

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования

«Национальный исследовательский университет ИТМО»

Факультет Программной Инженерии и Компьютерной Техники

Дисциплина: Информационные системы и базы данных

Лабораторная работа 4

Вариант 3452

Выполнил:

Гурьянов Кирилл Алексеевич

Группа: Р33302

Преподаватель:

Николаев Владимир Вячеславович

Санкт-Петербург

2023

Текст задания

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.

Изменяются ли планы при добавлении индекса и как?

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

Подробные ответы на все вышеперечисленные вопросы должны присутствовать в отчете (планы выполнения запросов должны быть нарисованы, ответы на вопросы - представлены в текстовом виде).

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

Таблицы: Н_ЛЮДИ, Н_СЕССИЯ.

Вывести атрибуты: Н_ЛЮДИ.ОТЧЕСТВО, Н_СЕССИЯ.УЧГОД.

Фильтры (AND):

а) Н_ЛЮДИ.ИД = 142095.

б) Н_СЕССИЯ.ЧЛВК_ИД > 100012.

Вид соединения: LEFT JOIN.

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

Таблицы: Н_ЛЮДИ, Н_ВЕДОМОСТИ, Н_СЕССИЯ.

Вывести атрибуты: Н_ЛЮДИ.ФАМИЛИЯ, Н_ВЕДОМОСТИ.ИД, Н_СЕССИЯ.ЧЛВК_ИД.

Фильтры (AND):

а) Н_ЛЮДИ.ИД > 163484.

б) Н_ВЕДОМОСТИ.ЧЛВК_ИД < 105590.

с) Н_СЕССИЯ.УЧГОД = 2008/2009.

Вид соединения: LEFT JOIN.

Запрос 1.

```
SELECT "Н_ЛЮДИ"."ОТЧЕСТВО", "Н_СЕССИЯ"."УЧГОД"  
FROM "Н_ЛЮДИ"  
LEFT JOIN "Н_СЕССИЯ" ON "Н_ЛЮДИ"."ИД" = "Н_СЕССИЯ"."ЧЛВК_ИД" WHERE  
"Н_ЛЮДИ"."ИД" = 142095 AND "Н_СЕССИЯ"."ЧЛВК_ИД" > 100012;
```

Индексы

Hash индекс для атрибуту ИД в таблице Н_ЛЮДИ позволит эффективно находить строки, где ИД равен определенному значению.

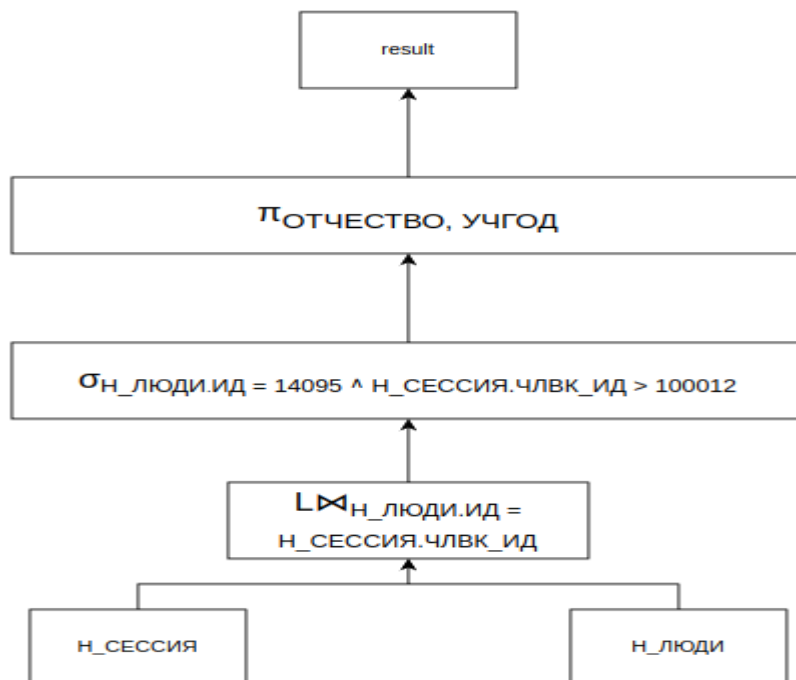
```
CREATE INDEX "Н_ЛЮДИ_ИД_ИНД" ON "Н_ЛЮДИ" USING HASH("ИД");
```

Btree индекс на атрибут ЧЛВК_ИД позволит эффективно находить строчки, где ЧЛВК_ИД сравнивается с каким-то значением.

