

Федеральное государственное автономное образовательное учреждение высшего образования

«Национальный исследовательский университет ИТМО»

Факультет ПИ и КТ

Лабораторная работа №4

по дисциплине: «Базы данных»

Вариант 311113567

Выполнил:

**Болорболд Аригуун**,

группа P3111

Преподаватель:

**Николаев Владимир Вячеславович**

Практик:

**Чупанов Аликылыч Алибекович**



Санкт-Петербург

2023

1. **Задание:**

Составить запросы на языке SQL (пункты 1-2).

Для каждого запроса предложить индексы, добавление которых уменьшит время выполнения запроса (указать таблицы/атрибуты, для которых нужно добавить индексы, написать тип индекса; объяснить, почему добавление индекса будет полезным для данного запроса).

Для запросов 1-2 необходимо составить возможные планы выполнения запросов. Планы составляются на основании предположения, что в таблицах отсутствуют индексы. Из составленных планов необходимо выбрать оптимальный и объяснить свой выбор. Изменятся ли планы при добавлении индекса и как?

Для запросов 1-2 необходимо добавить в отчет вывод команды EXPLAIN ANALYZE [запрос].

**Запросы:**

1. Сделать запрос для получения атрибутов из указанных таблиц, применив фильтры по указанным условиям:

Таблицы: Н\_ЛЮДИ, Н\_СЕССИЯ.

Вывести атрибуты: Н\_ЛЮДИ.ИМЯ, Н\_СЕССИЯ.ДАТА.

Фильтры (AND):

a) Н\_ЛЮДИ.ИД = 100865.

b) Н\_СЕССИЯ.ДАТА = 2012-01-25.

c) Н\_СЕССИЯ.ДАТА < 2002-01-04.

Вид соединения: RIGHT JOIN.

1. Сделать запрос для получения атрибутов из указанных таблиц, применив фильтры по указанным условиям:

Таблицы: Н\_ЛЮДИ, Н\_ОБУЧЕНИЯ, Н\_УЧЕНИКИ.

Вывести атрибуты: Н\_ЛЮДИ.ФАМИЛИЯ, Н\_ОБУЧЕНИЯ.ЧЛВК\_ИД, Н\_УЧЕНИКИ.НАЧАЛО.

Фильтры: (AND)

a) Н\_ЛЮДИ.ИД > 152862.

b) Н\_ОБУЧЕНИЯ.НЗК < 933232.

Вид соединения: LEFT JOIN.

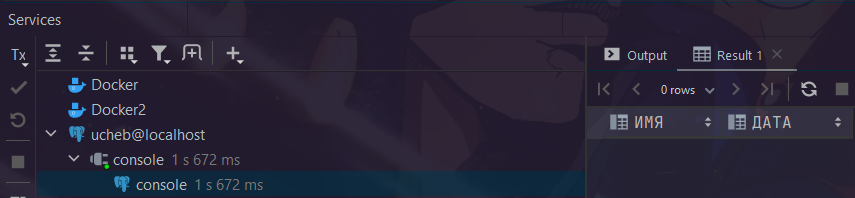
1. **Реализация запросов на SQL:**

SELECT Н\_ЛЮДИ.ИМЯ, Н\_СЕССИЯ.ДАТА FROM Н\_ЛЮДИ  
RIGHT JOIN Н\_СЕССИЯ ON Н\_ЛЮДИ.ИД = Н\_СЕССИЯ.ЧЛВК\_ИД  
WHERE Н\_ЛЮДИ.ИД = 100865  
AND Н\_СЕССИЯ.ДАТА = '2012-01-25'  
AND Н\_СЕССИЯ.ДАТА < '2002-01-04';

SELECT Н\_ЛЮДИ.ФАМИЛИЯ, Н\_ОБУЧЕНИЯ.ЧЛВК\_ИД, Н\_УЧЕНИКИ.НАЧАЛО FROM Н\_ЛЮДИ  
LEFT JOIN Н\_ОБУЧЕНИЯ ON Н\_ЛЮДИ.ИД = Н\_ОБУЧЕНИЯ.ЧЛВК\_ИД  
LEFT JOIN Н\_УЧЕНИКИ ON Н\_ОБУЧЕНИЯ.ЧЛВК\_ИД = Н\_УЧЕНИКИ.ЧЛВК\_ИД  
WHERE Н\_ЛЮДИ.ИД > 152862  
AND Н\_ОБУЧЕНИЯ.НЗК < '933232';

