**UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA**



**TITULACION DE INGENIERÍA EN CIENCÍAS DE LA COMPUTACIÓN**

**TEMA:**

**Catedra integradora**

**Tercer semestre.**

**Alumno:** Denis Alexander Cuenca Buele

**Tutor:** Ing. Nelson Piedra

SEGUNDO BIMESTRE

2021-2022

**Introducción:**

En el presente proyecto se aspira explicar el proceso de normailizacion de una base de datos “Movie dataset” la cual presenta e1errores de diseño los cuales imposibilitan el uso de la misma, motivo por el cual se aplicaran varias estrategias para la modificacion de esta base de datos, que nos permitan volverla un modelo que responda de forma eficiente y estable reglas de normalización y su uso en futuros procesos de implementación. En este documento se encuntra detallado el proceso para construir un modelo estable, detallando los campos exsitentes en las tablas, su errores, recomendaciones de diseño y aplicación de estos desde un enfoque práctico.

**Desarrollo:**

A continuacion se nos propone un modelo de base de datos “movie dataset” la misma que contiene informacion sobre varias películas como su nombre, presupuesto, cast, entre otras que se detallaran más adelante; la base de datos cuenta con una tabla unica “Movie” con un total de 24 atributos entre los cuales resaltan: una clave primaria, claves cadidatas, atributos multivaluados, atributos con formato JSON; todo esto en un mismo archivo csv.

A fin de crear un módelo sostenible y consistente se pretende la aplicación de relas de normalizacion sobre este primer estado de la base de datos, para lo cual se debe comprender a fondo el modelo con el que se trabajará, así pues se comienza por describir la única entidad que contiene el módelo, la tabla “Movie”.

La entidad “Movie” contiene los siguientes atributos:

MOVIE:

- index

-budget

-homepage

-id

-original language

-original title

-overview

-popularity

-realese date

-revenue

-runtime

-status

-tagline

-title

-vote average

-vote count

- genres

-keywords

-production companies

-production countries

-spoken languages

-cast

- crew

**Antes de aplicar Normalizacion:**

En la entidad “Movie” existen varios campos que no son compatibles con el concepto de modelo sostenible, según la normalizacion, por ejemplo existen atributos multivaluados a demas de una dependencia funcianal parcial en toda la tabla, lo cual se entra en conflicto con la segunda y tercera forma normal.

Los campos que presentan estos problemas en cuestion lon os siguientes:

**Campos Multivaluados (con sus respectivos atributos):**

**- genres:** id\_movie, genre

**-production companies:** id\_movie, id, company

**-production countries:** id\_movie, iso\_3166\_1, country

**-spoken languages:** Iso\_639₁, id\_movie, name

**-cast:** name, id, id\_movie

**- crew:** id\_movie, id\_crew, job, name, gender, credit\_id, department

**-Director:**  name, id, id\_movie

Los campos aquí expuestos representan un problema de primera forma normal, al ser campos que almacenan valores multivaluados, por lo cual se los separa en tablas difererntes con eperando así generar un mejor diseño, el resultado final se presenta a continuación donde se describe cada tabla con sus atrubutos:

**Dependencias funcionales de las nuevas tablas:**

A continuación se describe las dependencias funcionales de cada una de las nuevas tablas:

**- genres:**

id\_movie → genre

**-keywords:**

id\_movie, id → keyword

**-production companies:**

id\_movie, id → company name

**-production countries :**

id\_movie, iso\_3166\_1 → country

**-spoken languages :**

Iso\_639\_1, id\_movie → name

**- crew:**

id\_crew **→** name, gender

id\_movie, id\_crew, credit\_id **→** job, department

**-Director:**

id, id\_movie →director name

Siendo todas las tablas, a excepcion de crew dependencias funcionales completas las cuales responden al modelo que nos propone la segunda forma normal.

