

# Examen 1 – Base de Datos

## JSON

- Archivos en formato texto para el intercambio de información
- Independiente del lenguaje
- Fácil de entender y manipular
- Es soportado su formato en casi todas las tecnologías del backend

```
var employeeData = {  
  "employee_id": 12138,  
  "name": "Juan Otaegui",  
  "hire_date": "1/1/2013",  
  "location": "Ituzaingo, Buenos Aires",  
  "consultant": false,  
  "visited_days": [ 4,15,17,24 ]  
};
```

## Tipos de datos

- Strings (encomillados)
- Números (enteros y reales)
- Booleanos (true / false)
- Arrays
- Objetos
- Permite indicar NULL

## JSON vs XML

- Son texto plano
- Son autodescriptivos
- JSON es más liviano
  - JSON usa tipos para los objetos
  - JSON tiene menos sintaxis.
  - Las propiedad de JSON son inmediatamente accesibles en código JavaScript

## NoSQL

- Usa pocas restricciones de esquema  
No implementa el modelo relacional
- Son desestructurados
- Enfocado en distribución de datos
- Escala horizontal
- Acceso fácil por API
- No usa modelos de ACID

## Conceptos básicos

### ¿Qué es una base de datos?

- Una base de datos es una gran colección de datos relacionados (por pertenecer al mismo contexto) y almacenados para su posterior uso.
- Es una colección de datos lógicamente coherente con algún tipo de significado inherente.

**Dato:** Representación simbólica de un atributo cuantitativo o cualitativo

### **DBMS:** Database Management System

#### ¿Qué es un DBMS?

- Un **sistema de administración de datos** (DBMS, DataBase Management System) es una colección de programas **que permite a los usuarios crear y mantener una Base de Datos.**

El **DBMS** es un sistema de software de propósito general que facilita los procesos de **definición, construcción, manipulación y compartición** de bases de datos entre varios usuarios y aplicaciones.

- Autodescriptiva
- Abstracción
- Multiusuario
- **Concurrencia**

Papel fundamental en las transacciones

- **Aislamiento**
- Atomicidad
- Carga
- Copia de seguridad
- Reorganización
- Monitorización de rendimiento

### Ventajas:

- Control de redundancia
- Acceso seguro, eficiente y concurrente
- Interfaces de usuarios
- Administración uniforme de datos
- Integro
- Usa reglas para las acciones
- Backup y restore
- Implementa estándares
- Flexibilidad en mantenimiento
- Economías de escala

### Bases de datos vs Sistemas de archivos

- Posibilidad de realizar consultas a un lenguaje de 4ta generación (SQL)
- Control de redundancia.
- Control de Acceso
- Persistencia de objetos
- Backups y herramientas para migración.
- Accesos múltiples de usuarios. Concurrencia.
- Restricciones de integridad.
- Relaciones de los datos.
- Flexibilidad.
- Reduce tiempos-esfuerzos en desarrollos de sistemas.

### Normalización

#### 1er Forma

- Prohíbe atributos multivalor, compuestos y combinatorios
- Son atómicos
- Celda con único valor
- Clave primaria única
- Quita grupos repetidos

### Ejemplo no normalizado:

ID Cliente	Nombre	Teléfonos
1	Juan	12345, 67890

### En 1FN:

ID Cliente	Nombre	Teléfono
1	Juan	12345
1	Juan	67890

### 2da Forma

- Que este Normalizada en 1er Forma
- Todos los atributos que no son clave dependen completamente de la clave primaria (elimina dependencias parciales).

### Ejemplo 1FN pero no 2FN:

ID Pedido	ID Producto	Nombre Producto	Fecha Pedido
101	P01	Laptop	2024-11-20
101	P02	Monitor	2024-11-20

**Problema:** "Nombre Producto" depende solo de "ID Producto", no de la clave completa ("ID Pedido" + "ID Producto").

En 2FN: Dividimos en dos tablas:

1. Pedidos:

ID Pedido	Fecha Pedido
101	2024-11-20

2. Productos:

ID Producto	Nombre Producto
P01	Laptop
P02	Monitor

3. Detalles Pedido:

ID Pedido	ID Producto
101	P01
101	P02

3ra Forma

- Que este Normalizada en 2da Forma
- No tiene **dependencias transitivas**. (Un atributo no clave no debe depender de otro atributo no clave).

Ejemplo en 2FN pero no 3FN:

ID Cliente	Nombre Cliente	Ciudad	Código Postal
1	Juan	Buenos Aires	1001
2	Ana	Córdoba	5000

**Problema:** "Ciudad" depende de "Código Postal", no de "ID Cliente".

En 3FN: Dividimos en dos tablas:

1. Clientes:

ID Cliente	Nombre Cliente	Código Postal
1	Juan	1001
2	Ana	5000

2. Códigos Postales:

Código Postal	Ciudad
1001	Buenos Aires
5000	Córdoba

### Resumen de Reglas

1. **1FN**: Sin valores repetidos o multivaluados (Varios valores en mismas celdas).
2. **2FN**: Sin dependencias parciales (todo depende de la clave completa).
3. **3FN**: Sin dependencias transitivas (nada depende de atributos no clave).