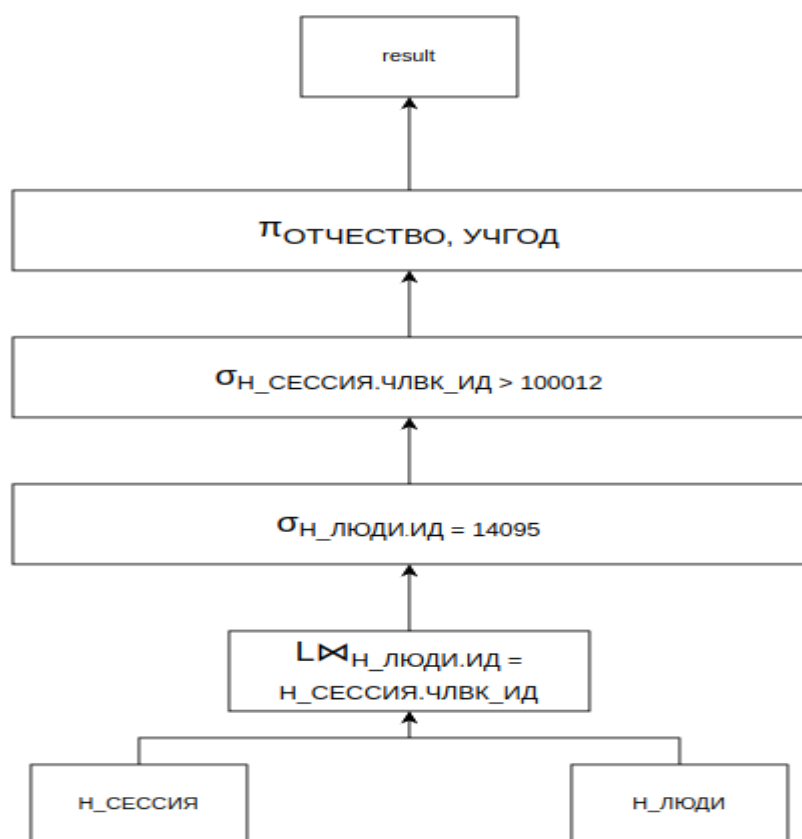
```
CREATE INDEX "Н_СЕССИЯ_ЧВКЛ_ИД_ИНД" ON "Н_СЕССИЯ" ("ИД");
```

Планы выполнения запросов

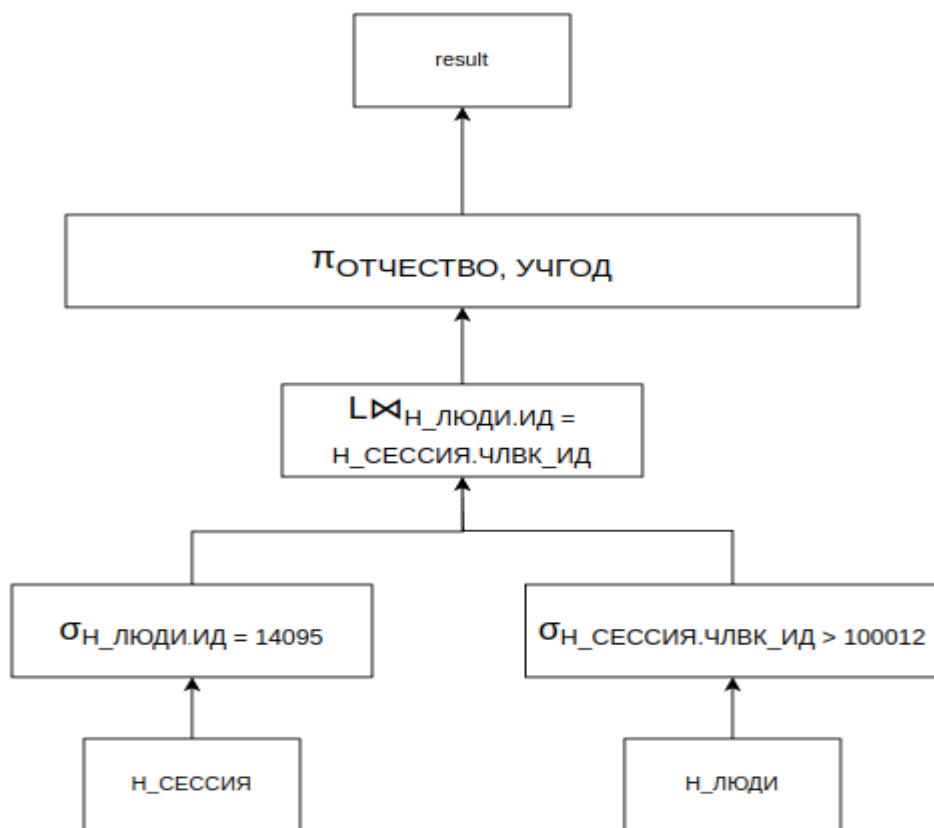
1)



