Национальный исследовательский ядерный университет «МИФИ»

(Московский Инженерно-Физический Институт)

Кафедра №42 «Криптология и кибербезопасность»

**Лабораторная работа №2-4**

**«**Индексы**»**

Тимин Александр Б21-515 (2024г.)

**1 Система индексов**

Начнем с разработки системы индексов для базы данных.

Сначала определим типичные запросы, которые могут использоваться в нашей базе данных. Основываясь на предоставленной ERD-схеме, определим типичные запросы:

* Получение информации о сериале:
  + Запросы по shows для поиска сериала по названию (title) или году выпуска (year).
  + Фильтрация сериалов по бюджету (budget).
* Работа с актерами:
  + Поиск актеров по имени (name).
  + Выборка всех сериалов, в которых участвовал определенный актер.
* Обработка отзывов:
  + Получение всех отзывов для конкретного сериала (show\_id).
  + Фильтрация отзывов по рейтингу (rating).
* Работа с сезонами и эпизодами:
  + Выбор всех сезонов конкретного сериала (show\_id).
  + Получение всех эпизодов конкретного сезона (season\_id).
* Связи актеров и ролей:
  + Фильтрация по актерам (actor\_id) или ролям (role) для определенного сериала (show\_id).

Теперь мы можем спроектировать индексы. Для каждого из перечисленных запросов создадим индексы, которые ускорят их выполнение.

Таблица shows:

* Индекс на title для поиска по названию

CREATE INDEX idx\_shows\_title ON series.shows(title);

* Индекс на year для фильтрации по году

CREATE INDEX idx\_shows\_year ON series.shows(year);

Таблица actors:

* Индекс на name для поиска по имени

CREATE INDEX idx\_actors\_name ON people.actors(name);

Таблица reviews:

* Индекс на show\_id для быстрого доступа к отзывам сериала

CREATE INDEX idx\_reviews\_show\_id ON series.reviews(show\_id);

* Индекс на rating для фильтрации по рейтингу

CREATE INDEX idx\_reviews\_rating ON series.reviews(rating);

Таблица seasons:

* Индекс на show\_id для получения сезонов сериала

CREATE INDEX idx\_seasons\_show\_id ON series.seasons(show\_id);

Таблица episodes:

* Индекс на season\_id для получения эпизодов сезона

CREATE INDEX idx\_episodes\_season\_id ON series.episodes(season\_id);

Таблица cast:

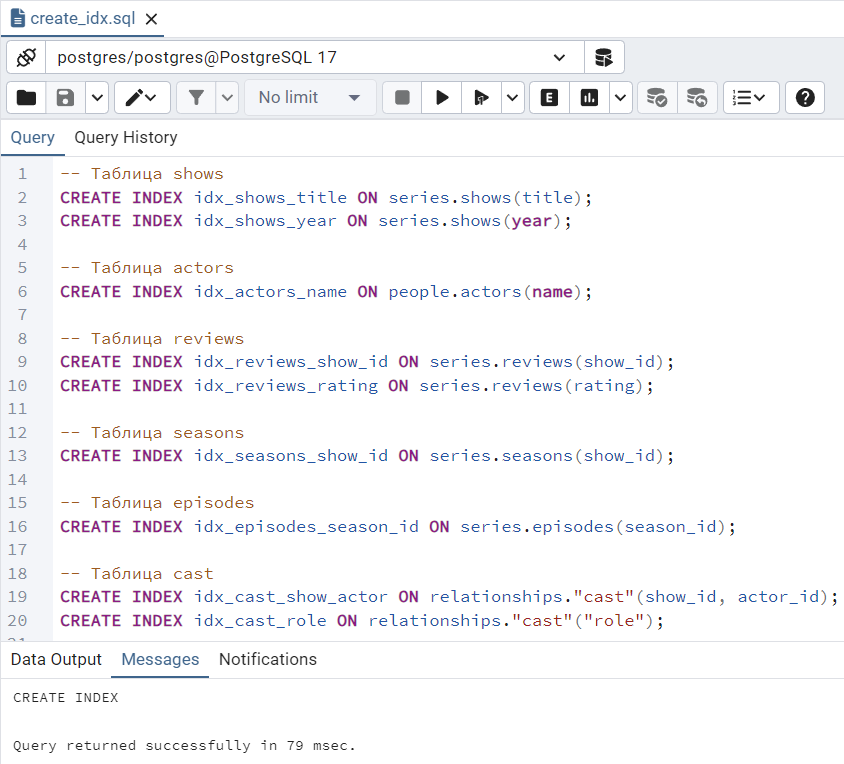
* Составной индекс на show\_id и actor\_id для связей сериалов и актеров

