

Data Book, Part 3

Lauren M. Harris

Department of Library and Information Science, Indiana University Purdue University-
Indianapolis

LIS-S 511: Database Design

Steven Thompson

December 10, 2023

Table of Contents

Part 1

Data Narrative

Data Dictionary

Business Rules

Entity Relationship Model (ERM) Components

Part 2

Relational schemas

Diagrams

Visio

phpMyAdmin

Normalized Table

Part 3

SQL Queries

Part 1

Data Narrative

Cultural institutions like museums, libraries, and galleries, provide various outreach programs for their users. Film programming is a common one for these institutions. For this reason, I chose to use Newfields, an art museum in Indianapolis, for this data book assignment. The public programming department has multiple film programs including the most popular and longest program “Summer Nights”. The program started in 1979 and involves showing films outside on the museum property from June to early August. It is a mix of stereotypical summer films like Action *Predator*, classic films like *The Godfather*, and cult favorites like *The Princess Bride*. Now the museum collaborates with other organizations like The Heartland Film Festival to show short films during the program run. Newfields also has programs like “Cereal Cinema”, showing children and family films along with cereal to eat for guests, and other one-off film programming related to the museum’s general programming and exhibits. The department should be able to track the movies being shown at the institution, who the leading actors are, the director, and the release date. Knowing the release date is important since Newfields does promote the movies, movie studios have their own rules about distribution after being made available physically or digitally.

The main users of the database would be the public programming department staff, especially the manager of film programming. If a guest requests information about a specific program but does not remember the date, staff would be able to find that information. For more popular films, knowing when the movies are released, and the studio information is important to know because it does cost more to get the distribution rights since Newfields charges a fee to see the films. Knowing movie release dates,

staff can also have anniversary viewings and already have future programming decided saving time and energy in the future. With streaming sites being a common source of entertainment, knowing if a film is available to see at home can influence attendance. Attendance is important when deciding what to show and at the very least breaking even on expenses related to programming.

Knowing if a film was an award winner is important for the program description on marketing materials or training volunteers to let guests know when they arrive. Some guests like to know trivial facts about the film and if it was critically acclaimed or just a niche film being shown. There are tables related to the Academy Awards, Golden Globes, and Screen Actors Guild (SAG) awards. There could be relationships where one film wins the Best Picture award and some films noted have won awards and have not been shown. Showing that relationship could give staff context to be creative with the programming when coming up with ideas.

Knowing which actors or directors a film, staff can make decisions to have diverse films to be shown. If they notice a potential bias of only certain actors or directors being shown, they can edit the list of movies before deadlines.

There are 11 tables: ACTOR, STUDIO, MOVIE, MOVIE_ACT, DIRECTOR, STREAMING, FILM_STREAM, NEWFIELDS_FP, AA_BP, GG_DMP, AND SAG_OC. The relationships between the tables would allow staff to create descriptions about the films, make decisions if a film will be shown given if it is available online, and studio contact information.

Data Dictionary

Table Name	Attribute Name	Contents	Type	Format	Range	Required	PK or FK	FK Referenced Table
ACTOR	ACTOR_FNAME	Performer's first name	VARCHAR			Y		
	ACTOR_LNAME	Performer's last name	VARCHAR			Y		
	ACTOR_INITAL	Performer's middle initial	CHAR(1)					
	ACTOR_DOB	Performer's date of birth	DATE	DD-MM-YYY				
	ACTOR_ID	Performer's code	CHAR(5)	ACTXX	ACT01-ACT99	Y	PK	
STUDIO	STUDIO_NAME	Studio's name	VARCHAR			Y		
	STUDIO_CITY	Studio's headquarter city	VARCHAR					
	STUDIO_STATE	Studio's headquarter state	CHAR(2)	XX				
	STUDIO_WEBSITE	Studio's website				N		
	STUDIO_ID	Studio's code	CHAR(4)	5XXX	5000-5999	Y	PK	
MOVIE	MOVIE_TITLE	Movie's title	VARCHAR			Y		
	MOVIE_GENRE	Movie's genre	VARCHAR					
	MOVIE_RELEASEYR	Movie's release year	YEAR	YYYY		Y		
	STUDIO_ID	Studio's code	CHAR(4)	5XXX	5000-5999	Y	FK	STUDIO
	DIRECTOR_ID	Director's code	CHAR(5)	DIRXX	DIR01-99			DIRECTOR
	MOVIE_ID	Movie's code	CHAR(4)	1XXX	1000-1999	Y	PK	

