

Base de Datos I

Trabajo Práctico Especial

2do Cuatrimestre 2017

1. Objetivo

El objetivo de este Trabajo Práctico Especial es aplicar los conceptos de SQL Avanzado (PSM, Triggers) vistos a lo largo del curso.

2. Descripción del Trabajo

El sitio Buenos Aires Data, <https://data.buenosaires.gob.ar/>, ofrece datasets de información pública del Gobierno de la Ciudad de Buenos Aires. Entre otros datasets, se encuentra el que contiene información sobre las distintas ediciones del festival de cine BAFIGI. En este trabajo utilizaremos los datos de dicho festival pero del año 2013.

Este dataset, al que llamaremos *Festival*, contiene distintas tablas en formato .csv. La **tabla bafici13-films.csv** contiene datos de las películas que participaron del BAFIGI 13 pero sus datos no están normalizados.

Se quiere entonces normalizar algunos datos de dicha tabla llevándolos a la siguiente estructura (no se muestran aquí todos los atributos de las entidades):

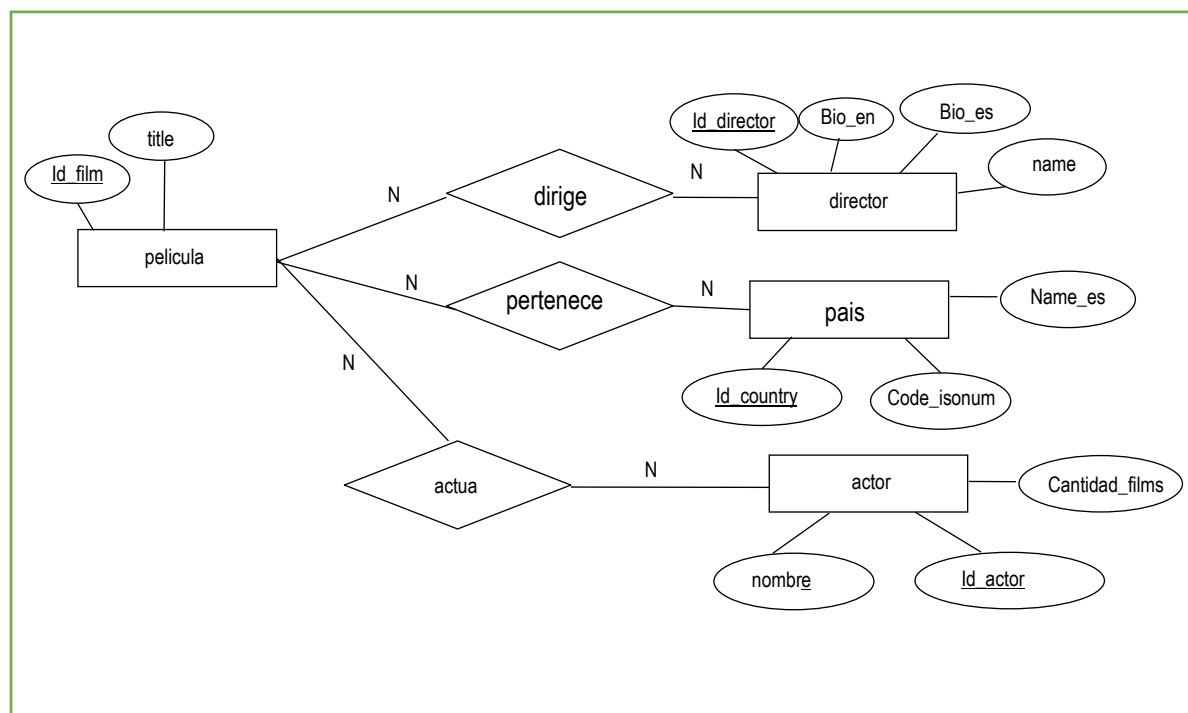


Figura 1: DER festival

Para esto, se provee a los alumnos con el script **festival.sql** que contiene los comandos para la creación de la tabla FILM, la cual deberá tener los mismos datos del archivo **bafici13-films.csv**, y los comandos para la creación de todas las tablas que surgen del mapeo del DER de la Fig 1.