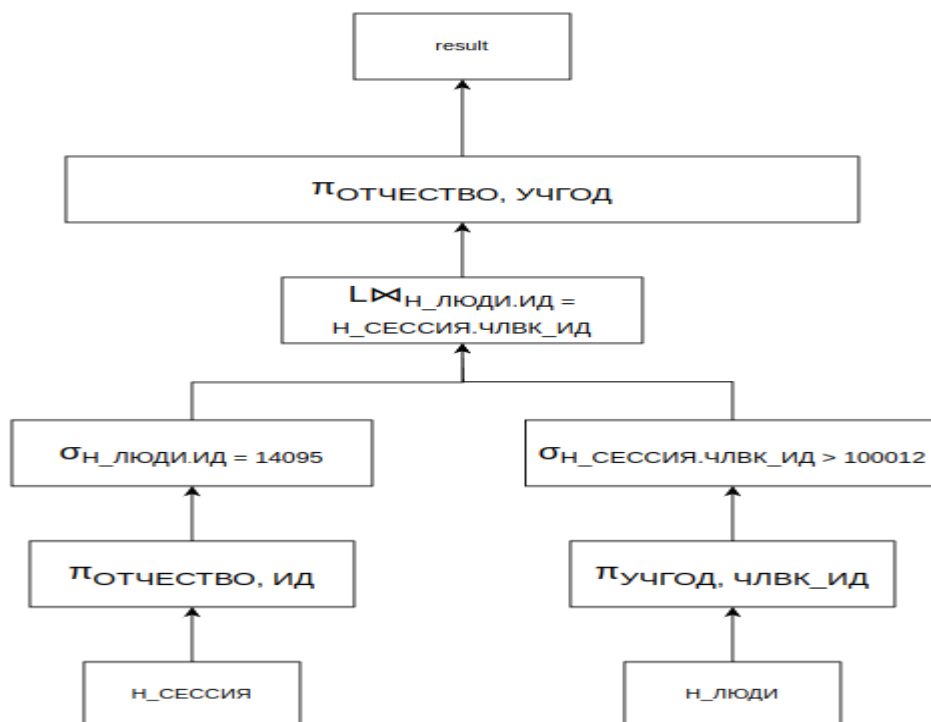
2)



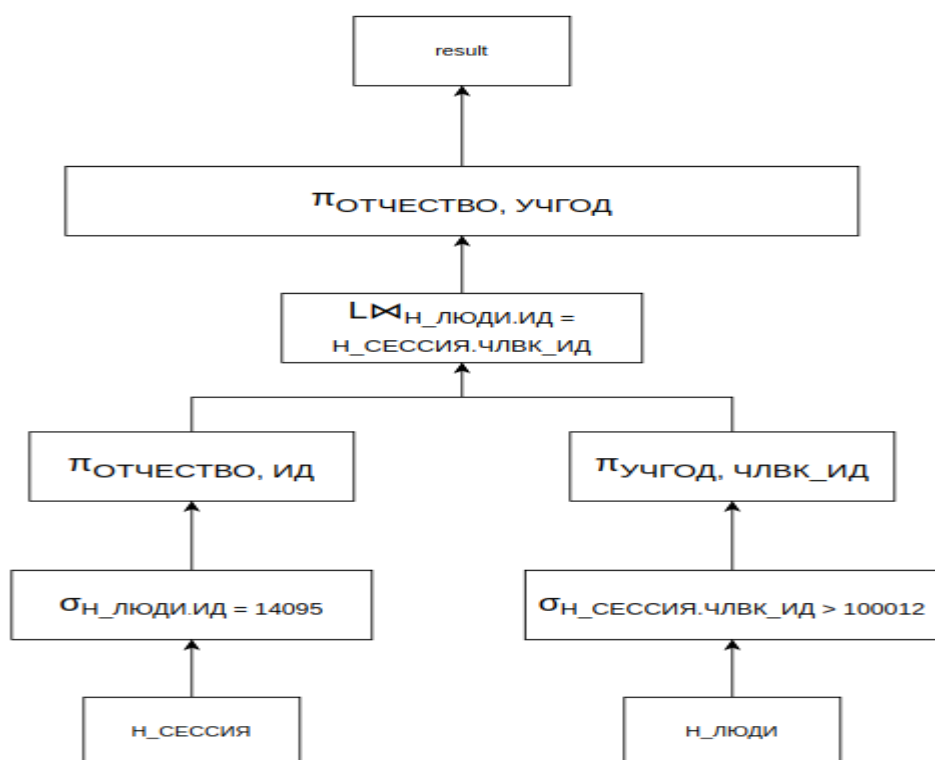
3)



4)



5)



Выбор оптимального плана

Исходя из составленных планов наиболее оптимальным кажется план 5, потому что сначала мы производим выборку строк (в таблице Н_СЕССИЯ выборка в целом происходит на равенство, что значительно уменьшает количество потенциальных строк для дальнейшей обработки). После этого стоит найти проекции по атрибутам, необходимым для соединения, а также дальнейшего вывода. В результате мы уменьшаем размеры промежуточных данных. После этого происходит соединение таблиц, которое уже не содержит полный список атрибутов обеих таблиц.