CREATE INDEX idx\_cast\_show\_actor ON relationships."cast"(show\_id, actor\_id);

* Индекс на role для фильтрации по ролям

CREATE INDEX idx\_cast\_role ON relationships."cast"("role");

Создадим эти индексы.

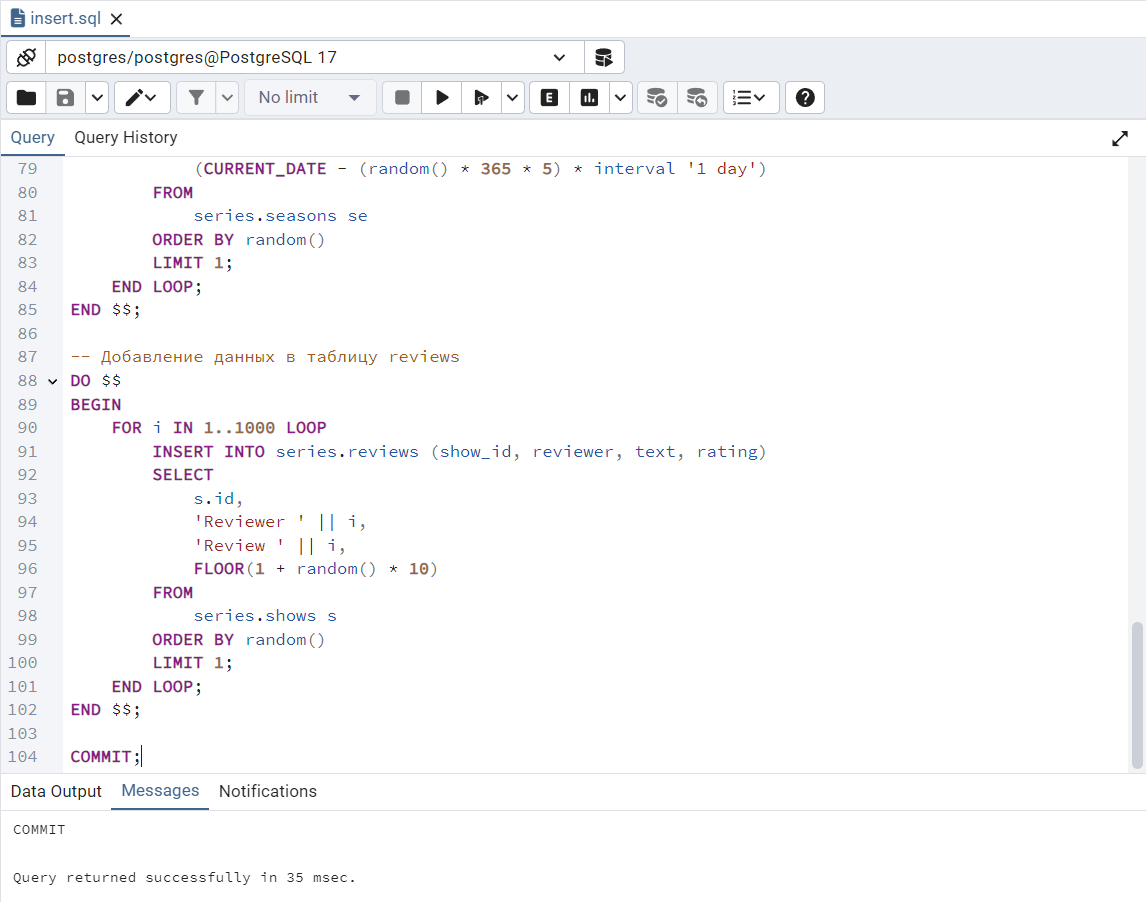


**2 Статистическая информация**

Данный пункт в ТЗ был помечен, как «дополнительное кармическое задание», так как я не верю в карму, решил его не выполнять😊