MOVIE_ACT	MOVIE_ID						FK	MOVIE
	ACTOR_ID		CHAR(5)	ACTXX	ACT01- ACT99		FK	ACTOR
DIRECTOR	DIRECTOR_FNAME	Director's first name	VARCHAR			Y		
	DIRECTOR_LNAME	Director's last name	VARCHAR			Y		
	DIRECTOR_INTIAL	Director's middle initial	CHAR(1)					
	DIRECTOR_DOB	Director's date of birth	DATE	DD-MM-YYY				
	DIRECTOR_ID	Director's code	CHAR(5)	DIRXX	DIR01- DIR99	Y	PK	
STREAMING	STREAM_NAME	Steaming site name	VARCHAR					
	STREAM_WEBSITE	Streaming site website	VARCHAR					
	STREAM_CITY	Streaming site's headquarter city	VARCHAR					
	STREAM_STATE	Streaming site's headquarter state	CHAR(2)	XX				
	STREAM_ID	Streaming site code	CHAR(5)	SIDXX	SID01- 99	Y	PK	
FILM_STREAM	MOVIE_ID	Movie's code	CHAR(4)	1XXX	1000- 1999	Y	FK	MOVIE
	MOVIE_TITLE	Movie's Title	VARCHAR			Y		
	DIRECTOR_ID	Director's code	CHAR(5)	DIRXX	DIR01- DIR99	Y	FK	DIRECTOR
	ACTOR_ID	Performer's code	CHAR(5)	ACTXX	ACT01- ACT99	Y	FK	ACTOR
	MOVIE_RELEASEYR	Movie's release year	YEAR	YYYY				

	FSTREAM_DATE	Date when movie is available to stream	DATE	YYYY		Y		
	STREAM_ID	Streaming site code	CHAR(5)	SIDXX	SID01-SID99	Y		STREAMING
NEWFIELDS_FP	MOVIE_ID	Movie's code	CHAR(4)	1XXX	1000-1999	Y	FK	MOVIE
	NEW_PDATE	The date Newfield's showed the movie	YEAR	YYYY		Y		
	NEW_PCODE	Newfield's program identifier	CHAR(3)	NXX	N01-N99	Y	PK	
AA_BP	MOVIE_ID	Movie's code	CHAR(4)	1XXX	1000-1999	Y	FK	MOVIE
	STUDIO_NAME	Studio's name	VARCHAR				FK	STUDIO
	MOVIE_RELEASEYR	Movie's release year	DATE	YYYY			FK	MOVIE
	DIRECTOR_ID	Director's code	CHAR(5)	DIRXX		Y	FK	DIRECTOR
	AWARD_ID	Academy Award Best Picture Award Code	CHAR(4)	AAXX	AA01-AA99	Y	PK	
GG_DMP	MOVIE_ID	Movie's code	CHAR(4)	1XXX	1000-1999	Y	FK	MOVIE
	STUDIO_NAME	Studio's name	VARCHAR				FK	STUDIO
	MOVIE_RELEASEYR	Movie's release year	YEAR	YYYY			FK	MOVIE
	DIRECTOR_ID	Director's code	CHAR(5)	DIRXX		Y	FK	DIRECTOR
	AWARD_ID	Golden Gold Drama Motion	CHAR(4)	GGXX	GG01-GG99	Y	PK	

