

Bases de datos.

PL2 Segunda parte.

Partes

1.

Cambios realizados en las tablas del modelo relacional

2.

Codigo SQL. Creación de tablas y carga de datos

Autores:

Hugo Valcarce García

Gonzalo González Silverio

Modificaciones:

Personal.

No hay ningun cambio necesario

Guionistas.

Quitamos los datos de año nacimiento y año muerte. Dichos datos ya se tienen en la tabla personal. Lo que importa es su nombre, ya que es la PK.

Actores.

Mismos cambios que en guionistas

Directores.

Mismos que en guionistas

Pagina Web.

Quitamos el dato tipo. Lo que nos interesa de una pagina web es su URL, lo que permite distinguir entre todas las posibles paginas webs

Película.

Ponemos el dato año como primary key y actualizamos el pk de películas para que contenga los primary keys de titulo y año, de esta forma podemos crear una PK que es mejor que la anterior, ya que es imposible que existan películas con un mismo titulo y año, de manera que así podemos distinguir entre cualquier tipo de pelicula. Quitamos el dato géneros ya que es su propia tabla, ya que una película puede poseer varios géneros, cada uno con su nombre, etc.

Criticas.

Indicamos el critico como PK de Criticas. Metemos el dato año de la pelicula a Criticas, para que se pueda usar como PK de criticas junto con el titulo de la película y el critico que ha hecho la critica, creando así la PK definitiva para Criticas, y como FK de película con el titulo de la película, haciendo así la FK de Criticas con el titulo y año de la película.

Caratulas.

Quitamos el dato URL de la pagina web de la caratula, quitando así su FK, ya que dicho dato se incluirá en la tabla de relación entre caratula y pagina web. Incluimos el año de la película para que sea PK de caratulas junto con el titulo de la película, y FK junto con titulo de la película.

Actua -> Actuan. (Actualizamos el nombre de la tabla)

Introducimos el dato año de la película. Creamos el PK de Actuan con el año de la película, el titulo de la película y el nombre del actor. Introducimos en la FK de películas junto con el titulo de la película el año de la película.

Escribe -> Guionizan. (Actualizamos el nombre de la tabla)

Introducimos el dato año de la película. Creamos el PK con los tres datos disponibles, siendo el nombre del guionista, el titulo de la película y el año de la película. Incluimos en el FK de películas el año de la película.

Creación de la tabla Alojadas como indica en el diagrama relacional.

Introducimos el dato fecha. Luego, ponemos los datos de la caratula, siendo su nombre. De las paginas web metemos su URL. Y luego, ponemos el año de película y titulo de película a la que pertenece la caratula. Con todos estos datos, excepto la fecha, creamos el PK de la tabla. También obtenemos el FK de caratulas, siendo con el nombre de la caratula y el titulo y año de la película. Por ultimo, creamos el ultimo FK

Creación de la tabla Géneros como indica en el diagrama relacional.

Cada genero tiene su propio nombre, por lo que se incluye como dato de la tabla. Además, incluimos el titulo y año de la película a la que pertenecerán los géneros de dicha película. Con esto, establecemos la PK de la tabla con todos esos tres datos. Y por último, creamos la FK de películas con su titulo y año.

Codigo SQL:

Hacemos la creación de dos esquemas:

pl2_Fin:

En este esquema creamos las tablas del modelo relacional, con todas las tablas y las relaciones estableciendo sus datos y sus tipos de datos, sus PK y FK con sus correspondientes restricciones. Ejemplo de una de las tablas:

```
CREATE TABLE IF NOT EXISTS pl2_Fin.Guionistas(  
    Nombre TEXT NOT NULL,  
    CONSTRAINT Guionistas_PK PRIMARY KEY (Nombre),  
    CONSTRAINT Personal_Guionistas_FK FOREIGN KEY (Nombre) REFERENCES pl2_Fin.Personal (Nombre) MATCH FULL  
    ON DELETE RESTRICT ON UPDATE CASCADE  
);
```

pl2_Temp:

En este esquema creamos las tablas temporales, en las cuales se van a cargar todos los datos de los archivos csv proporcionados. Solo se crean las siguientes tablas: Actúan, Directores, Guionizan, Criticas, Caratulas y Película. Nos aseguramos de poner los datos que aparecen en los archivos csv de cada una, y poniendo todo de tipo TEXT y sin ninguna restricción. Ejemplo de una de las tablas:

```
CREATE TABLE IF NOT EXISTS pl2_Temp.Pelicula (  
    Anno TEXT,  
    Titulo TEXT,  
    Generos TEXT,  
    Idioma TEXT,  
    Duracion TEXT,  
    Calificacion_MPA TEXT  
);
```

