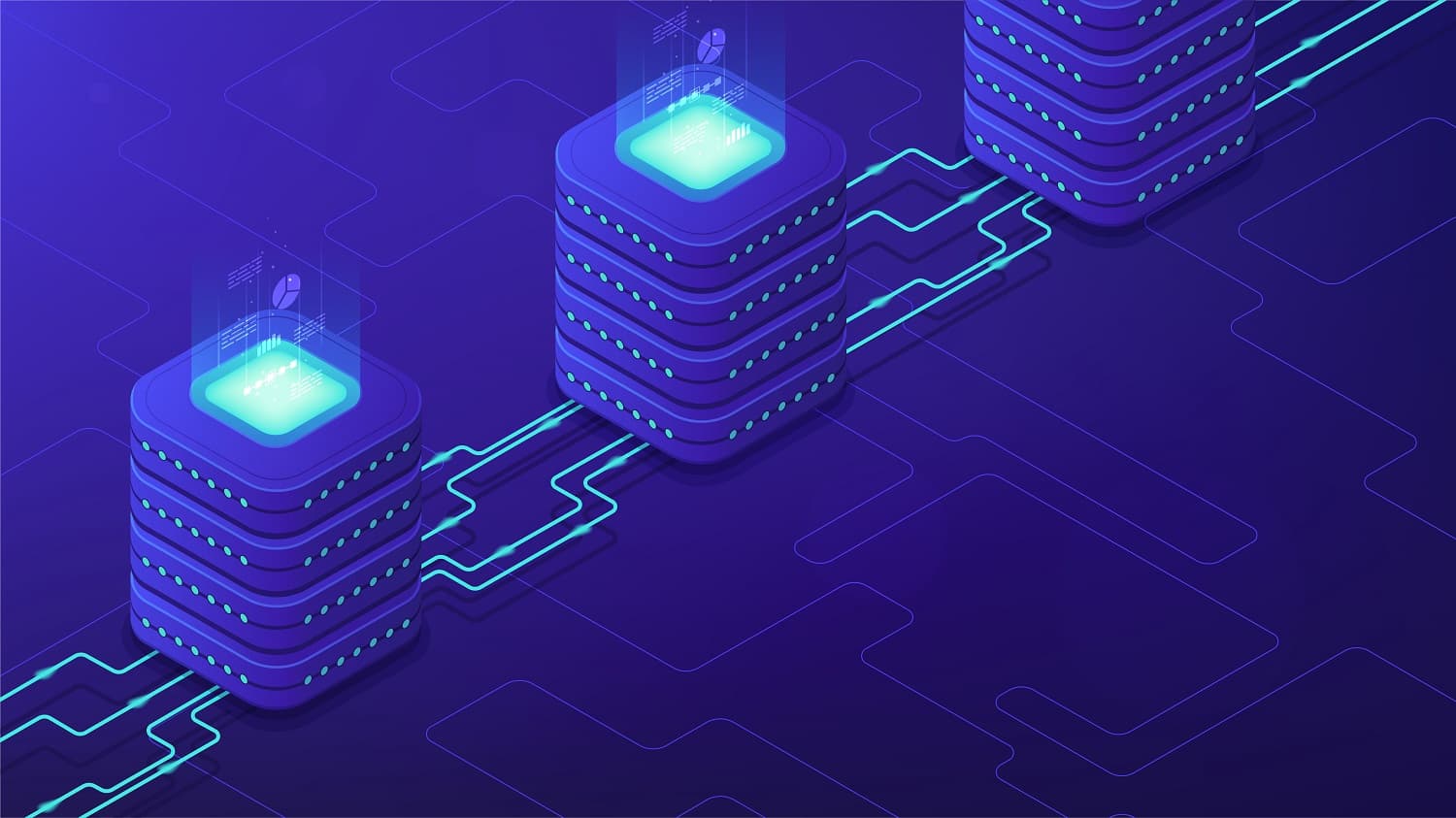
**UNIVERSIDAD TECNICA PARTICULAR DE LOJA**

**FUNDAMENTOS DE DATOS**

****

**Alumno:** Richard Ojeda

**Docente:** Nelson Piedra

**Periodo:**2021-2022

**Informe sobre el desarrollo del Proyecto Integrador**

**Tema:** Recolección y ordenamiento sobre la información interacción con el archivo de base de datos .csv

**Introducción**

Como primer punto de la actividad se comenzó consiguiendo el archivo que nos facilitara la información del base de datos con la que trabajamos en el presente proyecto.

