Ковальов Олександр Олексійович №22, **Зайнятість акторів театру**

Лабораторна робота №6

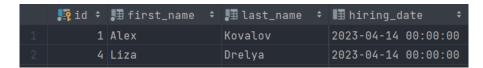
Вивчення основ реляційної алгебри (РА). Ознайомлення з основними принципами нормалізації таблиць у реляційній моделі бази даних.

Перевірити розроблену БД на відповідність вимог нормальних форм (НФ). Ненормалізована (нульова) нормальна форма (UNF).

Якщо БД знаходиться в нульовій формі, то вона не відповідає реляційній теорії. Наприклад, рядки пронумеровані, чого не повинно бути, так як порядок рядків і стовбців не враховується. Чисельний ідентифікатор не нумерує рядки, а дозволяє ідентифікувати кортеж, тому це не є помилкою (тим паче, ідентифікатори можуть йти не послідовно):

	🌇 id 🕏	, ≣ base_salary ≎	🎜 salary 🕏
1	1	3000	6000
2	4	6000	8000
3	5	10000	12000
4	6	2000	2000
5	7	0	0

Також, за реляційною моделлю, стовбці повинні мати назву, яка підходить за сенсом. Не можна називати їх «А» і «Б». Як підписувати правильно:



Перша нормальна форма.

БД повинна відповідати реляційній моделі. Окрім цього:

- В таблиці не повинно бути дублікатів:
- В кожній комірці зберігається атомарне значення відносно сенсу стовбця. Тобто, воно не повинно бути складеним;
- В стовбцях повинні зберігатися дані одного типу, з одного домену;
- Відсутні масиви і списки в будь-якому вигляді.

Головне правило першої форми:

- Призначення рядків зберігати дані кортежі, інформація з яких пов'язана між собою певним чином;
- Призначення стовбців зберігати структурну інформацію;
- Призначення комірок зберігати атомарне (окреме, не складене) значення.

Наприклад, є таблиця з контактами працівників:

Контакти працівників			
ПІБ Контакт			
Ковальов О. О.	123-456-789, 144-456-789		
Biden Joe	Домашній телефон +3809911122233, Робочий +3804455555555		
Жуковський Д. В.	+3805012312323		
Жуковський Д. В.	+3805012312323		

Ця таблиця не відповідає першій нормальній формі. По-перше, присутні дублікати — кортеж з контактами працівника «Жуковський Д. В.» повторюється двічі. По-друге, у працівника «Ковальов О. О.» в комірці знаходиться масив значень, а не атомарне значення. Така ж проблема з працівником «Віden Joe» - але, ще й присутній опис контактів, що потрібно винести в окрему колонку.

Ця ж таблиця, але у першій нормальній формі:

Контакти працівників			
ПІБ	Тип телефону		
Ковальов О. О.	123-456-789		
Ковальов О. О.	144-456-789		
Biden Joe	+3809911122233	Домашній	
Biden Joe +3804455555555		Робочий	
Жуковський Д. В.	+3805012312323		

Робоча база даних повністю відповідає вимогам першої нормальної форми, всі значення в таблицях атомарні.

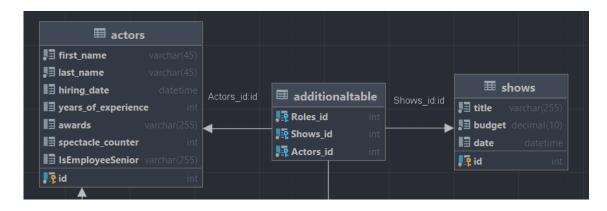
Друга нормальна форма.

- Таблиця повинна знаходитись у першій нормальній формі;
- Таблиця повинна мати ключ, простий або складений;
- Всі неключові стовбці (атрибути) таблиці повинні залежати від повного ключа (повного коли він складений).

В таблиці не повинно бути даних, які можна отримати, знаючи лише половину ключа, наприклад один атрибут з декількох ключових.