Como resultado de este trabajo, las tablas del esquema películas deben contener los datos apropiados tomados de **bafici13-films.csv**. Tal como se explicará en ítem siguiente, algunos de los datos deben importarse directamente desde los **csv** correspondientes, como es el caso de los directores, y para poder insertar el resto se deben crear un Trigger y una función PSQL.

3. Procedimiento

Las tablas disponibles en el script **festival.sql** son las siguientes:

- FILM
- PELÍCULA
- DIRECTOR
- DIRIGE
- PAÍS
- PERTENECE
- ACTOR
- ACTÚA

Revisar en el script los atributos que poseen y sus respectivos tipos de datos.

a) Importación de datos

La tabla FILM debe contener los mismos datos que la tabla **bafici13-films.csv**. Para ello deben importarse sus datos. Lo mismo ocurre con DIRECTOR y PAIS.

Para realizar esta importación estas tres tablas deben estar vacías. Es tarea de los alumnos realizar el borrado de las tablas antes de comenzar la importación.

A continuación se detallan las tablas a las que deben incorporarse datos importados y sus respectivas fuentes:

Tabla	Fuente
FILM	bafici13-films.csv
DIRECTOR	bafici13-directores.csv
PAIS	bafici13-paises.csv

b) Trigger

El campo *cantidad_films* de la tabla ACTOR contiene la cantidad de películas en las que participó el actor en el festival.

Se pide crear el Trigger **contador_actor** que al insertar en la tabla ACTÚA mantenga actualizado el campo *cantidad_films* de la tabla ACTOR con la cantidad de películas en las que participó el actor hasta el momento.

c) Funciones

Se pide crear la función PSQL **inserta_datos** que tomando la información de las tablas FILM, DIRECTOR y PAIS (que ya contienen datos importados en a)), inserte los datos apropiados en las tablas restantes.

- Tabla PELICULA

Esta tabla debe completar sus datos a partir de sus campos análogos en la tabla FILM.

- Tabla DIRIGE

Para guardar todos los directores de una película, la tabla FILM tiene varias columnas con el nombre *directorX*, donde X es un número entero entre 1 y 14. Dada una película, cada valor de *directorX* debe almacenarse en la tabla DIRIGE junto con el id de la película. Tener en cuenta cual es el valor que representa el NULL para los directores que aparecen en FILM.

Si el valor de *directorX* no aparece en la tabla DIRECTOR, esa tupla (id_film, id_director) no debe insertarse en DIRIGE.

- Tabla PERTENECE

Para almacenar todos los países a los que pertenece una película, la tabla FILM tiene varios atributos cuyos nombres son *id_countryX*, donde X es un número entero entre 1 y 8. Dada una película, cada valor de *id_countryX* debe almacenarse en la tabla PERTENECE junto con el id de la película. Tener en cuenta el valor con que se representa el NULL en los atributos *id_countryX*.

Si el valor de *id_countryX* no aparece en la tabla PAIS, esa tupla (id_film, id_country) no debe insertarse en PERTENECE.

- Tabla ACTOR

Cada película de la tabla FILM tiene en su atributo *cast_* los nombres de los actores que participaron de la película separados por coma (,).

Dado un film, se deben tomar cada uno de los nombres de los actores, y en caso de que no existieran ya en la tabla ACTOR, asignarle un id numérico único a cada uno e insertarlo en dicha tabla.

Los nombres de actores no pueden tener espacio en blanco ni al principio ni al final.

Para comparar dos nombres de actores no se tienen en cuenta mayúsculas y minúsculas. Es decir, si ya se insertó a 'Pedro FERNANDEZ', no debe insertarse por ejemplo, 'pedro fernandez' ni 'PEDRO FERNANDEZ'.