Добавление индексов приведет к более эффективной выборке строк по условиям =. При использовании индексов у нас не будет полного перебора всех строк, т.к. высокая селективность. Мы можем видеть, что под условие выборки ИД = 142095 подходит всего лишь одна строка, в то время как в целом в таблице Н_ЛЮДИ имеется 5118 записей.

При данном запросе индекс на ЧЛВК_ИД в таблице Н_СЕССИЯ не приведет к ускорению, т.к. очень низкая селективность. Мы можем видеть, что под условие ЧЛВК_ИД > 100012 подходит 3233 строк, в то время как всего в таблице Н_СЕССИЯ имеется 3752 записей. Однако индекс на ЧЛВК_ИД может быть полезен при ином значении ЧЛВК_ИД для сравнения (т.е. в ситуации, когда у нас будет большая селективность).

```
ucheb=> SELECT COUNT(*) FROM Н_ЛЮДИ WHERE ИД = 142095;
count
-----
      1
(1 строка)

ucheb=> SELECT COUNT(*) FROM Н_ЛЮДИ;
count
-----
    5118
(1 строка)

ucheb=> SELECT COUNT(*) FROM Н_СЕССИЯ WHERE ЧЛВК_ИД > 100012;
count
-----
    3233
(1 строка)

ucheb=> SELECT COUNT(*) FROM Н_СЕССИЯ;
count
-----
    3752
(1 строка)
```

Explain analyze

```
ucheb=> EXPLAIN ANALYZE SELECT "Н_ЛЮДИ"."ОТЧЕСТВО", "Н_СЕССИЯ"."УЧГОД"
FROM "Н_ЛЮДИ"
LEFT JOIN "Н_СЕССИЯ" ON "Н_ЛЮДИ"."ИД" = "Н_СЕССИЯ"."ЧЛВК_ИД" WHERE "Н_ЛЮДИ"."ИД" = 142095 AND "Н_СЕССИЯ"."ЧЛВК_ИД" > 100012;
QUERY PLAN
-----
Nested Loop  (cost=4.58..19.64 rows=2 width=30) (actual time=0.022..0.023 rows=0 loops=1)
-> Index Scan using "ЧЛВК_PK" on "Н_ЛЮДИ"  (cost=0.28..8.30 rows=1 width=24) (actual time=0.011..0.012 rows=1 loops=1)
    Index Cond: ("ИД" = 142095)
-> Bitmap Heap Scan on "Н_СЕССИЯ"  (cost=4.30..11.32 rows=2 width=14) (actual time=0.006..0.007 rows=0 loops=1)
    Recheck Cond: (("ЧЛВК_ИД" > 100012) AND ("ЧЛВК_ИД" = 142095))
    -> Bitmap Index Scan on "SYS_C003500_IFK" (cost=0.00..4.30 rows=2 width=0) (actual time=0.004..0.005 rows=0 loops=1)
        Index Cond: (("ЧЛВК_ИД" > 100012) AND ("ЧЛВК_ИД" = 142095))
Planning Time: 0.163 ms
Execution Time: 0.051 ms
(9 строк)
```

Запрос 2.

```
SELECT "Н_ЛЮДИ"."ФАМИЛИЯ", "Н_ВЕДОМОСТИ"."ИД", "Н_СЕССИЯ"."ЧЛВК_ИД"
FROM "Н_ЛЮДИ"
LEFT JOIN "Н_ВЕДОМОСТИ" ON "Н_ЛЮДИ"."ИД" = "Н_ВЕДОМОСТИ"."ЧЛВК_ИД"
LEFT JOIN "Н_СЕССИЯ" ON "Н_ЛЮДИ"."ИД" = "Н_СЕССИЯ"."ЧЛВК_ИД"
WHERE "Н_ЛЮДИ"."ИД" > 163484 AND "Н_ВЕДОМОСТИ"."ЧЛВК_ИД" < 105590 AND
"Н_СЕССИЯ"."УЧГОД" = '2008/2009';
```

Индексы

Btree индекс для атрибута ИД в таблице Н_ЛЮДИ для эффективной выборки строк, в которых ИД > 163484.

```
CREATE INDEX "Н_ЛЮДИ_ИД_ИНД" ON "Н_ЛЮДИ" ("ИД");
```

Btree индекс для атрибута ЧЛВК_ИД в таблице Н_СЕССИЯ для эффективной выборки строк, в которых ЧЛВК_ИД < 105590.