Таблиця повинна мати правильний ключ, за яким можна ідентифікувати кожен кортеж. Якщо таблиця не відповідає другій нормальній формі, можуть виникнути певні аномалії.

Робоча база даних повністю відповідає вимогам другої нормальної форми, всі таблиці мають ключі, за яким однозначно можна ідентифікувати певний кортеж.



Якщо б таблиці «Актори» і «Спектаклі» були одною — «Актори які приймають участь в шоу» — і в ній ключ був би складеним — назва шоу та прізвище актора, то ця таблиця не відповідала б вимогам другої нормальної форми. За назвою шоу можна було б дістати інформацію про її бюджет та дату, хоча це лише половина ключа, і навпаки з акторами. У такому випадку проводиться декомпозиція, що і було зроблено при проектуванні БД.

Без нормалізації, при видаленні шоу зникала б і інформація про акторів, що є аномалією видалення.

Третя нормальна форма.

В таблицях повинна бути відсутня транзитивна залежність — залежність неключових атрибутів від інших неключових стовбців. Неключові стовбці не мають імітувати роль ключа в таблиці. У такому випадку проводиться декомпозиція. Головне правило третьої нормальної форми — таблиця повинна мати лише правильні неключові атрибути.

Наприклад, є таблиця зі списком працівників. Первинним ключем є атрибут «Табельний номер». Також, є ПІБ працівників, посада, підрозділ, та опис підрозділу. ПІБ залежить від табельного номеру, посада та підрозділ також. А опис підрозділу залежить від підрозділу.

Працівники					
Табельний номер	ПІБ	Посада	Підрозділ	Опис підрозділу	
1	Дреля Є. В.	Художній керівник	Організатори	Управлінська позиція. До їхніх завдань входять такі обов'язки: контроль за виконанням посадових обов'язків колективом театру, підготовка та узгодження репертуару на квартал, рік та інші.	
2	Каркушевський В. Л.	Художній режисер	Режисери	Режисери разом із художніми керівниками підбирають актуальні п'єси для вистав, відповідають за підготовку репертуару. Крім того, вони зобов'язані брати участь у репетиціях, відпрацьовувати з акторським складом складні розуміння ділянки п'єси. Режисер вибирає артистів на ролі, а також контролює підготовку вбрання, реквізиту, декорацій.	
3	Biden Joe	Звукорежисер	Режисери	Режисери разом із художніми керівниками підбирають актуальні п'єси для вистав, відповідають за підготовку репертуару. Крім того, вони зобов'язані брати участь у репетиціях, відпрацьовувати з акторським складом складні розуміння ділянки п'єси. Режисер вибирає артистів на ролі, а також контролює підготовку вбрання, реквізиту, декорацій.	

У такому випадку, як було написано вище, проводиться декомпозиція. Повертаючись до прикладу, таблицю треба поділити на дві. Першою буде «словник» підрозділів, з ідентифікатором, назвою, та описом:

Підрозділи				
Ідентифікатор підрозділу	Підрозділ	Опис підрозділу		
1	Організатори	Управлінська позиція. До їхніх завдань входять такі обов'язки: контроль за виконанням посадових обов'язків колективом театру, підготовка та узгодження репертуару на квартал, рік та інші.		
2	Бухгалтерія	Ведення бухгалтерського та податкового обліку фінансово- господарської діяльності		
3	Режисери	Режисери разом із художніми керівниками підбирають актуальні п'єси для вистав, відповідають за підготовку репертуару. Крім того, вони зобов'язані брати участь у репетиціях, відпрацьовувати з акторським складом складні розуміння ділянки п'єси. Режисер вибирає артистів на ролі, а також контролює підготовку вбрання, реквізиту, декорацій.		

