

# Laboratorio

Módulo 01

# Para poder realizar el Laboratorio se recomienda:

- Leer los contenidos previos.



## Data Lake: Diseño de Solución

Simular el requerimiento de un banco ficticio que nos solicita implementar un Data Lake.

Nuestra propuesta debe incluir el diseño de la arquitectura y las tecnologías que se utilizaran para cada caso de uso.



## Caso 1: Descarga de Datos Relacionales

Lo primero que nos solicitan es cargar en el Data Lake datos de sistemas transaccionales (por ej. ERP o CRM).

En una ventana de tiempo de 6 horas, desde las 2 am hasta las 8 am, ejecutar la extracción de datos sin afectar la performance de los procesos existentes.

Mantener el formato tabular y permitir consultas mediante lenguaje SQL

**ORACLE®**

**DATABASE**



## Caso 2: Análisis de Sentimiento

Analizar las publicaciones en redes sociales (por ej. Twitter) donde el banco sea mencionado.

### Tipos de análisis:

- **Histórico:** entrenamiento de un modelo de Machine Learning que se ejecuta mensualmente.
- **Tiempo real:** disponibilizar los datos en un medio de rápido acceso para consultar las tendencias.



## Caso 3: Análisis de Navegación Web

**Clientes:** registrar los eventos de ingreso a Home Banking.

**No Clientes:** identificar cuál es su comportamiento dentro del sitio para ofrecer nuevos productos y servicios.

Como estos datos son de interés para más de un sector dentro del banco, se debe garantizar la disponibilidad de los mismos para ser consumidos en un período de 7 días.



## Caso 4: Gobierno de Datos

Por la naturaleza de su industria, el banco debe cumplir con las regulaciones del país en donde opera.

Se solicita disponer de un catálogo de datos que permita auditar las diferentes fuentes de información y un portal de administración de permisos para los diferentes usuarios del Data Lake.



# ¡Muchas gracias!

¡Sigamos trabajando!