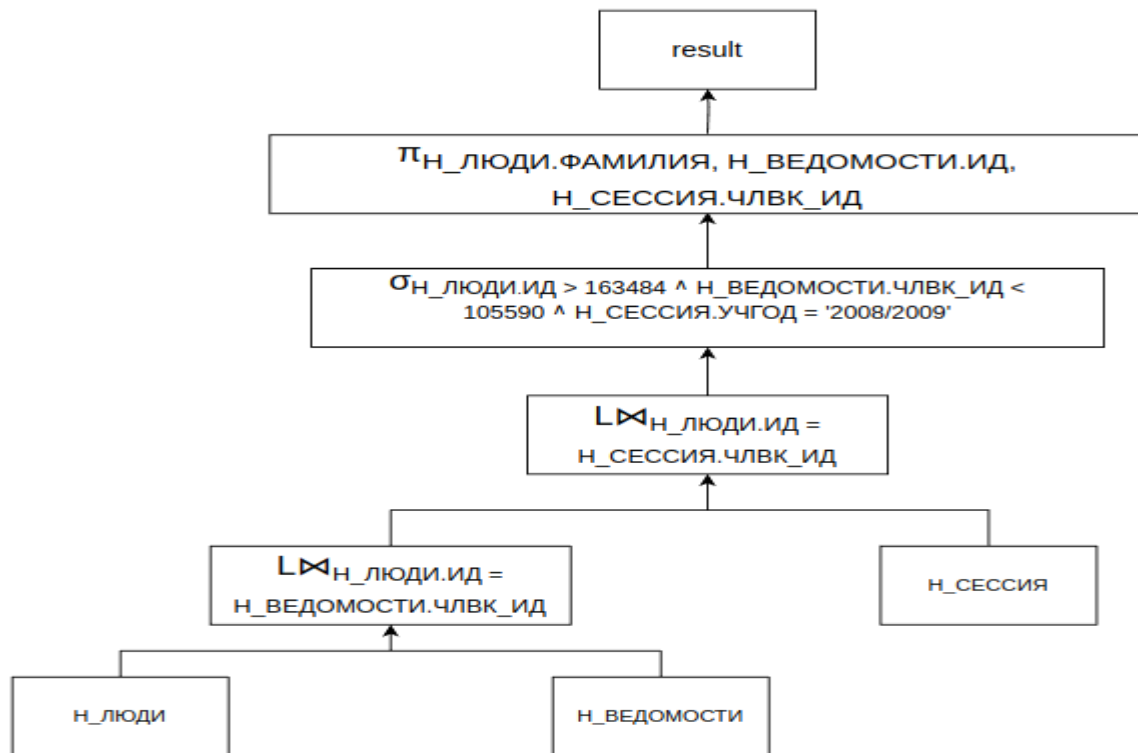
```
CREATE INDEX "Н_СЕССИЯ_ЧЛВК_ИД" ON "Н_СЕССИЯ" ("ЧЛВК_ИД");
```

Hash индекс для атрибута УЧГОД в таблице Н_СЕССИЯ для эффективной выборки строк, в которых УЧГОД = 2008/2009.