		Picture Award Code						
SAG_OC	MOVIE_ID	Movie's code	CHAR(4)	1XXX	1000-1999	Y	FK	MOVIE
	STUDIO_NAME	Studio's name	VARCHAR				FK	STUDIO
	MOVIE_RELEASEYR	Movie's release year	YEAR	YYYY			FK	MOVIE
	DIRECTOR_ID	Director's code	CHAR(5)	DIRXX		Y	FK	DIRECTOR
	AWARD_ID	Screen Actor Guild Original Cast Award Code	CHAR(5)	SAGXX	SAG01-SAG99	Y	PK	

Business Rules

1. ACTOR/MOVIE_ACT
Many ACTOR can star in many MOVIE_ACT
2. DIRECTOR/MOVIE
DIRECTOR can direct one or many MOVIE
3. STUDIO/MOVIE
STUDIO produces one or many MOVIE
4. NEWSFIELDS/MOVIE
NEWSFIELDS shows one MOVIE on program date
5. FILM_STREAM/MOVIE
FILM_STREAM streams one or more MOVIE
6. MOVIE/AA_BP
Only one MOVIE wins AA_BP for that year
7. MOVIE/GG_DMP
Only one MOVIE wins GG_DMP for that year
8. MOVIE/SAG_OC
Only one MOVIE wins SAG_OC for that year

Entity Relationship Model (ERM) Components

ENTITY	RELATIONSHIP	CONNECTIVITY	ENTITY
ACTOR	Stars in	M:N	MOVIE_ACT
DIRECTOR	Directs	1:M	MOVIE
STUDIO	Produces	1:M	MOVIE
NEWSFIELDS	Shows	1:1	MOVIE
FILM_STREAM	Streams	1:M	MOVIE
MOVIE	Wins	1:1	AA_BP
MOVIE	Wins	1:1	GG_DMP
MOVIE	Wins	1:1	SAG_OC

Part 2

Relational schemas

ACTOR (ACTOR_ID, ACTOR_FNAME, ACTOR_LNAME, ACTOR_INITIAL, ACTOR_DOB)

DIRECTOR (DIRECTOR_ID, DIRECTOR_FNAME, DIRECTOR_LNAME, DIRECTOR_INITIAL, DIRECTOR_DOB)

MOVIE (MOVIE_ID, STUDIO_ID, ACTOR_ID, DIRECTOR_ID, MOVIE_TITLE, MOVIE_GENRE, MOVIE_RELEASEYR)

MOVIE_ACT(MOVIE_ID, ACTOR_ID)

STUDIO (STUDIO_ID, STUDIO_NAME, STUDIO_CITY, STUDIO_STATE, STUDIO_WEBSITE)

FILM_STREAM (MOVIE_ID, DIRECTOR_ID, ACTOR_ID, MOVIE_RELEASEYR, MOVIE_TITLE, FSTREAM_DATE)

STREAMING (STREAM_ID, STREAM_NAME, STREAM_WEBSITE, STREAM-CITY, STREAM_STATE)

NEWSFIELDS_FP (NEW_PCODE, MOVIE_ID, NEW_PDATE)

