



“MATERIAL ETL”

UNIVERSIDAD TECNOLÓGICA DE SAN LUIS RIO COLORADO

JULIA ELIZABETH GARCIA HERRERA

EXTRACCION DE CONOCIMIENTO EN BASES DE DATOS

FIDEL OROZCO CASTRO

IDGS9-1

14 DE OCTUBRE DEL 2024

PROCESO ETL

El proceso ETL hace referencia al proceso de extracción, transformación y carga de datos con sus siglas en ingles. Este se utiliza para recopilar datos de diversas fuentes, transformarlos al tipo de dato correcto y cargarlos a su ubicación de destino.

Las fases son las siguientes:

- **Extracción:** Se busca recopilar datos utilizando diversas técnicas. Esto puede implicar la consulta en DB, lectura de archivos, consumir API etc.
- **Transformación:** Serie de transformaciones para que los datos sean útiles y coherentes pueden incluir limpieza de datos, conversión de tipos de datos, normalización, agregación etc.
- **Carga:** Los datos anteriormente transformados se cargan en un su destino, la carga puede ser programada para que sea periódica y automática para que los datos siempre estén actualizados.

TECNICAS DE EXTRACCION DE DATOS

Se utilizan para recopilar y procesar información relevante de diversas fuentes, como bases de datos, documentos y la web. Estas técnicas son clave en análisis de mercado, investigación científica y optimización de procesos empresariales. Algunas técnicas pueden ser:

- Consultas SQL.
- ODBC/JDBC.
- Lectura de archivos planos.
- Web scrapping.
- API.
- Captura de cambios.
- Análisis de registros.
- APIs de redes sociales.
- Protocolos de comunicación IoT.
- Procesamiento de correos.

TECNICAS DE TRANSFORMACION DE DATOS

Se utilizan para preparar y limpiar los datos extraídos anteriormente de diversas fuentes antes de cargarlos a su destino o utilizarlos en un análisis. Algunas técnicas podrían ser las siguientes:

- Generalización.
- Estandarización.
- Construcción de atributos.
- Agregación.
- Suavizado.
- Ingeniería de atributos.