

TABLA DE HECHOS Y DIMENSIÓN "COSTUMER"

TABLA DE HECHOS Y DIMENSIÓN "COSTUMER"

- 1. Identificar las Tablas Origen
- 2. Definir los Atributos de la Dimensión
- 3. Realizar las Uniones Necesarias (Transformación)
- 4. Crear la Tabla de Dimensión (Carga)
- 5. Validación y Mantenimiento
- 6. Comparación con Otras Dimensiones
- Conclusión
- 1. Preparación y Conexión
- 2. Extracción y Transformación
 - 2.1. Consulta de Extracción y Transformación
- 3. Carga en la Nueva Base de Datos en Estrella
 - 3.1. Creación de la Tabla de Dimensión
 - 3.2. Proceso de Carga
- 4. Procedimiento General para Otras Dimensiones y Hechos
- Conclusión

A continuación te explico, paso a paso, cómo montar la tabla de dimensión Customer (o Cliente) a partir de la base de datos Sakila. En líneas generales, el procedimiento es similar al de otras dimensiones: se extraen y transforman los datos de las tablas normalizadas (mediante un proceso ETL) para consolidarlos en una tabla que facilite el análisis. Sin embargo, cada dimensión puede tener sus particularidades según la cantidad de tablas de origen y la información que se desea incluir.

1. Identificar las Tablas Origen

Para la dimensión Customer en Sakila, las tablas principales son:

- **customer:** Contiene datos básicos del cliente (ID, nombre, apellido, email, dirección, etc.).
- **address:** Proporciona la información de la dirección (calle, distrito, código postal, teléfono, etc.).
- **city:** Relacionada con la dirección, contiene el nombre de la ciudad y un identificador.
- **country:** Indica el país asociado a la ciudad.

Cada una de estas tablas aporta atributos que, al combinarse, enriquecerán la dimensión Customer.

2. Definir los Atributos de la Dimensión

Decide qué campos serán útiles para el análisis. Un ejemplo de atributos para la dimensión Customer (que podríamos llamar **DimCustomer**) podría ser:

- **Customer_Key:** Clave sustituta (surrogate key) para la dimensión.

- **customer_id:** Identificador natural del cliente, obtenido de la tabla customer.
- **store_id:** Identificador de la tienda a la que pertenece el cliente.
- **first_name y last_name:** Nombres y apellidos.
- **email:** Correo electrónico.
- **address:** Dirección completa (puede ser la concatenación de los campos de la tabla address, como address y address2).
- **district:** Distrito o barrio.
- **city:** Nombre de la ciudad (extraído de la tabla city).
- **country:** Nombre del país (extraído de la tabla country).
- **postal_code:** Código postal.
- **phone:** Número de teléfono.
- **create_date:** Fecha de creación del registro del cliente (útil para análisis temporales).
- **active:** Estado del cliente (activo o inactivo).
- **last_update:** Fecha de última actualización (opcional, para auditoría).

La selección de atributos depende de los requerimientos analíticos de tu Data Warehouse.

3. Realizar las Uniones Necesarias (Transformación)

Para consolidar la información, deberás unir las tablas de la siguiente manera:

1. **customer y address:**

Se unen mediante el campo `address_id` en la tabla customer con el campo `address_id` en la tabla address.

2. **address y city:**

Se unen a través del campo `city_id`.

3. **city y country:**

Se unen mediante el campo `country_id`.

Un ejemplo de consulta SQL para obtener los datos consolidados sería:

```
SELECT
    c.customer_id,
    c.store_id,
    c.first_name,
    c.last_name,
    c.email,
    CONCAT(a.address, ' ', IFNULL(a.address2, '')) AS address,
    a.district,
    ci.city,
    co.country,
    a.postal_code,
    a.phone,
    c.create_date,
    c.active,
    c.last_update
FROM customer c
JOIN address a ON c.address_id = a.address_id
```

```
JOIN city ci ON a.city_id = ci.city_id
JOIN country co ON ci.country_id = co.country_id;
```

Esta consulta extrae la información combinada de las tablas originales y te permite visualizar cómo se agruparán los datos en la dimensión.