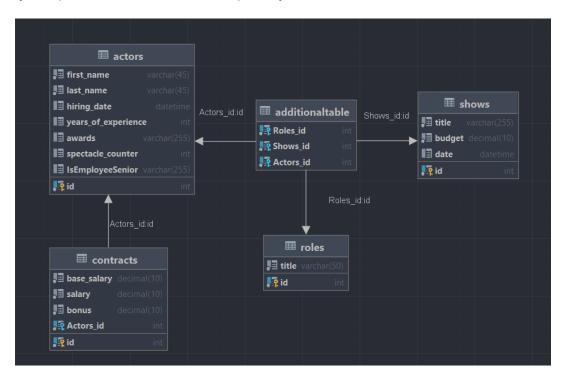
Таблиця «Працівники» повинна виглядати так:

Працівники				
Табельний номер ПІБ Посада			Підрозділ	
1	Дреля Є. В.	Художній керівник	1	
2	Каркушевський В. Л.	Художній режисер	3	
3	Biden Joe	Звукорежисер	3	

Також, доцільно додати словник «Посада-Підрозділ».

Якщо не нормалізувати за третьою нормальною формою, може виникнути така ж аномалія видалення – при видаленні працівників зникне інформація про підрозділ та його опис. Але ключовою тут є аномалія модифікації – при зміні підрозділу, опис залишався б попереднім, що неприпустимо.

Робоча база даних повністю відповідає вимогам третьої нормальної форми. Декомпозиція була проведена ще на стадії проектування.



Нормальна форма Бойса-Кодда (BCNF).

- Таблиця повинна знаходитись у третій нормальній формі;
- Ключові атрибути складеного ключа не повинні залежати від неключових атрибутів. Вимоги BCNF актуальні лише для таблиць зі складеним ключем. Головне правило нормальної форми Бойса-Кодда: частина складеного первинного ключа не повинна залежати від неключового атрибуту.

Наприклад, маємо таблицю «Куратори напрямків проектів» з інформацією про номер проекту, напрямок, та відповідного куратора. Проекти можуть повторюватись, напрямки також, а от комбінація «Проект-Відділ» ні, тому вона є первинним ключем (помічена зеленим). Дана таблиця є нормалізованою до третьої нормальної форми — виконуються вимоги усіх попередніх та нема транзитивних залежностей. Припустимо, є вимога — куратор може керувати лише одним напрямком, але в багатьох проектах. У такому випадку, ми можемо однозначно визначити половину ключа, а саме напрямок, за ПІБ куратора.

Проект	Відділ	Куратор	
1	Режисура	Каркушевський В. Л.	
1	Актори	Діковський А. М.	
2	Актори	Жуковський Д. В.	
2	Суфлери	Ковальов О. О.	
2	Організатори	Дреля Є. В.	
3	Актори	Жуковський Д. В.	

Проект-Відділ-Куратор

У цьому випадку також проводиться декомпозиція. Таблицю треба розділити на 2. Таблиця «Куратори»:

Куратори

ID	Куратор	Відділ
1	Каркушевський В. Л.	Режисура
2	Діковський А. М.	Актори
3	Жуковський Д. В.	Актори
4	Ковальов О. О.	Суфлери
5	Дреля Є. В.	Організатори

Таблиця відповідності кураторів та проектів:

Проекти

Проект	ID Куратора
1	1
1	2
2	3
2	4
2	5
3	1

До робочої бази даних нормальна форма Бойса-Кодда не застосовується, так як в ній немає складених ключів.

Четверта нормальна форма.

У таблицях повинні бути відсутні нетривіальні багатозначні залежності. Ця форма не застосовується, якщо в таблиці менше трьох стовбців.

Приклад:

Є атрибути A, B, C. B і C ніяк не пов'язані між собою, але окремо один від одного залежать від A, і для кожного значення A є багато значень B і C.

Отримані залежності:

$$A \rightarrow B$$

$$A \rightarrow C$$

Наприклад, є таблиця з курсами, викладачами та аудиторіями. Припустимо, що замість текстових даних в таблиці знаходяться ідентифікатори з відповідних таблиць «словників», а текстові дані на скріншоті лише для зручності. Є обмеження — курс може викладатись лише певними викладачами та в певних аудиторіях, а викладачу все одно в якій аудиторії його викладати.

