

Санкт-Петербургский государственный университет
Факультет прикладной математики – процессов управления

Курсовая работа
по дисциплине “Базы данных и сетевые технологии”
на тему “База данных Кинематографа Голливуда”

Выполнил:

Студент группы 21.Б07-ПУ

Бударов Алексей Андреевич

Руководитель:

Митрофанов Евгений Павлович

Предметная область и описание работы

Данная работа представляет собой пример базы данных, созданной как с целью отслеживания статистических данных различных аспектов кинематографа Голливуда (что может быть полезно работникам индустрии), так и с целью ознакомления, создания источника знаний для искушенных фанатов кинофильмов.

Схема базы данных

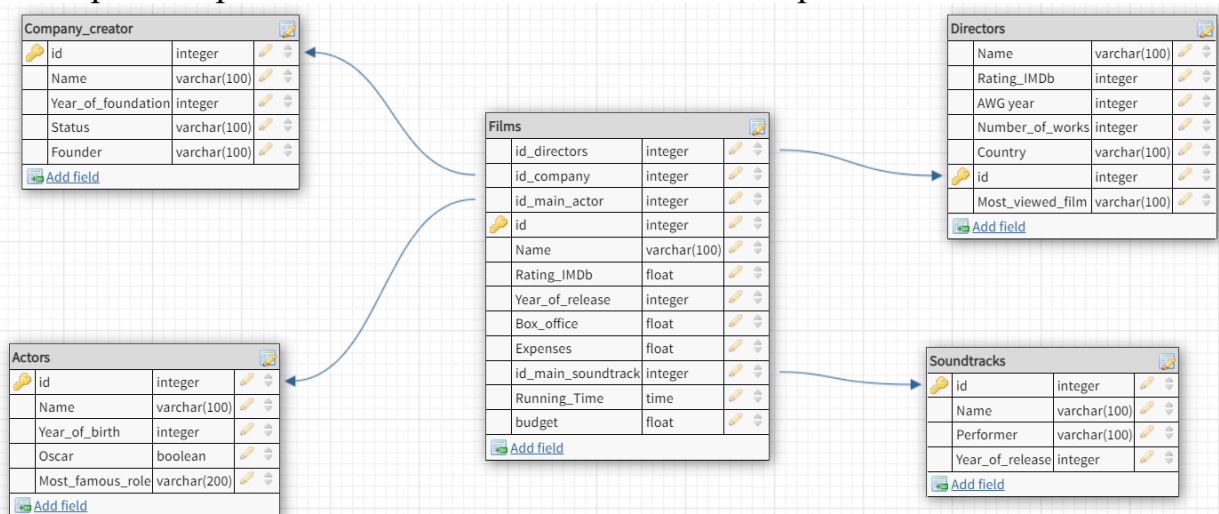
Поля, требующие дополнительного пояснения:

Actors – таблица, содержащая информацию об актерах.

- 'Oscar' boolean NOT NULL – Наличие оскара у актера
- 'Most_famous_role' character varying(200) NOT NULL UNIQUE – Самая известная роль актера. «Фильм – роль»

Company_creator – таблица, содержащая информацию о кинокомпаниях Голливуда.

- 'Status' character varying(30) – статус компании (Exists, Does not exist, Bought by...)
- 'Country' character varying(100) EXCLUDE USING gist (country WITH < >) – страна, в которой зарегистрирована компания. В базе данных рассматриваются компании только одной страны



Отношения 1:m (между Actors и Films, Films и Directors, Company_creator и Films)

Отношение m:m (Между Company_creator и Actors)

Простые запросы

1.Выделение самых прибыльных фильмов (За доход от проката фильма на всех странах и киноплощадках берется разница между кассовым сбором и бюджетом фильма), вышедших в прокат после 2000-го года для дальнейшего выявления востребованности актеров, режиссеров, и остального рабочего состава продакшена.

```
SELECT Name, Box_office - budget AS Income FROM Films
WHERE Year_of_release >= 2000
ORDER BY INCOME DES
```

