

UNIVERSIDAD DON BOSCO
Dirección de Educación a Distancia
Ingeniería en Ciencias de la Computación



Datawarehouse y Minería de Datos
Ing. Karens Medrano

Grupo Teórico I
Desafío Práctico I

Nombre: Oscar Alexander Guevara Rodríguez
Carnet: GR222756

Documentación de proyecto ETL y modelo estrella

A continuación, se detalla el diseño del modelo dimensional de estrella, así como también el proceso de extracción, transformación y carga para poblar el Data Warehouse. El objetivo principal es consolidar los datos de ventas de la base de datos transaccional Chinook.

Tabla de hechos (FactSales)

En este caso, la tabla FactSales es el centro del modelo y su propósito es almacenar las ventas o medidas.

Dimensiones

Se provee el contexto, de “quién, qué y cuándo”, con cada una de las dimensiones con atributos principales para filtrar, agrupar y describir datos

- **DimCustomer:** Contiene atributos de clientes que realizaron compras. Da el contexto de "quién". También se unificó FirstName y LastName en un solo campo FullName
- **DimProduct:** Consolida la información de las pistas musicales, que originalmente se encontraba dispersa. Esta dimensión provee el contexto de “qué” se vendió.
- **DimDate:** Es una dimensión generada para proveer el contexto de "cuándo" ocurrió cada venta.

Proceso ETL

El proceso ETL se construyó utilizando SQL Server Integration Services (SSIS) para automatizar el flujo de datos desde la base de datos de origen “Chinook” hasta el Data Warehouse “DWChinook”.

- **Tareas Específicas:**

Cargar DimCustomer

Extracción: Se extraen todos los datos de la tabla “Customer”

Transformación: Se utiliza una concatenación entre FirstName y LastName para un nuevo campo llamado FullName

Carga: Los datos transformados se cargan en la tabla DimCustomer del DW.

Cargar DimProduct

Extracción: Se utiliza un OLE DB Source con un SQL Statement para la extracción de datos. Realizando un JOIN entre las tablas Track, Album, Artist, Genre y MedioType, consolidando toda la información en una sola fuente.

Transformación: No requerida.

Carga: Los datos consolidados se cargan en DimProduct

Poblar DimDate

Se utiliza para crear una tabla de fechas, de forma completa y estática, también se ejecuta un T-SQL script que utiliza una CTE Recursiva para generar un registro en un rango de (2000-2030). Almacenando y calculando el DateKey de cada fecha en formato YYYYMMDD.

Cargar FactSales

Se extraen los datos de Invoice e InvoiceLine, y se unen para obtener el detalle de cada venta.

Transformación: se utilizan tres transformaciones de Lookup:

1. El primer Lookup toma el InvoiceDate y lo busca en DimDate obteniendo el DateKey Correspondiente
2. Toma el CustomerId y lo busca en DimCustomer para obtener el CustomerKey.
3. Toma el TrackId y lo busca en DimProduct para obtener el ProductKey

Carga de datos: El flujo de datos se carga en la tabla FactSales de DWChinook.

SQL Statements utilizados:

Total ventas por cliente:

```
SELECT
    c.FullName,
    SUM(fs.TotalAmount) AS TotalSales
FROM
    FactSales AS fs
JOIN
    DimCustomer AS c ON fs.CustomerKey = c.CustomerKey
GROUP BY
    c.FullName
ORDER BY
    TotalSales DESC;
```

Total ventas por género:

```
SELECT
    p.Genre,
    SUM(fs.TotalAmount) AS TotalSales
FROM
    FactSales AS fs
JOIN
    DimProduct AS p ON fs.ProductKey = p.ProductKey
GROUP BY
    p.Genre
ORDER BY
    TotalSales DESC;
```

Total ventas por género musical

```
SELECT
    p.Genre,
    SUM(fs.TotalAmount) AS TotalSales
FROM
    FactSales AS fs
JOIN
    DimProduct AS p ON fs.ProductKey = p.ProductKey
GROUP BY
    p.Genre
ORDER BY
    TotalSales DESC;
```

Total ventas por artista

```
SELECT
    p.ArtistName,
    SUM(fs.TotalAmount) AS TotalSales
FROM
    FactSales AS fs
JOIN
    DimProduct AS p ON fs.ProductKey = p.ProductKey
GROUP BY
    p.ArtistName
ORDER BY
    TotalSales DESC;
```

Total ventas por país

```
SELECT
    c.Country,
    SUM(fs.TotalAmount) AS TotalSales
FROM
    FactSales AS fs
JOIN
    DimCustomer AS c ON fs.CustomerKey = c.CustomerKey
GROUP BY
    c.Country
ORDER BY
    TotalSales DESC;
```

Imágenes del proceso