1. **Уменьшение выполнения времени 1 запроса:**

Время выполнения исходного запроса: 1,672 с



**Индексы, добавления которых уменьшит время выполнения запроса:**

1. **На таблице Н\_ЛЮДИ:**

Индекс на атрибуте Н\_ЛЮДИ.ИД (B-tree) ускорит фильтрацию с условием «Н\_ЛЮДИ.ИД = 100865».

1. **На таблице Н\_СЕССИЯ:**

Индекс на атрибуте Н\_СЕССИЯ.ЧЛВК\_ИД (B-tree) ускорит соединение таблиц по атрибуту ЧЛВК\_ИД.

Индекс на атрибуте Н\_СЕССИЯ.ДАТА (B-tree) ускорит фильтрацию с условиями «Н\_СЕССИЯ.ДАТА = 2012-01-25» и «Н\_СЕССИЯ.ДАТА < 2002-01-04».

Добавление индексов на указанные столбцы позволит ускорить поиск нужных записей по фильтрам в запросе.

**Возможные планы выполнения запроса без индексов:**

**План 1:**

* Полный скан таблицы Н\_ЛЮДИ;
* Полный скан таблицы Н\_СЕССИЯ с применением фильтра по условию «Н\_СЕССИЯ.ДАТА = 2012-01-25».
* Полный скан таблицы Н\_СЕССИЯ с применением фильтра по условиям «Н\_СЕССИЯ.ДАТА < 2002-01-04».
* Соединение таблиц с использованием Nested Loops Join по атрибуту ЧЛВК\_ИД.
* Фильтрация результата соединения по условию «Н\_ЛЮДИ.ИД = 100865».

**План 2:**

* Полный скан таблицы Н\_ЛЮДИ с применением фильтра по условию «Н\_ЛЮДИ.ИД = 100865».
* Полный скан таблицы Н\_СЕССИЯ с применением фильтра по условию «Н\_СЕССИЯ.ДАТА = 2012-01-25».
* Полный скан таблицы Н\_СЕССИЯ с применением фильтра по условиям «Н\_СЕССИЯ.ДАТА < 2002-01-04».
* Соединение таблиц с использованием Nested Loops Join по атрибуту ЧЛВК\_ИД.

**Оптимальный план:** второй, так как фильтрация данных выполняется до соединения таблиц, что уменьшает количество строк для обработки, а следовательно, и время выполнения запроса.

**При добавлении индексов планы выполнения запросов изменятся:**

* Вместо полного скана таблиц будет использоваться индексный скан.
* Nested Loops Join станет быстрее благодаря индексу на атрибуте ЧЛВК\_ИД.

1. **План выполнения 1 запроса:**

Nested Loop (cost=4.68..50.32 rows=1 width=21) (actual time=0.125..0.126 rows=0 loops=1)

-> Index Scan using "ЧЛВК\_PK" on "Н\_ЛЮДИ" (cost=0.28..8.30 rows=1 width=17) (actual time=0.041..0.042 rows=1 loops=1)

Index Cond: ("ИД" = 100865)

-> Bitmap Heap Scan on "Н\_СЕССИЯ" (cost=4.39..42.01 rows=1 width=12) (actual time=0.080..0.081 rows=0 loops=1)

Recheck Cond: ("ЧЛВК\_ИД" = 100865)

Filter: (("ДАТА" < '2002-01-04 00:00:00'::timestamp without time zone) AND ("ДАТА" = '2012-01-25 00:00:00'::timestamp without time zone))

