

## **1. Capacidad del clúster**

Para aumentar la capacidad del clúster, hay que contar con un sistema distribuido y escalable horizontalmente. El aumento de almacenamiento se realiza añadiendo nuevos nodos de almacenamiento y procesamiento.

## **2. Capas de arquitectura necesarias**

Cómo mínimo, la estructura necesita seis capas principales:

- Ingesta: se encarga de recoger la información y transferirla al sistema.
- Almacenamiento: es un repositorio distribuido que almacena la información en bruto.
- Procesamiento: transforma los datos para añadirlos al dataset final.
- Almacén de datos: contiene los datos ya estructurados y normalizados.
- Analítica: aplica métodos de análisis y de aprendizaje automático.
- Visualización: proporciona acceso interactivo a los datos.

## **3. Uso de las características ACID**

Se deben cumplir, ya que al trabajar con Data Warehouse, existiendo un almacén de datos y realizando operaciones en el día a día, se debe asegurar que las transacciones son seguras y confiables.

## **4. Uso de subsistema OLTP**

Se requiere el uso del subsistema OLTP, ya que se realizan gestiones diarias de transacciones, tales como contrataciones, pagos, etc; y se debe garantizar que estas se realicen en tiempo real.

## **5. Uso de subsistema OLAP**

También sería necesario utilizar este subsistema para la creación de informes, análisis y consultas que se realizan con cierta frecuencia en las juntas.

## **6. Almacén de datos**

El almacén de datos será una parte imprescindible, ya que hay que guardar información histórica proveniente de distintas fuentes y permitir consultas rápidas e implementar un sistema Lakehouse para unificar almacenamiento estructurado y no estructurado.

## **7. Estrategia de procesamiento**

Se pueden añadir de forma conjunta el procesamiento por lotes, para toda la información almacenada y el análisis de esta; y procesamiento en streaming, para trabajar los datos constantes que se añaden. Este último podría ser mejor que el procesamiento en tiempo real ya que el streaming es más utilizado para el flujo constante de datos y el de tiempo real es más empleado para los análisis que no se realizan tan a menudo.

## **8. Modelos predictivos**

Los modelos predictivos se necesitan para la personalización de ofertas y el análisis de comportamiento de los clientes, permitiendo, por ejemplo, conocer momentos críticos en el que se pueden dar lugar a muchas bajas y cambiar el servicio antes de que suceda.