Курси-Викладачі-Аудиторії

Курс	Викладач	Аудиторія	
SQL	Діковський А. М.	101	
SQL	Діковський А. М.	202	
WinForms	Каркушевський В. Л.	303	
WinForms	Каркушевський В. Л.	404	
JavaScript	Жуковський Д. В.	303	

У такому випадку можуть виникнути аномалії — наприклад, при видаленні з розкладу зникає аудиторія, хоча вона повинна там бути. Для вирішення цієї проблеми потрібно провести декомпозицію, і створити дві таблиці — Курс-Викладач та Курс-Аудиторія.

П'ята нормальна форма.

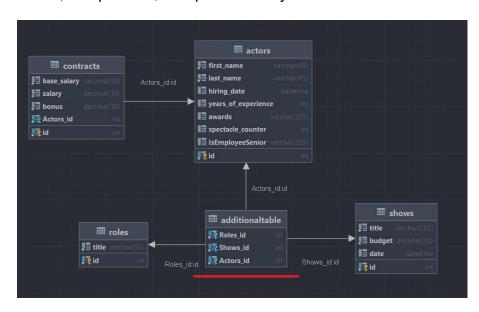
Зміна відношення знаходиться у п'ятій нормальній формі (проекційно-сполучувальній формі), коли кожна нетривіальна залежність з'єднання в ній визначається потенціальним ключом або ключами цього відношення.

Є таблиці, які при декомпозиції на 2 таблиці можуть втратити певні дані, але при декомпозиції на 3 можна запобігти втраті даних.

Декомпозиція без втрат – процес розбиття одної таблиці на декілька, за умови, що у випадку з'єднання таблиць, які були отримані в результаті декомпозиції, буде формуватись та ж інформація, що й у початковій таблиці до декомпозиції.

Щоб виконати вимоги п'ятої нормальної форми, треба виконати декомпозицію без втрат.

Прикладом може слугувати таблиця для зв'язку ролей, шоу та акторів з робочої БД. Є зв'язок багато до багатьох, між ролями, акторами та шоу:



Наприклад, вміст таблиці такий:

Актор	Спектакль	Роль	
Діковський А. М.	Hamlet	Hamlet	
Жуковський Д. В.	Hamlet	Tiresias	
Жуковський Д. В.	Oedipus Rex	George Pig	
Ковальов О. О.	Long Day's Journey Into Night	Mary Tyrone	
Діковський А. М.	Long Day's Journey Into Night	Oedipus	
Діковський А. М.	Death of a Salesman	Honey	

Нормалізувати до п'ятої форми неможливо, не знаючи предметної області. Припустимо, що було проведене дослідження. Результати такі:

- Діковський А. М. може грати лише роль "Hamlet";
- Діковський А. М. може грати у великій кількості вистав;
- Жуковський Д. В. може грати в будь-якому спектаклі, окрім Death of Salesman;
- У актора Ковальов О. О. контракт заключений лише на одну виставу;

Без нормалізації до п'ятої форми, можна легко внести не ті дані, та викликати певну аномалію. Для нормалізації потрібно виконати декомпозицію без втрат. У даному випадку, таблицю треба поділити на 3: «Актор-Спектакль», «Спектакль-Роль», «Актор-Роль». Якщо об'єднання (Inner Join) отримується без втрат відносно початкової таблиці, то нормалізація проведена успішно.

Робоча таблиця не потребує нормалізації до п'ятої форми, так як немає жорстких вимог за предметною областю.

П'ята форма – остання, яку можна отримати шляхом проведення декомпозиції.

Доменно-ключова нормальна форма.

Обмеження домену – це обмеження щодо використання значень певного атрибуту лише з певного заданого домену (набору значень).

Обмеження ключа — це обмеження, яке полягає в тому, що деякий атрибут або комбінація атрибутів є потенційним ключем.

Кожне обмеження на таблицю повинно бути логічним наслідком обмежень доменів і обмежень ключів, які накладаються на цю таблицю.

Відношення знаходиться в ДКНФ, якщо воно не має аномалій модифікації.