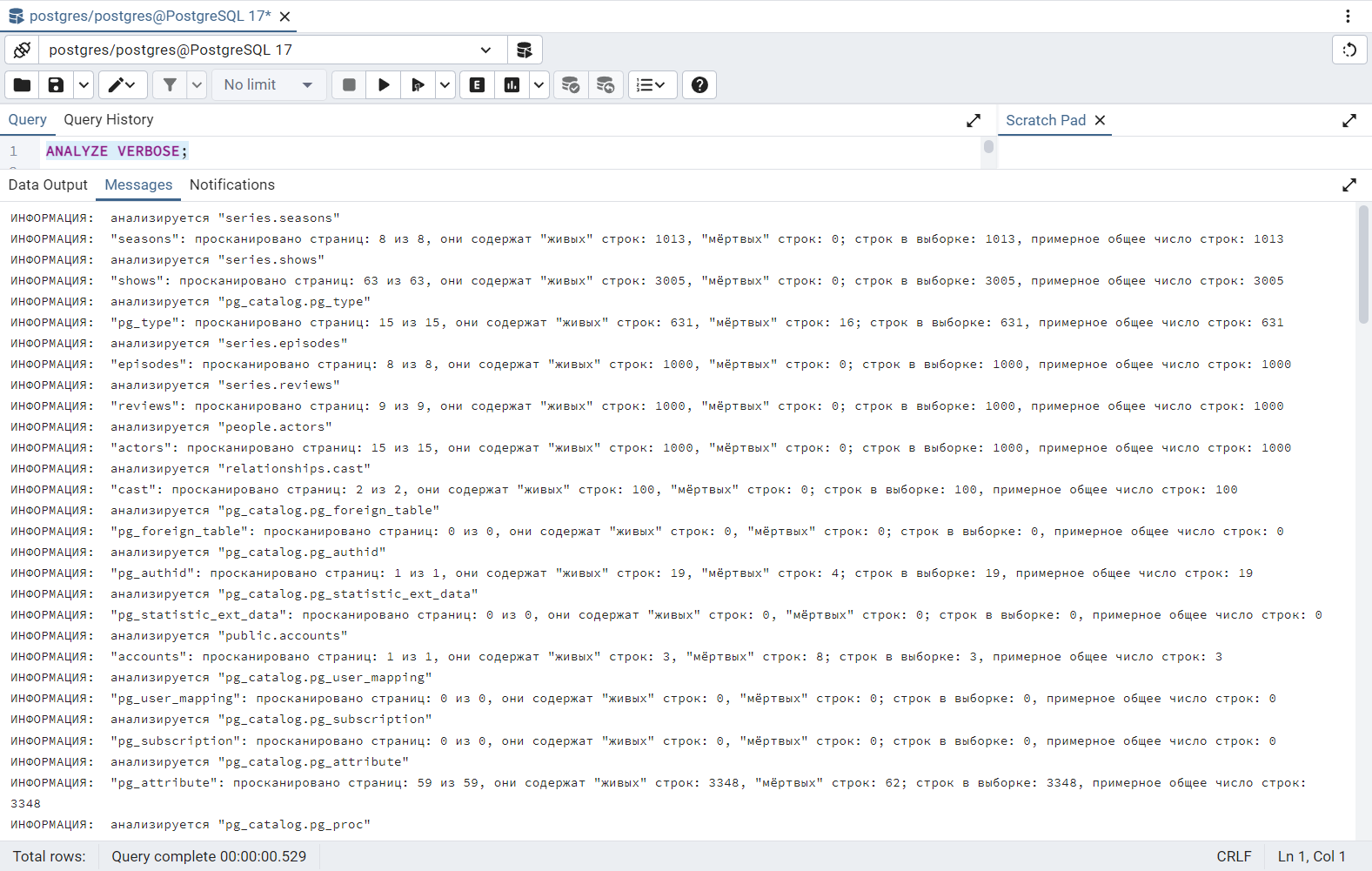
**3 Много новых данных**

С помощью скрипта добавил кучу записей в БД (ссылка на скрипт – в приложении). Запросы не слишком оптимизированы, но я думаю, это не большая проблема.