Rows Removed by Filter: 15

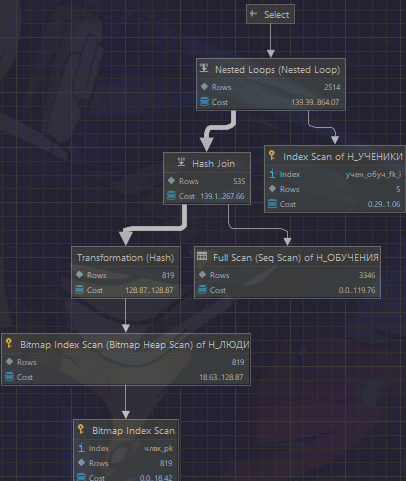
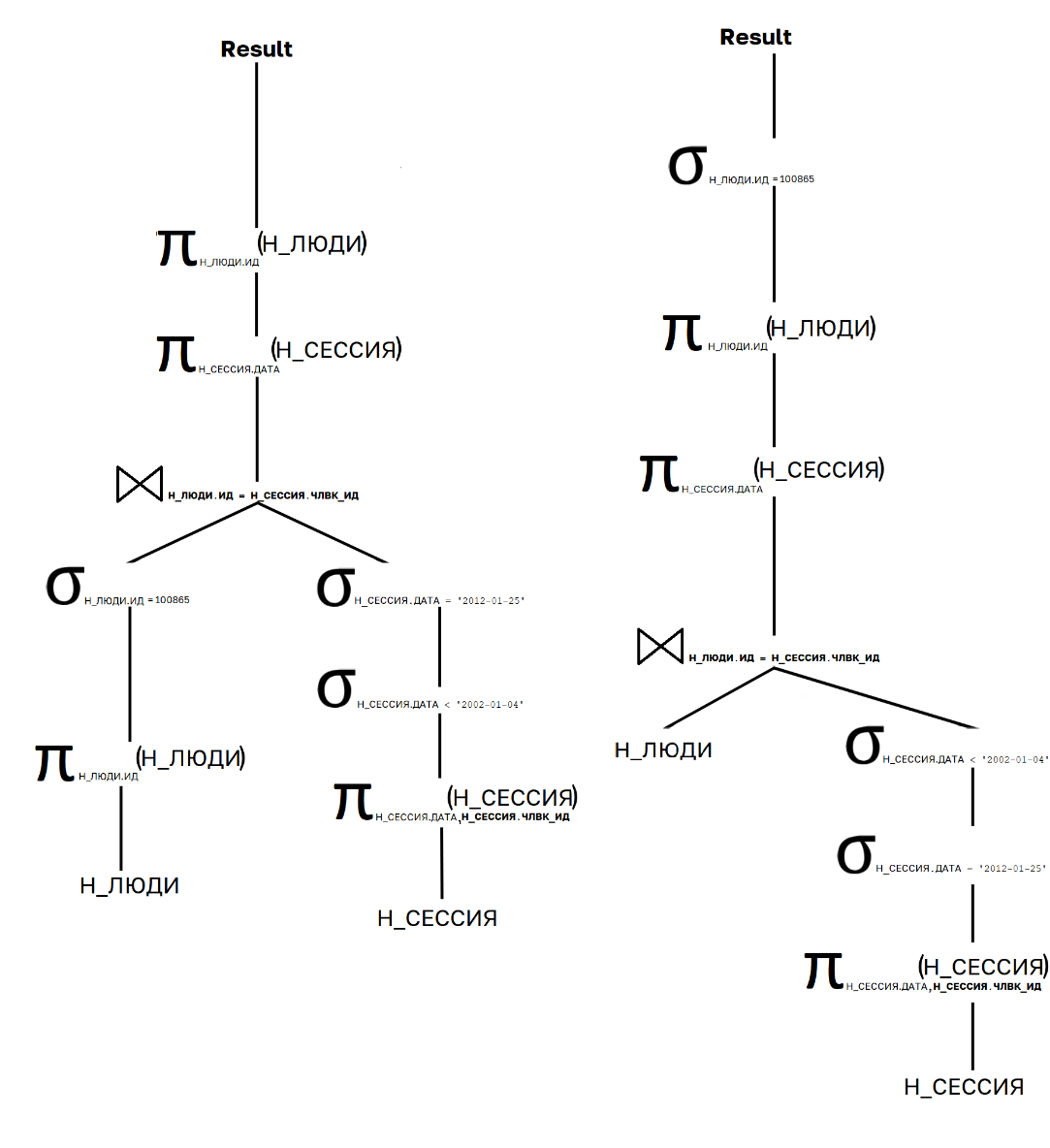
Heap Blocks: exact=14

