MODULO 08 - EJERCICIO 01-A

ALEXIS YURI M.

A partir del caso planteado, describe qué tipo de integración sería mas adecuada, qué tareas y herramientas aplicarías, y qué beneficios esperas obtener. Fundamenta tu respuesta con conceptos clave vistos en clase.

1. Análisis del Caso.

Problema de datos: La organización enfrenta un problema de silos de información. Los departamentos de ventas, marketing y logística operan con bases de datos aisladas y sin conexión , lo que impide una visión integral y coordinada de las operaciones.

Riesgos: La falta de integración genera varios riesgos críticos para la operación y la toma de decisiones:

- Demoras en la generación de reportes: La información está dispersa, lo que dificulta y retrasa la consolidación de datos para los informes.

- Errores y duplicación de información: Los datos pueden estar duplicados o ser inconsistentes entre los diferentes sistemas.

- Decisiones basadas en datos inconsistentes: La información poco fiable conduce a análisis incorrectos y a la toma de decisiones ineficaces.

- Falta de visión unificada del negocio: No se puede obtener una vista completa de los clientes o procesos.

2. Elección del Tipo de Integración.

El tipo de integración más adecuado para esta problemática es la consolidación de datos.

Dado que el objetivo principal de la empresa es generar reportes y tomar decisiones basadas en datos consistentes, se elige la consolidación porque permite reunir la información en un único repositorio centralizado, como un Data Warehouse. Esto permite realizar análisis históricos y obtener una vista unificada y confiable, sin la necesidad de actualizaciones en tiempo real que podrían requerir enfoques más complejos como el streaming. La consolidación también es ideal para mejorar la calidad de los datos a través de procesos de limpieza y validación antes del almacenamiento.

La virtualización de datos permite acceder a la información de múltiples fuentes sin moverla, lo que es ideal para obtener una visión unificada en tiempo casi real. Sin embargo, el principal problema de la organización no era la necesidad de inmediatez, sino la dificultad para generar reportes, que generalmente requieren un proceso de limpieza, validación y consolidación de datos históricos.

3. Identificación de Tareas.

Para llevar a cabo la integración, se aplicará el proceso ETL (Extracción, Transformación y Carga), que es el enfoque clásico en la consolidación.

Extracción: Se extraerán los datos desde las bases de datos aisladas de ventas, marketing y logística. Dado que el objetivo es la generación de reportes, las extracciones se realizarán por lotes (batch) , probablemente de forma diaria o semanal, según la frecuencia de actualización requerida para los informes.

Transformación: Esta es la etapa más crítica. Los datos se limpiarán, validarán y normalizarán para asegurar la coherencia. Se realizarán tareas de integración semántica para unificar la interpretación de campos que tienen nombres diferentes pero el mismo significado (ej. cliente\_id en ventas y user\_id en marketing). Se eliminarán los registros duplicados y se aplicarán reglas para corregir errores o inconsistencias. Los datos se enriquecerán, si es necesario, con información de otras fuentes para generar un perfil más completo.

Carga: Los datos transformados se cargarán en un Data Warehouse central. La carga se realizará de forma incremental para solo añadir los nuevos registros o los cambios, optimizando así el proceso.

Validaciones de calidad de datos: Se aplicarán validaciones de integridad referencial para asegurar que las relaciones entre tablas sean correctas. Se harán validaciones de formato y tipo de datos (ej. unificar formatos de fecha). Se detectarán y gestionarán los valores atípicos o incorrectos.

4. Identificación de Herramientas.

Para esta solución, se pueden utilizar plataformas de integración de datos con capacidades ETL. Ejemplos de herramientas adecuadas son:

Talend Data Fabric: Una plataforma que ofrece funcionalidades de ETL, calidad y orquestación de datos.

Pentaho: Una suite de código abierto para inteligencia de negocios y ETL.

AWS Glue: Si la infraestructura está en la nube de AWS, es una excelente opción serverless para ETL.

Microsoft Azure Data Factory: Para entornos de Microsoft Azure, permite crear pipelines visuales para mover y transformar datos.

5. Propuesta de Solución y Beneficios Esperados

La organización tiene un problema de silos de información que genera demoras, datos inconsistentes y decisiones erróneas. Se necesita una arquitectura de datos que permita consolidar la información en una vista unificada.

La propuesta consiste en implementar una arquitectura de consolidación de datos utilizando el enfoque ETL. Se diseñará un Data Warehouse centralizado al que se cargarán periódicamente los datos limpios y transformados de las áreas de ventas, marketing y logística. Para ello, se utilizará una plataforma de integración de datos que automatice el proceso de extracción, transformación y carga.

Beneficios esperados:

- Visión unificada de la información: Al consolidar los datos en un solo lugar, la organización obtendrá un panorama completo de sus operaciones, eliminando las duplicidades e inconsistencias.

- Mejora en la calidad de los datos: Los procesos de validación y limpieza durante la etapa de transformación garantizarán que los datos sean precisos y confiables.

- Mayor agilidad analítica: Los equipos de negocio podrán acceder a los datos consolidados de manera más rápida y eficiente, lo que permitirá generar reportes y análisis con mayor agilidad.

- Toma de decisiones más informada: Con datos de alta calidad y una visión completa, las decisiones estarán basadas en evidencia sólida y confiable, lo que reducirá los riesgos operacionales.