

# Manejo de Datos Proyecto

Profesor: Ricardo López López  
Ayudante: Gibran Aguilar Zuñiga  
Ayudante: Rogelio Alcántar Arenas

Fecha de entrega: 26 de enero

## 1. Introducción

Las películas del Universo cinematográfico de Marvel (MCU por sus siglas en inglés) son una serie de películas de superhéroes estadounidenses, basadas en personajes que aparecen en publicaciones de Marvel Comics. Las cintas han estado en producción desde 2007, y desde ese entonces Marvel Studios ha producido más de veinte películas y se tienen otras en etapas de producción.

La idea de este proyecto es crear una pequeña base de datos para almacenar información de estas películas y ejemplificar el uso de las bases de datos en la elaboración de consultas y pequeños reportes desde el lenguaje de programación Python.

## 2. Problemática

El problema general es crear una base de datos y poblarla (o llenar sus tablas) con datos tomados de un archivo de texto y realizar algunas consultas.

Dicho archivo contiene un conjunto de datos con información de las películas del MCU. Estos datos se encuentran almacenados en el archivo `mcu.txt`. Los datos fueron capturados de forma manual por distintas personas, por lo que pueden contener pequeños errores de captura, tales como:

- Títulos o nombres escritos de distinta forma, por ejemplo: Robert Downy Jr., Robert Downy Junior o Robert Downy JR.
- Pequeñas faltas de ortografía, por ejemplo Spyder-Man en lugar de Spider-Man.
- Datos duplicados, por ejemplo: Dos renglones con la misma película o escritas de forma distinta.

Antes de pensar en la base de datos se deben *limpiar* los datos. Es decir eliminar los errores descritos anteriormente junto con los renglones duplicados.

Después de limpiar los datos ahora si podemos abordar la base de datos. En esta ocasión el experto en bases de datos nos ha proporcionado el diagrama del esquema de la base de datos. Ver Figura 2. Sin embargo es parte de nuestro trabajo crear la base de datos siguiendo dicho diagrama.

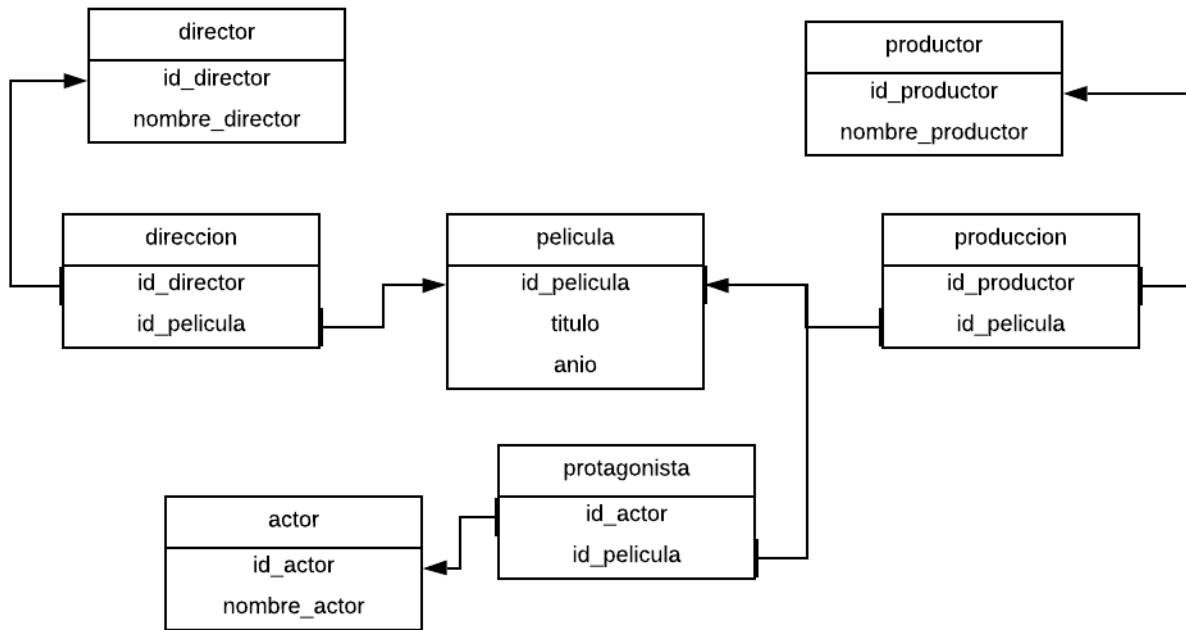


Figura 1: Diagrama de base de datos.

Al terminar de crear la base de datos debemos poblarla con los datos del archivo de texto. También es parte de nuestro trabajo obtener los datos necesarios para llenar todas y cada una de las tablas creadas.

Cada renglón del archivo contiene los datos de una película, por ejemplo: título, director, protagonistas, etc. Se deben tomar los datos necesarios para agregarse en cada tabla evitando tener registros repetidos.

Por último se tienen que realizar las siguientes consultas sobre la base de datos obtenida:

- Obtener el título y el año de todas las películas estrenadas entre los años 2015 y 2020.
- Obtener el título y el año de todas las películas dirigidas por Jon Watts ordenadas

alfabéticamente.

- Obtener el título y el productor(es) de todas las películas que contengan la palabra 'Man' en su título.
- Obtener el título, el año, y el director(es) de todas las películas en las que Benedict Cumberbatch es protagonista.
- Obtener el título, el año, director y productor de todas las películas que fueron producidas únicamente por Kevin Feige.

### 3. Entregables

Recuerda que la entrega del proyecto final es en equipos de tres integrantes. Como resultado de la solución de tu proyecto debes entregar:

- El archivo `datos_limpios.txt`, el cual contiene los datos después de eliminar los errores de captura ya descritos.
- El script de SQL con el esquema de la base de datos que construiste.
- El script(s) de Python con todas tus clases y funciones utilizados para poblar la base de datos.
- El script(s) de Python donde realizas todas las consultas pedidas.
- Un archivo `readme.txt` con los datos de los miembros del equipo.

¡Suerte!