**ENTIDAD MOVIE:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Columna | Dominio | Opcional | Multivaluado | Proposito | Tipo | comentario |
| id | int(10) | FALSE | FALSE | Identificar las pelicula de forma única | Clave primaria |  |
| Vote Average | double | FALSE | FALSE | Promedio de los votos recolectados |  |  |
| Vote Count | int(8) | FALSE | FALSE | Total de votos recolectados |  |  |
| home Page | varchar(20) | FALSE | FALSE | Página oficial de la película |  | El valor introducido es una url |
| revenue | double | TRUE | FALSE | Ganancia recolectada por la película |  |  |
| budget | double | TRUE | FALSE | Presupuesto destinado para la producción de la película |  |  |
| status | varchar(10) | FALSE | FALSE | Estatus de producción de la película |  |  |
| tagline | varchar(20) | TRUE | FALSE | Lema de la película |  |  |
| original language | varchar(20) | FALSE | FALSE | Lenguaje original en el que la película fue grabada |  |  |
| original title | varchar(20) | FALSE | FALSE | Titulo seleccionado para la película |  |  |
| runtime | double | TRUE | FALSE | Duración de la proyección de la película |  |  |
| release date | date | FALSE | FALSE | Fecha de salida del juego |  | Valor introducido en el formato: D/M/A |
| keywords | varchar(200) | TRUE | FALSE | Palabras que describen las peliculas |  | Campo multivaluado |
| overview | Varchar(500) | TRUE | FALSE | Sinopsis de la trama principal de la película |  |  |
| popuarity | Int(15) | TRUE | FALSE | Popularidad en puntos que obtuvo la película |  |  |
| title | Varchar(30) | TRUE | FALSE | Título de la película |  |  |
| Caste | Varchar(300) | TRUE | FALSE | Nombre de la trabajadores |  |  |

**Entidad Production\_companies:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Columna | tipo de dato | Opcional | Multivaluado | Propósito | tipo | comentario |
| id | Int(10) | FALSE | FALSE | Identificar de forma única la columna | Clave primaria |  |
| name | Varchar(20) | FALSE | FALSE | Nombre de la compañía productora del película |  |  |
| Id(Movie) | int(10) | FALSE | FALSE | Identificar las pelicula de forma única | Clave primaria |  |

**Entidad Production\_countries:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Columna | tipo de dato | Opcional | Multivaluado | Propósito | tipo | comentario |
| name | Varchar(15) | FALSE | FALSE | Nombre del país donde se produjo la película. |  |  |
| iso\_3166\_1 | Varchar(2) | FALSE | FALSE | Código de abreviación de un país | Clave primaria |  |
| Id(Movie) | int(10) | FALSE | FALSE | Identificar las pelicula de forma única | Clave primaria |  |

**Entidad Spoken\_languages:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Columna | tipo de dato | Opcional | Multivaluado | Propósito | tipo | comentario |
| name | Varchar(20) | FALSE | FALSE | Nombre del lenguaje usado en la película. |  |  |
| Iso\_639\_1 | Varchar(15) | FALSE | FALSE | Código de abreviación de un lenguaje | Clave primaria |  |
| Id(Movie) | int(10) | FALSE | FALSE | Identificar las pelicula de forma única | Clave primaria |  |

**Entidad crew:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Columna | tipo de dato | Opcional | Multivaluado | Propósito | tipo | comentario |
| name | Varchar(40) | FALSE | FALSE | Nombre del trabajador |  |  |
| Gender | Int(1) | FALSE | FALSE | Genero del trabajador |  | Atributo identificado con 2 para hombre y 1 para mujer |
| Credit\_id | Varchar(20) | FALSE | FALSE | Identificador del crédito del trabajador |  |  |
| Id | Int(5) | FALSE | FALSE | Identifica de forma única a cada trabajador | Clave primaria |  |
| Id(Movie) | int(10) | FALSE | FALSE | Identificar las pelicula de forma única | Clave primaria |  |