Carga de datos:

Una vez creadas todas estas tablas, empezamos la carga de datos. Primero introducimos en las tablas temporales los datos de los archivos csv empleando un copy:

```
\copy pl2_Temp.Directores from 'c:/PL2_files/directores.csv' with (format csv, delimiter E'\t', NULL '\N', encoding 'UTF8')  
\copy pl2_Temp.Actuan from 'c:/PL2_files/actuan.csv' with (format csv, delimiter E'\t', NULL '\N', encoding 'UTF8')  
\copy pl2_Temp.Guionizan from 'c:/PL2_files/guionizan.csv' with (format csv, delimiter E'\t', NULL '\N', encoding 'UTF8')  
\copy pl2_Temp.Pelicula from 'c:/PL2_files/peliculas.csv' with (format csv, delimiter E'\t', NULL 'NULL', encoding 'UTF8')  
\copy pl2_Temp.Criticas from 'c:/PL2_files/criticas.csv' with (format csv, delimiter E'\t', NULL '\N', encoding 'UTF8')  
\copy pl2_Temp.Caratulas from 'c:/PL2_files/caratulas.csv' with (format csv, delimiter E'\t', NULL '\N', encoding 'UTF8')
```

Es importante indicar que puede haber posibles datos nulos, para ello se indica, con NULL o \N como se ve en la imagen.

Una vez hemos cargado los datos en las tablas temporales, empezamos a meter los datos en las tablas finales realizando las consultas necesarias.

Personal.

A la tabla personal debemos meterle el nombre del personal, su año de nacimiento y de muerte. Los datos los obtenemos de las tablas temporales de Actúan, Guionizan y Directores, que son los que los tienen desde el principio. Para ello realizamos 3 INSERTS. Dado que existe la posibilidad de que haya datos duplicados, empleamos un distinct para poder evitarlo. Ejemplo de uno de los inserts:

```
INSERT INTO pl2_Fin.Personal (Nombre, Anno_Nacimiento, Anno_Muerte)
SELECT
    distinct Guionista_Nombre, Fecha_nac::INT ,Fecha_muer ::INT
FROM
    pl2_Temp.Guionizan
WHERE Guionista_Nombre NOT IN (SELECT Nombre FROM pl2_Fin.Personal);
```

Actores. Guionistas. Directores.

En las tablas actores, guionistas y directores solo nos interesa sus nombres, por lo que hacemos un unico INSERT para obtener sus nombres en cada tabla de sus tablas temporales correspondientes, siendo Actúan, Guionizan, y Directores, una vez mas asegurando de que no haya ningun dato duplicado. Ejemplo:

```
INSERT INTO pl2_Fin.Guionistas (Nombre)
SELECT
    distinct (Guionista_Nombre)
FROM
    pl2_Temp.Guionizan;
```

Película.

En la tabla película hay muchos datos que introducir. Los mas básicos se hacen en un simple INSERT. Metemos su año, título, etc. Nos interesa que no existan películas que no tengan mismo año y título, porque puede ser una misma película duplicada, por lo que hacemos un distinct en esos dos datos. La tabla películas posee el FK de directores, por lo que tenemos que hacer un JOIN, donde el director que queremos relacionar tenga mismo año y título de película en la que trabaja. Es decir, si queremos relacionar el director "D" en una película de Año "A" y Título "T", debemos asegurarnos de que ese director haya dirigido esa misma película, que el título y año de la película en la que dirige el director D sea los mismos que la película en la que le queremos relacionar. Los datos de la película los obtenemos de la tabla temporal Película, y para relacionar el director debemos consultar su nombre desde su tabla temporal Directores. Ejemplo del insert.

```
INSERT INTO pl2_Fin.Pelicula (Anno, Titulo, Idioma,Duracion,Calificacion_MPA,Director_Nombre)
SELECT
    distinct ON (p.Anno::int, p.Titulo)
    p.Anno::int,
    p.Titulo,
    p.Idioma::char(2),
    LEFT (Duracion, length(Duracion)-5)::int,
    p.Calificacion_MPA,
    d.Director_Nombre
FROM
    pl2_Temp.Pelicula p
LEFT JOIN pl2_Temp.Directores d ON (p.Anno=d.Pelicula_Anno) AND (p.Titulo=d.Pelicula_Titulo);
```

Actuan. Guionizan.

