

Creación de una base de datos de Staging

Estudiante:

Juan Camilo Valencia Gutiérrez

C.C 1.001.391.213

Docente:

Antonio Jesús Valderrama Jaramillo

BASES DE DATOS II

Institución Universitaria Digital de Antioquia

Septiembre 2025

INTRODUCCIÓN

En el mundo actual, donde los datos son el activo más valioso para las organizaciones, la calidad y la integridad de la información son fundamentales para la toma de decisiones acertadas. Sin embargo, a medida que las fuentes de datos se multiplican y los volúmenes de información crecen exponencialmente, garantizar la consistencia y la fiabilidad de los datos se convierte en un desafío cada vez mayor.

A continuación, se pretende mostrar la extracción como primer paso del modelo staging, una arquitectura de base de datos diseñada para abordar los retos relacionados con la integración, transformación y limpieza de datos provenientes de múltiples fuentes. El modelo staging actúa como una capa intermedia entre las fuentes de datos y el almacén de datos principal, proporcionando un entorno controlado y aislado para la preparación y validación de los datos antes de su carga en el sistema.

El presente trabajo tiene como objetivo analizar en profundidad la extracción de los datos dada una base de datos llamada “Jardinería” y sus beneficios en el proceso de gestión de datos. Se explorará cómo esta arquitectura permite a las organizaciones consolidar y armonizar datos de diversas fuentes, aplicar reglas de calidad y transformaciones necesarias, y garantizar la integridad y consistencia de la información antes de su integración en el almacén de datos principal.

Además, se examinarán las mejores prácticas y técnicas utilizadas en la extracción de los datos incluyendo el diseño de la estructura de datos, la definición de flujos de trabajo de procesamiento de datos. A través de este trabajo, se busca proporcionar una comprensión sólida del modelo

staging haciendo énfasis en la selección de los datos que queremos extraer y su impacto en el siguiente paso para la transformación de estos.

OBJETIVOS

1. Establecer conexiones seguras y eficientes con las fuentes de datos heterogéneas.
2. Definir y documentar los requisitos y reglas de extracción para cada fuente de datos.
3. Desarrollar procesos de extracción optimizados para minimizar el impacto en los sistemas operacionales.
4. Implementar mecanismos de monitoreo y registro de auditoría para el proceso de extracción.
5. Asegurar la integridad y consistencia de los datos extraídos antes de su carga en el área de staging.
6. Garantizar que el material extraído es una fuente confiable para luego hacer la transformación de los datos.

PLANTEAMIENTO DE PROBLEMA

En un escenario empresarial moderno, las organizaciones se enfrentan al desafío de recopilar y procesar datos de diversas fuentes heterogéneas, tanto internas como externas. Estas fuentes pueden incluir sistemas operacionales, bases de datos relacionales, archivos planos, fuentes de datos en la nube, entre otros. Cada una de estas fuentes puede tener su propio formato, estructura y convenciones de datos, lo que dificulta la integración y el análisis de la información.

El problema radica en la necesidad de extraer datos de estas fuentes de manera eficiente, confiable y escalable, asegurando la calidad y la integridad de los datos extraídos. La extracción de información es el primer paso crítico en un proceso ETL, y su correcta implementación es fundamental para garantizar la precisión y consistencia de los datos a lo largo de todo el flujo de trabajo.

ANÁLISIS DE PROBLEMA

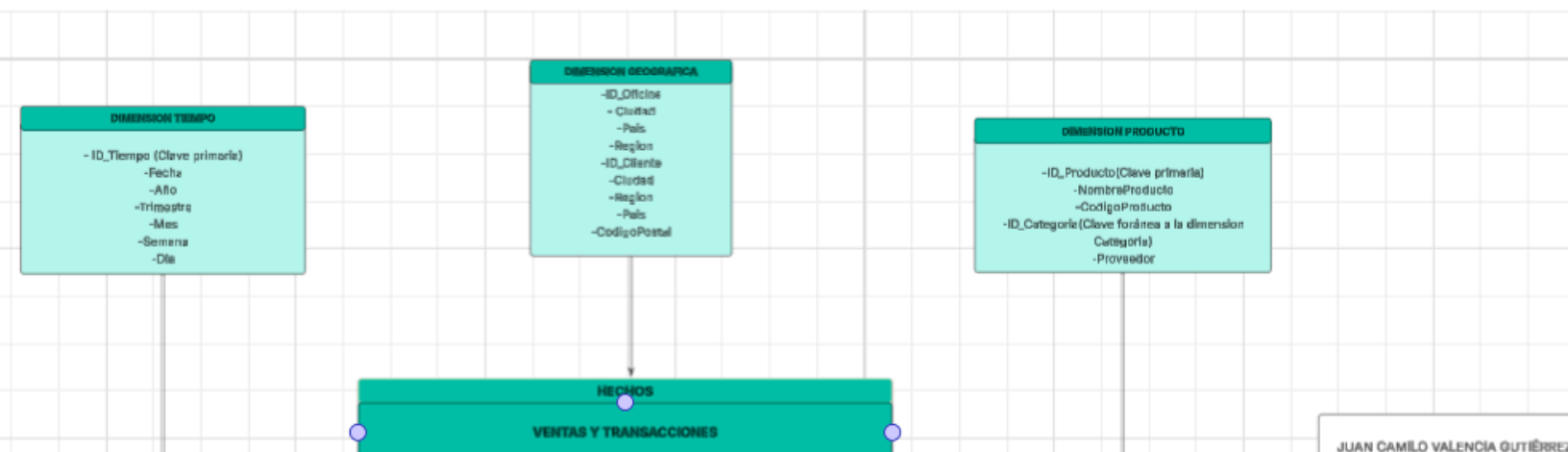
El problema principal radica en la necesidad de extraer datos de manera eficiente, confiable y escalable desde diversas fuentes de datos heterogéneas, garantizando la calidad, integridad y consistencia de los datos extraídos.

Algunos de los desafíos identificados son:

1. Conectividad y acceso a fuentes de datos heterogéneas: Cada fuente puede requerir mecanismos de conexión, autenticación y protocolos de acceso diferentes, dificultando la extracción uniforme.

2. Rendimiento y escalabilidad: A medida que los volúmenes de datos crecen, es crucial asegurar que el proceso de extracción sea eficiente y escalable, evitando enredarnos con excesos de información innecesaria y minimizando el impacto en los sistemas operacionales.
3. Integridad y consistencia de los datos: Los datos extraídos deben mantenerse íntegros y consistentes, evitando duplicaciones, pérdidas o corrupción durante la extracción.
4. Programación y monitoreo: Se requiere establecer programaciones, flujos de trabajo, mecanismos de monitoreo y registro de auditoría para la extracción y para eso usamos la herramienta “Project Integration Services”.

DISEÑO MODELO ENTIDAD-RELACIÓN CORREGIDO



ENLACE DISEÑO CORREGIDO

https://lucid.app/lucidchart/7cc3bb17-e62f-4dbe-aced-666282db973c/edit?invitationId=inv_29a17ea1-2f6a-4294-b6f9-f998a388fd18

Objetivos Actualizados diagrama estrella

- Proporcionar información precisa y oportuna para respaldar la toma de decisiones estratégicas y operativas relacionadas con las ventas.
- Mejorar el análisis y la generación de informes permitiendo a los usuarios generar informes detallados, realizar análisis y explorar los datos desde diferentes perspectivas.
- Identificar tendencias y patrones para identificar patrones, tendencias y oportunidades de ventas que pueden ser difíciles de detectar en sistemas operacionales dispersos.

- Respalda la planificación y la previsión para la planeación de la asignación de recursos y la fijación de objetivos de ventas.
- Monitorear el desempeño para facilitar la identificación de áreas de mejora.
- Mejorar la satisfacción del cliente e identificar oportunidades para mejorar la satisfacción.

Descripción del análisis realizado a los datos Jardinería y cómo estos se trasladaron a la base de datos Staging

Una vez proporcionada la base de datos “Jardinería” el primer paso fue hacer un análisis de datos para realizar el modelo estrella en un Datamart. Ya teniendo claro esto, procedemos entonces a extraer los datos más relevantes que nos proporcionarán la respuesta a las preguntas que se deben responder en la tabla de hechos. Se identifica que algunos datos relevantes como el tiempo se extraen individualizados y otros se deben agrupar para formar futuras relaciones.

El proceso de traslado se hace a través de la herramienta de Visual Studio llamada Project Integration Services, que nos proporciona una interfaz que permite conectarnos a la base de datos para así extraer, mediante un flujo de datos, la información que necesitamos de manera muy técnica, ya que podemos visualizar cómo quedarán nombrados y catalogados los nuevos datos en el staging.

BIBLIOGRAFÍA

MashaMSFT. (s/f). *SQL Server technical documentation - SQL Server*. Microsoft.com.

Recuperado el 17 de septiembre de 2025, de

<https://learn.microsoft.com/en-us/sql/sql-server/?view=sql-server-ver17>".