4. Crear la Tabla de Dimensión (Carga)

Una vez que tengas claros los atributos y la consulta para extraer los datos, el siguiente paso es crear la tabla de dimensión en tu Data Warehouse. **Por ejemplo:**

```
CREATE TABLE DimCustomer (
  Customer_Key INT AUTO_INCREMENT PRIMARY KEY,
  customer_id INT,
  store_id INT,
  first_name VARCHAR(45),
  last_name VARCHAR(45),
  email VARCHAR(50),
  address VARCHAR(100),
  district VARCHAR(50),
  city VARCHAR(50),
  country VARCHAR(50),
  postal_code VARCHAR(20),
  phone VARCHAR(20),
  create_date DATETIME,
  active BOOLEAN,
  last_update DATETIME
);
```

Posteriormente, con un proceso ETL, cargas los datos extraídos y transformados (mediante la consulta anterior) en esta tabla de dimensión.

5. Validación y Mantenimiento

- **Validación:** Verifica que los datos cargados en DimCustomer sean consistentes y que todas las relaciones se hayan realizado correctamente.
- **Actualizaciones:** Define la periodicidad y el método (carga completa o incremental) para actualizar la dimensión conforme los datos de origen cambien.

6. Comparación con Otras Dimensiones

El proceso descrito (extracción, transformación y carga) es el mismo o muy similar para otras dimensiones del modelo dimensional (por ejemplo, DimFilm, DimStore, DimTime).

- Particularidades:
 - Algunas dimensiones pueden requerir la unión de más de dos o tres tablas (como DimFilm, que puede combinar datos de film, language, film_category y category).
 - La definición de atributos puede variar según el análisis requerido.

- En algunas dimensiones se utiliza una clave sustituta (surrogate key) para asegurar la consistencia en el Data Warehouse.

En esencia, la metodología ETL es uniforme, pero cada dimensión puede tener su propia complejidad dependiendo de la cantidad y variedad de datos involucrados.

Conclusión

Montar la tabla de dimensión Customer en Sakila implica:

1. **Identificar las tablas fuente:** customer, address, city y country.
2. **Definir los atributos relevantes** que conformarán la dimensión.
3. **Realizar las uniones necesarias** para consolidar los datos.
4. **Crear la tabla de dimensión** en el Data Warehouse y cargarla mediante un proceso ETL.
5. **Validar y mantener** la dimensión a lo largo del tiempo.

Ejemplo detallado de cómo realizar la extracción de datos desde las tablas normalizadas de Sakila y cargarlos en una nueva base de datos con un esquema en estrella. El proceso consta de tres fases: Extracción, Transformación y Carga (ETL).

1. Preparación y Conexión

Primero, asegúrate de tener acceso tanto a la base de datos original (Sakila) como a la nueva base de datos (por ejemplo, llamada DataWarehouse) en la que implementarás el modelo estrella. Puedes utilizar herramientas de línea de comandos (como MySQL o PostgreSQL según corresponda) o herramientas ETL específicas (por ejemplo, Talend, Pentaho, scripts en Python, etc.).

2. Extracción y Transformación

El objetivo es extraer la información de las tablas normalizadas y transformarla en el formato requerido por cada tabla de dimensión o de hechos del esquema estrella.

Ejemplo: Dimensión Customer (Cliente)

Las tablas involucradas son:

- **customer**
- **address**
- **city**
- **country**

2.1. Consulta de Extracción y Transformación

Puedes escribir una consulta SQL que una todas estas tablas para obtener una vista “plana” de los datos de cliente. Por ejemplo:

```
SELECT
```

```

c.customer_id,
c.store_id,
c.first_name,
c.last_name,
c.email,
CONCAT(a.address, ' ', IFNULL(a.address2, '')) AS address,
a.district,
ci.city,
co.country,
a.postal_code,
a.phone,
c.create_date,
c.active,
c.last_update
FROM customer c
JOIN address a ON c.address_id = a.address_id
JOIN city ci ON a.city_id = ci.city_id
JOIN country co ON ci.country_id = co.country_id;

```

Esta consulta:

- **Une** la tabla `customer` con `address` usando el campo `address_id`.
- **Conecta** `address` con `city` mediante `city_id` y posteriormente `city` con `country` a través de `country_id`.
- **Transforma** los datos, por ejemplo, concatenando los campos de dirección para obtener un único campo que represente la dirección completa.