-> Bitmap Index Scan on "SYS\_C003500\_IFK" (cost=0.00..4.39 rows=15 width=0) (actual time=0.046..0.047 rows=15 loops=1)

Index Cond: ("ЧЛВК\_ИД" = 100865)

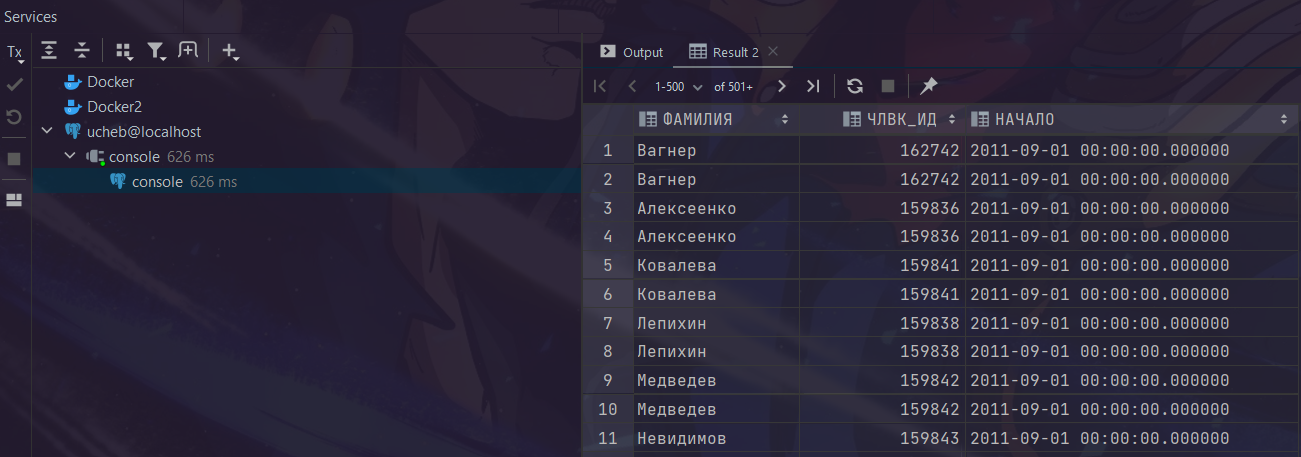
Planning Time: 0.861 ms

Execution Time: 0.207 ms

****

1. **Уменьшение выполнения времени 2 запроса:**

Время выполнения исходного запроса: 626 мс





Вербовка???

**Индексы, добавления которых уменьшит время выполнения запроса:**

1. **На таблице Н\_ЛЮДИ:**

Индекс на атрибуте Н\_ЛЮДИ.ИД (B-tree) ускорит фильтрацию с условием «Н\_ЛЮДИ.ИД > 152862».

1. **На таблице Н\_ОБУЧЕНИЯ:**

Индекс на атрибуте Н\_ОБУЧЕНИЯ.НЗК (B-tree) ускорит фильтрацию с условием «Н\_ОБУЧЕНИЯ.НЗК < 933232».

1. **На таблице Н\_УЧЕНИКИ:**

Возможно добавление индекса на атрибуте Н\_УЧЕНИКИ.ЧЛВК\_ИД (B-tree), чтобы ускорить соединение таблиц по атрибуту ЧЛВК\_ИД, но это вряд ли будет способствовать уменьшению времени выполнения запроса и

Добавление индексов на указанные столбцы позволит ускорить поиск нужных записей по фильтрам в запросе.

**Возможные планы выполнения запроса без индексов:**

**План 1:**

* Полный скан таблицы Н\_ЛЮДИ с применением фильтра по условию «Н\_ЛЮДИ.ИД > 152862».
* Полный скан таблицы Н\_ОБУЧЕНИЯ с применением фильтра по условию «Н\_ОБУЧЕНИЯ.НЗК < 933232».
* Соединение таблиц с использованием Nested Loops Join по атрибуту ЧЛВК\_ИД.
* Полный скан таблицы Н\_УЧЕНИКИ.
* Соединение таблиц с использованием Nested Loops Join по атрибуту ЧЛВК\_ИД.