Выполним актуализацию статистических данных:

ANALYZE VERBOSE;



Вывод, касаемый таблиц, в которые добавлялись данные:

ИНФОРМАЦИЯ: анализируется "series.seasons"

ИНФОРМАЦИЯ: "seasons": просканировано страниц: 8 из 8, они содержат "живых" строк: 1013, "мёртвых" строк: 0; строк в выборке: 1013, примерное общее число строк: 1013

ИНФОРМАЦИЯ: анализируется "series.shows"

ИНФОРМАЦИЯ: "shows": просканировано страниц: 63 из 63, они содержат "живых" строк: 3005, "мёртвых" строк: 0; строк в выборке: 3005, примерное общее число строк: 3005

ИНФОРМАЦИЯ: анализируется "series.episodes"

ИНФОРМАЦИЯ: "episodes": просканировано страниц: 8 из 8, они содержат "живых" строк: 1000, "мёртвых" строк: 0; строк в выборке: 1000, примерное общее число строк: 1000

ИНФОРМАЦИЯ: анализируется "series.reviews"

ИНФОРМАЦИЯ: "reviews": просканировано страниц: 9 из 9, они содержат "живых" строк: 1000, "мёртвых" строк: 0; строк в выборке: 1000, примерное общее число строк: 1000

ИНФОРМАЦИЯ: анализируется "people.actors"

ИНФОРМАЦИЯ: "actors": просканировано страниц: 15 из 15, они содержат "живых" строк: 1000, "мёртвых" строк: 0; строк в выборке: 1000, примерное общее число строк: 1000

ИНФОРМАЦИЯ: анализируется "relationships.cast"

ИНФОРМАЦИЯ: "cast": просканировано страниц: 2 из 2, они содержат "живых" строк: 100, "мёртвых" строк: 0; строк в выборке: 100, примерное общее число строк: 100

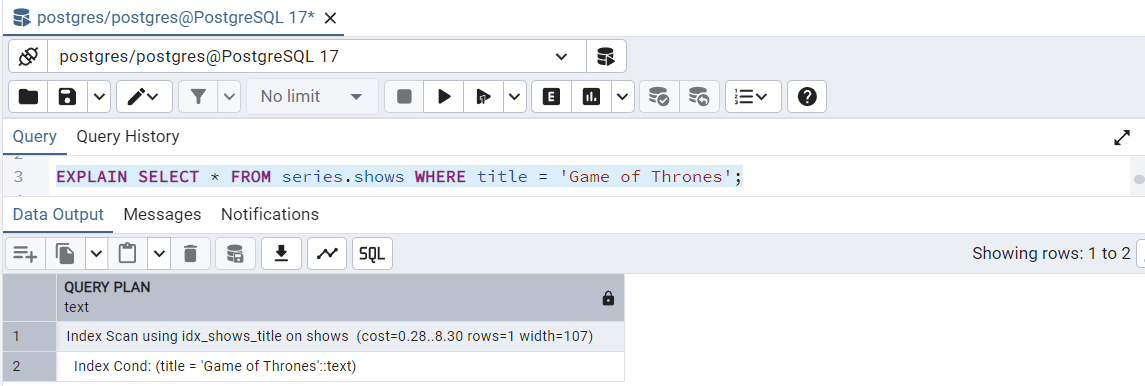
Вывод представляет собой информацию о процессе сканирования и анализа таблиц в базе данных. Каждая строка соответствует отдельной таблице в базе данных, и в ней указаны:

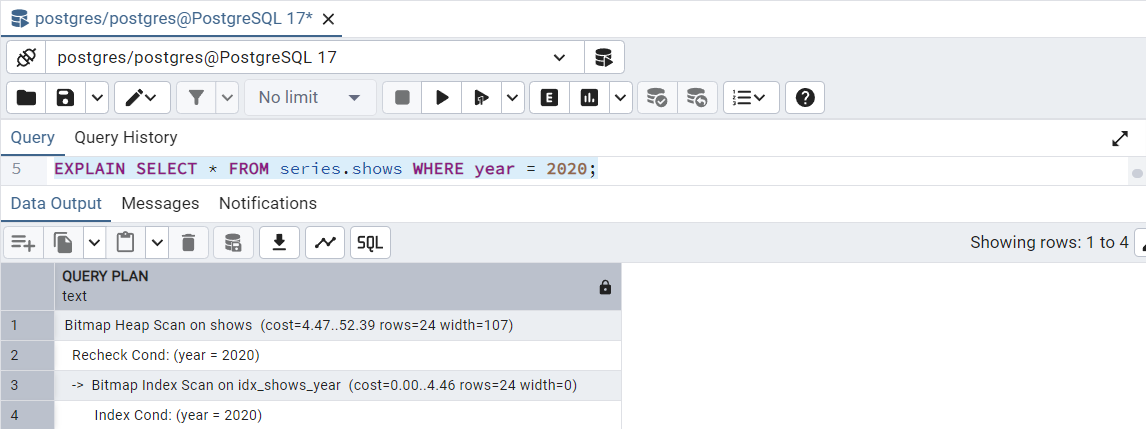
* Название таблицы
* Статус сканирования: указано, сколько страниц таблицы было просканировано
* Количество живых строк
* Количество мертвых строк
* Количество строк в выборке: это количество строк, которое было обработано в текущем запросе или анализе
* Примерное общее число строк: это общее количество строк в таблице, включая как живые, так и мертвые строки

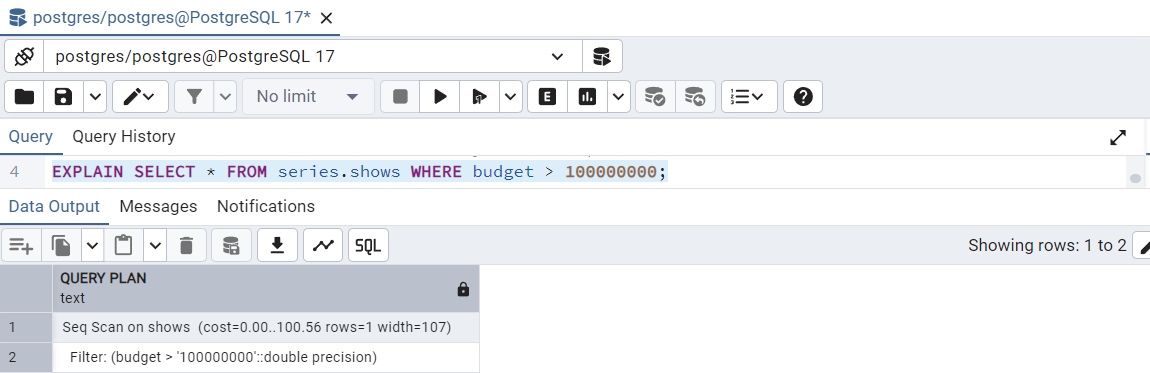
Живые строки — это данные, которые всё ещё актуальны и могут быть использованы для выполнения запросов или операций. Мертвые строки — это данные, которые были удалены, но ещё не освобождены системой