Este resultado representa la transformación de los datos para la dimensión Customer.

3. Carga en la Nueva Base de Datos en Estrella

3.1. Creación de la Tabla de Dimensión

Primero, crea la tabla destino en tu Data Warehouse. Por ejemplo, para la dimensión Customer:

```

CREATE TABLE DimCustomer (
  Customer_Key INT AUTO_INCREMENT PRIMARY KEY,
  customer_id INT,
  store_id INT,
  first_name VARCHAR(45),
  last_name VARCHAR(45),
  email VARCHAR(50),
  address VARCHAR(100),
  district VARCHAR(50),
  city VARCHAR(50),
  country VARCHAR(50),
  postal_code VARCHAR(20),
  phone VARCHAR(20),
  create_date DATETIME,
  active BOOLEAN,
  last_update DATETIME
);

```

3.2. Proceso de Carga

Usa la consulta de extracción y transformación para insertar los datos en la nueva tabla. Un ejemplo usando SQL (en MySQL) es:

```
INSERT INTO DimCustomer (  
    customer_id,  
    store_id,  
    first_name,  
    last_name,  
    email,  
    address,  
    district,  
    city,  
    country,  
    postal_code,  
    phone,  
    create_date,  
    active,  
    last_update  
)  
SELECT  
    c.customer_id,  
    c.store_id,  
    c.first_name,  
    c.last_name,  
    c.email,  
    CONCAT(a.address, ' ', IFNULL(a.address2, '')) AS address,  
    a.district,  
    ci.city,  
    co.country,  
    a.postal_code,  
    a.phone,  
    c.create_date,  
    c.active,  
    c.last_update  
FROM Sakila.customer c  
JOIN Sakila.address a ON c.address_id = a.address_id  
JOIN Sakila.city ci ON a.city_id = ci.city_id  
JOIN Sakila.country co ON ci.country_id = co.country_id;
```

Este comando:

- **Extrae** y **transforma** los datos mediante la misma consulta de antes.
- **Carga** directamente los resultados en la tabla `DimCustomer` de la nueva base de datos.

4. Procedimiento General para Otras Dimensiones y Hechos

El proceso para crear las demás dimensiones (por ejemplo, DimFilm, DimStore, DimTime) y la tabla de hechos (por ejemplo, FactRental) es muy similar:

1. **Identifica las tablas fuente** relevantes en Sakila.

2. **Define los atributos** que serán necesarios para el análisis.
3. **Crea una consulta SQL** que una y transforme los datos de las tablas origen (usando JOIN y funciones de transformación según sea necesario).
4. **Crea la tabla destino** en la nueva base de datos en estrella.
5. **Inserta los datos** en la tabla destino usando un comando INSERT INTO ... SELECT ... similar al ejemplo anterior.

Cada dimensión puede tener sus particularidades según la cantidad de tablas involucradas o transformaciones específicas, pero la metodología ETL se mantiene uniforme.

Conclusión

Para trasladar los datos de Sakila a un Data Warehouse con un modelo en estrella, se debe:

1. **Extraer** los datos de las tablas normalizadas (por ejemplo, customer, address, city, country).
2. **Transformar** estos datos aplicando uniones y concatenaciones para obtener una vista “plana” adecuada para análisis.
3. **Cargar** los datos en la nueva tabla de dimensión mediante sentencias SQL que inserten los datos transformados.

Este ejemplo con la dimensión Customer puede adaptarse de forma similar a otras dimensiones y a la tabla de hechos según los requerimientos analíticos.