- Tabla ACTUA

Cada uno de los id de los actores de las películas debe almacenarse en junto con el id de la película en la tabla ACTUA.

No se garantiza que las tablas descriptas anteriormente estén vacías al ejecutar la función. Es importante entonces, que la función **inserta_datos** evite las excepciones producidas por ejemplo, por duplicación de datos.

Se recomienda que la función **inserta_datos**, se factorice en tantas funciones como sea necesario.

Ejemplo

Si ejecutáramos **inserta_datos** únicamente sobre los cuatro primeros films de la tabla FILM (id_film 1, 2, 3, y 4), las tablas del esquema películas afectadas por esta función, deben quedar:

PELICULA

id_film [PK] integer	title text	title_es text	title_en text	title_orig text	year_ integer	plot text	synopsis_es text	synopsis_en text	tagline text	duration text	color text	id_youtub text	url_ticke text
1	"No"				2012	"NULL"	"En plena d	"During the	"NULL"	"117"	"NULL"	"0P8KKKJJC	
2	"La parte del león "		"The Lion's She		1978	"NULL"	"(...) La h	"(...) The st	"NULL"	"85"	"NULL"		"https:
3	"La playa del amor"		"The Beach of I		1979	"NULL"	"(...) Cacho	"(...) Cacho	"NULL"	"86"	"NULL"		"https:
4	"La discoteca del amor		"The Disco of I		1980	"NULL"	"(...) Con el	"(...) Featur	"NULL"	"90"	"NULL"		"https:

PERTENECE

id_film [PK] integer	id_country [PK] integer
1	46
1	75
2	13
3	13
4	13

DIRIGE

id_director [PK] integer	id_film [PK] integer
5	2
5	3
5	4
8	1

ACTOR

id_actor integer	nombre text	cantidad_films integer
1	Gael Garcia Bernal	1
2	Alfredo Castro	1
3	Antonia Zegerz	1
4	Luis Gnecco	1
5	Marcial Tagle	1
6	Julio de Grazia	1
7	Ulises Dumont	1
8	Arturo Maly	3
9	Julio Chaves	1
10	Luisina Brando	1
11	Cacho Castaña	2
12	Ricardo Darín	2
13	Carlos del Burgo	2
14	Mónica Gonzaga	1
15	Tincho Zabala	1

ACTUA

id_actor integer	id_film integer
1	1
2	1
3	1
4	1
5	1
6	2
7	2
8	2
9	2
10	2
8	3
11	3
12	3
13	3
14	3
8	4
11	4
12	4
13	4
15	4

Para resumir, el **flujo** del trabajo es el siguiente:

- i. Correr el script **festival.sql** para crear todas las tablas
- ii. Asegurarse que las tablas DIRECTOR y PAIS estén vacías.
Importar los datos en las tablas según lo indicado en 2 a).
En la figura 3 las tablas de este paso aparecen en **anaranjado**.
- iii. Asegurarse que las tablas FILM, esté vacía.
Importar los datos en las tablas según lo indicado en 2 a).
En la figura 3 las tablas de este paso aparecen en **anaranjado**.