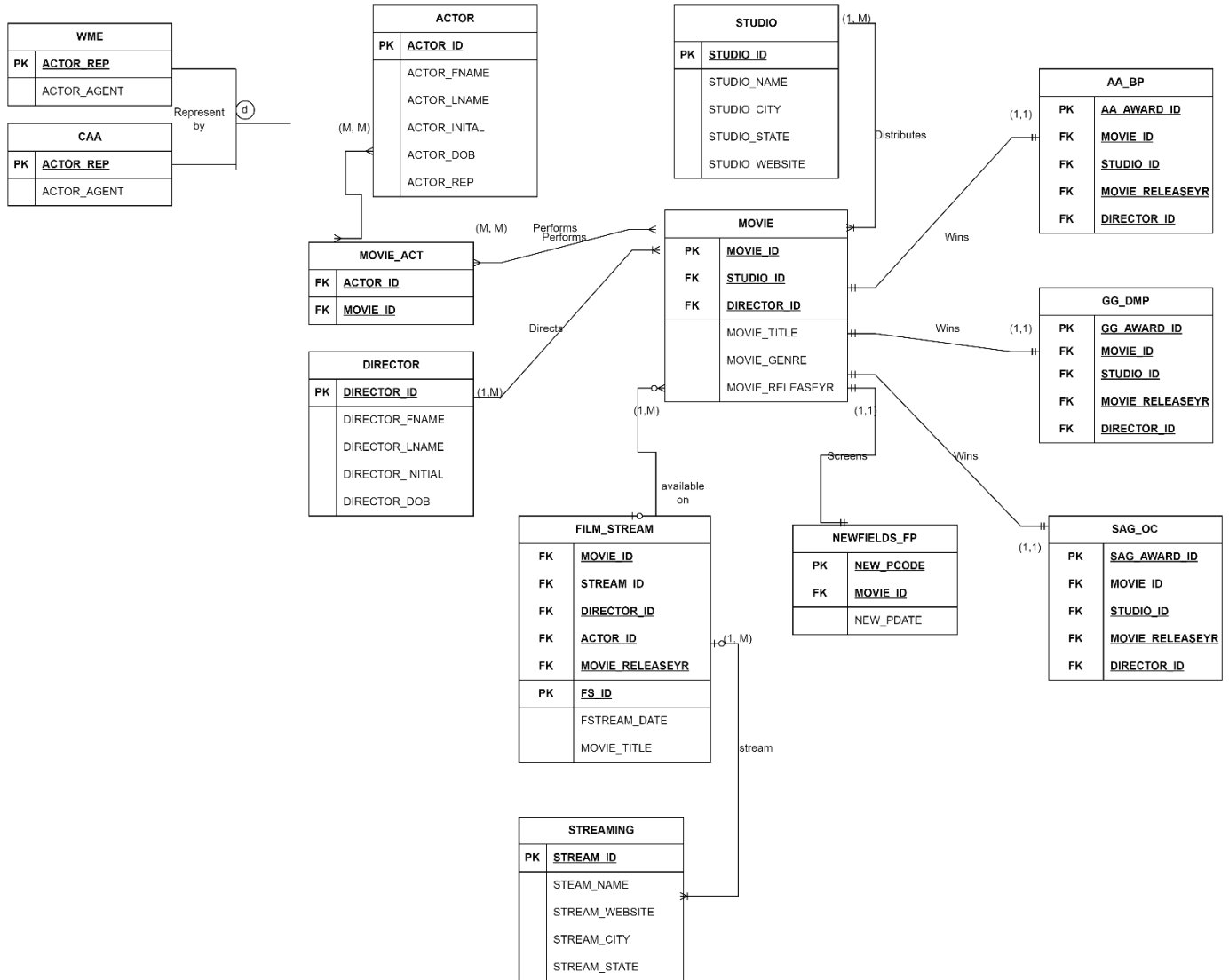
AA_BIP (AA_AWARD_ID, MOVIE_ID, STUDIO_ID, MOVIERELEASE_YR, DIRECTOR_ID, AWARD_YR)

GG_DMP (GG_AWARD_ID, MOVIE_ID, STUDIO_ID, MOVIERELEASE_YR, DIRECTOR_ID, AWARD_YR)