**Desarrollo**

1. Realizamos la descarga de la fuente de la información que se encuentra ubicado en un repositorio de GitHub de autoría del usuario“[*rashida048*](https://github.com/rashida048)”, el archivo csv se encontraba con el nombre movie\_dataset.csv, el presente archivo contiene toda la información que se divide en nombre, su productora, su director, su personal, fecha de lanzamiento, toda esta información nos ayudara con el informe del proyecto.
2. Mediante el uso de la herramienta Excel procedimos a manejar la data y a ordenar los datos y colocar en sus respectivos campos.
3. Una vez que tenemos todo organizado procedemos a establecer varios campos del archivo, se obtuvieron los siguientes campos:

* index
* budget
* genres
* homepage
* id
* keywords
* original\_language
* original\_title
* overview
* popularity
* production\_companies
* production\_countries
* release\_date
* revenue
* runtime
* spoken\_languages
* status
* tagline
* title
* vote\_average
* vote\_count
* cast
* crew
* director

1. Una vez que pudimos identificar los varios campos, realizamos el planteamiento de entidades para toda esta data. Teniendo en cuenta que la empresa de estas entidades no podría cumplir el criterio que estamos solicitando, según el número de entidades se iban a establecer de acuerdo con el número de campos en los que está organizada la data, debido a que, en ciertos campos la data no

atómica y nos da como resultado que no se dé el cumplimiento de uno de los principios solicitados. Por consiguiente, revisamos cada campo estableciendo las siguientes entidades:

* MOVIE

Index

IdMovie

Budget

Homepage

Keywords

Original\_title

Overview

Popularity

Release\_date

Revenue

Runtime

Tagline

TitleMovie

Vote\_average

Votecount

* GENRE

IdGenre

GenreName

* ORIGINALLANGUAGE

Id\_OriginalLanguage

OrigLanguageName

* PRODUCTIONCOMPANY

Id\_ProductionCompany

ProdCompanyNa

* CREW

IdCrew

GenderCrew

Credit\_idCrew

CrewName

* PRODUCTIONCOUNTRY

Id\_ProductionCountry

ISO\_3166\_1

ProdCountryNam

* SPOKENLANGUAGE

Id\_SpokenLanguage

ISO\_639\_1

SpokenLangName

* DEPARTMENTCREW

IdDepartment

DepartcrewName

* JOBCREW

IdJob

JobcrewName

* DIRECTOR

IdDirector

DirectorName

* STATUS

IdStatus

StatusName

* CAST

IdCast

CastName

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Property** | **Domain** | **Optional** | **MultiValuated** | **Type** | **Comment** |
| Index | Int (7000) | Not Null | No | Primary Key | Número en el catalogo |
| IdMovie | Varchar (26) | Not Null | No | Primary Key | Id de cada película |
| Budget | Int (7000) |  | si |  | Presupuesto de película |
| Homepage | Varchar (300) |  | si |  | Página Web |
| Keywords | Varchar (200) |  | si |  | Palabras claves |
| Original\_title | Varchar (300) | Not Null | no |  | Titulo original |
| Overview | Varchar (500) |  | si |  | Descripción |
| Popularity | Int |  | si |  | Popularidad |
| Release\_date | Date |  | si |  | Fecha de estreno |
| Revenue | Int |  | si |  | Ingresos de la película |
| Runtime | Int |  | si |  | Duración de la película |
| Tagline | Varchar (200) |  | si |  | Frases icónicas |
| TitleMovie | Varchar (120) | Not Null | no |  | Titulo Película (inglés) |
| Vote\_average | Int |  | si |  | Valoración de la película |
| Votecount | Int |  | si |  | Numero de personas que votaron |

**Entidad *GENRE***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Property** | **Domain** | **Optional** | **MultiValuated** | **Type** | **Comment** |
| IdGenre | Varchar (30) | Not Null | No | Primary Key | ID del genero |
| GenreName | Varchar (15) | Not Null | Yes |  | Nombre del Género |