**План 2:**

* Полный скан таблицы Н\_ЛЮДИ с применением фильтра по условию «Н\_ЛЮДИ.ИД > 152862».
* Полный скан таблицы Н\_ОБУЧЕНИЯ с применением фильтра по условию «Н\_ОБУЧЕНИЯ.НЗК < 933232».
* Соединение таблиц с использованием Hash Join по атрибуту ЧЛВК\_ИД.
* Полный скан таблицы Н\_УЧЕНИКИ.
* Соединение таблиц с использованием Hash Join по атрибуту ЧЛВК\_ИД.

**Оптимальный план:** первый, так как использование Nested Loops Join предпочтительнее по сравнению в Hash Join-ом в случаях, когда обрабатываемые наборы данных небольшие, и когда заранее известно, что результаты фильтрации будут содержать мало строк.

**При добавлении индексов планы выполнения запросов изменятся:**

* Вместо полного скана таблиц будет использоваться индексный скан.
* Nested Loops Join станет быстрее благодаря индексу на атрибутам Н\_ Heap Blocks: exact=44 ОБУЧЕНИЯ.ЧЛВК\_ИД и Н\_УЧЕНИКИ.ЧЛВК\_ИД.

1. **План выполнения 2 запроса:**

Nested Loop Left Join (cost=139.39..864.07 rows=2514 width=28) (actual time=1.139..5.935 rows=2190 loops=1)

-> Hash Join (cost=139.10..267.66 rows=535 width=20) (actual time=1.115..2.722 rows=803 loops=1)

Hash Cond: ("Н\_ОБУЧЕНИЯ"."ЧЛВК\_ИД" = "Н\_ЛЮДИ"."ИД")

-> Seq Scan on "Н\_ОБУЧЕНИЯ" (cost=0.00..119.76 rows=3346 width=4) (actual time=0.013..1.627 rows=3347 loops=1)

Filter: (("НЗК")::text < '933232'::text)

Rows Removed by Filter: 1674

-> Hash (cost=128.87..128.87 rows=819 width=20) (actual time=0.486..0.487 rows=827 loops=1)

Buckets: 1024 Batches: 1 Memory Usage: 52kB

-> Bitmap Heap Scan on "Н\_ЛЮДИ" (cost=18.63..128.87 rows=819 width=20) (actual time=0.079..0.324 rows=827 loops=1)

Recheck Cond: ("ИД" > 152862)

-> Bitmap Index Scan on "ЧЛВК\_PK" (cost=0.00..18.42 rows=819 width=0) (actual time=0.052..0.052 rows=827 loops=1)

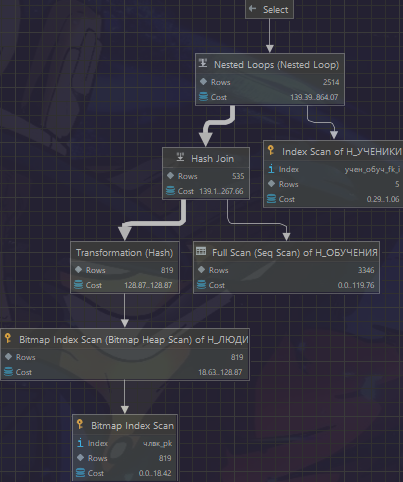
Index Cond: ("ИД" > 152862)