**4 Используются ли индексы?**

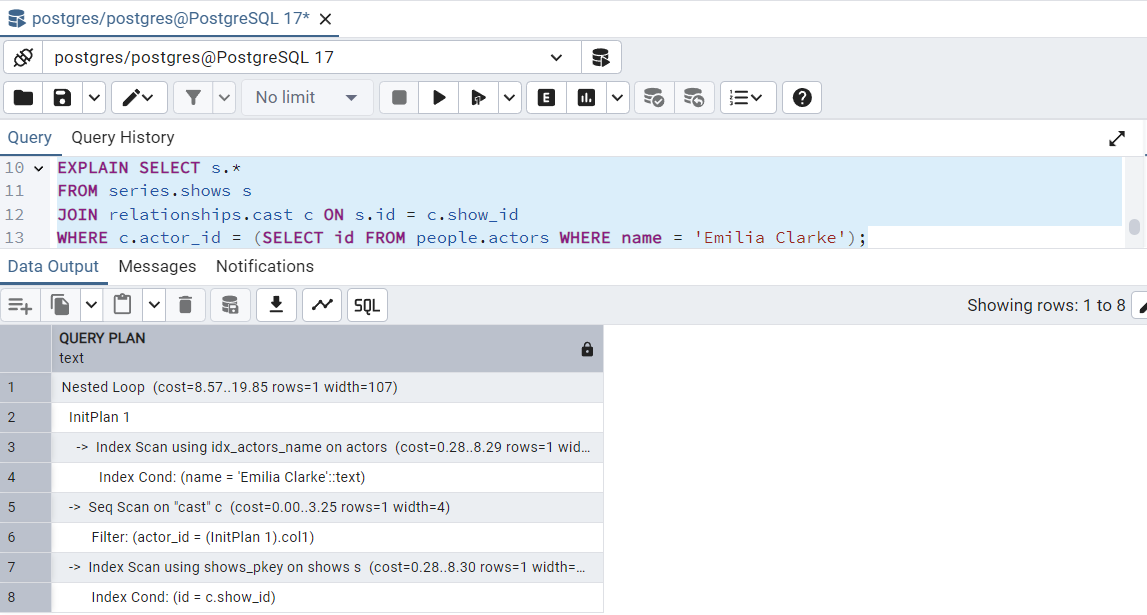
Проверим, используются ли индексы.