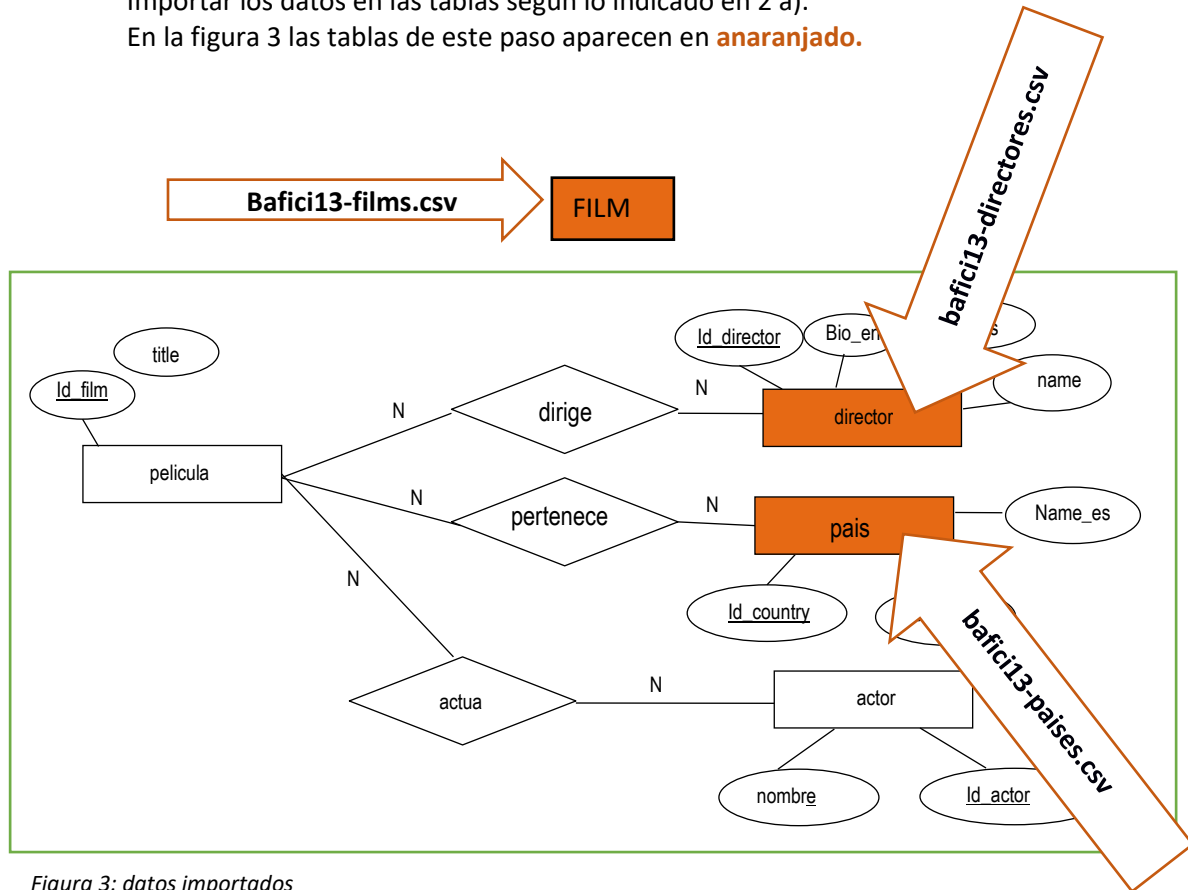


Figura 3: datos importados

- iv. A partir del contenido de la tabla FILMS, DIRECTOR y PAIS, cargar los datos del resto de las tablas, marcadas en **verde** en la figura 4, utilizando la función **inserta_datos()** y el trigger **contador_actor**. Tener en cuenta que en este caso las tablas pueden no estar vacías.