**Entidad *ORIGINALLANGUAGE***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Property** | **Domain** | **Optional** | **MultiValuated** | **Type** | **Comment** |
| Id\_OriginalLanguage | Varchar (30) | Not Null | No | Primary Key | Id Asignado al Idioma Original |
| OrigLanguageName | Varchar (15) | Not Null | Yes |  | Nombre del Idioma Original |

**Entidad *PRODUCTIONCOMPANY***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Property** | **Domain** | **Optional** | **MultiValuated** | **Type** | **Comment** |
| Id\_ProductionCompany | Varchar (30) | Not Null | No | Primary Key | Id Asignado a la Compañía Productora |
| ProdCompanyName | Varchar (15) | Not Null | Yes |  | Nombre de la Compañía Productora |

**Entidad *PRODUCTIONCOUNTRY***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Property** | **Domain** | **Optional** | **MultiValuated** | **Type** | **Comment** |
| Id\_ProductionCountry | Varchar (20) | Not Null | No | Primary Key | Id País de la Productora |
| ISO\_3466\_1 | Varchar (5) | Not Null | Yes |  | Abreviatura del País |
| ProdCountryName | Varchar (50) | Not Null | Yes |  | Nombre del País |

**Entidad *SPOKENLANGUAGE***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Property** | **Domain** | **Optional** | **MultiValuated** | **Type** | **Comment** |
| Id\_SpokenLanguage | Varchar (20) | Not Null | No | Primary Key | Id Idioma Hablado |
| ISO\_639\_1 | Varchar (5) | Not Null | Yes |  | Abreviatura del Idioma |
| SpokenLangName | Varchar (20) | Not Null | Yes |  | Nombre del Idioma |

**Entidad *STATUS***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Property** | **Domain** | **Optional** | **MultiValuated** | **Type** | **Comment** |
| IdStatus | Varchar (20) | Not Null | No | Primary Key | Id Estatus |
| StatusName | Varchar (15) | Not Null | Yes |  | Nombre del Estatus |

**Entidad *CAST***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Property** | **Domain** | **Optional** | **MultiValuated** | **Type** | **Comment** |
| IdCast | Varchar (20) | Not Null | No | Primary Key | Id Elenco |
| CastName | Varchar (50) | Not Null | Yes |  | Nombre del Elenco |

**Entidad *CREW***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Property** | **Domain** | **Optional** | **MultiValuated** | **Type** | **Comment** |
| IdCrew | Varchar (30) | Not Null | No | Primary Key | Id al Personal |
| GenderCrew | Int | Not Null | Yes |  | Genero del Personal |
| Credit\_idCrew | Varchar (30) | Not Null | Yes |  | Id de crédito del Personal |
| CrewName | Varchar (90) | Not Null | Yes |  | Nombre del Personal |

***DEPARTMENTCREW***

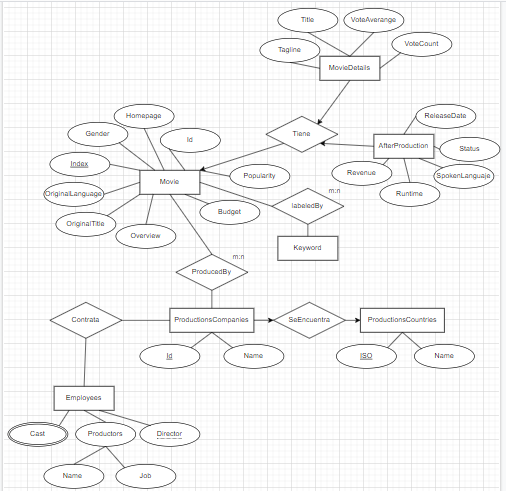
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Property** | **Domain** | **Optional** | **MultiValuated** | **Type** | **Comment** |
| IdDepartment | Varchar (20) | Not Null | No | Primary Key | Id Departamento del Elenco |
| DepartcrewName | Varchar (15) | Not Null | Yes |  | Nombre Departamento del Elenco |