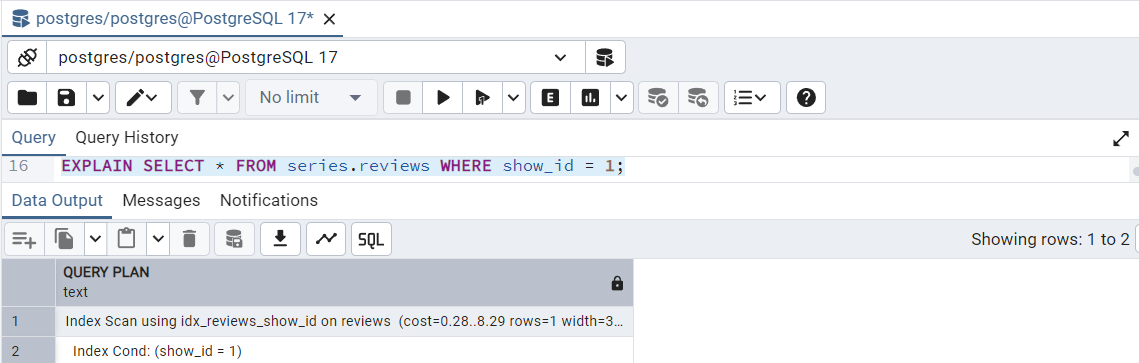




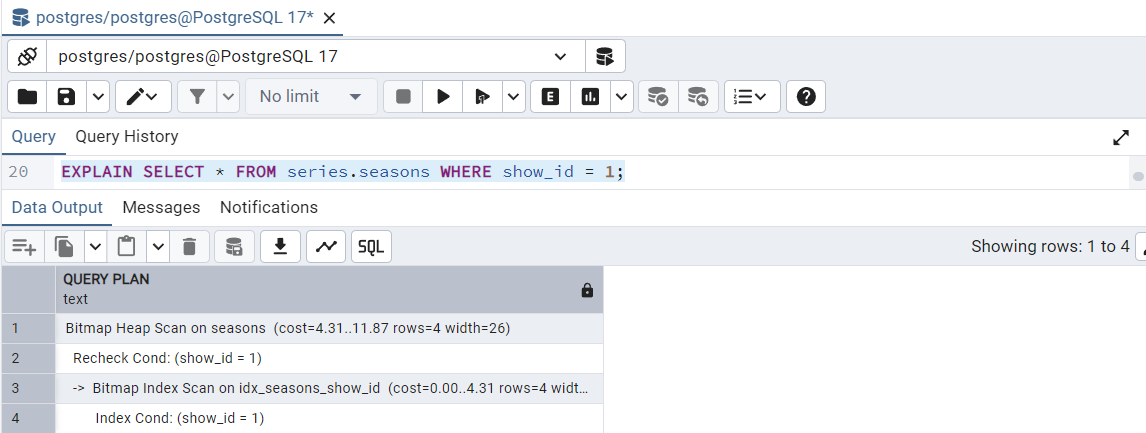


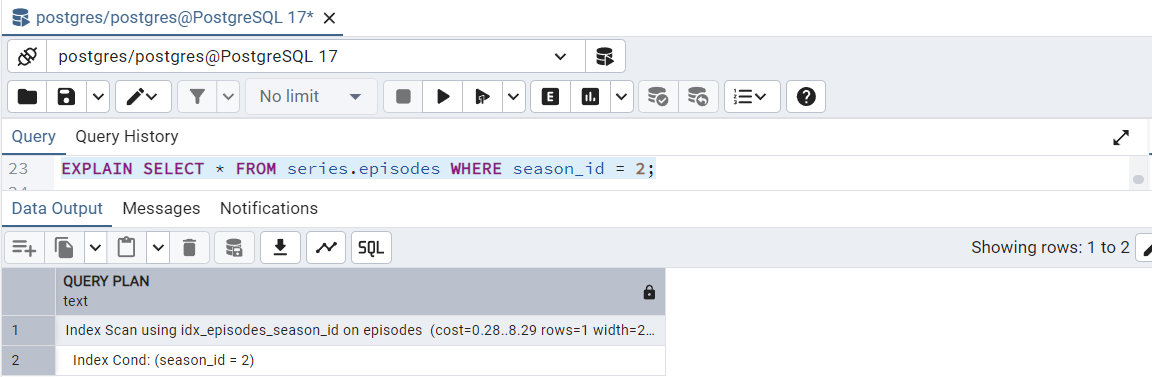












При анализе с помощью команды EXPLAIN, можно увидеть, что созданные индексы используются для ускорения выполнения данных запросов (запросы будут представлены в скрипте по ссылке из приложения).

**5 -----**

Данный пункт в ТЗ был помечен, как «дополнительное кармическое задание», так как я не верю в карму, решил его не выполнять😊

**6 -----**

Данный пункт в ТЗ был помечен, как «дополнительное кармическое задание», так как я не верю в карму, решил его не выполнять😊

**Заключение**

В ходе выполнения работы была разработана система индексов для базы данных, с учетом схемы данных и типичных запросов, характерных для БД сериалов.

Для проверки эффективности предложенных индексов было добавлено значительное количество данных в таблицы, после чего была выполнена команда ANALYZE с ключевым словом VERBOSE, что позволило актуализировать статистику и улучшить выбор планов выполнения запросов. Это обеспечило точность в оценке использования индексов для ускорения выполнения запросов.

С помощью команды EXPLAIN были построены планы выполнения различных запросов, что позволило убедиться, что предложенные индексы действительно используются и позволяют существенно ускорить выполнение типичных операций с данными.

Приложение

* [ERD схема](https://github.com/KATEHOK/DBS-labs-7/blob/main/1/pics/ERD.png);
* [скрипт создания индексов](https://github.com/KATEHOK/DBS-labs-7/blob/main/4/script/create_idx.sql);
* [скрипт вставки данных](https://github.com/KATEHOK/DBS-labs-7/blob/main/4/script/insert.sql);
* [скрипт запросов с EXPLAIN](https://github.com/KATEHOK/DBS-labs-7/blob/main/4/script/queries.sql);
* [отчет (docx)](https://github.com/KATEHOK/DBS-labs-7/blob/main/4/report.docx);
* [отчет (pdf)](https://github.com/KATEHOK/DBS-labs-7/blob/main/4/report.pdf).