

TRABAJO ACADEMICO Nro. 9

Escuela Profesional:	Ingeniería de Sistemas	Asignatura:	Gestión de Datos e Información II
Ciclo:	V	Turno:	Mañana/Tarde
Semestre Académico:	2022-1		
Docentes:	Ing. Gustavo Coronel Castillo	Fecha:	Del 30/05/2022 al 03/06/2022

Integrando Datos 2

I. INTRODUCCION.....	2
Presentación	2
Fase de Extracción.....	2
Fase de Transformación	3
Proceso de Carga	3
Aplicaciones de los procesos ETL	4
II. OBJETIVO.....	5
III. METODOLOGIA Y ACTIVIDADES.....	5
IV. RECURSOS	5
V. OBSERVACIONES	6
VI. REQUERIMIENTO A RESOLVER	6

I. INTRODUCCION

Presentación

Los procesos ETL son una parte de la integración de datos, pero es un elemento importante cuya función completa el resultado de todo el desarrollo de la cohesión de aplicaciones y sistemas.

La palabra ETL corresponde a las siglas en inglés de:

- Extraer: **extract**
- Transformar: **transform**
- Cargar: **load**

Con ello, queremos decir que todo proceso ETL consta precisamente de estas tres fases: extracción, transformación y carga. Vamos a definir en qué consisten cada una de estas fases.

Fase de Extracción

Para llevar a cabo de manera correcta el proceso de extracción, primera fase de los procesos ETL, hay que seguir los siguientes pasos:

- Extraer los datos desde los sistemas de origen.
- Analizar los datos extraídos obteniendo un chequeo.
- Interpretar este chequeo para verificar que los datos extraídos cumplen la pauta o estructura que se esperaba. Si no fuese así, los datos deberían ser rechazados.
- Convertir los datos a un formato preparado para iniciar el proceso de transformación.

Además, uno de las prevenciones más importantes que se deben tener en cuenta durante el proceso de extracción sería el exigir siempre que esta tarea cause un impacto mínimo en el sistema de origen. Este requisito se basa en la práctica ya que, si los datos a extraer son muchos, el sistema de origen se podría ralentizar e incluso colapsar, provocando que no pudiera volver a ser utilizado con normalidad para su uso cotidiano.

Fase de Transformación

La fase de transformación de los procesos ETL aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. Estas directrices pueden ser declarativas, pueden basarse en excepciones o restricciones, pero, para potenciar su pragmatismo y eficacia, hay que asegurarse de que sean:

- Declarativas.
- Independientes.
- Claras.
- Inteligibles.
- Con una finalidad útil para el negocio.

Proceso de Carga

En esta fase, los datos procedentes de la fase anterior (fase de transformación) son cargados en el sistema de destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones diferentes.

Existen dos formas básicas de desarrollar el proceso de carga:

1. **Acumulación simple:** esta manera de cargar los datos consiste en realizar un resumen de todas las transacciones comprendidas en el período de tiempo seleccionado y transportar el resultado como una única transacción hacia el data warehouse, almacenando un valor calculado que consistirá típicamente en un sumatorio o un promedio de la magnitud considerada. Es la forma más sencilla y común de llevar a cabo el proceso de carga.
2. **Rolling:** este proceso sería el más recomendable en los casos en que se busque mantener varios niveles de granularidad. Para ello se almacena información resumida a distintos niveles, correspondientes a distintas agrupaciones de la unidad de tiempo o diferentes niveles jerárquicos en alguna o varias de las dimensiones de la magnitud almacenada (por ejemplo, totales diarios, totales semanales, totales mensuales, etc.).

Sea cual sea la manera elegida de desarrollar este proceso, hay que tener en cuenta que esta fase interactúa directamente con la base de datos de destino y, por eso, al realizar esta operación se aplicarán todas las restricciones que se hayan definido en ésta. Si están bien definidas, la calidad de los datos en el proceso ETL estará garantizada.

Aplicaciones de los procesos ETL

Gracias a los procesos ETL es posible que aplicarlos para:

- Mover datos desde una o múltiples fuentes.
- Reformatear esos datos y los limpie, cuando sea necesario.
- Cargar en otro lugar como una base de datos, un data mart o un data warehouse.
- Una vez alojados en un destino, esos datos se puedan analizar.
- Cuando ya están cargados en su ubicación definitiva, se empleen en otro sistema operacional, para apoyar un proceso de negocio.

No obstante, las herramientas ETL no tienen por qué utilizarse sólo en entornos de Data Warehousing o construcción de un Data Warehouse, sino que pueden ser útiles para multitud de propósitos, como, por ejemplo:

- Tareas de Bases de datos: que también se utilizan para consolidar, migrar y sincronizar bases de datos operativas.
- Migración de datos entre diferentes aplicaciones por cambios de versión o cambio de aplicativos.
- Sincronización entre diferentes sistemas operacionales (por ejemplo, entre el entorno ERP y la web de ventas).
- Consolidación de datos: sistemas con grandes volúmenes de datos que son consolidados en sistemas paralelos, ya sea para mantener históricos o para llevar a cabo procesos de borrado en los sistemas originales.
- Interfases de datos con sistemas externos: como el envío de información a clientes o proveedores. También servirían para la recepción, proceso e integración de la información recibida.
- Interfases con sistemas Frontoffice: serían interfases de subida/bajada con sistemas de venta.
- Otros cometidos: como la actualización de usuarios a sistemas paralelos o la preparación de procesos masivos (tipo mailings o newsletter).

II. OBJETIVO

Diseñas e implementar un proceso ETL para cargar datos en la base dimensional desarrollada en la tarea académica nro. 8.

El origen de datos es la base de datos PUBS.

III. METODOLOGIA Y ACTIVIDADES

Se asume que ya tiene implementada la base de datos PUBS y la base de datos dimensional elaborada en la tarea académica nro. 8.

Debe seguir los siguientes pasos:

1. Defina el origen de datos.
2. Defina el destino de datos.
3. Implemente las transformaciones necesarias.
4. Ejecute el proceso completo.
5. Verifique que funciona correctamente.
6. Recoger evidencias.
7. Elaborar informe que incluya conclusiones y recomendaciones.

IV. RECURSOS

1. Acceso a internet.
2. Computadora personal con Windows 10 o superior
3. Microsoft SQL Server correctamente instalado.
4. Microsoft SQL Server Management Studio
5. Visual Studio
6. Plataforma Blackboard para la entrega de su solución.

V. OBSERVACIONES

1. Se debe crear una **carpeta de trabajo** con el nombre "**\GDI2\GPO##TA09**", en donde grabara todos los archivos de este trabajo académico. Donde **##** representa el número de grupo, por **01**, **02**, **03**, y así sucesivamente.
2. Para finalizar, debe empaquetar la carpeta **GPO##TA09** en un archivo **RAR** o **ZIP**, este archivo empaquetado es el que debe subir a la plataforma Blackboard en la sección correspondiente.

VI. REQUERIMIENTO A RESOLVER

Crear el proceso ETL para cargar datos al modelo dimensional desarrollado en la tarea académica 8.



El proceso ETL que debe desarrollar debe cargar datos de la base de datos **PUBLS** a la base de datos dimensional **PUBSDM**.

Forman parte de la entrega:

- Script para crear la base de datos PUBLS.
- Archivo Erwin del diseño dimensional.
- Script para crear la base de datos dimensional PUBSDM.
- Proyecto SSIS.
- Informe
- Otros que considere necesarios