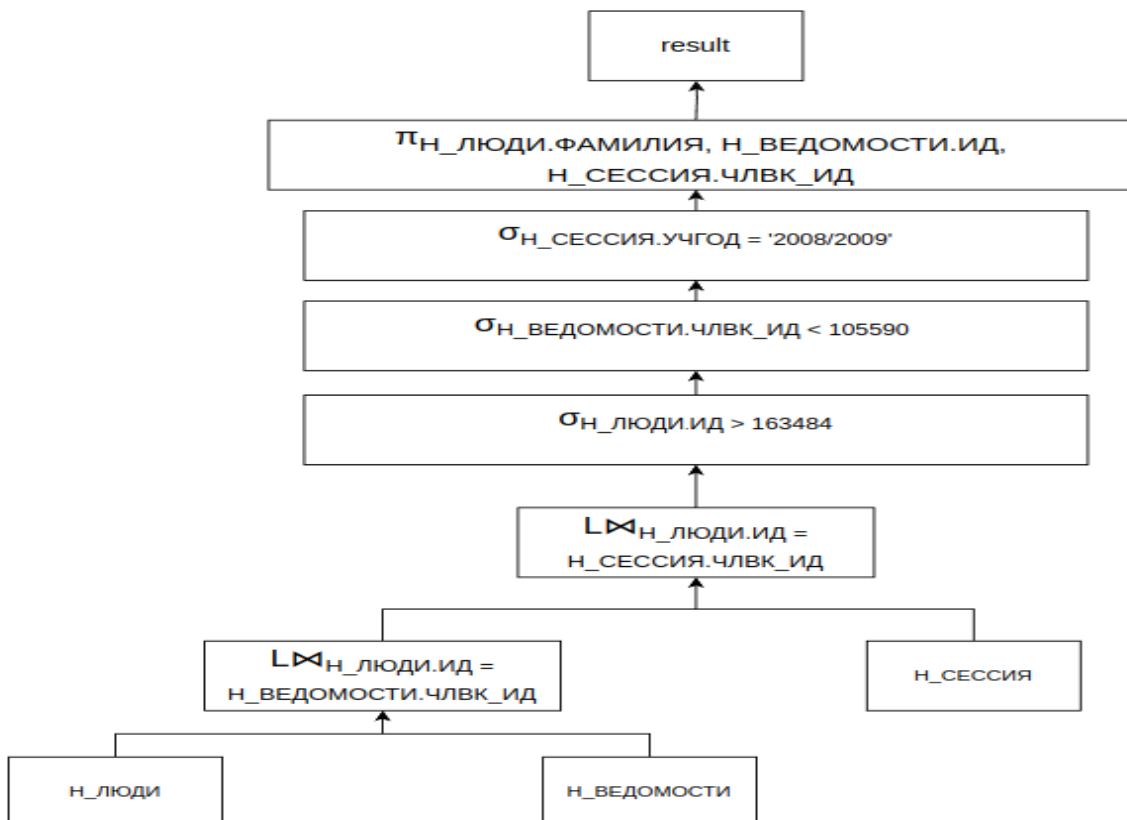
```
CREATE INDEX "Н_СЕССИЯ_ИНД" ON "Н_СЕССИЯ" USING HASH("УЧГОД");
```

Планы выполнения запросов

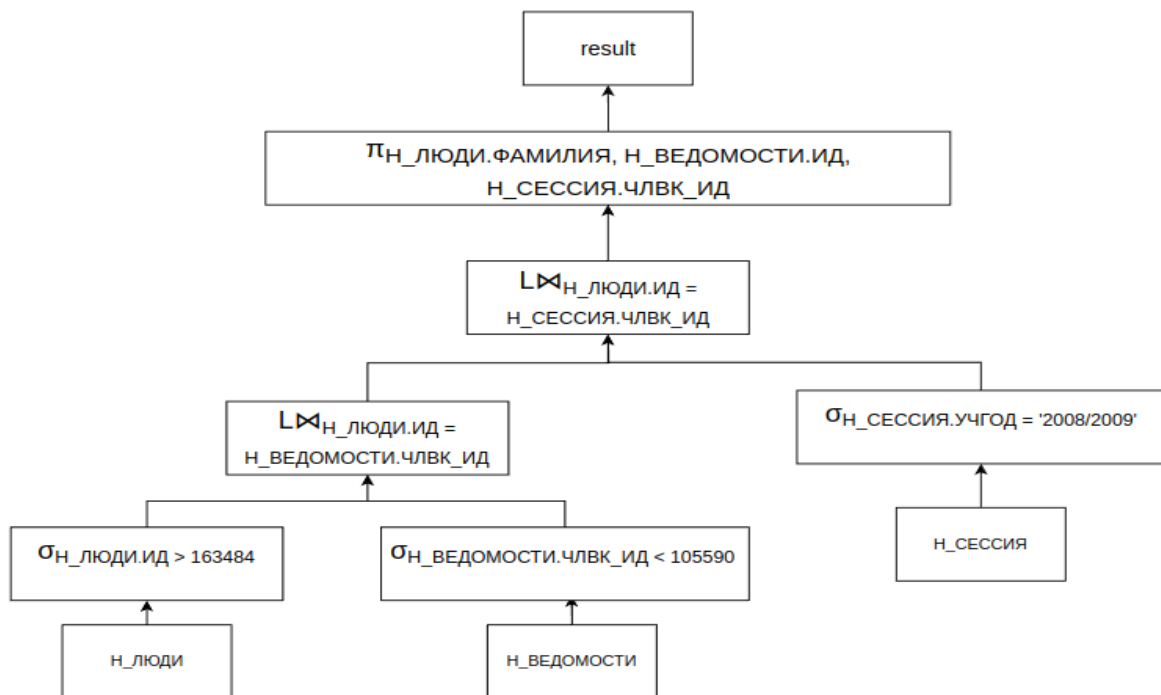
1)