Ці правила діють приблизно так: не можна просто видалити категорію з таблиці категорій, якщо з цією категорією пов'язані, наприклад, продукти з таблиці продуктів. Перш ніж видаляти категорію, необхідно виконати попередні дії в таблиці продуктів (наприклад, поле, що відповідає за ID категорії цього товару, потрібно зробити NULL).

В робочій таблиці для виконання умов ДКНФ встановлення обмеження цілісності зовнішнього ключа: каскадне оновлення та видалення записів (для всіх зв'язків):

	Schema: TheaterA	Actors		\Rightarrow
Column	Referenced Column	Foreign Key Option	ons —	
id		On Update:	CASCADE	~
base_salary salary		On Delete:	CASCADE	~
bonus				
✓ Actors_id	id	Skip	in SQL generation	n

Шоста нормальна форма.

Шоста нормальна форма (6NF) була введена під час роботи з хронологічними базами даних.

Хронологічна база даних – це база, яка може зберігати як поточні дані, так й історичні дані, тобто дані, що належать до минулих періодів часу. Однак така база може зберігати і дані, що стосуються майбутніх періодів часу.

У процесі проектування хронологічних баз даних виникають деякі особливі проблеми, які можна вирішити за допомогою горизонтальної та вертикальної декомпозиції.

Шоста нормальна форма запроваджує таке поняття як «Декомпозиція до кінця», тобто. максимально можлива декомпозиція таблиць.

Однак, якщо в хронологічних базах даних така нормалізація може бути корисною, оскільки вона дозволяє боротися з надлишковістю, то в нехронологічних базах даних нормалізація таблиць до шостої нормальної форми призведе до значного зниження продуктивності. Крім цього, така нормалізація зробить роботу з базою даних дуже складною за рахунок багаторазового збільшення кількості таблиць.

Тому шосту нормальну форму в реальному світі не використовують, більше того, важко навіть уявити собі ситуацію, за якої виникала б необхідність нормалізувати базу даних до шостої нормальної форми. Практичного застосування шостої нормальної форми, мабуть, просто немає.

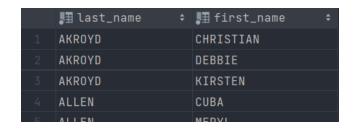
6NF означає, що кожне відношення складається з ключа-кандидата плюс не більше ніж один інший (ключовий або неключовий) атрибут. Наприклад, якщо «елемент» ідентифікується за допомогою ProductCode, а іншими атрибутами є Description і Price, то схема 6NF складатиметься з двох відносин (* позначає ключ у кожному):

Написати на мові SQL запити, які реалізують операції реляційної алгебри. 1. Операція об'єднання.

Результатом операції об'єднання є відношення, що містить усі кортежі, що належать одному з двох або обом відносинам. На відміну від об'єднання множин, результатом об'єднання відносин має стати відношення, а не набір різнорідних кортежів. Відношення повинні бути визначені однією схемою. Результуюче відношення містить усі рядки операндів, за винятком повторюваних.

Приклад: обрати ПІБ акторів з робочої БД та БД Sakila.

```
SELECT last_name, first_name FROM TheaterActors.actors
UNION
SELECT last_name, first_name FROM sakila.actor
ORDER BY last name;
```

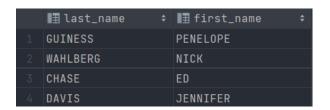


2. Операція перетину.

Результатом операції перетину є відношення, яке містить кортежі, що належать до обох відношень. Для операції перетину необхідні дві умови, що й для об'єднання: сумісність за типом і замкнутість.

Приклад: обрати ПІБ акторів з робочої БД та БД Sakila, які є в обох таблицях одночасно.

```
SELECT last_name, first_name FROM TheaterActors.actors AS MainActors
INNER JOIN sakila.actor AS SakilaActors
USING (last name, first name);
```

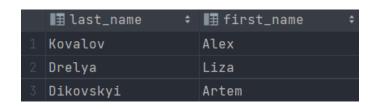


3. Операція різниці.

