Big Data → NoSQL

Sistema tradicional → SQL

Datos rápidos, grandes o complejos que son imposible de procesar

Volumen, velocidad, variedad, veracidad

Ejemplos bigdata

apuestas deportivas, predicciones meteorológicas, trading, recomendaciones, astronomía, genética, medicina.

**Concepto 1 :** Escalabilidad horizontal y escalabilidad vertical

Escalabilidad horizontal → Añadir clusters al sistema. Coste lineal

Muchos pequeños servidores con capacidad normal

Escalabilidad vertical → Ampliar la memoria con más RAM (caro, problemas logísticos)

Un gran servidor con mucha capacidad.

**Concepto 2 :** Divide y vencerás

**Concepto 3 :** SQL y NoSQL

SQL → Base de datos relacional | Funcionan en una única máquina | Asegura la consistencia

NoSQL → Tablas aisladas | Pueden funcionar en muchas máquinas | No hay estructura de datos | No asegura la consistencia

**Tipos de datos**

Estructurados → CSV, base de datos (SQL)

No estructurados → Datos que crecen de forma específica → Word

Semiestructurales, XML, JSON Son datos que tienen la estructura y formato, aunque pueden variar.

Teorema de Cap → Los datos deben tener consistencia, disponibilidad y tolerancia a particiones pero nunca las 3 a la vez

NoSQL

Modelo de datos

* Modelo clave-valor # Solo se pueden hacer consultas sobre la clave
* Modelo documental # JSON (MongoDB)
* Modelo orientado a columnas # El más parecido a una base de datos relacionales. Puedes crear 10.000 columnas y solo usar 10 ya que está optimizado. Puede crear grupos de columnas. (Cassandra, Hbase, Amazon SimpleDB)
* Modelo de grafos # Consultas de relaciones. Se utiliza cuando hay más relaciones que usuarios, como Instagram. (Neo4j)

Data Lake → Es un entorno de datos compartido. La información se almacena en multitud de ficheros planos. Tiene todos los datos de la empresa.

**Concepto 4** : On premise - Cloud

On premise → En tus propios servidores | servidores físicos | se escala comprando servidores | Costo inicial alto + Coste de obsolescencia | No se paga por uso

Cloud → En la nube de terceros | No se dispone de servidores físicos en propiedad | Fácil de escalar | No requiere un gran coste inicial | No hay obsolescencia

**Concepto 5** : Docker

Docker → Evolución de máquina virtual. + ligero. Cuando instalas Docker instalas todo el entorno (aplicaciones, librerías, etc). Esto quiere decir que si en mi servidor Docker tengo Java16, se instalará también esa versión de Java en el Docker de quien se lo instale.

**Concepto 6** : Procesamiento en lotes - Procesamiento en tiempo real

Procesamiento en lotes → mucha cantidad cada x tiempo | No requiere interactuar con los usuarios | tareas repetitivas | gran cantidad de datos

Procesamiento en tiempo real → pocos datos todo el rato.

**Concepto 7** : Diferencia entre estadística o análisis IA

Estadística → Se formulan hipótesis

IA → Se buscan correlaciones o patrones

El camino hacia una empresa analítica

Business Intelligence (BI) Seguimiento y respuesta

Analítica de Datos Inteligencia Artificial (AI) Predecir y Actuar

Datos en crudo → Datos procesados → Informes y Cuadros de mando Predefinidos → Informes Ad HOC OLAP → Predicción genérica → Modelo Predictivo →