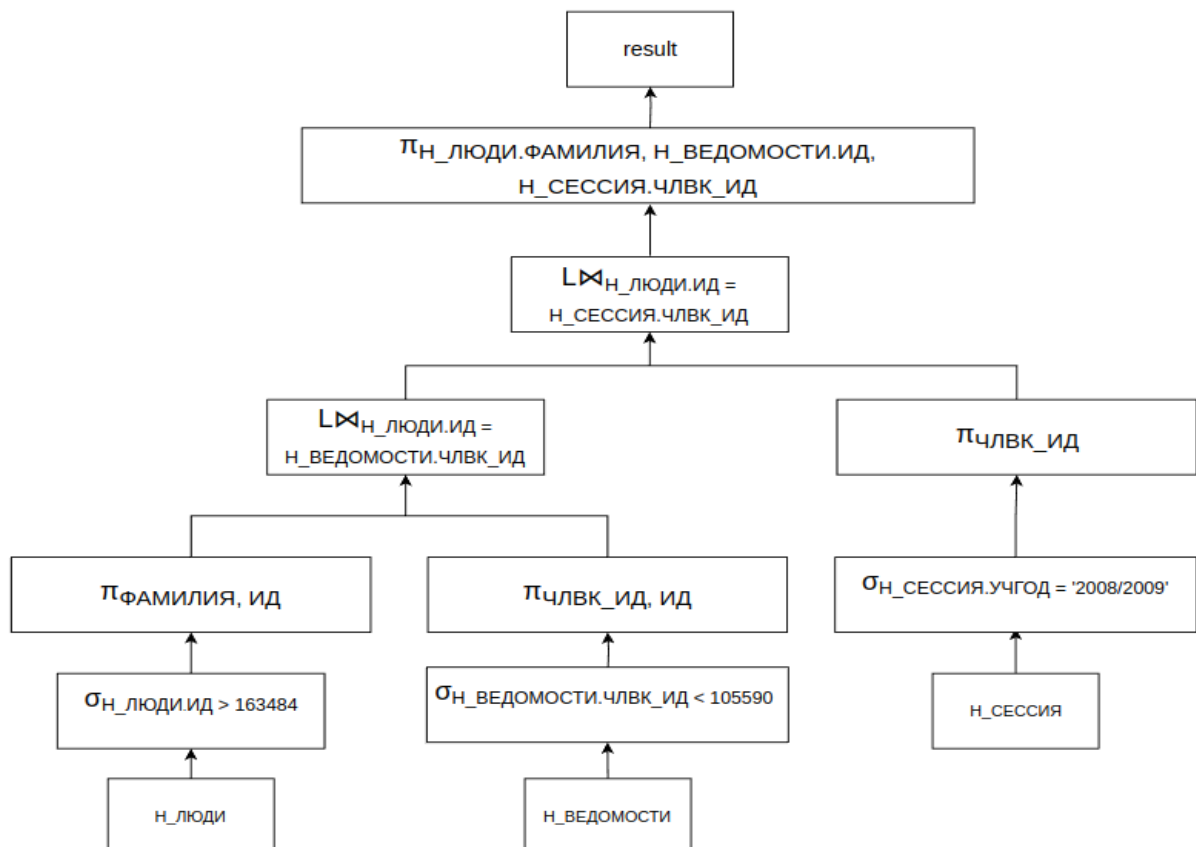
2)



3)



4)



Выбор оптимального плана

Оптимальным планом является план 4. За счет применения выборки на нижнем уровне плана, мы отсекаем, в теории, довольно большое количество строк (в нашем примере мы в целом отсекаем все строки, т.к. строки, удовлетворяющие условиям $ИД > 163484$ и $ЧЛВК_ИД < 105590$ в принципе отсутствуют). После чего уменьшаем объем данных за счет проекции и уже относительно небольшой (относительно первоначального объема) объем данных соединяем через LEFT JOIN. Аналогично происходит с таблицей Н_ВЕДОМОСТИ, после чего все данные соединяются в последнем LEFT JOIN (самый верхний уровень на плане запросов).

Добавление индексов приведет к более эффективной выборке строк по условиям $>$. При использовании индексов у нас, возможно, не будет полного перебора всех строк (если под условие выборки подходит небольшое количество записей в БД, т.е. высокая селективность). В нашем запросе, к сожалению, количество удовлетворяющих строк под условия сравнения ИД в таблицах Н_ВЕДОМОСТИ и Н_ЛЮДИ равно 0. При выборке УЧГОД из Н_СЕССИЯ индексы могут оказаться эффективными, если в таблице будет достаточное количество записей, чтобы оптимизатор решил использовать индексы. А также было небольшое количество подходящих под условие данных. При запросах на выборку строк с $УЧГОД = '2008/2009'$ индексы не применяются по причине низкой селективности. Хотя количество строчек, подходящее под условие $= '2008/2009'$ относительно невелико.

```

uscheb=> SELECT COUNT(*) FROM Н_ЛЮДИ WHERE ИД > 163484;
count
-----
0
(1 строка)

uscheb=> SELECT COUNT(*) FROM Н_ВЕДОМОСТИ WHERE ЧЛВК_ИД <105590;
count
-----
0
(1 строка)

uscheb=> SELECT УЧГОД, COUNT(УЧГОД) FROM Н_СЕССИЯ GROUP BY УЧГОД;
УЧГОД | count
-----+-----
2005/2006 | 347
2010/2011 | 500
2006/2007 | 412
2003/2004 | 310
2001/2002 | 122
2004/2005 | 381
2008/2009 | 424
2011/2012 | 209
2002/2003 | 236
2009/2010 | 437
2007/2008 | 374
(11 строк)

uscheb=> SELECT COUNT(*) FROM Н_ЛЮДИ;
count
-----
5118
(1 строка)

uscheb=> SELECT COUNT(*) FROM Н_ВЕДОМОСТИ;
count
-----
222440
(1 строка)