Son dos INSERT básicos, metemos los datos necesarios en la tabla desde la tabla temporal Actuan y Guionizan. Estos INSERTS no tiene mucho misterio, son solo introducir datos. Ejemplo de uno de ellos:

```
INSERT INTO pl2_Fin.Actuan (Papel, Pelicula_Titulo, Pelicula_Anno, Actor_Nombre)
SELECT
    Papel,
    Pelicula_Titulo,
    Pelicula_Anno::int,
    Actor_Nombre
FROM
    pl2_Temp.Actuan;
```

Caratulas.

En Caratulas no hay mucha diferencia. En un solo INSERT metemos los datos necesarios de la tabla temporal Caratulas, pero nos tenemos que asegurar una vez mas de que no haya caratulas duplicadas, por lo que realizamos un distinct en el titulo, año de película y nombre de la caratula. Gracias a ello, nos podemos asegurar de que no haya ninguna caratula duplicada. Ejemplo:

```
INSERT INTO pl2_Fin.Caratulas (Tamaño, Nombre, Pelicula_Titulo, Pelicula_Anno)
SELECT
    DISTINCT ON (Pelicula_Titulo,Pelicula_Anno,Nombre)
    Tamaño,
    Nombre,
    Pelicula_Titulo,
    Pelicula_Anno::int
FROM
    pl2_Temp.Caratulas;
```

Géneros.

Para los géneros tenemos que obtener sus nombres, y el titulo y año de la película a la que pertenecen. Estos datos los extraemos de la tabla temporal Películas. Dado que son varios géneros en una misma película, debemos asegurarnos de separar cada uno de ellos a partir de la coma que los separa, así obteniendo cada genero por separado y en su respectiva película a la que pertenece. Ejemplo:

```
INSERT INTO pl2_Fin.Generos (Genero, Pelicula_Anno, Pelicula_Titulo)
SELECT
    distinct
    regexp_split_to_table(Generos, ','),
    Anno::int,
    Titulo
FROM
    pl2_Temp.Pelicula;
```

Pagina_Web.

En la pagina web solo nos interesa saber sus URL de cada pagina. Estos URL los podemos encontrar en las tablas temporales de Caratulas y Criticas, por lo que debemos hacer dos INSERT para cada tabla. Dado que la URL es, un poco larguilla, hacemos un Split para solo obtener la parte que nos interesa, que es la parte que va antes de la tercera '/'. Esto lo hacemos en ambos INSERT. Ejemplo de uno de los INSERT:

```
INSERT INTO pl2_Fin.Pagina_Web (PW_URL)
SELECT
    distinct split_part(Pagina_Web_URL,'/',1) || '/' || split_part (Pagina_Web_URL,'/',3)
FROM
    pl2_Temp.Caratulas
WHERE split_part(Pagina_Web_URL,'/',1) || '/' || split_part (Pagina_Web_URL,'/',3) NOT IN (SELECT PW_URL FROM pl2_Fin.Pagina_Web);
```

Alojadas.

No tiene mucha dificultad. Guardamos todos los datos necesarios usando la tabla temporal Caratulas. Para evitar los datos duplicados, hacemos un distinct del titulo y año de la película, la URL, y el nombre de la caratula. De esta forma no tendremos relaciones exactamente iguales. Y como antes, queremos la URL corta, por lo que hacemos el mismo split que en Paginas web. Ejemplo:

```
INSERT INTO pl2_Fin.alojadas(Fecha, Peliculas_Anno_Caratulas, Pelicula_Titulo_Caratulas, PaginaWeb_URL, Nombre_Caratula)
SELECT
    DISTINCT ON(Pelicula_Titulo, Pelicula_Anno, Pagina_Web_URL, Nombre)
    Fecha_Aloj::TIMESTAMP,
    Pelicula_Anno::int,
    Pelicula_Titulo,
    split_part(Pagina_Web_URL,'/',1) || '/' || split_part (Pagina_Web_URL,'/',3),
    Nombre
FROM
    pl2_Temp.Caratulas;
```

Criticas.

Mas o menos igual que con Alojadas. Introducimos los datos de la tabla temporal Criticas. Para evitar datos duplicados, hacemos un distinct con los datos Critico, titulo y año de la película. De esta forma, se puede evitar tener dos Criticas duplicadas. Mismo Split para las URL. Ejemplo:

```
INSERT INTO pl2_Fin.Criticas (Pelicula_Anno, Pelicula_Titulo, Critico, Puntuacion, Texto, Pagina_Web_URL)
SELECT
    DISTINCT ON (Critico,Pelicula_Titulo,Pelicula_Anno)
    Pelicula_Anno::int,
    Pelicula_Titulo,
    Critico,
    Puntuacion::numeric(2,1),
    Texto,
    split_part(Pagina_Web_URL,'/',1) || '/' || split_part (Pagina_Web_URL,'/',3)
FROM
    pl2_Temp.Criticas;
```