**Director:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Columna | tipo de dato | Opcional | Multivaluado | Propósito | tipo | comentario |
| name | Varchar(40) | FALSE | FALSE | Nombre del Director |  |  |
| id | int(10) | FALSE | FALSE | Número de identificacion de director | Clave primaria |  |
| Id(Movie) | int(10) | FALSE | FALSE | Identificar las pelicula de forma única | Clave primaria |  |

**Genres:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Columna | tipo de dato | Opcional | Multivaluado | Propósito | tipo | comentario |
| Genre | Varchar(20) | FALSE | FALSE | Nombre del género |  |  |
| id | int(10) | FALSE | FALSE | Número de identificacion para el género | Clave primaria |  |
| Id(Movie) | int(10) | FALSE | FALSE | Identificar las pelicula de forma única | Clave primaria |  |

**Department:**

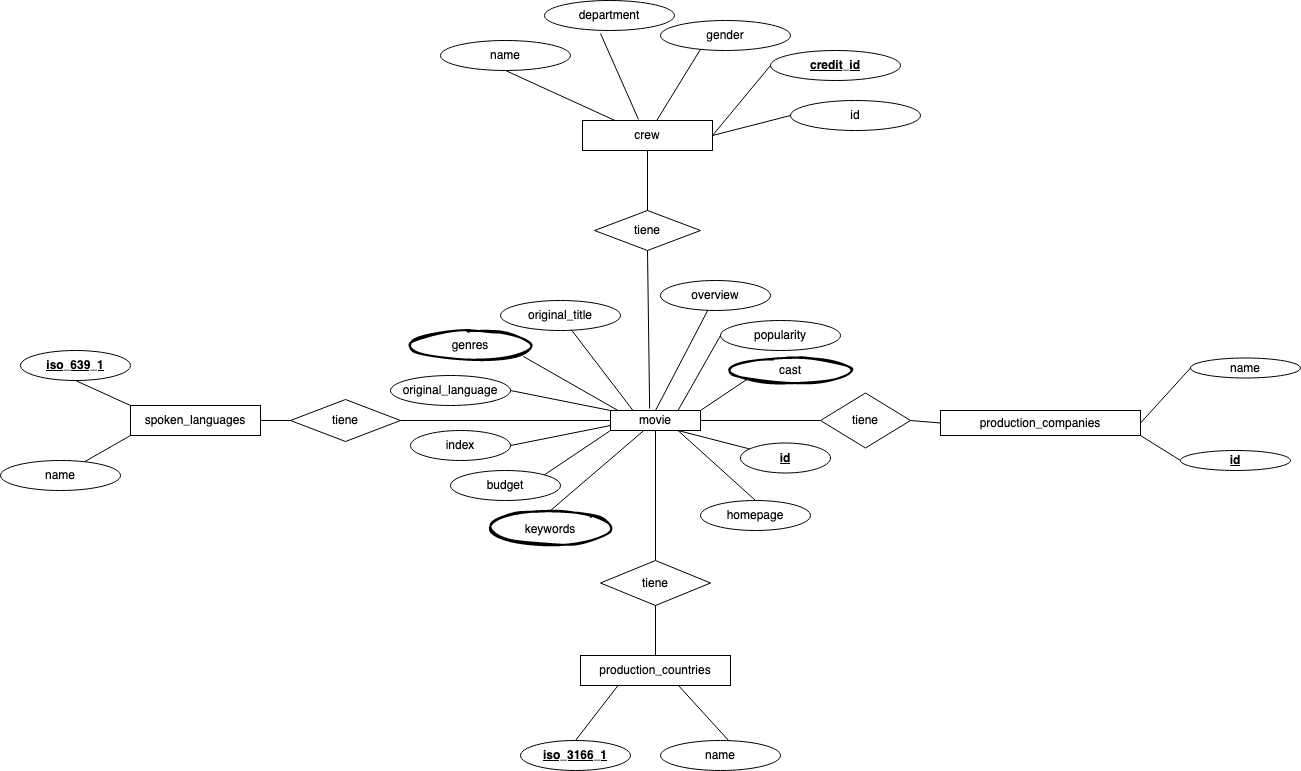
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Columna | tipo de dato | Opcional | Multivaluado | Propósito | tipo | comentario |
| name | Varchar(20) | FALSE | FALSE | Nombre del departamento de trabajo |  |  |
| id | int(10) | FALSE | FALSE | Número de identificacion de departamento | Clave primaria |  |
| Id(Crew) | int(10) | FALSE | FALSE | Identificar el personal de forma única | Clave primaria |  |