```

```

uscheb=> EXPLAIN ANALYZE SELECT * FROM Н_СЕССИЯ WHERE УЧГОД = '2008/2009';
QUERY PLAN

-----
Seq Scan on "Н_СЕССИЯ" (cost=0.00..117.90 rows=424 width=106) (actual time=0.043..0.671 rows=424 loops=1)
  Filter: (("УЧГОД")::text = '2008/2009'::text)
  Rows Removed by Filter: 3328
Planning Time: 0.377 ms
Execution Time: 0.712 ms
(5 строк)

```

```

ucheb=> EXPLAIN ANALYZE SELECT COUNT(*) FROM Н_ЛЮДИ WHERE ИД > 163484;
                                         QUERY PLAN

-----
Aggregate  (cost=8.30..8.31 rows=1 width=8) (actual time=0.011..0.012 rows=1 loops=1)
->  Index Only Scan using "ЧЛВК_РК" on "Н_ЛЮДИ"  (cost=0.28..8.29 rows=1 width=0) (actual
time=0.006..0.007 rows=0 loops=1)
      Index Cond: ("ИД" > 163484)
      Heap Fetches: 0
Planning Time: 0.190 ms
Execution Time: 0.044 ms
(6 строк)

ucheb=> EXPLAIN ANALYZE SELECT COUNT(*) FROM Н_ВЕДОМОСТИ WHERE ЧЛВК_ИД <105590;
                                         QUERY PLAN

-----
Aggregate  (cost=7.28..7.29 rows=1 width=8) (actual time=0.011..0.012 rows=1 loops=1)
->  Index Only Scan using "ВЕД_ЧЛВК_FK_IFK" on "Н_ВЕДОМОСТИ"  (cost=0.29..7.28 rows=1 width=0) (actual time=0.005..0.005 rows=0 loops=1)
      Index Cond: ("ЧЛВК_ИД" < 105590)
      Heap Fetches: 0
Planning Time: 0.195 ms
Execution Time: 0.051 ms
(6 строк)

```

Explain analyze

```

ucheb=> EXPLAIN ANALYZE SELECT "Н_ЛЮДИ"."ФАМИЛИЯ", "Н_ВЕДОМОСТИ"."ИД", "Н_СЕССИЯ"."ЧЛВК_ИД" FROM "Н_ЛЮДИ"
LEFT JOIN "Н_ВЕДОМОСТИ" ON "Н_ЛЮДИ"."ИД" = "Н_ВЕДОМОСТИ"."ЧЛВК_ИД" LEFT JOIN "Н_СЕССИЯ" ON "Н_ЛЮДИ"."ИД" = "Н_СЕССИЯ"."ЧЛВК_ИД"
WHERE "Н_ЛЮДИ"."ИД" > 163484 AND "Н_ВЕДОМОСТИ"."ЧЛВК_ИД" < 105590 AND "Н_СЕССИЯ"."УЧГОД" = '2008/2009';
                                         QUERY PLAN

-----
Nested Loop  (cost=4.99..54.60 rows=1 width=24) (actual time=0.005..0.006 rows=0 loops=1)
  Join Filter: ("Н_ВЕДОМОСТИ"."ЧЛВК_ИД" = "Н_ЛЮДИ"."ИД")
->  Nested Loop  (cost=4.71..53.20 rows=1 width=12) (actual time=0.005..0.005 rows=0 loops=1)
      ->  Index Scan using "ВЕД_ЧЛВК_FK_IFK" on "Н_ВЕДОМОСТИ"  (cost=0.29..7.28 rows=1 width=8) (actual time=0.004..0.004 rows=0 loops=1)
          Index Cond: ("ЧЛВК_ИД" < 105590)
      ->  Bitmap Heap Scan on "Н_СЕССИЯ"  (cost=4.42..45.90 rows=2 width=4) (never executed)
          Recheck Cond: ("ЧЛВК_ИД" = "Н_ВЕДОМОСТИ"."ЧЛВК_ИД")
          Filter: (("УЧГОД")::text = '2008/2009'::text)
          ->  Bitmap Index Scan on "SYS_C003500_IFK"  (cost=0.00..4.42 rows=18 width=0) (never executed)
              Index Cond: ("ЧЛВК_ИД" = "Н_ВЕДОМОСТИ"."ЧЛВК_ИД")
      ->  Index Scan using "ЧЛВК_РК" on "Н_ЛЮДИ"  (cost=0.28..1.40 rows=1 width=20) (never executed)
          Index Cond: (("ИД" = "Н_СЕССИЯ"."ЧЛВК_ИД") AND ("ИД" > 163484))
Planning Time: 1.626 ms
Execution Time: 0.061 ms
(14 строк)

```