**Entidad *JOBCREW***

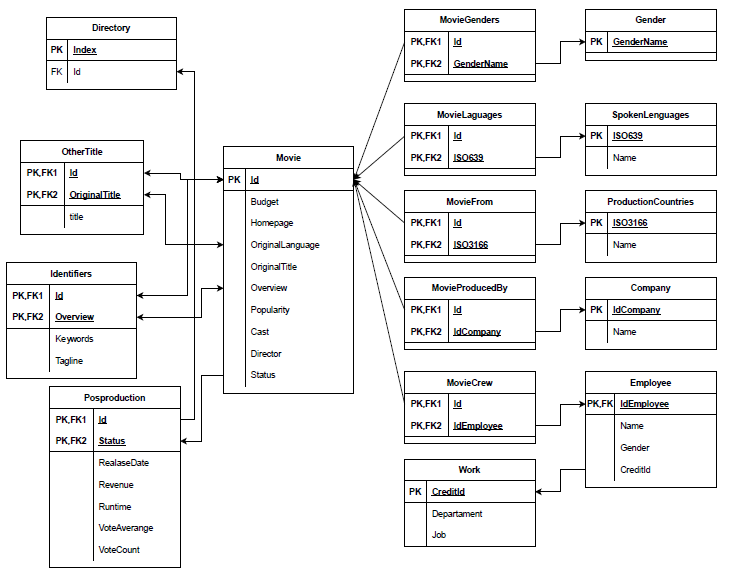
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Property** | **Domain** | **Optional** | **MultiValuated** | **Type** | **Comment** |
| IdJob | Varchar (30) | Not Null | No | Primary Key | Id Trabajo del Personal |
| JobcrewName | Varchar (30) | Not Null | Yes |  | Nombre del Trabajo del Personal |

**Entidad *DIRECTOR***

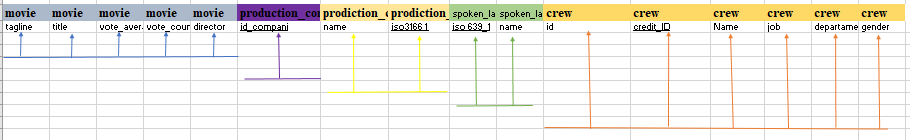
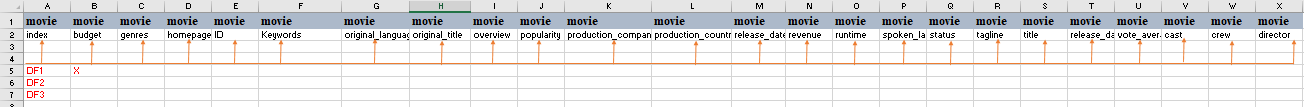
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Property** | **Domain** | **Optional** | **MultiValuated** | **Type** | **Comment** |
| IdDirector | Varchar (30) | Not Null | No | Primary Key | Id al director |
| DirectorName | Varchar (30) | Not Null | Yes |  | Nombre del director |

**MODELO RELACIONAL**

**Modelo conceptual**

****

**Dependencia funcional**



Como podemos observar en el grafico anterior, se exponen todas nuestras las dependencias funcionales de la tabla general con todos los campos, consecutivamente procedemos a describe de la forma más detallada cada dependencia funcional.

IdMovie → {keyword, index, homepage, genres, budgets, original\_languaje, original\_title, overview, popularity, director, vote\_count, vote\_average, tagline, status, cast} • Id\_company

→ name • Iso\_639\_1 → name • Iso\_3166\_1 → name • Credit\_id → job, department • Id\_Crew → name, gender

**Normalización**

Para poder continuar debemos saber que es normalización, la normalización es un proceso el cual permite organizar los datos, lo que incluye el análisis y creación de tablas, así como establecer sus relaciones entre las mismas, con el fin de eliminar redundancia, dependencias incoherentes y optimizar el uso de memoria del ordenador.

**1FN**

Para la primera regla formal comenzamos apartando todos los campos que no son atómicos y los apartamos para la posibilidad de colocarlos en otra tabla, algunos de ellos a la final se establecieron como una tabla en particular con sus propios campos y la mayoría se las relaciono con una clave primaria que actuaría como clave foránea para las demás, se decidió tomar como clave primaria al campo “id” esto debido a que es único para cada película además de ser un dato que no se puede confundir.