Оптимизация:

```
CREATE INDEX films_name_id ON films(Name)
WHERE Year_of_release >=2000;
```

	name text	income numeric
1	The King_s Speech	412374317.00
2	Django Unchained	326074373.00
3	Once Upon a Time in Hollywood	287617598.00
4	Knives Outs	272897920.00
5	Seven Pounds	114748929.00
6	There Will Be Blood	51100000.00
7	Men of Honor	50343495.00
8	The Ghost Writer	17300000.00
9	Lucky	-772517.00

2.Выделение актеров старше 62 лет, имеющих Оскар, для отслеживания тенденций самой известной в мире кинопремьеры.

```
SELECT Name,Year_of_birth, most_famous_role
FROM Actors
WHERE (Oscar = True)AND(Year_of_birth > 1960)
```




Оптимизация:

```
CREATE INDEX actors_name_id ON actors(Name)
WHERE Year_of_birth > 1960;
```

	name character varying (100)	year_of_birth smallint	most_famous_role character varying (200)
1	William Bradley Pitt	1963	Fight Club-Tyler Durden
2	Leonardo DiCaprio	1974	Titanic - Jack Dawson

3. Вывод 10-ти самых востребованных режиссеров исходя из среднего рейтинга и количества работ для формирования актуального списка лидирующих работников индустрии, что будет полезно киностудиям Голливуда

```
SELECT name, Rating_IMDb, number_of_works FROM Directors
WHERE (Rating_IMDb < 100) AND (number_of_works > 15)
ORDER BY Rating_IMDb ASC, number_of_works DESC
LIMIT 10
```


	name character varying (100) 	rating_imdb real 	number_of_works smallint 
1	Steven Spielberg	5	32
2	George Tillman Jr.	7	17
3	Martin Scorsese	9	29
4	David Fincher	20	10
5	Woody Allen	32	18
6	Clint Eastwood	38	29
7	Roman Polanski	59	17

4. Подбор существующих на данное время компаний, чтобы оценить возможных работодателей для режиссеров и актеров

```
SELECT Name
FROM Company_creator
WHERE Status = 'Exist'
```

Оптимизация

```
CREATE INDEX company_name_id ON company_creator(name)
WHERE Status = 'Exist';
```

	name character varying (100) 
1	Babelsberg Studios
2	Universal Pictures
3	A Band Apart
4	Metro-Goldwyn-Mayer
5	Malpaso Productions
6	Divide and Conquer
7	Warner Bros. Entertainment

Средние запросы

1. Вывод фильмов одного из знаковых режиссеров Голливуда (Квентина Тарантино в частности) для сравнения и выявления зависимости рейтинга его работ, окупаемости и оценок критиков

```
SELECT films.name,directors.name,films.year_of_release, films.rating_imdb,
films.box_office - films.budget AS Income
FROM films
JOIN Directors ON Directors.id = films.id_directors AND Directors.name =
'Quentin Tarantino'
BY Rating_IMDb DESC, Income DESC
```

	name text	name character varying (100)	year_of_release smallint	rating_imdb real	income numeric
1	Pulp Fiction	Quentin Tarantino	1994	8.9	205000000.00
2	Once Upon a Time in Hollywood	Quentin Tarantino	2019	7.6	287617598.00

2. Вывод компании и основного состава, работающих над созданием фильмов, что позволит оценить затраты кинокомпании на рабочий коллектив в будущем в зависимости от темпа инфляции США

```
SELECT company_creator.name, films.name, actors.name, directors.name,
soundtracks.performer, films.budget
FROM films
JOIN Company_creator ON Company_creator.id = films.id_company
JOIN Actors ON Actors.id = films.id_main_actor
JOIN Directors ON Directors.id = films.id_directors
JOIN Soundtracks ON Soundtracks.id = films.id_main_soundtrack AND
soundtracks.name = 'Author_s music'
ORDER BY budget DESC
```

