

Universidad Tecnológica de Chihuahua
Tecnologías de la Información



**Universidad Tecnológica
de Chihuahua**

Diagnostico Unidad 2: Preparación de datos

Alumno:

Jatzel Israel Cruz Castruita

Grupo:

IDGS91N

Materia:

Extracción de Conocimiento en Bases de Datos

Docente:

Enrique Mascote

¿Qué tipo de dato no es estructurado?

- A) CSV
- B) JSON
- C) Texto libre
- D) Tabla SQL

Un esquema “estrella” incluye:

- A) Varias tablas de hechos
- B) Una tabla de hechos y varias de dimensiones
- C) No usa dimensiones
- D) Solo tablas de dimensiones

¿Cuál técnica no sirve para imputación de valores faltantes?

- A) Media aritmética
- B) Regresión
- C) Valor constante
- D) Análisis de clusters

¿Qué parámetro de base de datos mejora el rendimiento de consultas grandes?

- A) Indexación
- B) Codificación de texto
- C) Interfaz gráfica
- D) Formato PDF

Respuesta Corta

Define brevemente qué es “outlier”.

Valor que se encuentra significativamente alejado de los demás datos en un conjunto.

Menciona dos fuentes comunes de datos semiestructurados.

JSON y XML

¿Para qué sirve particionar tablas en un data warehouse?

Mejora el rendimiento de consultas y facilita el manejo de grandes volúmenes de datos