**Job**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Columna | tipo de dato | Opcional | Multivaluado | Propósito | tipo | comentario |
| name | Varchar(20) | FALSE | FALSE | Nombre del trabajo |  |  |
| id | int(10) | FALSE | FALSE | Número de identificacion de puesto | Clave primaria |  |
| Id(Crew) | int(10) | FALSE | FALSE | Identificar el personal de forma única | Clave primaria |  |

**Diagrama Entidad Relación:**

Avontinuación se presenta el resultado de separar cada una de las tablas y se presenta la relación existente entre ellas, notece que por el estilo de flecha la cardinalidad es de muchos a muchos (m:n) entre la mayoria de las entidades



En el diagrma, todas las entidades, a excepcion de department y job, tienen relacion con la entidad Movie, pues es esa la entidad de donde surgen, la relación de movie con keywords es de muchos a muchos, con ello se podrá guardar por una pelicula varias palabras clave, y una palabra clave será capaz de describir a varias peliculas; similar es el caso para las entidades genres, director, cast y spoken languages. Production companies y movie comparten la misma cardinalidad muchos a muchoas en donde una compañía será capaz de producir varias peliculas y una pelicula producida por muchas compañias; lo mismo sucede con la entidad production cuontrie, La columna crew conservo un formato JSON en donde se detallavan a forma de clave-valor los campos que ha de contener una tablas individual, en caso que separace la columna como un entidad, en esta tabla las persona pertenecientes al crew son capaces de pertenecer a uno o más pepartementos, y estas mismas personas podran ejercer una o mas funciones de trabajo en la pelicuala, por lo que estas son estidades que comparten una cardinalidad de muchos a muchos con la entidad crew.

**Tercera forma normal:**

No debe existir dependencias funcionales trancitivas, partiendo de esto se pantes crear desde la tabla movie y crew crear la conexión que permita establecer la tercera forma norma entre las tablas que necesiten ser conectadas llevando a la tabla padre los id de las tablas hijas.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Columna | Dominio | Opcional | Multivaluado | Proposito | Tipo | comentario |
| id | int(10) | FALSE | FALSE | Identificar las pelicula de forma única | Clave primaria |  |
| Vote Average | double | FALSE | FALSE | Promedio de los votos recolectados |  |  |
| Vote Count | int(8) | FALSE | FALSE | Total de votos recolectados |  |  |
| home Page | varchar(20) | FALSE | FALSE | Página oficial de la película |  | El valor introducido es una url |
| revenue | double | TRUE | FALSE | Ganancia recolectada por la película |  |  |
| budget | double | TRUE | FALSE | Presupuesto destinado para la producción de la película |  |  |
| status | varchar(10) | FALSE | FALSE | Estatus de producción de la película |  |  |
| tagline | varchar(20) | TRUE | FALSE | Lema de la película |  |  |
| original language | varchar(20) | FALSE | FALSE | Lenguaje original en el que la película fue grabada |  |  |
| original title | varchar(20) | FALSE | FALSE | Titulo seleccionado para la película |  |  |
| runtime | double | TRUE | FALSE | Duración de la proyección de la película |  |  |
| release date | date | FALSE | FALSE | Fecha de salida del juego |  | Valor introducido en el formato: D/M/A |
| overview | Varchar(500) | TRUE | FALSE | Sinopsis de la trama principal de la película |  |  |
| popuarity | Int(15) | TRUE | FALSE | Popularidad en puntos que obtuvo la película |  |  |
| title | Varchar(30) | TRUE | FALSE | Título de la película |  |  |
| id\_genres | int(10) | FALSE | FALSE | Número de identificacion de genero | Clave foranea |  |
| Id  production  companies | Int(10) | FALSE | FALSE | Identificar de forma única la columna | Clave foranea |  |
| Id  production countries | Varchar(2) | FALSE | FALSE | Código de abreviación de un país | Clave foranea |  |
| Id spoken languages | Varchar(15) | FALSE | FALSE | Código de abreviación de un lenguaje | Clave foranea |  |
| Id crew | Int(5) | FALSE | FALSE | Identifica de forma única a cada trabajador | Clave foranea |  |
| Id director | int(10) | FALSE | FALSE | Número de identificacion de director | Clave foranea |  |
| Id cast | int(10) | FALSE | FALSE | Número de identificacion para personal | Clave foranea |  |