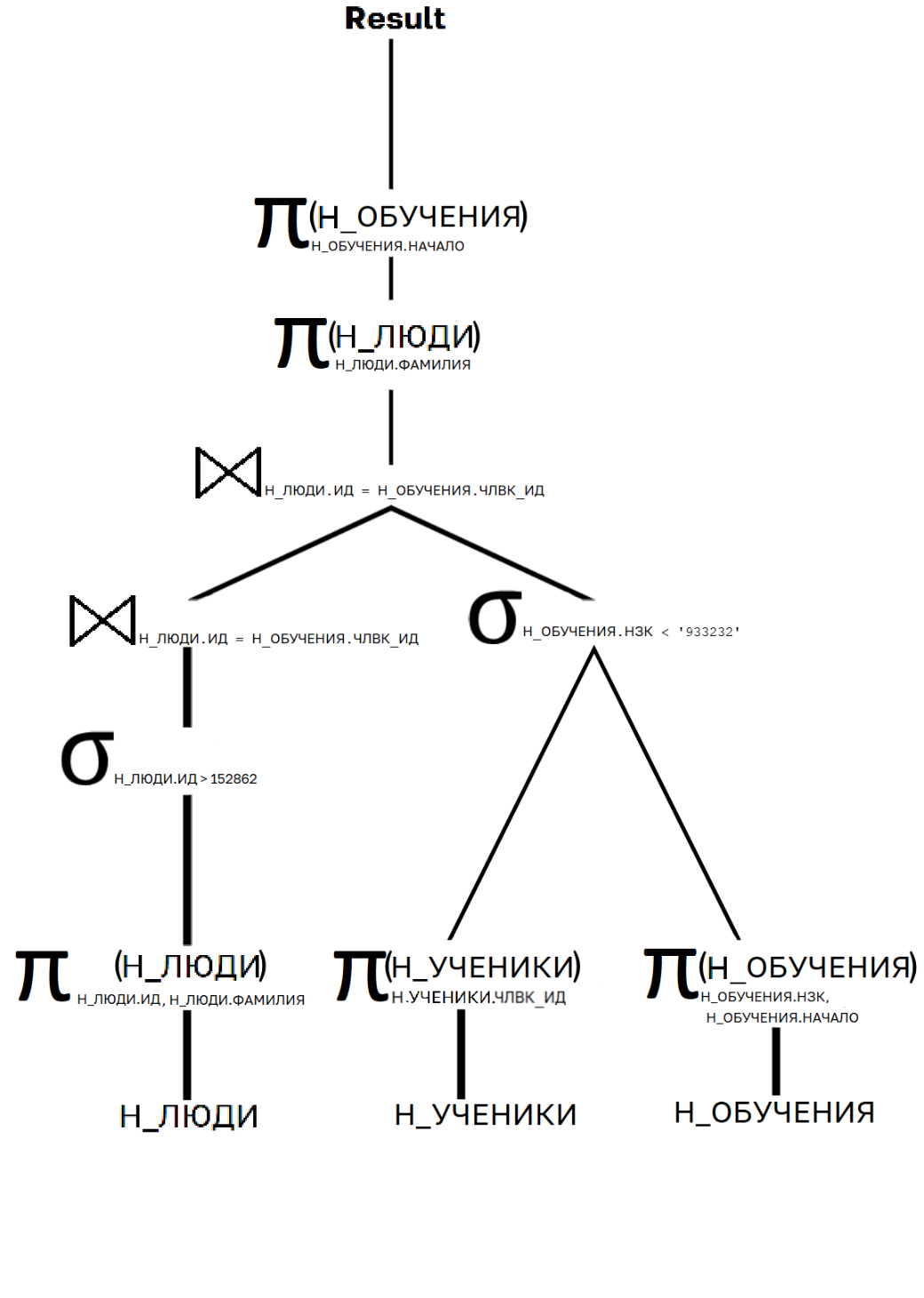
-> Index Scan using "УЧЕН\_ОБУЧ\_FK\_I" on "Н\_УЧЕНИКИ" (cost=0.29..1.06 rows=5 width=12) (actual time=0.002..0.003 rows=3 loops=803)

Index Cond: ("ЧЛВК\_ИД" = "Н\_ОБУЧЕНИЯ"."ЧЛВК\_ИД")

Planning Time: 1.391 ms

Execution Time: 6.140 ms





1. **Вывод:**

В ходе выполнения лабораторной работы я освоил тему реляционной алгебры, планы выполнения запроса и составления деревьев планирования. Также изучил индексы в pgSQL, который является одним из двух методов оптимизации скорости выполнения запроса. В целом, здесь заканчивается курс базы данных, но кроме PostgreSQL-а есть множество других тип базы данных, поэтому обязательно буду заниматься их изучением.