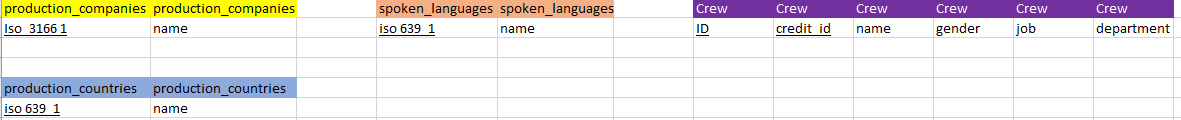
En la imagen se muestra el resultado del proceso llevado a cabo en Excel para determinar la primera forma normal.



**2FN**

Para la segunda regla formal de normalización se determinaron los datos que dependían de la clave primaria, se separaron varios campos de la que sería la entidad fuerte o principal, sin embargo, aún quedaban muchos datos que formarían parte de esta, por ello se tomaron en cuenta de otros datos como los campos que no podían ser nulos, los campos adicionales y campos que se determinarían solo a cierto estado de la película.

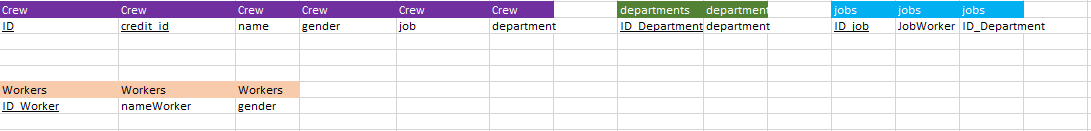
En la imagen se muestra el desarrollo de la segunda forma normal en Excel para el cual fue necesario determinar las dependencias funcionales y con ello sacamos las tablas resultantes.

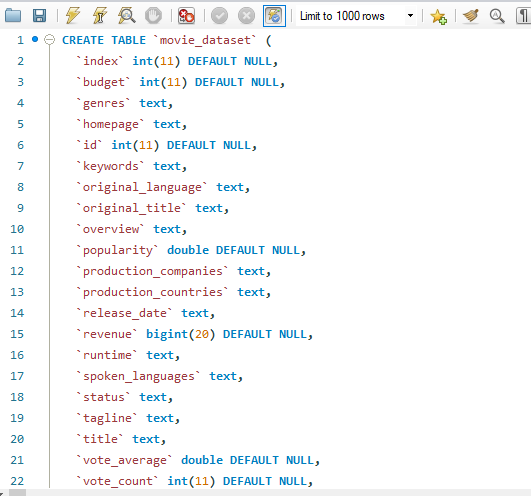


**3FN**

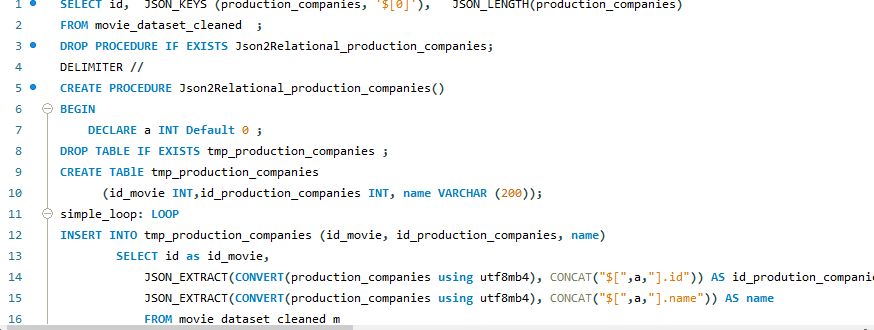
Al culminar la segunda regla normal se pasó a la tercera regla normal revisando los campos que no son transitivos y aislándolos para crear una nueva tabla para cada uno de estos datos, con esto se pudo llegar a completar las entidades necesarias y los campos que deberían ir en cada una de ellas.

En la imagen se aprecian los resultados de la tercera forma normal con un proceso llevado a cabo en Excel y con el cual se establecieron los campos transitivos y las tablas resultantes.