Результатом операції віднімання є відношення, яке містить усі кортежі, що належать першому відношенню і не належать другому відношенню. Умови ті ж самі: сумісність за типом і замкнутість.

Приклад: обрати ПІБ акторів з робочої БД, яких немає в БД Sakila.

```
SELECT last_name, first_name FROM TheaterActors.actors AS A
WHERE NOT EXISTS (SELECT last_name, first_name
FROM sakila.actor AS B
INNER JOIN theateractors.actors
USING (last_name, first_name)
WHERE A.last name = B.last name AND A.first name = B.first name);
```



4. Операція декартового добутку.

Результатом є відношення, що містить усі можливі кортежі з двох відносин.

Для добутку потрібна властивість замкнутості, тобто результатом є не просто безліч пар кортежів, а множина цілих кортежів. У реляційній алгебрі кожна пара кортежів замінюється одним завдяки зчепленню (зчеплення – це об'єднання у сенсі теорії множин, а чи не реляційної алгебри).

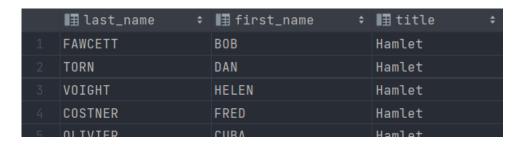
Приклад: обрати всіх кандидатів на ролі серед акторів, які вільні.

```
SELECT DISTINCT actors.last_name, actors.first_name, r.title FROM actors

LEFT JOIN additionaltable AS a ON actors.id = a.Actors_id

CROSS JOIN roles AS r

WHERE a.Roles id IS NULL;
```



5. Операція вибірки (фільтрація, селекція, обмеження).

Результатом вибірки є відношення, що містить усі кортежі з вихідного відношення, які відповідають певній умові.

Приклад: обрати акторів, у яких прізвище починається на літеру D.

```
SELECT last_name, first_name, awards
FROM actors
WHERE last_name LIKE 'D%'
ORDER BY last name;
```



6. Операція проекції (вертикальна підмножина).

Проекція є операцією, коли з відношення виділяються атрибути лише з зазначених доменів, тобто з таблиці вибираються лише потрібні стовпці. Якщо вийде кілька однакових кортежів, то в результуючому відношенні залишається лише один екземпляр подібного кортежу.

Приклад: обрати спектаклі, та дати, коли вони відбудуться.

```
SELECT DISTINCT title, date
FROM shows;
```

	.⊞ title ÷	;	I date	‡
1	Hamlet		2023-04-14 00:00:00	
2	Long Day's Journey Into Night		2023-04-14 00:00:00	
3	Who's Afraid of Virginia Woolf?		2023-04-14 00:00:00	
4	Death of a Salesman		2023-04-14 00:00:00	
5	Oedinus Rex		2023-04-14 00:00:00	

7. Операція з'єднання (сполучення).

Результат з'єднання – це відношення, кортежі якого є поєднанням двох кортежів, що належать двом початковим відносинам і мають спільні значення для одного або декількох атрибутів (загальне значення в результуючому відношенні з'являється лише один раз).

```
SELECT actors.last_name, actors.first_name, r.title
FROM actors
INNER JOIN additionaltable AS a ON actors.id = a.Actors_id
INNER JOIN roles r on a.Roles_id = r.id;
```

	I ∄ last_name	‡	■ first_name	‡	I≣ title
1	Kovalov		Alex		Hamlet
2	Drelya		Liza		Gertrude
3	Kovalov		Alex		George
4	Dikovskyi		Artem		Honey
5	Kovalov		Alex		Linda Loman
6	Drelya		Liza		0edipus
7	Dikovskyi		Artem		Tiresias

8. Операція ділення.

Результатом ділення двох відносин (бінарного та унарного) є відношення, що містить усі значення атрибуту першого бінарного відношення, які відповідають усім значенням унарного відношення.

