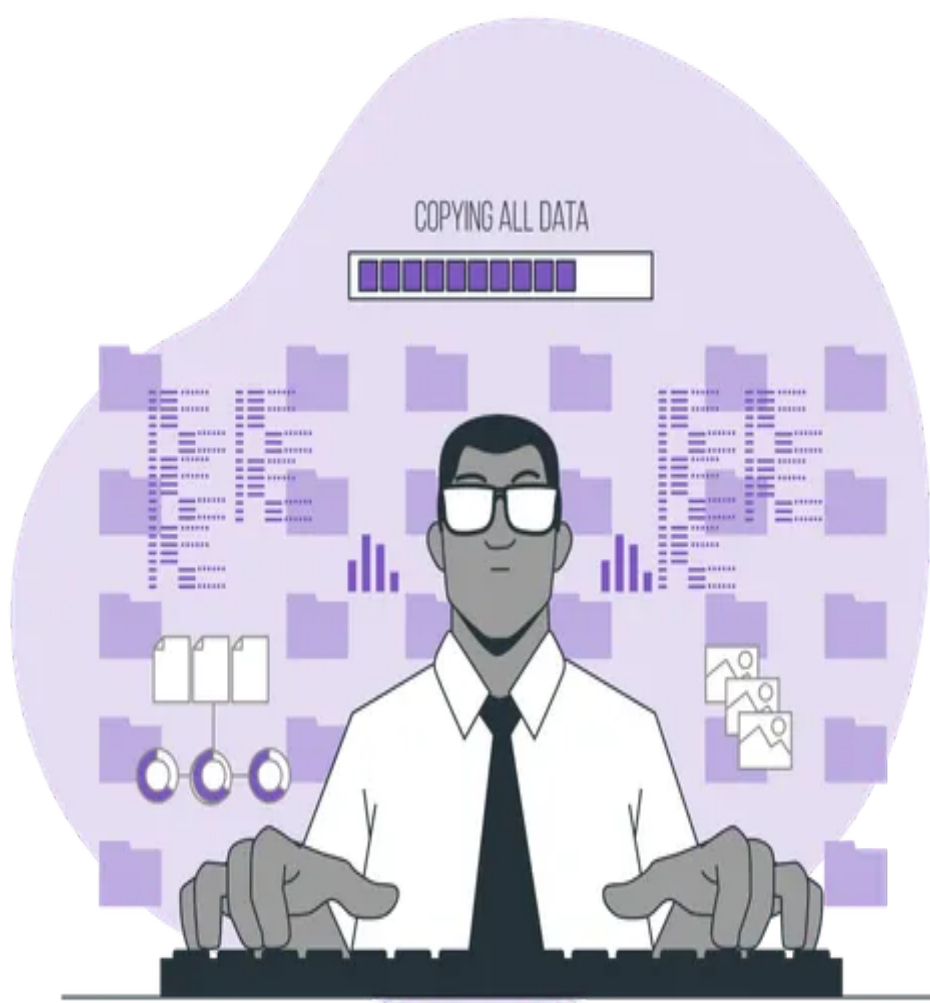


# Sistemas de Big Data



Nombre: Victoria Jiménez Martín

Módulo: Sistemas de Big Data

Curso: Especialización de Inteligencia Artificial y Big Data

# Índice

## Apartado 1: Prediseño un sistema para Big Data

<b>Crea un documento en el que explicas cómo sería el sistema a emplear para resolver las necesidades Big Data del supuesto práctico. Deberás:</b>	<b>3</b>
Indicar qué habrá que hacer para ir aumentando la capacidad del cluster según se reciben nuevos datos.	3
Indicar qué capas de la arquitectura Big Data necesitarán estar presentes como mínimo en el sistema a crear.	3
Indicar si alguna parte del sistema necesitará cumplir con las características ACID.	3
Indicar si será necesario un subsistema OLTP.	3
Indicar si será necesario un subsistema OLAP.	4
Indicar si habrá un almacén de datos.	4
Indicar qué estrategia de procesamiento habrá que emplear para poder crear el cuadro de mandos que quiere la junta directiva.	4
Indicar si será necesario crear modelos predictivos a partir de los datos.	4

# **Apartado 1: Prediseño un sistema para Big Data**

Crea un documento en el que explicas cómo sería el sistema a emplear para resolver las necesidades Big Data del supuesto práctico.

**Deberás:**

**Indicar qué habrá que hacer para ir aumentando la capacidad del cluster según se reciben nuevos datos.**

Para poder aumentar la capacidad del clúster al recibir los datos, se puede implementar un sistema escalable (como por ejemplo, Hadoop o Spark) que permitirán la adición de nodos para manejar los 100TB de datos anuales.

**Indicar qué capas de la arquitectura Big Data necesitarán estar presentes como mínimo en el sistema a crear.**

- Almacenamiento con HDFS (Hadoop Distributed File System) o un data lake.
- Procesamiento distribuido con herramientas como Spark o MapReduce para manipulación y análisis.
- Integración con sistemas OLTP para transacciones en tiempo real y OLAP para análisis.
- Gestión y manipulación de datos con herramientas como Hive, Pig o Spark SQL.
- Herramientas de visualización como Tableau, Power BI o herramientas personalizadas.

**Indicar si alguna parte del sistema necesitará cumplir con las características ACID.**

El subsistema OLTP que maneje transacciones con el cliente o las partes del sistema que manejan transacciones críticas o actualizaciones en tiempo real podrían necesitar garantías ACID. Por ejemplo, si se realizan transacciones financieras o se actualizan datos sensibles, se requerirá ACID. Con ello, garantizamos la consistencia, atomicidad, aislamiento y durabilidad de las transacciones.

**Indicar si será necesario un subsistema OLTP.**

Si necesitaremos un OLTP (Procesamiento Transaccional en Línea) debido a que es subsistema que maneja transacciones en tiempo real, como registros de usuarios, ventas instantáneas, etc.

## **Indicar si será necesario un subsistema OLAP.**

Si necesitaremos un subsistema OLAP (Procesamiento Analítico en Línea) debido a que será esencial para realizar consultas analíticas complejas sobre grandes conjuntos de datos. Ayudará a analizar los datos y proporcionar información analítica a la junta directiva.

## **Indicar si habrá un almacén de datos.**

Sí, será necesario un almacén de datos para almacenar datos procesados y estructurados, facilitando así consultas rápidas y análisis. Esto puede lograrse mediante herramientas como data warehouses o data lakes para almacenar los 500TB actuales y los 100TB adicionales anuales.

## **Indicar qué estrategia de procesamiento habrá que emplear para poder crear el cuadro de mandos que quiere la junta directiva.**

Para generar un cuadro de mandos (dashboard) para la junta directiva, se requerirá un proceso de ETL (Extract, Transform, Load) para actualizar los datos una semana antes de cada reunión.

Para poder visualizar estos datos, podríamos usar herramientas de visualización interactiva como podrían ser Tableau o Power BI para crear un cuadro de mandos que permita filtros y actualizaciones en tiempo real.

## **Indicar si será necesario crear modelos predictivos a partir de los datos.**

Si se busca aprovechar al máximo los datos disponibles, una opción sería la creación de modelos predictivos a partir de los datos para identificar patrones de comportamiento de los clientes y sugerir ofertas personalizadas. Esto implicaría utilizar técnicas de machine learning o análisis predictivo para generar modelos que puedan predecir tendencias futuras o tomar decisiones basadas en datos históricos.