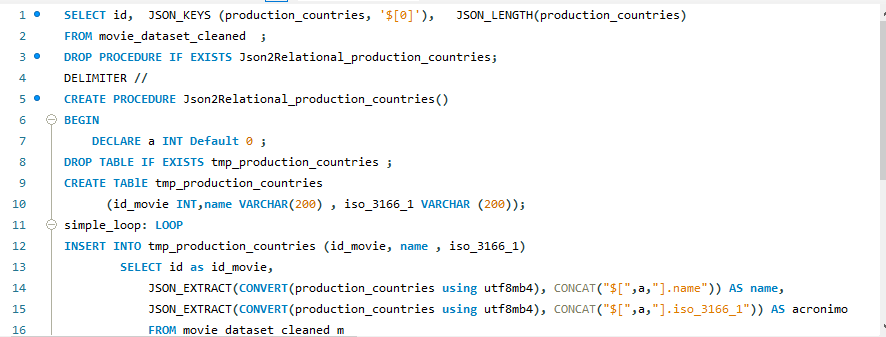


Creacion de la base de datos

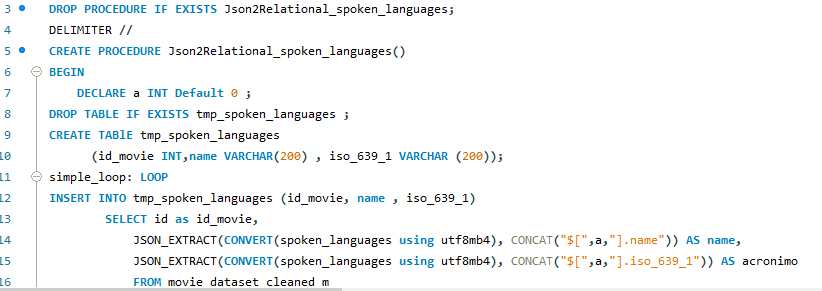
Creacion de la Talbal Produccion\_ companies



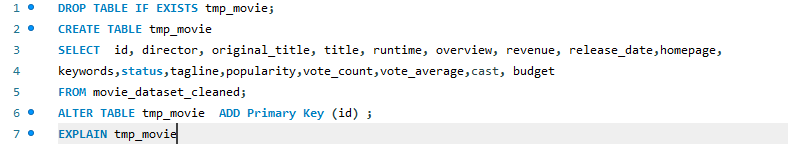
**Creacion de la tabla Production\_countries**



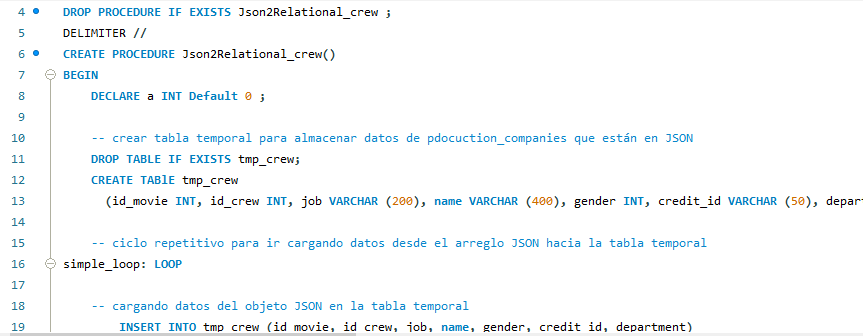
**Creacion de la tabla SPOKEN\_LENGUAGES**



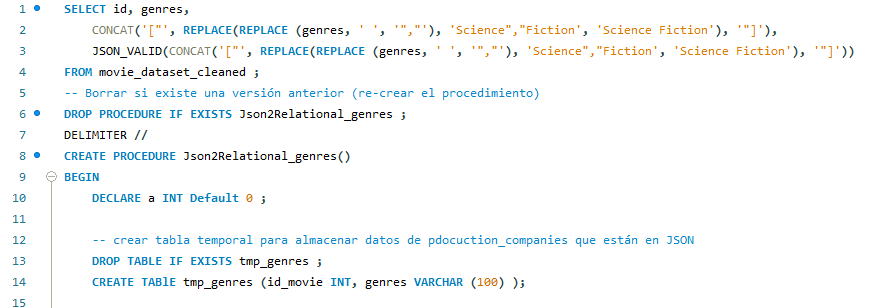
Creacion de la tabla Movie



Creacion de la tabla CREW



Cracion de la tabla Genres



**CONCLUSIONES**

El informe de catedra integradora que hemos realizados nos ha ayudado a conocernos de una forma muy importante y a destacar la importación de la materia Fundamentos de Datos y otros temas abordados en el presente informe, otro aspecto a recalcar es el uso de nuevas herramientas de trabajo que nos ayudaron a resolver la problemática del proyecto, de esta manera cada vez vamos adentrando más y más en mundo de la programación