## Machine Learning

Machine Learning es una rama de la inteligencia artificial que se centra en desarrollar algoritmos que permitan a las computadoras **aprender automáticamente a partir de datos** sin ser programadas explícitamente. El objetivo es crear modelos que puedan hacer predicciones o tomar decisiones basadas en patrones detectados en los datos. Se enfoca en **aprender y predecir**.

- ¿Como definir aprendizaje? (rae.es)
  - *Adquirir conocimiento por el estudio, la experiencia o siendo instruido (recibir instrucciones).*
    - Concebir algo por meras apariencias
  - *Ser consciente por información o por observación.*
  - Fijar algo en la memoria.
- En términos de computadoras...
- ¿Qué podemos decir?
- Podemos resumir diciendo que con ML o DM existe aprendizaje si con mayor tiempo (y datos) se logran mejores predicciones.

Tiene

- Automatización
- Predicción
- Entrenamiento
- Algoritmos comunes

### **Tipos de Aprendizaje en ML:**

- **Supervisado:**
  - Se entrena el modelo con datos etiquetados (entrada y salida conocidas).
  - Ejemplo: Clasificación de correos como "spam" o "no spam".
- **No supervisado:**
  - No hay etiquetas; el objetivo es encontrar patrones o estructuras en los datos.
  - Ejemplo: Agrupación de clientes según comportamiento.
- **Aprendizaje por refuerzo:**
  - Un agente aprende a través de prueba y error, recibiendo recompensas o castigos.
  - Ejemplo: Un robot aprendiendo a caminar.

Big Data

- Es un concepto “esquivo” que representa una cantidad de datos digitales:
  - Incómodos para almacenar
  - Incómodos para transportar
  - Pero sobre todas las cosas incómodos para **Analizar**.
- Se tratan de volúmenes tan altos que sobrepasan la mayoría de las tecnologías actuales.
- Generan desafíos para desarrollar nuevas generaciones de herramientas.

### **Data mining**

Data Mining es el proceso de **explorar y analizar grandes conjuntos de datos** para descubrir patrones, relaciones y tendencias útiles. Utiliza métodos estadísticos, matemáticos y algoritmos informáticos. Se enfoca en **descubrir patrones y extraer conocimiento**.

- Data mining es definido como el proceso de descubrimiento de patrones en los datos.
- El proceso debe ser automático (usualmente semiautomático)
- Los patrones descubiertos deben ser útiles en el sentido de poder sacar un provecho a través de su conocimiento.

Características

- **Exploración de datos:** Se enfoca en encontrar patrones en datos históricos.
- **Enfoque descriptivo:** Más orientado a describir lo que ocurrió en lugar de predecir lo que sucederá.
  - **Interdisciplinar:** Combina estadística, aprendizaje automático y bases de datos.
- **Técnicas comunes:**
  - Reglas de asociación (ej.: Market Basket Analysis).
  - Clasificación.
  - Clustering (agrupación).
  - Análisis de series temporales.

Aspecto	Machine Learning	Data Mining
Objetivo	Crear modelos predictivos o sistemas autónomos.	Descubrir patrones y relaciones en datos.
Enfoque	Predicción y toma de decisiones.	Análisis descriptivo e interpretación.
Interacción humana	Minimiza la intervención humana.	Requiere más intervención para analizar resultados.
Metodología	Entrena modelos que aprenden automáticamente.	Utiliza técnicas estadísticas y exploratorias.
Uso de datos	Generalmente datos dinámicos y en tiempo real.	Datos históricos o estáticos.

Ejemplos:

- **Machine Learning:**



- Un sistema de recomendación de películas (Netflix).
- Un modelo de predicción de enfermedades basado en datos médicos.
- **Data Mining:**
  - Descubrir que los clientes que compran pañales también suelen comprar cerveza (análisis de canasta de mercado).
  - Identificar patrones en fraudes financieros.

## Pregunta de examen

Hacer DER y MR

Hacer consultas SQL basado en puntos anteriores

Explicar cuando es 1FN, 2FN, 3FN

Las siguientes características de los sistemas NoSQL. Seleccione la opción (o las opciones) correctas(s)

- a) **Se enfoca** principalmente en datos estructurados
- b) **Utiliza pocas restricciones de esquema**
- c) **No implementa el modelo relacional**
- d) Todas las opciones anteriores son correctas los sistemas

¿Propiedades que da DBMS cuál o cuáles son?

- a) **Aislamiento**
- b) **Concurrencia**
- c) Detectar Patrones (**Es de Data Mining**)
- d) Todas las opciones anteriores son correctas