos 1](#_Toc191138410)

[3.2 Explicación tecnológica 1](#_Toc191138411)

[3.3 Diseño del Data Warehouse 1](#_Toc191138412)

[4. Resultados 1](#_Toc191138413)

[5. Usos y utilización 1](#_Toc191138414)

[6. Conclusión 1](#_Toc191138415)

# Resumen

En esta primera entrega de la asignatura se tiene como objetivo general crear una serie de programas/herramientas que permitan manejar los datos en su formato original y poder ejecutar Extracción, Transformación y Carga de datos en el Data Warehouse. Además, se propondrá un diseño en estrella con sus tablas de dimensiones y de hechos para posteriormente poder mostrar una serie de datos extraídos tal y como se ha diseñado.

# Introducción

El objetivo de este trabajo, como bien se ha comentado en el apartado anterior, es crear una serie de programas/herramientas que permitan el manejo de datos relacionados con ciberataques, implementado un proceso de ETL (Extracción, Transformación y Carga) en un Data Warehouse.

Para ello, se ha trabajado con un dataset en formato CSV denominado “Cyber Security Attacks”, el cual contiene información sobre 40000 ciberataques. Este conjunto de datos cuenta con 25 columnas, cada una representando distintas características de los ataques, como direcciones IP, tipo de protocolo, severidad, alertas generadas, entre otros.

El objetivo final es construir un Data Warehouse que facilite la consulta y el análisis de estos datos, además de permitirnos responder algunas preguntas claves cómo:

\*PONER

PREGUNTAS\*

Para lograrlo, se ha seguido un enfoque estructurado basado en lo visto en las clases de la asignatura.

# Desarrollo

## Procesamiento de datos

El primer paso en la creación del Data Warehouse ha sido el procesamiento de los datos, asegurando que estén limpios y estructurados antes de cargarlos en la base de datos.

Pasos realizados:

1. *Carga del dataset*

* Se ha importado el archivo CSV en Pandas para su análisis inicial.

*2. Limpieza de los datos*

* Identificación y eliminación de los valores nulos o inconsistentes.

Se define como valor inconsistente a cualquier dato dentro del dataset que no sigue el formato esperado, presenta contradicciones o es incoherente en el contexto del análisis de ciberataques.

* Reemplazo de los datos faltantes con valores por defecto.
* Conversión de los tipos de datos (fechas, números, cadenas).

*3. Transformaciones aplicadas*

## Explicación tecnológica

## Diseño del Data Warehouse

# Resultados

# Usos y utilización

# Conclusión

1. Resumen
2. Introducción
3. Desarrollo
   1. Procesamiento de datos
   2. Explicación tecnológica
   3. Diseño del DW
4. Resultados
   1. Para qué puede servir
5. Conclusión