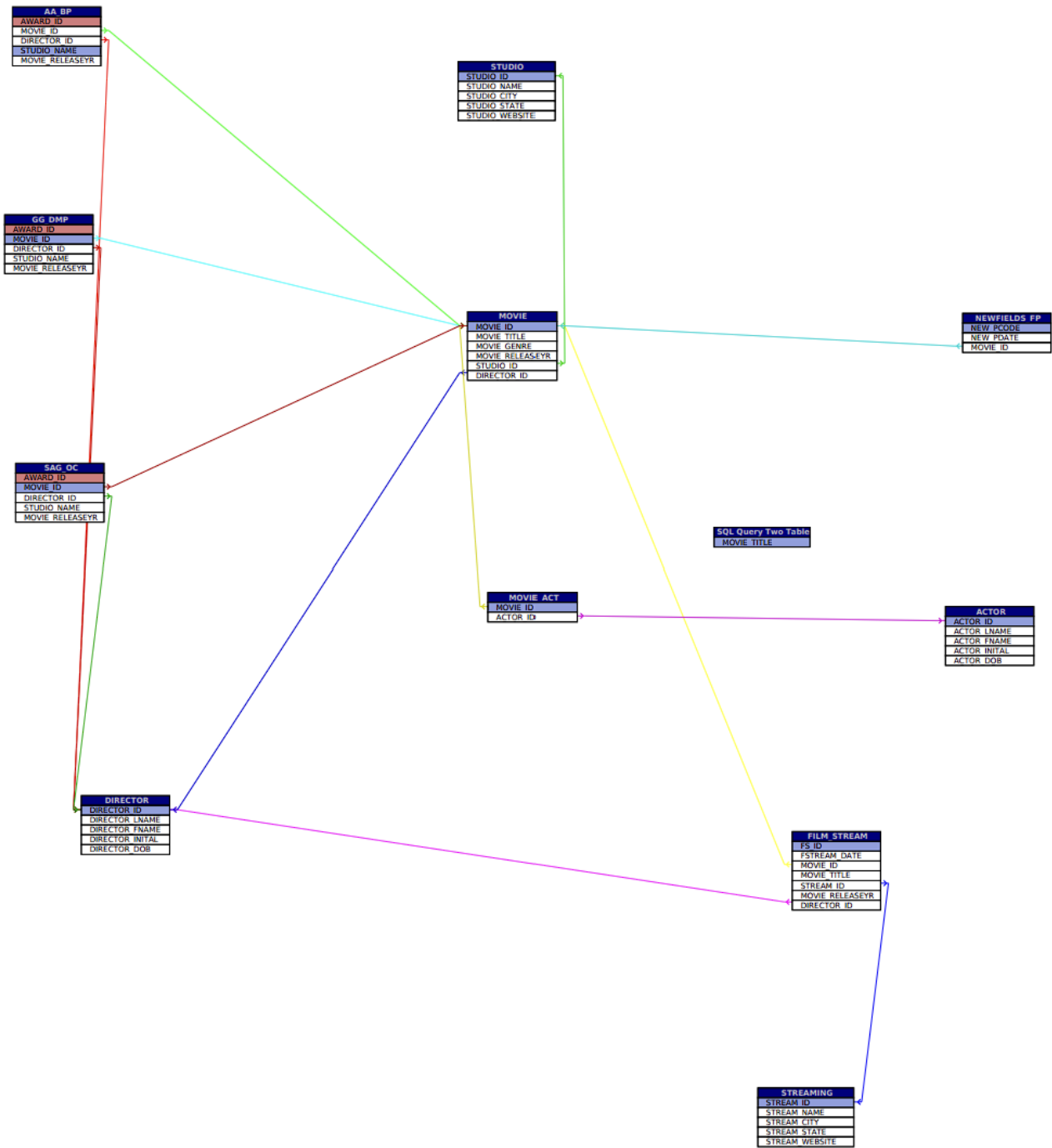
SAG_OC (SAG_AWARD_ID, MOVIE_ID, STUDIO_ID, MOVIERELEASE_YR, DIRECTOR_ID, AWARD_YR)

Diagrams

App.diagrams.net



PhpMyAdmin

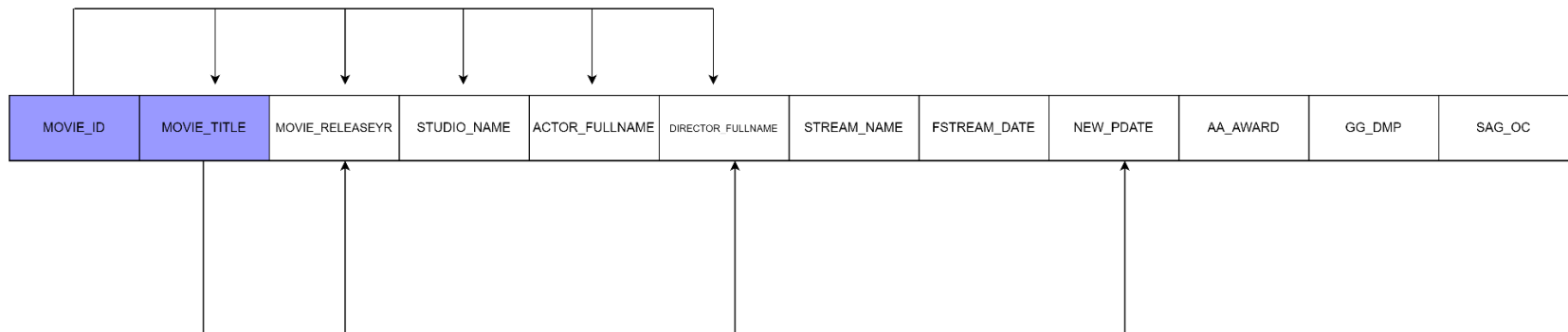


Normalized Table

1NF

MOVIE_ID	MOVIE_TITLE	MOVIE_RELEASEYR	STUDIO_NAME	ACTOR_FULLNAME	DIRECTOR_FULLNAME	STREAM_NAME	FSTREAM_DATE	NEW_PDATE	AA_AWARD	GG_DMP	SAG_OC
10001	Gone With The Wind	1939	Metro-Goldwyn-Mayer Studios, Inc	Gable, Clark	Fleming, Victor	Max	2022	1978			
1001	Gone With The Wind	1939	Metro-Goldwyn-Mayer Studios, Inc	Leigh, Vivien				1978			
1001	Gone With The Wind	1939	Metro-Goldwyn-Mayer Studios, Inc	Howard, Leslie				1978			
1002	Streetcar Named Desire, A	1951	Warner Bros. Pictures	Leigh, Vivien	Kazan, Elia			1999			
1002	Streetcar Named Desire, A	1951		Brando, Marlon							
1003	Ben Hur	1959	Metro-Goldwyn	Heston, Charlton	Wyler, William			1979	1959	1959	

			-Mayer Studios, Inc								
1004	Rocky	1976	United Artists Corporation	Stallone, Sylvester	Avildsen, John G.	Max		1987	1976	1976	
1004	Rocky	1976	United Artists Corporation	Shire, Talia		Amazon Prime		1987			
1005	The Godfather	1972	Paramount Pictures	Brando, Marlon	Coppola, Francis F.		2014		1972		



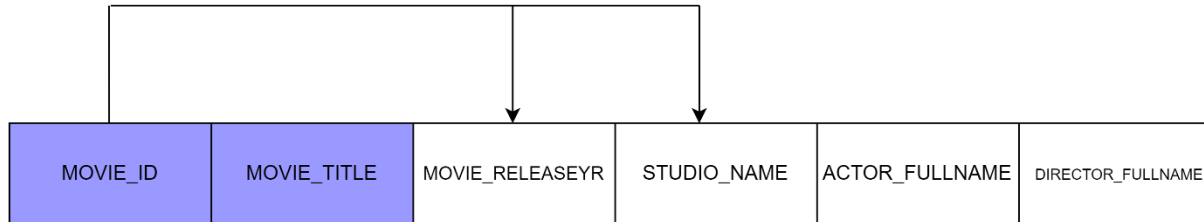
MOVIE (**MOVIE_ID**, **MOVIE_TITLE**, MOVIE_RELEASEYR, STUDIO_NAME, ACTOR_FULLNAME, DIRECTOR_FULLNAME, STREAM_NAME, FSTREAM_DATE, NEW_PDATE, AA_AWARD, GG_DMP, SAG_OC)