**Entidad crew:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Columna | tipo de dato | Opcional | Multivaluado | Propósito | tipo | comentario |
| name | Varchar(40) | FALSE | FALSE | Nombre del trabajador |  |  |
| Gender | Int(1) | FALSE | FALSE | Genero del trabajador |  | Atributo identificado con 2 para hombre y 1 para mujer |
| Credit\_id | Varchar(20) | FALSE | FALSE | Identificador del crédito del trabajador |  |  |
| Id | Int(5) | FALSE | FALSE | Identifica de forma única a cada trabajador | Clave primaria |  |
| Id(Movie) | int(10) | FALSE | FALSE | Identificar las pelicula de forma única | Clave primaria |  |
| Id\_ department | int(10) | FALSE | FALSE | Número de identificacion de puesto | Clave foranea |  |
| Id\_job | int(10) | FALSE | FALSE | Número de identificacion de puesto | Clave Foranea |  |

**Concluciones:**

**-** El uso de las formas normales es fundamental en la creacion de una base de datos, donde la informacion está simpre expuesta a errores de almacenamiento, y se corre el riesgo de perderla o hacer una mal uso de esta por malas practicas.

- La base de datos\_movi dataset tenia errores de diseño los cuales imposibilitan el uso coorecto de los datos almacenados, volviendola para algunos fines una base de datos inutil, almenos como se presento en una primera etapa, sin enbargo tras la aplicación de formas normales y otras modificaciones que nos permitan su mejor uso movi dataset se funcional y estable.

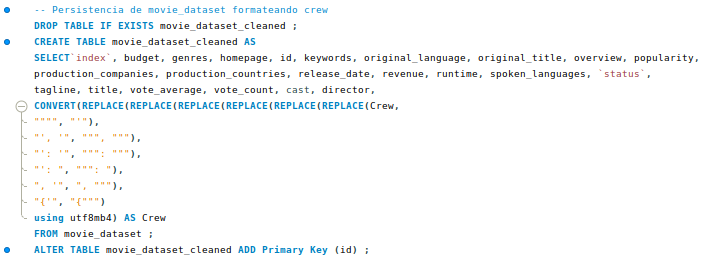
**Anexos:**

**Aplicacion en Mysql:**

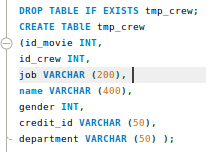
A continuacion se presenta el código escrito en mysql para la creacion de una base de datos fisica con las especificaciones que ya fueron dictadas:

**Pasos previos:**

Limpiar la tabla movie en los campos que contienen un formato JSON, para obtener un resultado adecuado en su futuro uso:



**- Creación de las tablas:**

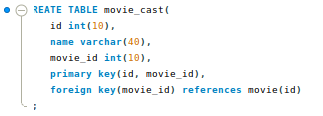
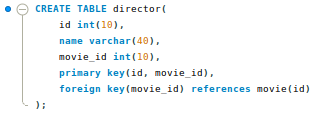


