**Caso Ficticio: Empresa**

**Situación actual:**

* La empresa tiene múltiples áreas (Ventas, Finanzas, Marketing, Operaciones).
* Cada área gestiona su propia base de datos en silos.
* Hay duplicidad de datos (por ejemplo, clientes registrados en varias áreas con diferentes identificadores).
* No existe trazabilidad clara del origen de los datos.
* Los reportes son manuales y tardan días en consolidarse.
* No hay una fuente única de verdad (Single Source of Truth).

**Diagnóstico**

**Problemas identificados:**

1. **Silos de información**: Cada área trabaja de forma aislada, dificultando la integración.
2. **Falta de trazabilidad**: No se puede rastrear el origen ni la transformación de los datos.
3. **Duplicidad de datos**: Múltiples versiones de la misma entidad (cliente, producto, etc.).
4. **Baja gobernanza**: No hay reglas claras sobre calidad, acceso y uso de los datos.
5. **Tiempo de respuesta lento**: Los reportes requieren mucho esfuerzo manual.

**Solución Propuesta: Data Lake + Data Warehouse + Data Marts**

| **Criterio** | **Data Lake** | **Data Warehouse** | **Data Mart** |
| --- | --- | --- | --- |
| **Flexibilidad** | Alta (estructura libre) | Media (estructurado) | Alta (enfocado por área) |
| **Gobernanza** | Media (requiere herramientas) | Alta (modelo bien definido) | Alta (controlado por dominio) |
| **Implementación** | Rápida (para ingestión masiva) | Media (requiere modelado) | Rápida (si se basa en DW) |
| **Perfil de usuario** | Científicos de datos, TI | Analistas, BI | Usuarios de negocio |

Diagrama

Fuentes de datos ( ERP, EXCEL, Bases

Herramientas de Vizualización BI

Data Lake AZURE

Herramientas ETL

Data Wharehose bajo gobierno de datos

Data Mart - Ventas

Data Mart - Finanzas