	name character varying (100)	name text	name character varying (100)	name character varying (100)	performer character varying (100)	?column? numeric
1	Babelsberg Studios	The Ghost Writer	Ewan Gordon McGregor	Roman Polanski	Alexander Desplat	48735000.00000
2	Arnold Kopelson Productions	Seven	William Bradley Pitt	David Fincher	Howard Shore	35739000.00000
3	Warner Bros. Entertainment	Goodfellas	Robert Anthony De Niro Jr.	Martin Scorsese	Pete Townshend	27075000.00000
4	Paramount Vantage	There Will Be Blood	Daniel Day-Lewis	Paul Thomas Anderson	Johnny Greenwood	27075000.00000
5	Malpaso Productions	Unforgiven	Clint Eastwood	Clint Eastwood	Lennie Niehaus	15595200.00000
6	Metro-Goldwyn-Mayer	Manhattan	Diane Keaton	Woody Allen	George Gershwin	9747000.00000
7	Divide and Conquer	Lucky	Harry Dean Stanton	David Lynch	Elvis Kuehn	1871902.68600

3. Вывод режиссеров и продолжительности их работ для отслеживания тенденций роста или падения длительности фильмов с течением развития кинематографа.

```
SELECT directors.name, films.name, films.year_of_release, films.running_time
FROM films
JOIN Directors ON Directors.id = films.id_directors
WHERE Running_time
BETWEEN '02:00' AND '03:00'
AND
box_office > 62300000.00
ORDER BY Year_of_release ASC, Running_time DESC
```

	name character varying (100)	name text	year_of_release smallint	running_time time without time zone
1	Clint Eastwood	Unforgiven	1992	02:11:00
2	Quentin Tarantino	Pulp Fiction	1994	02:34:00
3	David Fincher	Seven	1995	02:07:00
4	David Fincher	Men of Honor	2000	02:09:00
5	Paul Thomas Anderson	There Will Be Blood	2007	02:38:00
6	Quentin Tarantino	Once Upon a Time in Hollywood	2019	02:41:00

Сложные запросы

1. Вычислить средний рейтинг актеров, как средний рейтинг фильмов, в которых они принимал участие и вывести наилучшего

```
SELECT Temp.name, (Temp.TotalRating/count(films.id_main_actor)) as
AVG_Rating, count(films.id_main_actor) as COUNT_
from
(SELECT actors.name,actors.id, sum(films.Rating_iMDB) as TotalRating
from actors JOIN films
ON films.id_main_actor = actors.id
GROUP BY actors.id, actors.name) as Temp JOIN films
ON films.id_main_actor = Temp.id
Group by Temp.name,Temp.TotalRating
ORDER BY AVG_Rating DESC
LIMIT 1
```

	name character varying (100)	avg_rating double precision	count_ bigint
1	Liam Neeson	9	1

2. Вычислить количество фильмов, в которых принимал участие актер

```
SELECT actors.name, COUNT(films.id_main_actor) as COUNT_
from actors RIGHT JOIN films
ON films.id_main_actor = actors.id
GROUP BY actors.name
ORDER BY COUNT_ DESC
```

	name character varying (100)	count_ bigint
1	William Bradley Pitt	5
2	Leonardo DiCaprio	4
3	Clint Eastwood	1
4	Harry Dean Stanton	1
5	Liam Neeson	1
6	Robert Anthony De Niro Jr.	1
7	John Travolta	1
8	Diane Keaton	1
9	Ewan Gordon McGregor	1
10	Daniel Day-Lewis	1

3. Найти самую прибыльную кинокомпанию, а также вывести ее доход за весь период работы

```
SELECT Temp.name, max(Temp.TotalIncome) as Max_income
from
(SELECT company_creator.name,company_creator.id, sum(films.box_office -
films.budget) as TotalIncome
from company_creator JOIN films
ON films.id_company = company_creator.id
GROUP BY company_creator.id, company_creator.name) as Temp
Group by name
ORDER BY max_income DESC
LIMIT 1
```

	name character varying (100)	max_income numeric
1	A Band Apart	492617598.00