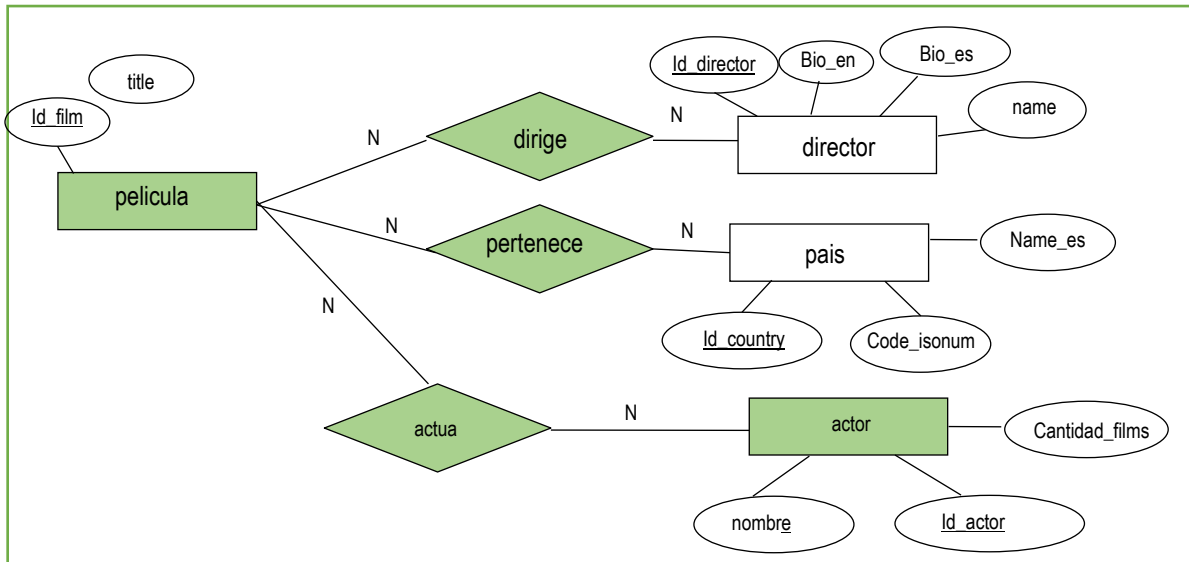


Figura 4: datos que debe insertar la función `inserta_datos()`.

- v. Si se cuenta con un archivo **bafici13-films.csv** con nuevos datos, debe ser posible incorporarlos al esquema repitiendo el proceso a partir del paso iii.

4. Modalidad

El Trabajo Práctico estará disponible en Campus a partir del 9/11/2017, indicándose allí mismo, la fecha de entrega.

Se incluye junto con el enunciado:

- a) Script **festival.sql** con los comandos para crear las tablas.
- b) Archivos de datos:
 - **Bafici13-filmsPRUEBA.csv**: Archivo de muestra para ejecutar el proceso por primera vez.
 - **Bafici13-filmsTEST.csv**: Archivo para ejecutar el proceso por segunda vez y testear el correcto funcionamiento.
 - **Bafici13-directores.csv**
 - **Bafici13-paises.csv**

El TP deberá realizarse en grupos de 3 alumnos y entregarse a través de la plataforma Campus ITBA hasta la fecha allí indicada.

5. Entregables

Los alumnos deberán entregar los siguientes documentos

- El script sql **importación.sql** con los comandos utilizados para insertar los datos en las tablas
- El script sql **funciones.sql** con los códigos de todas las funciones y el trigger.
- Un informe que debe contener:

- El rol de cada uno de los participantes del grupo. Si bien en el TP deben estar involucrados todos los integrantes, se debe asignar un rol de supervisión de cada una de las tareas. Mínimamente los roles son: encargado del informe, encargado de las funciones, encargado del trigger, encargado del funcionamiento global del proyecto y encargado de investigación. Pueden asignarse más roles en caso de requerirse.
- Todo lo investigado para realizar el TP.
- Las dificultades encontradas y cómo se resolvieron

El informe debe tener como máximo 3 páginas.

6. Evaluación

La evaluación del trabajo se llevará a cabo teniendo en cuenta los parámetros establecidos en la rúbrica asociada a la actividad en Campus.

Se tendrá en cuenta que las consultas, más allá del funcionamiento (lo cual es fundamental), sean genéricas.

Los docentes ejecutarán el proceso usando los conjuntos de datos entregados (detallados en 4b)) y usando otros conjuntos con la misma estructura.

El informe deberá estar completo y sin faltas de ortografía.

En caso de que el trabajo no cumpliera los requisitos básicos para ser aprobado, los alumnos serán citados en la fecha de recuperatorio para defenderlo y corregir los errores detectados.