Іншими словами, якщо є дві таблиці, які поєднані певним чином, то операція ділення – процес отримання однієї з таблиць, використовуючи результативну та одну з початкових.

Приклад: для кожного актора, у якого немає ролей, підібраний список ролей які можна зіграти. Задача: отримати список невлаштованих акторів.

```
SELECT PossibleRoles.id, PossibleRoles.last_name, PossibleRoles.first_name FROM (SELECT DISTINCT actors.id, actors.last_name, actors.first_name, r.title FROM actors

LEFT JOIN additionaltable AS a ON actors.id = a.Actors_id

CROSS JOIN roles AS r

WHERE a.Roles_id IS NULL) AS PossibleRoles

GROUP BY PossibleRoles.id, PossibleRoles.last_name, PossibleRoles.first_name HAVING COUNT(PossibleRoles.title) = COUNT(DISTINCT PossibleRoles.title);
```

	🛂 id 🕏	.⊞ last_name	∰ first_name ÷
1	6	GUINESS	PENELOPE
2	7	WAHLBERG	NICK
3	8	CHASE	ED
4	9	DAVIS	JENNIFER
5	10	LOLLOBRIGIDA	JOHNNY
6	11	NICHOLSON	BETTE
7	12	MOSTEL	GRACE
8	13	JOHANSSON	MATTHEW
9	14	SWANK	J0E
10	16	CAGE	ZER0
11	17	BERRY	KARL
12	18	WOOD	UMA
13	19	BERGEN	VIVIEN

9. Операція розширення.

За допомогою операції розширення відношення створюється нове відношення, що містить новий атрибут, значення якого отримані за допомогою скалярних обчислень.

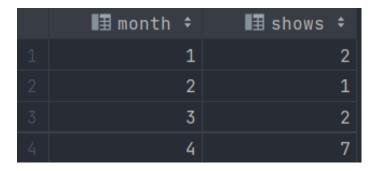
Приклад: визначити, скільки років пройшло з моменту найму актора.

	II last_name ÷	፟∰ first_name ÷	■ hiring_date ÷	■ YearsFromHiring ÷
1	Kovalov	Alex	2023-04-14 00:00:00	0
2	Drelya	Liza	2023-04-14 00:00:00	0
3	Dikovskyi	Artem	2023-04-14 00:00:00	0
4	GUINESS	PENELOPE	2020-07-24 00:00:00	3
5	WAHLBERG	NICK	2021-01-01 00:00:00	2
6	CHASE	ED	2020-01-17 00:00:00	3
7	DAVIS	JENNIFER	2013-12-05 00:00:00	9

10. Операція підбиття підсумків.

Операція підведення підсумків використовується для виконання групових обчислень. Приклад: вивести кількість шоу у кожному місяці.

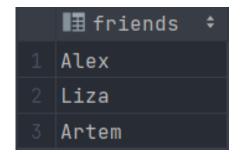
```
SELECT MONTH(date) AS month, COUNT(title) AS shows FROM shows
GROUP BY month
ORDER BY month;
```



11. Операція перейменування.

Операція перейменування призначена для зміни імені атрибута відносин. Приклад: всі актори, які грають ролі— є друзями, інші ні. Вивести друзів.

```
SELECT DISTINCT first_name AS friends
FROM actors
INNER JOIN additionaltable a on actors.id = a.Actors_id
INNER JOIN roles r on a.Roles id = r.id
```



Предметна область (22 варіант):

Зайнятість акторів театру. Комерційний директор театру організує залучення акторів та укладання контрактів. Щороку театр здійснює постановку різних вистав. Кожна вистава має певний бюджет. Для участі у конкретних постановках у певних ролях залучаються актори. З кожним із акторів укладається персональний контракт на певну суму. Кожен актор має стаж роботи, деякі з них удостоєні різних звань. В рамках одного спектаклю на ту саму роль залучається кілька акторів. Договір визначає базову зарплату актора, а за підсумками реально відіграних вистав актору призначається премія. У базі даних слід зберігати інформацію протягом кількох років.

Діаграма прецедентів:

