**Введение**

Социальные сети - это один из самых популярных и влиятельных видов коммуникации в современном мире. Они позволяют людям обмениваться мнениями, интересами, новостями и развлечениями с миллионами других пользователей. Для социальных сетей очень важно хранение информации, так как она является основой их работы и ценностью для их клиентов. Без надежного и безопасного хранения информации социальные сети не смогут предоставлять свои услуги на высоком уровне и защищать данные своих пользователей от утечек и злоупотреблений.

База данных (бд) является одним из ключевых компонентов для социальной сети, так как она хранит и обрабатывает различные данные о пользователях, сообществах, постах, лайках, комментариях и т.д. Подбор подходящей базы данных и архитектуры - это важный этап разработки любого проекта или приложения, так как от этого зависит эффективность, надежность и безопасность работы с данными.

Современные системы управления базами данных (СУБД) — это комплексы программных средств, которые позволяют создавать, хранить, обрабатывать и предоставлять доступ к большим объемам структурированной или неструктурированной информации. СУБД могут поддерживать различные модели данных, то есть способы представления и организации данных в базе. Реляционные и нереляционные модели данных — это два основных подхода к организации данных в базах данных, которые имеют разные принципы, преимущества и недостатки. Выбор между ними зависит от специфики и играет ключевую роль в проектировании архитектуры бд.

**1 Постановка задачи**

Главная особенность социальных сетей — огромное количество сложной и взаимосвязанной информации, создаваемой пользователями информации. Для поддержания работоспособности и удобства использования приложения очень важно разработать надежную архитектуру бд.

Целью данного курсового проекта является создание программного обеспечения для социальной сети.

Задача проекта: совершенствование практических навыков в использовании и разработке современных информационных систем, разработка архитектуры приложения, создание базы данных и интерфейса, выполнение тестирования готового продукта.

Функционально должны быть выполнены следующие задачи:

˗ разработка приложения для взаимодействия с БД;

˗ регистрация и авторизация пользователей;

˗ обеспечение социальной функциональности;

˗ анализ работы сервиса

˗ мониторинг сети администратором;

Должны быть выполнены следующие требования:

- база данных должна быть спроектирована в СУБД PostgreSQL.

- доступ к данным должен осуществляться только через соответствующие процедуры;

- должен быть проведен импорт данных из JSON файлов, экспорт данных в формат JSON;

- необходимо протестировать производительность базы данных (на таблицах, содержащих не менее 100 000 строк) и внести изменения в структуру в случае необходимости;

- Применить технологию базы данных согласно выбранной теме: подробно описать применяемые системные пакеты, утилиты или технологии; показать применение указанной технологии в базе данных.

Серверная часть приложения должна быть реализовано с использованием фреймворка ASP.NET Core и языка C#, клиентская часть должна быть реализована с использованием фреймворка React и языка JavaScript.

Разрабатываемое приложение должно содержать форму авторизации, ленту постов, страницу профиля и подписок, панель модератора. Пользователь может подписываться на других пользователей, создавать посты, комментировать, оставлять лайки, делиться публикациями и оставлять жалобы. Модератор может просматривать статистику приложения и посты с жалобами и удалять их, блокировать пользователей.

**2 Разработка модели данных**

Для хранения данных в бд была разработана структура взаимосвязанных таблиц с помощью задания отношений. Отношения в реляционной БД - это логические связи между данными, хранящимися в разных таблицах. Отношения позволяют объединять данные из разных источников и извлекать нужную информацию. Отношения могут быть трех типов: один ко одному, один ко многим и много ко многим. Тип отношения зависит от того, сколько строк в одной таблице соответствует строкам в другой таблице.

Из-за особенностей требований для разработки социальной сети, для реализация функционала требуется большое количество отношений много ко многим, которые будут отображать такие связи как лайки, которые пользователи ставят публикациям.

Диаграмма базы данных, спроектированной в ходе разработки приложения представлена на рисунке 2.1.

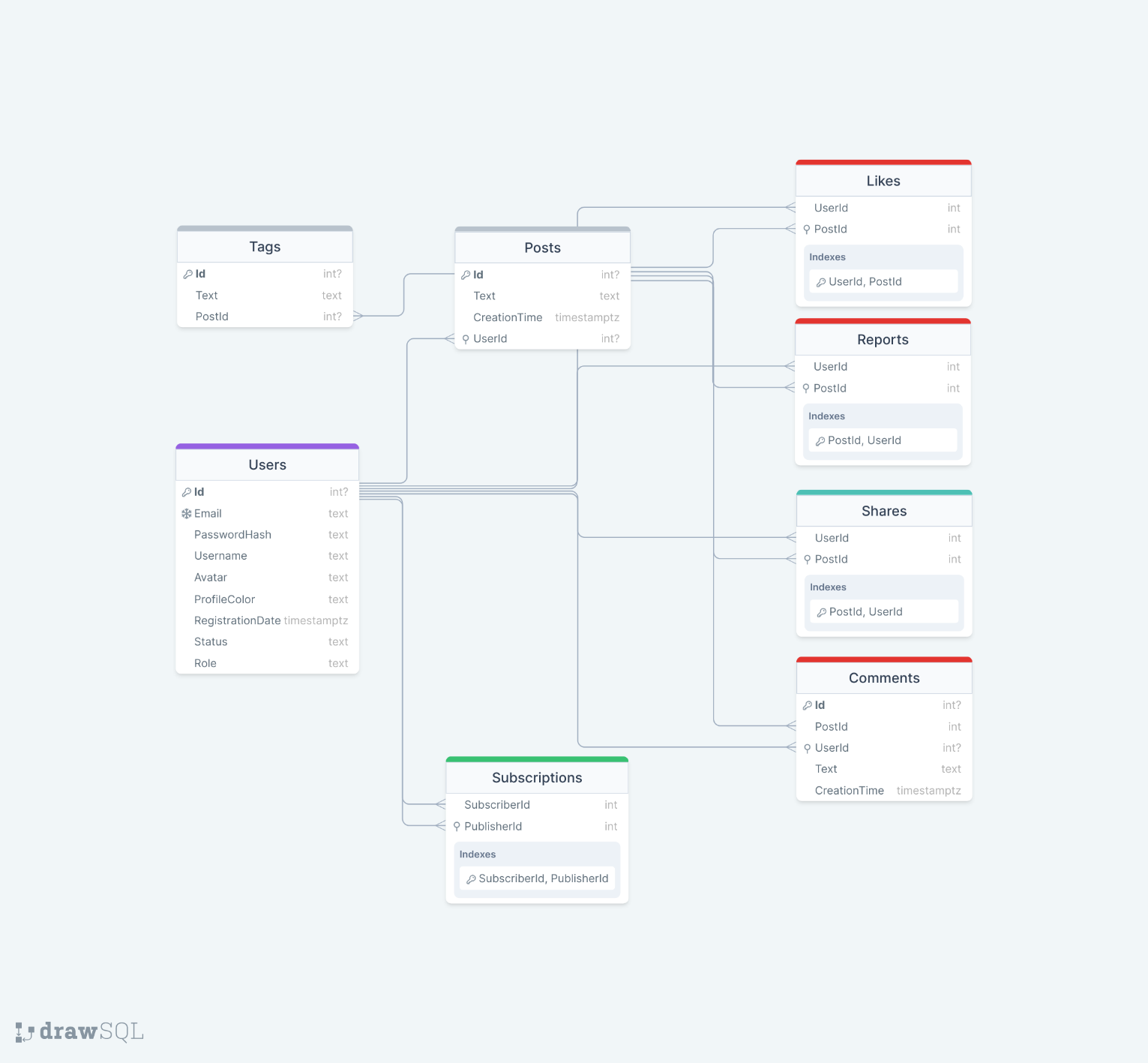


Рисунок 2.1 – Диаграмма базы данных

База данных состоит из 8 таблиц: Users, Posts, Comments, Likes, Shares, Reports, Tags, Subscriptions.

Таблица Students хранит список пользователей, состоит из столбцов (таблица 3.1):

Таблица 3.1 – Столбцы таблицы Users

|  |  |  |
| --- | --- | --- |
| Наименование | Описание | Тип |
| Id | Идентификатор пользователя, первичный ключ | int |
| Email | Почта пользователя, not null, unique | text |
| Username | Псевдоним пользователя, not null | text |
| PasswordHash | Зашифрованный пароль пользователя, not null | text |
| Avatar | Аватар пользователя, not null | text |
| ProfileColor | Цвет оформления профиля пользователя, not null | text |
| RegistrationDate | Дата регистрации пользователя, not null | timestamp with time zone |
| Status | Статус пользователя, not null | text |
| Role | Роль пользователя, not null | text |

Таблица Posts хранит список публикаций, состоит из столбцов (таблица 3.2):

Таблица 3.2 – Столбцы таблицы Posts

|  |  |  |
| --- | --- | --- |
| Наименование | Описание | Тип |
| Id | Идентификатор публикации, первичный ключ | int |
| Text | Текст публикации, not null | text |
| CreationTime | Дата создания публикации, not null | timestamp with time zone |
| UserId | Идентификатор автора публикации, внешний ключ на таблицу Users | int |

Таблица Comments хранит список комментариев к публикациям, состоит из столбцов (таблица 3.3):

Таблица 3.3 – Столбцы таблицы Comments

|  |  |  |
| --- | --- | --- |
| Наименование | Описание | Тип |
| Id | Идентификатор комментраия, первичный ключ | int |
| Text | Текст комментария, not null | text |
| CreationTime | Дата создания публикации, not null | timestamp with time zone |
| UserId | Идентификатор автора комментария, внешний ключ на таблицу Users | int |
| PostId | Идентификатор публикации, которой принадлежит комментарий, внешний ключ на таблицу Posts | int |

Таблица Likes хранит список лайков у публикаций, состоит из столбцов (таблица 3.4):

Таблица 3.4 – Столбцы таблицы Likes

|  |  |  |
| --- | --- | --- |
| Наименование | Описание | Тип |
| UserId | Идентификатор автора лайка, внешний ключ на таблицу Users, часть составного первичного ключа | int |
| PostId | Идентификатор публикации, которой принадлежит лайк, внешний ключ на таблицу Posts, часть составного первичного ключа | int |

Таблица Shares хранит список людей, которые поделились публикацией, состоит из столбцов (таблица 3.5):

Таблица 3.5 – Столбцы таблицы Shares

|  |  |  |
| --- | --- | --- |
| Наименование | Описание | Тип |
| UserId | Идентификатор пользоавтеля, который поделился, внешний ключ на таблицу Users, часть составного первичного ключа | int |
| PostId | Идентификатор публикации, которой поделились, внешний ключ на таблицу Posts, часть составного первичного ключа | int |

Таблица Reports хранит список жалоб на публикацим, состоит из столбцов (таблица 3.6):

Таблица 3.6 – Столбцы таблицы Reports

|  |  |  |
| --- | --- | --- |
| Наименование | Описание | Тип |
| UserId | Идентификатор автора жалобы, внешний ключ на таблицу Users, часть составного первичного ключа | int |
| PostId | Идентификатор публикации, на которую пожаловались, внешний ключ на таблицу Posts, часть составного первичного ключа | int |

Таблица Tags хранит список тэгов публикаций, состоит из столбцов (таблица 3.7):

Таблица 3.7 – Столбцы таблицы Tags

|  |  |  |
| --- | --- | --- |
| Наименование | Описание | Тип |
| Id | Идентификатор тега, первичный ключ | int |
| Text | Наименование тэга, not null | text |
| PostId | Идентификатор публикации, которой принадлежат тэги, внешний ключ на таблицу Posts | int |

Таблица Subscriptions хранит список подписок пользователей, состоит из столбцов (таблица 3.8):

Таблица 3.8 – Столбцы таблицы Subscriptions

|  |  |  |
| --- | --- | --- |
| Наименование | Описание | Тип |
| SubscriberId | Идентификатор подписчика, внешний ключ на таблицу Users, часть составного первичного ключа | int |
| PublisherId | Идентификатор пользоавтеля, на которого подписался подпищик, внешний ключ на таблицу Users, часть составного первичного ключа | int |

**3 Разработка необходимых объектов**

При разработке курсового проекта были созданы следующие объекты: табличные пространства, таблицы, представления, индексы, пользователи, процедуры, триггеры.

3.1 Табличные пространства

Табличное пространство - это концепция, которая позволяет администраторам баз данных определять места в файловой системе, где хранятся файлы, представляющие объекты базы данных, такие как таблицы, индексы и т.д Табличные пространства в PostgreSQL позволяют контролировать расположение файлов на диске и оптимизировать производительность базы данных.

В PostgreSQL Табличные пространства являются неотъемлемой частью кластера базы данных и не могут быть перемещены, подключены или отключены от другого кластера без специальных процедур.

Создание табличного пространства, а также базы данных, которая будет использоавть это пространство представлено на рис. 3.1.1.

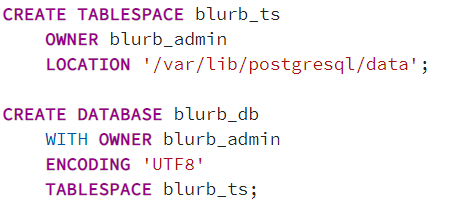


Рисунок 3.1.1 – Создание табличного пространства и базы данных

3.2 Таблицы

Таблицы в базах данных - это совокупность связанных данных, хранящихся в структурированном виде. Они состоят из столбцов и строк, которые представляют атрибуты и значения объектов базы данных Таблицы позволяют организовать данные логически и избежать избыточности. Для связи таблиц между собой используются ключи - уникальные идентификаторы строк в таблицах.

Для генерации первичных ключей используется IDENTITY. IDENTITY - это свойство столбца в таблице, которое позволяет автоматически генерировать уникальные значения для этого столбца при вставке новых строк. Использование identity упрощает работу с базой данных, так как не нужно заботиться о том, чтобы предоставлять уникальные значения для ключевых столбцов при добавлении данных.

Также для упрощения работы с бд используется опция ON DELETE CASCADE. это опция, которая позволяет автоматически удалять связанные строки из других таблиц при удалении строки из одной таблицы.

Создание таблиц представлено на рисунках 3.2.1, 3.2.2, 3.2.3, 3.2.4.

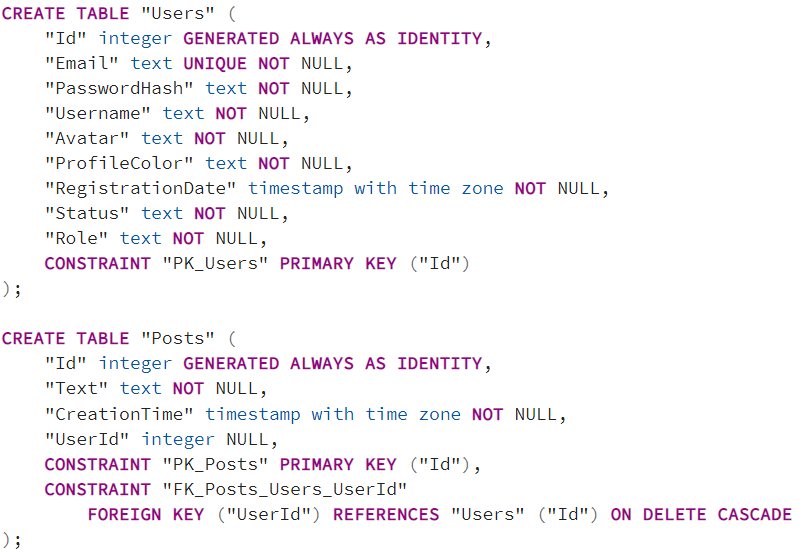


Рисунок 3.2.1 – Создание таблиц Users и Posts

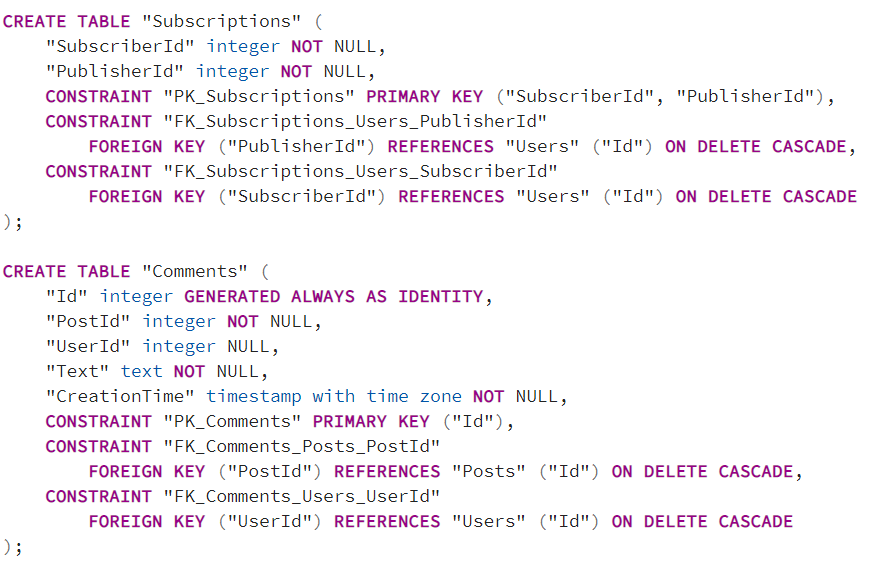


Рисунок 3.2.2 – Создание таблиц Subscriptions и Comments

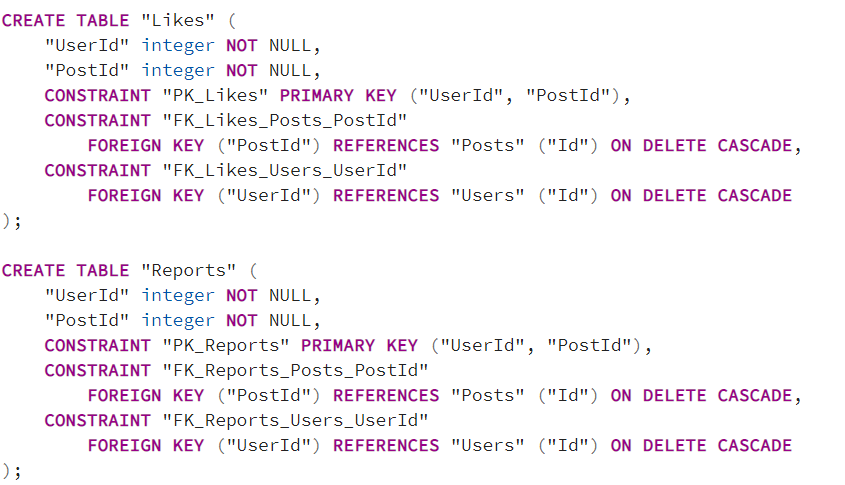


Рисунок 3.2.3 – Создание таблиц Likes и Reports

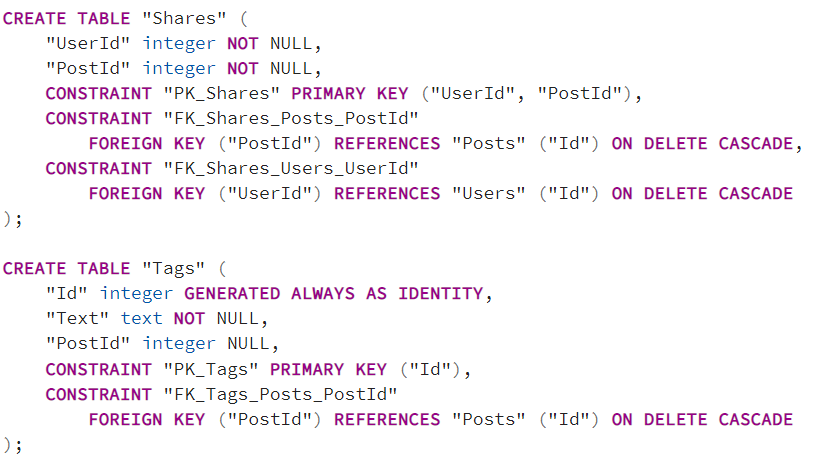


Рисунок 3.2.4 – Создание таблиц Shares и Tags

3.3 Роли

Роль - это сущность, которая может владеть объектами и иметь определённые права в базе.

В PostgreSQL нет формального различия между ролями и пользователями. Оба термина используются взаимозаменяемо для обозначения сущностей, которые могут подключаться к базе данных и выполнять действия с ней.

Однако для удобства можно считать, что роль - это более общее понятие, которое включает в себя как индивидуальных пользователей, так и группы пользователей. Пользователь - это роль с атрибутом LOGIN, который позволяет ей аутентифицироваться в системе. Группа - это роль без атрибута LOGIN, которая служит для объединения других ролей по общим правам доступа.

VALID UNTIL означает, что пароль будет действителен до указанной даты и времени.

Создание ролей приведено на рисунке 3.3.1.

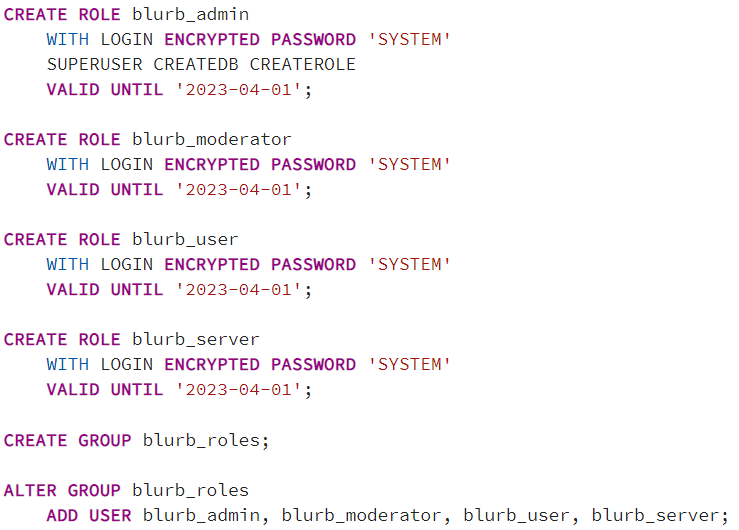


Рисунок 3.3.1 – Создание ролей

Привилегии, выдаваемые ролям, определяют, какие действия могут быть выполнены в отношении определенных объектов базы данных (например, таблиц, представлений и т.д.) и контролируют доступ пользователей к этим объектам.

Выдача ролей представлена на рисунке 3.3.2.



Рисунок 3.3.2 – Выдача привелегий для ролей

Все роли приложения могут выполнять базовые операции для добавления и обновления данных в таблицах, всем ролям также выдается доступ на функции для выборки, так как они не содержат приватной информации. Ролям администратора и модератора выдаются привилегии на процедуры для удаления данных, а администратор имеет права менять роли пользователей. Специальной роли blurb\_server выдаются привилегии на сами таблицы, что может быть полезно при использовании приложения с ORM и автоматической генерацией SQL кода при разработке с использованием code first approach.

3.4 Процедуры и функции

Процедура в СУБД - это программный объект, блок кода, который может быть вызван многократно, который позволяет выполнять определенные действия с данными в базе данных. Функция в СУБД - это программный объект, который возвращает значение. Процедуры и функции позволяют уменьшить количество кода, а также обеспечить его повторное использование, ускорить запросы к базе данных, ограничить доступ к данным.

* + 1. Выборка данных из таблиц

Процедуры в PostgreSQL не могут возвращать таблицы или результирующме наборы, поэтому для выборки данных используются функции. Всего было разработано 16 функций для выборки различной информации из таблиц и представлений: get\_users\_summary, get\_users\_summary\_by\_registration\_date, get\_posts\_summary, get\_trending\_posts, get\_reported\_posts, get\_comments\_on\_post, get\_comments\_from\_user, get\_subscriptions\_for\_publisher, get\_subscriptions\_of\_user, get\_post\_likes, get\_user\_likes, get\_post\_reports, get\_user\_reports, get\_post\_shares, get\_user\_shares, get\_post\_tags, get\_popular\_tags.

Пример создания функций в PostgreSQL представлен на рисунке 3.4.1.1 и 3.4.1.2.

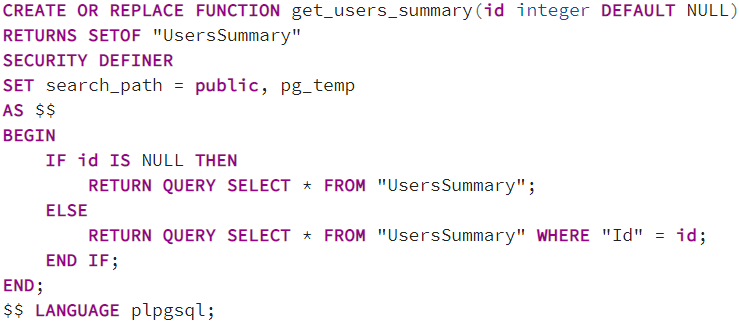


Рисунок 3.4.1.1 – Создание функции get\_users\_summary

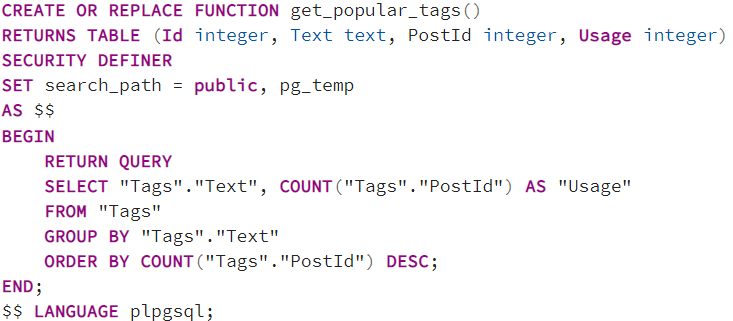


Рисунок 3.4.1.2 – Создание функции get\_popular\_tags

По умолчанию в PostgreSQL функции и процедуры выполняются с привилегиями той роли, которая вызвала процедуры, а это значит, что пользователи не смогут получить доступ к таблице. Чтобы обойти это ограничение, все функции и процедуры были созданы с опцией SECURITY DEFINER, что позволяет вызывать процедуры с теми привилегиями, которые есть у владельца процедуры. Чтобы избежать уязвимостей, связанных с временными таблицами используется свойство search\_pathб которое устанавливается в желаемую схему и временную схему по умолчанию

* + 1. Добавление данных в таблицы

Данные в таблицу будут добавляться, изменяться и удаляться при помощи процедур. Для добавления данных были созданы следующие процедуры: add\_user, add\_post, add\_comment, add\_tag, add\_tags, add\_like, add\_likes, add\_report, add\_reports, add\_share, add\_shares, add\_subscription.

Создание процедуры для добавления пользователя и для добавления тэгов к публикации представлено на рисунке 3.4.2.1 и 3.4.2.2.

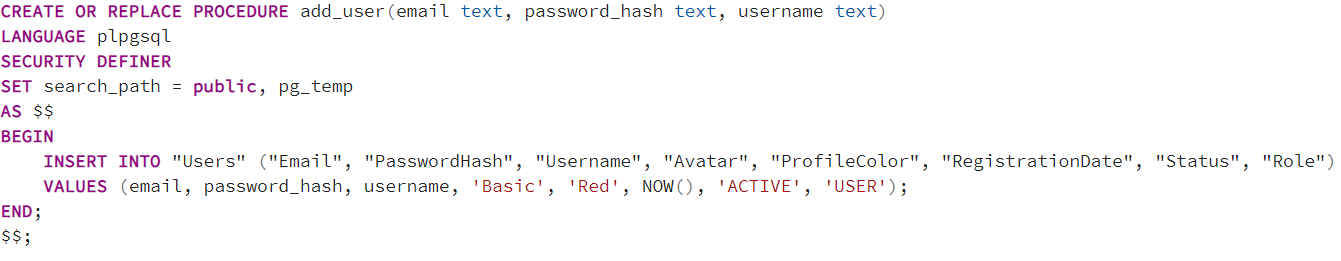


Рисунок 3.4.2.1 – Создание процедуры add\_user

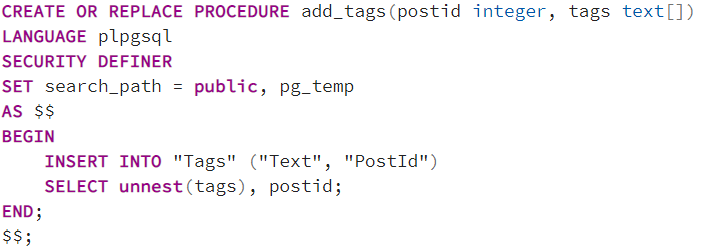


Рисунок 3.4.2.2 – Создание процедуры add\_tags

Процедура add\_tags использует возможность PostgreSQL работать с массивами значений, что позволяет за один вызов процедуры добавить несколько строк в таблицу.

* + 1. Изменение данных в таблицах

Процедуры, отвечающие за обновление данных: update\_user, update\_user\_status, update\_user\_role, update\_user\_password\_hash, update\_user\_email, update\_post, update\_comment.

Создание процедуры для обновления профиля и пароля пользователя представлено на рисунке 3.4.3.1 и 3.4.3.2.

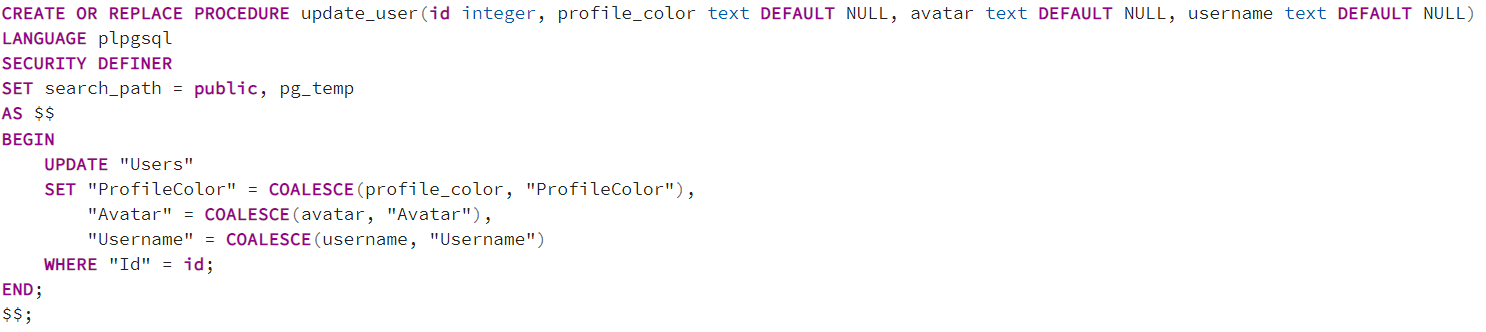


Рисунок 3.4.3.1 – Создание процедуры update\_user

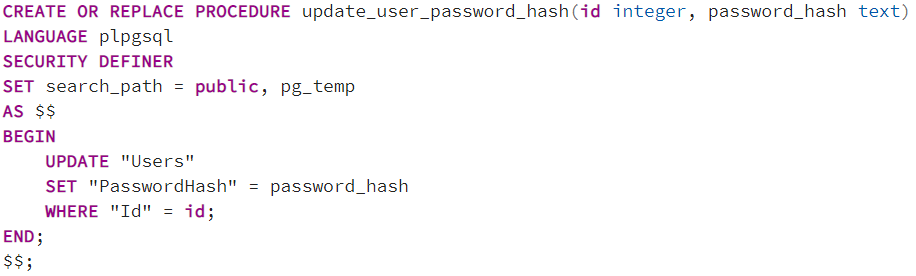


Рисунок 3.4.3.2 – Создание процедуры update\_user\_password\_hash

Процедура update\_user использует значения параметров процедуры по умолчанию и функцию COALESCE, которая выбирает первое значение из набора, которое не является NULL, чтобы дать возможность обновить только определенные настройки профиля.

* + 1. Удаление данных из таблиц

Для удаления данных были разработаны следующие процедуры: delete\_user, delete\_post, delete\_comment, delete\_tag, delete\_tags, delete\_like, delete\_likes, delete\_report, delete\_reports, delete\_share, delete\_shares, delete\_subscription.

Создание процедуры для удаления пользователя и публикации представлено на рисунке 3.4.4.1 и 3.4.4.2.

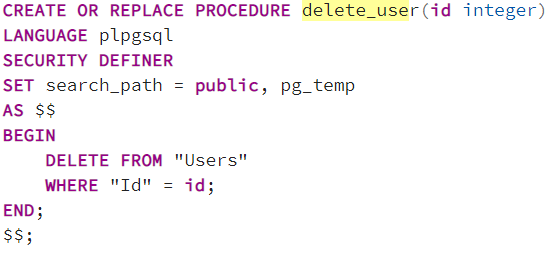


Рисунок 3.4.4.1 – Создание процедуры delete\_user

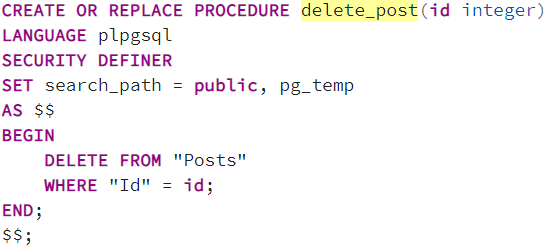


Рисунок 3.4.4.2 – Создание процедуры delete\_post

* + 1. Заполнение определённых таблиц на 100000 строк

На рисунке 3.5.5.1 представлен скрипт, заполняющий таблицу Likes более 100000 строками. Генерация случайных чисел позволяет создать различное количество лайков для постов в диапазоне от 1000 до 2000 лайков на пост.

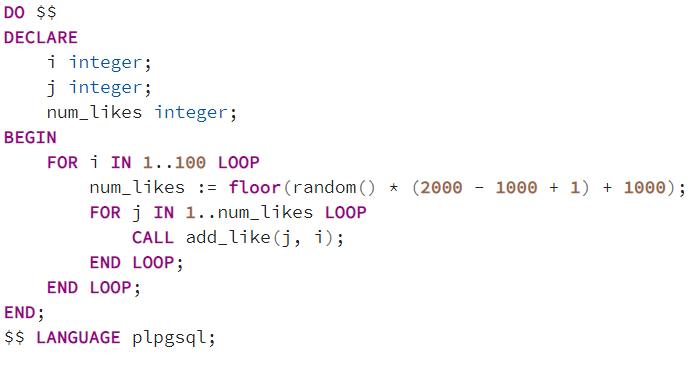


Рисунок 3.6.6.1 – Скрипт, заполняющий таблицу Likes

* + 1. Импорт данных таблицы в формате JSON

На рисунке 3.4.6.1 представлена процедура, выполняющая вставку данных в таблицу Posts из формата JSON.

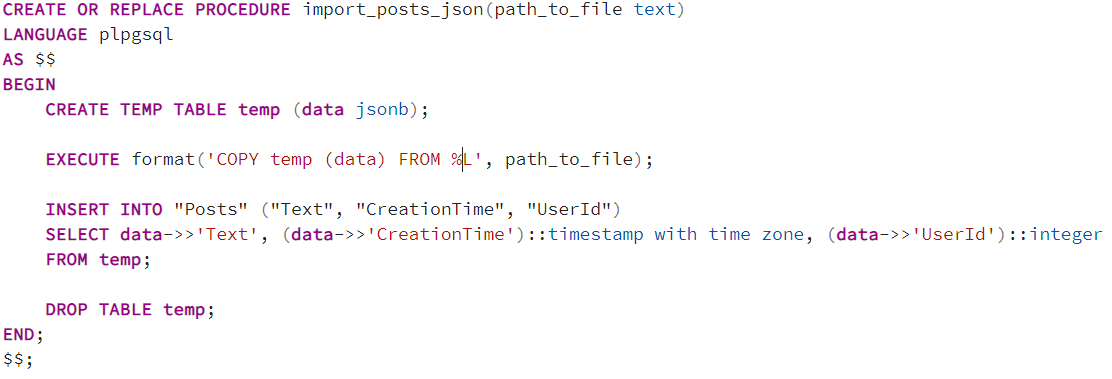


Рисунок 3.4.6.1 – Процедура, выполняющая импорт данных в формате JSON

* + 1. Экспорт данных таблицы в формате JSON

На рисунке 3.4.7.1 представлена процедура, выполняющая экспорт данных из таблицы Posts в формат JSON.

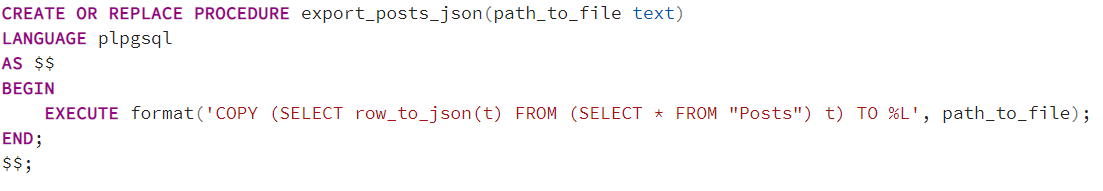


Рисунок 3.4.7.1 – Процедура, выполняющая экспорт данных в формате JSON

3.5 Триггеры

Триггеры SQL - это триггеры баз данных, которые позволяют вам указать движку SQL выполнить определенный код при наступлении определенного события или даже перед его наступлением. В PostgreSQL исполняемый код описывается путем создания функции, которая возвращает значение типа TRIGGER.

Для обеспечения сохранности данных все операции по редактированию профилей, публикаций и комментариев логгируются с помощью триггеров. Были созданы 3 таблицы, функции и триггера для лоогирования данных

На рисунках 3.5.1, 3.5.2 и 3.5.3 представлено создание триггеров для логгирования.



Рисунок 3.5.1 – Создание user\_update\_trigger



Рисунок 3.5.2 – Создание процедуры comment\_update\_trigger



Рисунок 3.5.3 – Создание процедуры post\_update\_trigger

* 1. Представления

Представление - это виртуальная таблица, которая представляет собой поименованный запрос, который будет подставлен как подзапрос при использовании представления. В отличие от обычных таблиц, представление не является самостоятельной частью набора данных, хранящихся в базе данных.

Преимуществами представлений являются возможность гибкой настройки данных, а также сокрытии приватной информации. Были разработаны представления для таблицы Users и Posts, которые содержат вычисляемую информацию, а также исключают электронную почту и пароль пользователя.

Создание процедуры для удаления пользователя и публикации представлено на рисунке 3.4.4.1 и 3.4.4.2.



Рисунок 3.6.1 – Создание процедуры UsersSummary

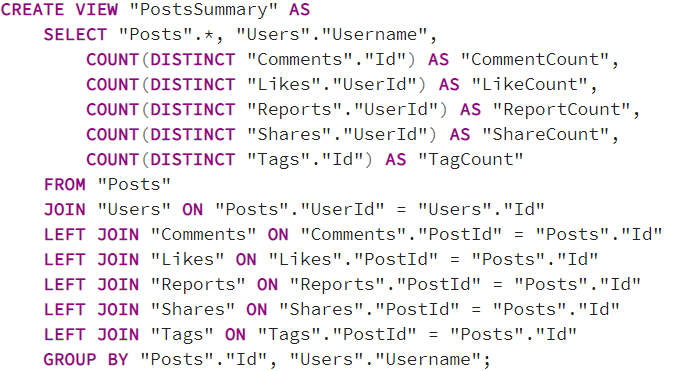


Рисунок 3.6 .2 – Создание процедуры PostsSummary

* 1. Индексы

Индекс в базе данных - это структура данных, которая улучшает скорость поиска данных в таблице. Индексы создаются для одного или нескольких столбцов таблицы и обеспечивают быстрый доступ к строкам таблицы, которые соответствуют определенным критериям поиска. Индексы могут существенно ускорить выполнение запросов к базе данных, но также могут замедлить операции вставки, обновления и удаления, так как индексы должны быть обновлены при каждой такой операции.

PostgreSQL предоставляет несколько типов индексов: B-Tree, Hash, GiST, SP-GiST, GIN и BRIN. Каждый тип индекса использует разные алгоритмы, которые лучше всего подходят для разных типов запросов1.

B-Tree: B-Tree индексы могут обрабатывать запросы на равенство и диапазон для данных, которые могут быть отсортированы в некотором порядке. Оптимизатор запросов PostgreSQL будет рассматривать использование B-Tree индекса всякий раз, когда индексируемый столбец участвует в сравнении с использованием одного из этих операторов: <, <=, =, >= и >1.

Hash: Hash индексы хранят 32-битный хеш-код, полученный из значения индексируемого столбца. Следовательно, такие индексы могут обрабатывать только простые сравнения на равенство1.

GiST: GiST (Generalized Search Tree) индексы не являются одним видом индекса, а скорее инфраструктурой, в рамках которой могут быть реализованы многие различные стратегии индексации. Соответственно, конкретные операторы, с которыми может быть использован GiST индекс, варьируются в зависимости от стратегии индексации (класса оператора)1.

SP-GiST: SP-GiST (Space-Partitioned GiST) индексы предоставляют инфраструктуру для создания масштабируемых и эффективных по памяти некластеризованных индексов1.

GIN: GIN (Generalized Inverted Index) индексы оптимизированы для обработки запросов с операторами @>, <@ и &&1.

BRIN: BRIN (Block Range Index) индексы хранят метаданные о диапазонах страниц данных в таблице и могут обеспечить значительное ускорение запросов для очень больших таблиц.

По умолчанию используются B-Tree индексы, потому что они подходят для большинства задач. Часто столбцов со строковыми данными может быть использован Hash индекс, но они подходят только для операций типа =, что неподходит под требования для созданных таблиц.

Создание индексов представлено на рисунке 3.7.1.

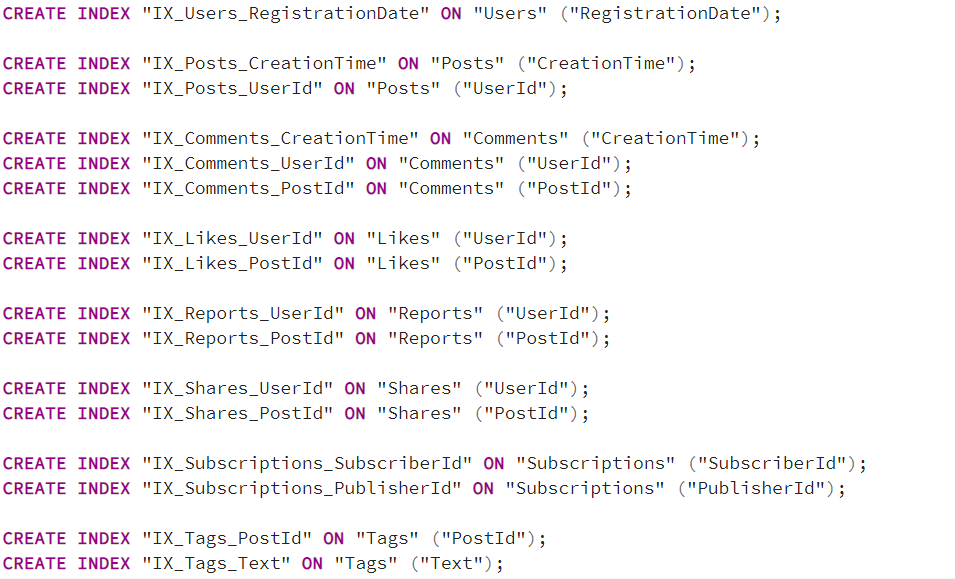


Рисунок 3.7.1 – Создание индексов

**4 Сравнение SQL и NoSQL решений**

NoSQL - это подход к проектированию баз данных, который позволяет хранить и запрашивать данные за пределами традиционных структур, используемых в реляционных базах данных. Одним из самых популярных NoSQL решений сегодня является MongoDB.

MongoDB является документо-ориентированной системой, в которой центральным понятием является документ. Документ можно представить как объект, хранящий некоторую информацию. Документная база данных – это тип нереляционных баз данных, предназначенный для хранения и запроса данных в виде документов в формате, подобном JSON. Гибкий, полуструктурированный, иерархический характер документов и документных баз данных позволяет им развиваться в соответствии с потребностями приложений.

4.1 Табличные пространства

Табличное пространство - это концепция, которая позволяет администраторам баз данных определять места в файловой системе, где хранятся файлы, представляющие объекты базы данных, такие как таблицы, индексы и т.д Табличные пространства в PostgreSQL позволяют контролировать расположение файлов на диске и оптимизировать производительность базы данных.

**5 Тестирование производительности**

Производительность базы данных является критически важным фактором для обеспечения быстрого и эффективного доступа к данным. Ключевым в обеспечении произваодительности бд является правильное использование индексов. Правильное использование индексов может значительно улучшить производительность базы данных и обеспечить быстрый и надежный доступ к данным.

Для тестирования производительности была выбрана таблица Likes, так как лайки являются самым часто используемым действием взаимодействия в социальной сети. После заполнения таблицы с используем генерации случайных значений в таблице находится 145335 строк, представлено на рисунке 6.1.

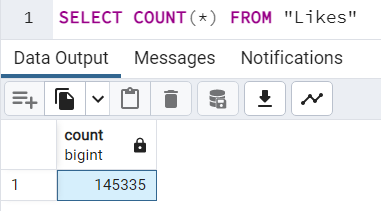


Рисунок 6.1 – Число строк в таблице

Для проверки эффективности использования индексов выполним типичную задачу для подобной таблицы - подсчет количества лайков на определенной публикации и оценим план запроса и его стоимость, представлено на рисунке 6.2 и 6.3.

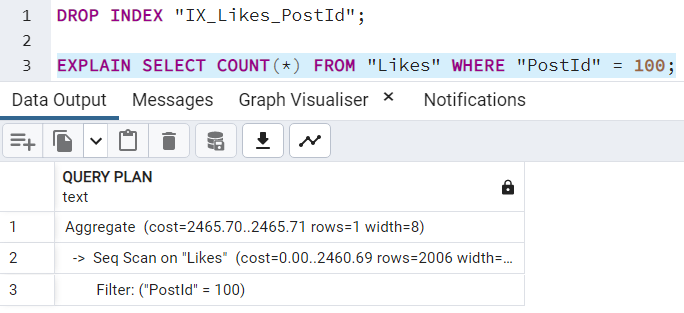


Рисунок 6.2 – План запроса без использования индекса

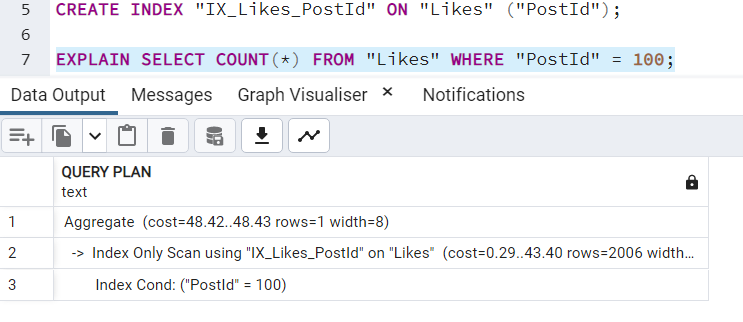


Рисунок 6.3 – План запроса с использованием индекса

Использование индекса существенно понижает стоимость агрегации, что увеличивает производительность запроса.

Также важным является использование составных первичных ключей, что оптимально для запросов где используется фильтрация по обоим столбцам, например при ситуации, когда нужно узнать поставил ли пользователь лайк, представлено на рисунке 6.4.

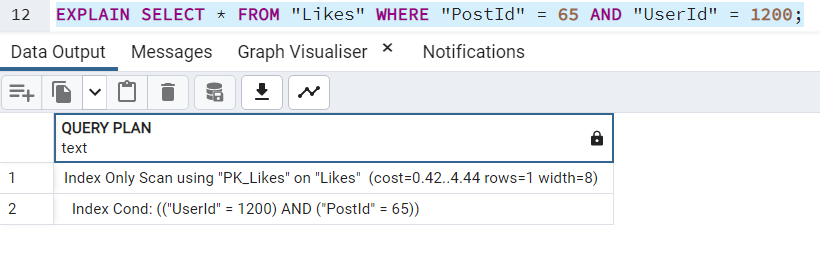


Рисунок 6.4 – Запрос с фильтром по обоим столбцам