Dependencies

MOVIE_TITLE (MOVIE_TITLE, MOVIE_RELEASEYR, DIRECTOR_FULLNAME, NEW_PDATE)

There are values like ACTOR_FULLNAME that has repeating values and for NEWP_DATE has nulls because not every movie has been screened at Newfields. To get the table to 2NF I removed values that repeated like ACTOR_FULLNAME and null values like NEW_PDATE.

2NF

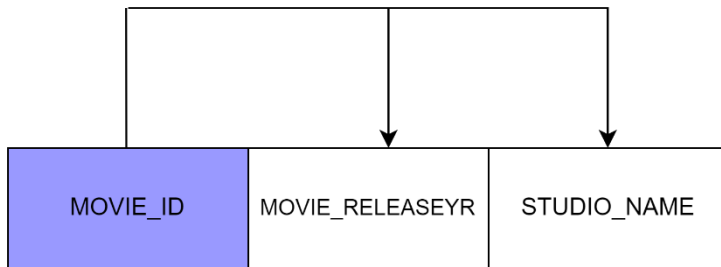


MOVIE (**MOVIE ID, MOVIE TITLE**, MOVIE_RELEASEYR, STUDIOO_NAME, ACTOR_FULLNAME, DIRECTOR_FULLNAME)

MOVIE_ID	MOVIE_TITLE	MOVE_RELEASEYR	STUDIO_NAME	ACTOR_FULLNAME	DIRECTOR_FULLNAME
10001	Gone With The Wind	1939	Metro-Goldwyn-Mayer Studios, Inc	Gable, Clark	Fleming, Victor
1001	Gone With The Wind	1939	Metro-Goldwyn-Mayer Studios, Inc	Leigh, Vivien	Fleming, Victor
1001	Gone With The Wind	1939	Metro-Goldwyn-Mayer Studios, Inc	Howard, Leslie	Fleming, Victor
1005	The Godfather	1972	Paramount Pictures	Brando, Marlon	Coppola, Francis F.

Even in 2NF there were repeats for actors so to minimize that I removed the actors and only kept MOVIE_ID, MOVIE_RELEASEYR, and STUDIO_NAME.

3NF



MOVIE (**MOVIE ID**, MOVIE_RELEASEYR, STUDIO_NAME)

MOVIE_ID	MOVE_RELEASEYR	STUDIO_NAME
10001	1939	Metro-Goldwyn-Mayer Studios, Inc
1004	1976	United Artists Corporation
1005	1972	Paramount Pictures

There is no dependencies nor repeating values in the table.

SQL Queries

Questions

1. How many movies has Paramount Pictures produced?
2. How many movies of the genre drama has Newfields shown?
3. What is the average film release date to when Newfields showed the film?

SQL Queries

1.

```
SELECT COUNT(STUDIO_ID)
FROM MOVIE
WHERE STUDIO_ID = '5007';
```
2.

```
SELECT MOVIE_TITLE
FROM MOVIE
JOIN NEWFIELDS_FP
ON MOVIE.MOVIE_ID=NEWFIELDS_FP.MOVIE_ID
WHERE MOVIE_GENRE = 'drama';
```
3.

```
SELECT MOVIE_ID, NEW_PCODE, NEW_PDATE FROM NEWFIELDS_FP
WHERE
(SELECT AVG(MOVIE_RELEASEYR) FROM MOVIE AS AVGRELEASEYR);
```