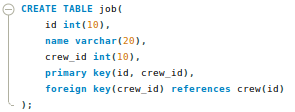
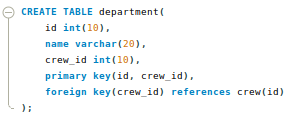


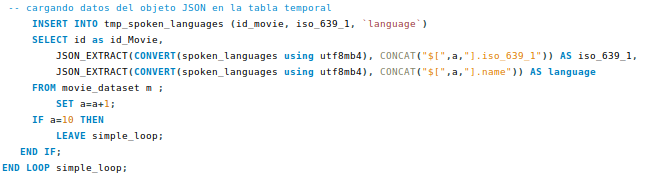


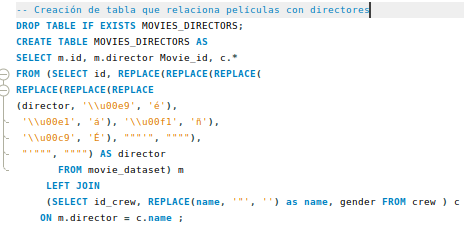


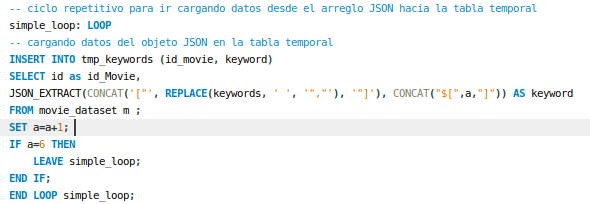
****

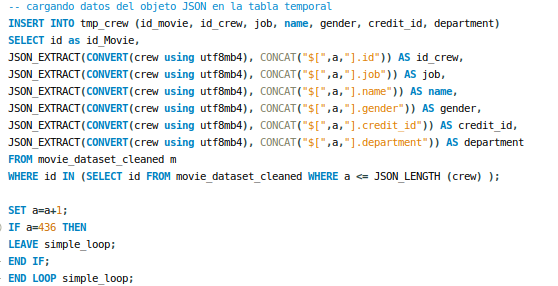
****

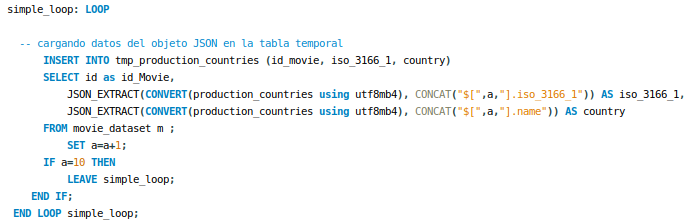
****

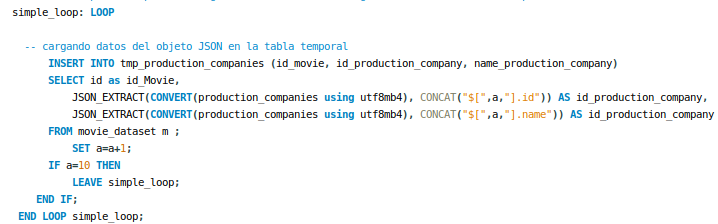
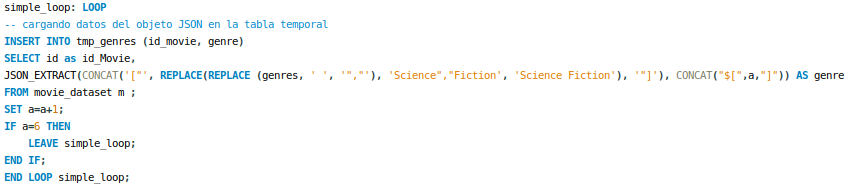
 **Cargar Datos en las tablas:**

****

****

****

****

****