

# UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

## Ingeniería en Desarrollo y Gestión de Software



### Extracción de Conocimientos de Bases de Datos

#### II.4. Modelo de Datawarehouse (40%)

**IDGS91N**

PRESENTAN:

Giselle Cantú Chávez

NOMBRE DEL DOCENTE:

Ing. Luis Enrique Mascote Cano

Chihuahua, Chih., 29 de noviembre de 2025

## Índice

<b>Introducción .....</b>	<b>3</b>
<b>¿Qué es SSIS? (Visión general).....</b>	<b>3</b>
<b>Arquitectura y Componentes.....</b>	<b>3</b>
<b>Características Clave (3–5 puntos) .....</b>	<b>5</b>
<b>Ventajas .....</b>	<b>5</b>
<b>Desventajas .....</b>	<b>5</b>
<b>Proceso de Implementación DW (Resumen).....</b>	<b>6</b>
<b>Diagrama Alto Nivel (ETL con SSIS) .....</b>	<b>8</b>
<b>Retos Comunes .....</b>	<b>9</b>
<b>Recomendaciones Prácticas .....</b>	<b>10</b>
<b>Conclusiones .....</b>	<b>10</b>
<b>Fuentes de Apoyo.....</b>	<b>11</b>

## Introducción

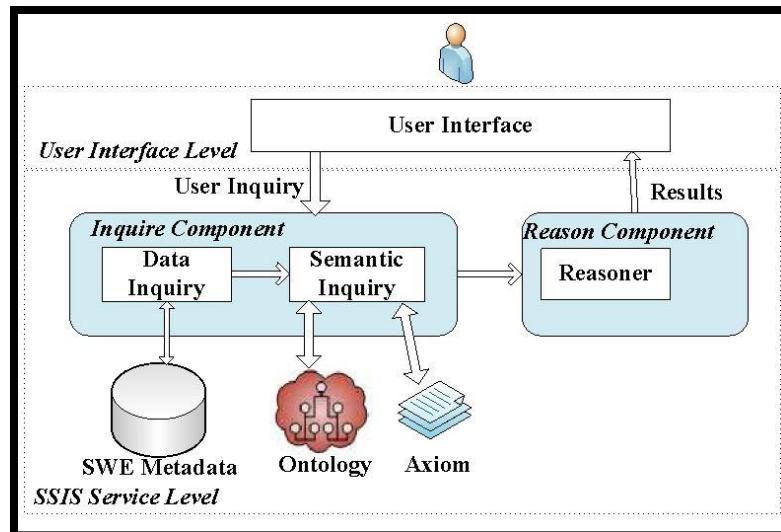
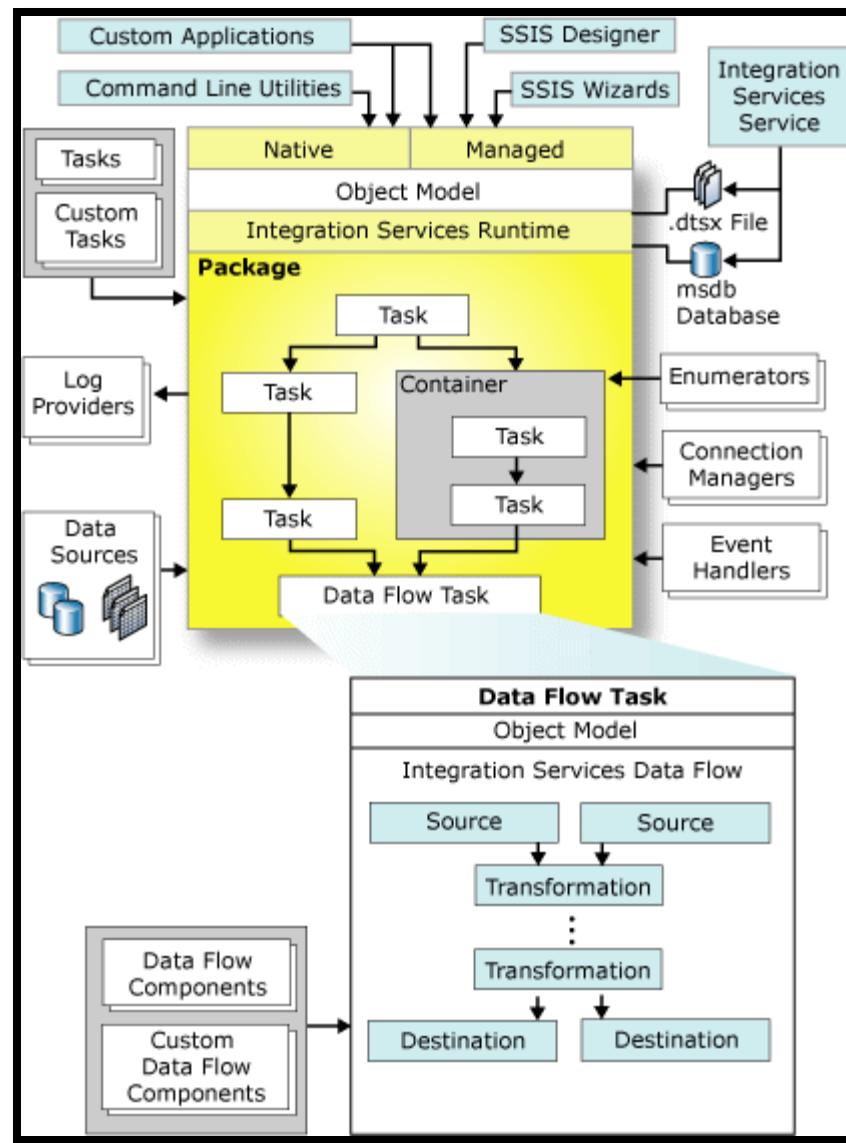
SSIS es la herramienta ETL oficial de Microsoft para construir Data Warehouses en entornos empresariales. Su propósito es integrar datos de distintas fuentes, transformarlos y cargarlos en un almacén central que permita análisis confiables, automatizados y escalables. En esta presentación describo sus características clave, ventajas, desventajas y el proceso típico para implementar un flujo ETL profesional.

### ¿Qué es SSIS? (Visión general)

- Herramienta ETL de Microsoft incluida en SQL Server.
- Permite extraer, transformar y cargar datos de múltiples formatos.
- Se usa para automatizar procesos de integración y alimentar Data Warehouses.
- Facilita flujos visuales mediante paquetes (.dtsx).
- Ideal para proyectos donde se requiere estabilidad, rendimiento y conexión nativa con SQL Server.

### Arquitectura y Componentes

- **SSIS Runtime:** motor que ejecuta paquetes y controla flujos.
- **Data Flow Engine:** procesa transformaciones y movimientos de datos.
- **Control Flow:** secuencia general de tareas del ETL.
- **Data Flow Tasks:** actividades de extracción, transformación y carga.
- **SSIS Catalog (SSISDB):** repositorio para despliegue, auditoría y seguridad.



### Características Clave (3–5 puntos)

- **Integración visual:** diseño de flujos mediante interfaz gráfica sin tener que programar todo.
- **Conectores nativos:** Excel, SQL Server, Oracle, archivos planos, APIs, etc.
- **Transformaciones avanzadas:** merge, lookup, derived columns, pivot, calidad de datos.
- **Automatización total:** ejecución programada con SQL Server Agent.
- **Escalabilidad:** maneja grandes volúmenes con particiones y buffers optimizados.

### Ventajas

1. **Integración perfecta con SQL Server**, sin configuraciones externas.
2. **Muy buen rendimiento** en cargas masivas.
3. **Gran cantidad de conectores y transformaciones listas para usar.**
4. **Automatización sólida** mediante SQL Server Agent.
5. **Monitoreo profesional** con SSISDB (logs, auditorías, versionamiento).

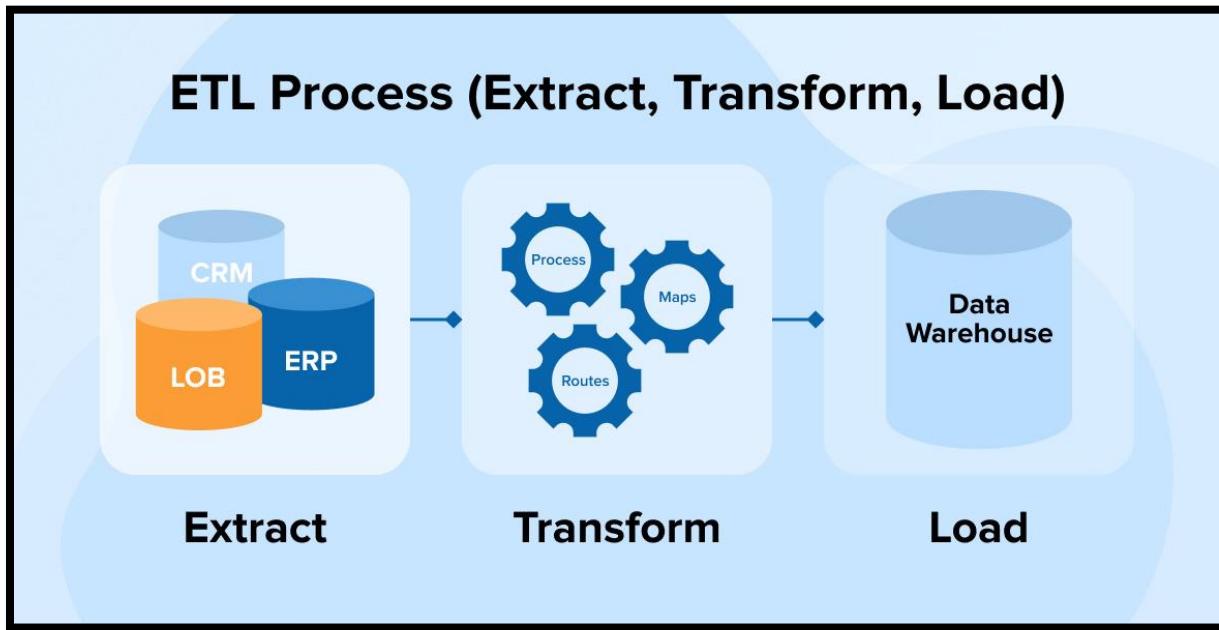
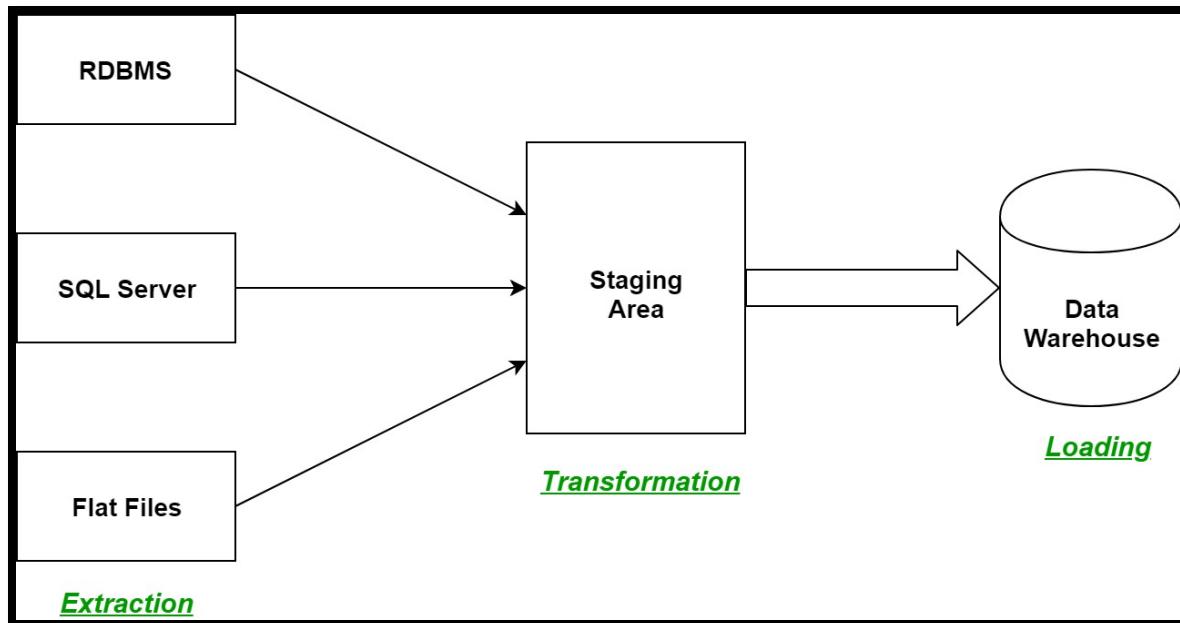
### Desventajas

1. Depende del ecosistema Windows/SQL Server.
2. No es la mejor opción para arquitecturas 100% en la nube.
3. Curva de aprendizaje inicial para flujos complejos.
4. Mantenimiento de paquetes grandes puede volverse pesado.

5. Menos flexible para pipelines basados en microservicios modernos.

### Proceso de Implementación DW (Resumen)

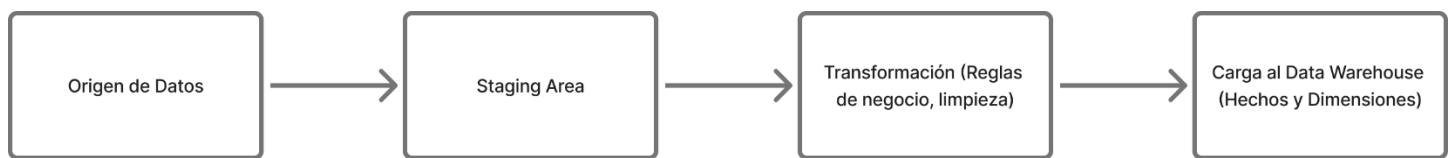
- **Extracción:** obtener datos desde sistemas transaccionales, Excel, APIs o bases externas.
- **Staging:** área temporal para almacenar datos sin procesar.
- **Transformación:** limpieza, unificación, validación y reglas de negocio.
- **Carga (Load):** inserción en tablas dimension, hechos o modelos estrella/copo de nieve.



## Diagrama Alto Nivel (ETL con SSIS)

**Origen → Staging → Transformación → DW (Hechos y Dimensiones)**

- Se extraen datos desde múltiples fuentes.
- Se almacenan temporalmente para validación y limpieza.
- SSIS aplica transformaciones, joins, cálculos y reglas.
- Finalmente se llena el Data Warehouse.



## Retos Comunes

- Datos inconsistentes o incompletos.
- Cambios en la estructura de las fuentes (columnas nuevas/eliminadas).
- Grandes volúmenes que requieren particiones o cargas incrementales.
- Manejo de errores y registros fallidos.
- Necesidad de monitoreo constante en producción.

## Recomendaciones Prácticas

- Definir un **staging robusto** antes de procesar datos.
- Implementar **auditorías** y logs desde el inicio.
- Separar paquetes por módulos (extract, transform, load).
- Preferir cargas incrementales para eficiencia.
- Documentar las reglas de negocio del ETL.

## Conclusiones

SSIS es una herramienta de integración empresarial completa, estable y altamente escalable, ideal para Data Warehouses que dependen del ecosistema SQL Server. Su arquitectura basada en flujos visuales, automatización y transformaciones avanzadas lo convierte en una solución sólida para organizaciones que requieren procesos ETL confiables. Aunque no siempre es la mejor opción en entornos cloud-native, su rendimiento y compatibilidad lo siguen posicionando como una de las herramientas más utilizadas en la industria.

## Fuentes de Apoyo

Microsoft. (2024). *Integration Services (SSIS) documentation*. Microsoft Learn.

<https://learn.microsoft.com/en-us/sql/integration-services>

Microsoft. (2023). *SSIS packages and projects*. Microsoft Learn. <https://learn.microsoft.com/en-us/sql/integration-services/packages>

Ullah, W., Khan, Q., & Kim, D. (2023). *Efficient ETL pipeline strategies for data warehouse environments*. International Journal of Data Engineering, 12(4), 55–72.

<https://doi.org/10.30436/ijde.2023.12.4.55>

Basu, A. (2022). *Designing data pipelines: ETL patterns and best practices*. O'Reilly Media.

<https://www.oreilly.com/library/view/designing-data-pipelines>

Kimball, R., & Ross, M. (2023). *The Data Warehouse Toolkit* (4th ed.). Wiley.

<https://www.wiley.com/en-us/The+Data+Warehouse+Toolkit>