**Федеральное государственное автономное образовательное учреждение**

**высшего образования**

**"Национальный исследовательский университет**

**"Высшая школа экономики"**

Московский институт электроники и математики им. А.Н. Тихонова НИУ ВШЭ

Департамент компьютерной инженерии

**Курс: Проектный семинар "Python в науке о данных"**

**ОТЧЕТ**

**по домашней работе № 1**

**Студенты:**

Пахомова Алина Евгеньевна

Агафонова Екатерина Александровна

Алешина Алиса Дмитриевна

**Группа:** БИВ216

**Дата сдачи:** 27.04.2022

МОСКВА 2022

**Оглавление**

[Введение 3](#_Toc102238500)

[Создание информационной основы для приложений. 4](#_Toc102238501)

[Первая нормальная форма (1НФ). 4](#_Toc102238502)

[Вторая нормальная форма (2НФ) 4](#_Toc102238503)

[Третья нормальная форма (3НФ) 5](#_Toc102238504)

[Итог 6](#_Toc102238505)

# Введение

Согласно ТЗ создаваемые приложения должны обладать следующими функциональными возможностями:

1. Управление специализированной базой данных
2. Формирование отчетов для специализированной базы данных

# Создание информационной основы для приложений

## Первая нормальная форма (1НФ)

* В отношении нет одинаковых кортежей.
* Кортежи не упорядочены, т.е. их можно менять местами без потери смысла отношения.
* Атрибуты не упорядочены и различаются по наименованию.

Пример:

* userId – идентификационный номер пользователя
* movieId - идентификационный номер фильма
* rating – оценка пользователя по шкале от 0 до 5
* title – название фильма
* year – год выпуска фильма
* genre 1, 2, 3, 4, 5 – жанры фильмов

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| userId | movieId | rating | title | year | genre 1 | genre 2 | genre 3 | genre 4 | genre 5 |
| 1 | 1 | 4 | Toy Story | 1995 | Adventure | Animation | Children | Comedy | Fantasy |
| 1 | 2 | 4 | Jumanji | 1995 | Adventure | Children | Fantasy | 0 | 0 |
| 1 | 3 | 4 | Grumpier Old Men | 1995 | Comedy | Romance | 0 | 0 | 0 |
| 1 | 4 | 5 | Waiting to Exhale | 1995 | Comedy | Drama | Romance | 0 | 0 |
| 2 | 5 | 5 | Father of the Bride Part II | 1995 | Comedy | 0 | 0 | 0 | 0 |
| 2 | 6 | 5 | Heat | 1995 | Action | Crime | Thriller | 0 | 0 |
| 3 | 7 | 0,5 | Sabrina | 1995 | Comedy | Romance | 0 | 0 | 0 |
| 3 | 8 | 0,5 | Tom and Huck | 1995 | Adventure | Children | 0 | 0 | 0 |

Таблица 1. Исходное отношение userId movieId.

## Вторая нормальная форма (2НФ)

Исходное отношение декомпозируем на два отношения —movieId userId

|  |  |  |
| --- | --- | --- |
| userId | movieId | rating |
| 1 | 1 | 4 |
| 1 | 3 | 4 |
| 1 | 6 | 4 |
| 1 | 47 | 5 |
| 2 | 122882 | 5 |
| 2 | 131724 | 5 |
| 3 | 31 | 0,5 |
| 3 | 527 | 0,5 |

Таблица 1. userId, movieId, rating

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| movieId | title | year | genre 1 | genre 2 | genre 3 | genre 4 | genre 5 |
| 1 | Toy Story | 1995 | Adventure | Animation | Children | Comedy | Fantasy |
| 2 | Jumanji | 1995 | Adventure | Children | Fantasy | 0 | 0 |
| 3 | Grumpier Old Men | 1995 | Comedy | Romance | 0 | 0 | 0 |
| 4 | Waiting to Exhale | 1995 | Comedy | Drama | Romance | 0 | 0 |
| 5 | Father of the Bride Part II | 1995 | Comedy | 0 | 0 | 0 | 0 |
| 6 | Heat | 1995 | Action | Crime | Thriller | 0 | 0 |
| 7 | Sabrina | 1995 | Comedy | Romance | 0 | 0 | 0 |
| 8 | Tom and Huck | 1995 | Adventure | Children | 0 | 0 | 0 |

Таблица 2. movieId, title, year, genre 1, genre 2, genre 3, genre 4, genre 5

## 

## Третья нормальная форма (3НФ)

В данном случае 2НФ и 3НФ совпадают, так как соблюдены все условия.

movieId userId

|  |  |  |
| --- | --- | --- |
| userId | movieId | rating |
| 1 | 1 | 4 |
| 1 | 3 | 4 |
| 1 | 6 | 4 |
| 1 | 47 | 5 |
| 2 | 122882 | 5 |
| 2 | 131724 | 5 |
| 3 | 31 | 0,5 |
| 3 | 527 | 0,5 |

Таблица 1. userId, movieId, rating

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| movieId | title | year | genre 1 | genre 2 | genre 3 | genre 4 | genre 5 |
| 1 | Toy Story | 1995 | Adventure | Animation | Children | Comedy | Fantasy |
| 2 | Jumanji | 1995 | Adventure | Children | Fantasy | 0 | 0 |
| 3 | Grumpier Old Men | 1995 | Comedy | Romance | 0 | 0 | 0 |
| 4 | Waiting to Exhale | 1995 | Comedy | Drama | Romance | 0 | 0 |
| 5 | Father of the Bride Part II | 1995 | Comedy | 0 | 0 | 0 | 0 |
| 6 | Heat | 1995 | Action | Crime | Thriller | 0 | 0 |
| 7 | Sabrina | 1995 | Comedy | Romance | 0 | 0 | 0 |
| 8 | Tom and Huck | 1995 | Adventure | Children | 0 | 0 | 0 |

Таблица 2 movieId, title, year, genre 1, genre 2, genre 3, genre 4, genre 5

# Итог

Таким образом, исходное отношение было нормализовано — приведено к третьей нормальной форме за счет декомпозиции в два справочника — movieId, userId