

ДЗ 2. Оптимизация запросов

- Для генерации написана команда и генератор анкет. Для запуска сидов нужно выполнить:
\$ make seed
- После генерации имеем базу с русскими именами содержащую более миллиона записей:

PostgreSQL » postgres » social_network » public » Select: users

Select: users

Select data Show structure Alter table New item

Select Search Sort Limit 20 Text length 100 Action Select

SELECT * FROM "users" LIMIT 20 (0,000 s) Edit

	id	name	surname	email	birthday	biography
<input type="checkbox"/> Modify	d09703ca-660a-11ee-ab5a-0242ac130003	Эвелина	Лазарева	Lazareva.Evelina1647@gmail.com	2003-08-17	Consequatur facilis vero reiciendis qui modi. Dignissimos error c
<input type="checkbox"/> edit	d0970686-660a-11ee-862d-0242ac130003	Игорь	Козлов	Kozlov.Igor1630@gmail.com	1972-06-02	Cupiditate rerum aliquam asperiores sed officiis. Fugit cumque c
<input type="checkbox"/> edit	d0971040-660a-11ee-9168-0242ac130003	Созон	Бычков	Bychkov.Sozon1610@gmail.com	1990-03-30	Magnam ratione voluptatum consequuntur voluptate et. Officiis c
<input type="checkbox"/> edit	d0971bb2-660a-11ee-87ef-0242ac130003	Ираида	Жвалова	Zhalova.Iraida1631@gmail.com	1975-06-07	Consectetur nemo magnam error sint tenetur. Quo at corporis la
<input type="checkbox"/> edit	d0972e7c-660a-11ee-9732-0242ac130003	Анаит	Парфенчикова	Parfencikova.Anait1649@gmail.com	1982-10-31	Iste aut voluptatibus architecto error est. Voluptas nam ut facere
<input type="checkbox"/> edit	d0972ea4-660a-11ee-a28e-0242ac130003	Снежана	Нарицина	Naricina.Snezana1637@gmail.com	2020-07-08	Consequatur incidunt placeat hic similique atque. Repellat repre
<input type="checkbox"/> edit	d097343a-660a-11ee-a072-0242ac130003	Ляйсан	Заволокина	Zavolokina.Ljaysan1619@gmail.com	2018-10-21	Modi vel maiores consequatur cumque nisi. Dicta ut in quaerat c
<input type="checkbox"/> edit	d09739da-660a-11ee-b9d0-0242ac130003	Юлиан	Плакидин	Plakidin.Julian1654@gmail.com	1977-05-28	Est perspicatis ut quo expedita sint. Voluptas qui sit praesentiur
<input type="checkbox"/> edit	d097461e-660a-11ee-87ef-0242ac130003	Камила	Остратова	Ostratova.Kamila1657@gmail.com	1985-08-19	Temporibus hic et eius asperiores laborum.
<input type="checkbox"/> edit	d09746c8-660a-11ee-af3d-0242ac130003	Поликарп	Белезянский	Belezjanskij.Polikarp1638@gmail.com	1984-12-02	Ea aspernatur cupiditate nostrum dicta soluta. Dolorum totam vr
<input type="checkbox"/> edit	d097483a-660a-11ee-a1b1-0242ac130003	Дания	Афанасова	Afanasova.Danija1639@gmail.com	2015-05-06	Accusantium voluptatibus non facilis rem commodi. Cumque ulli
<input type="checkbox"/> edit	d0974c9a-660a-11ee-9665-0242ac130003	Лев	Янковский	Jankovskij.Lev1636@gmail.com	2013-10-04	Adipisci est culpa dolore assumenda aut. Sit ut officiis cumque s
<input type="checkbox"/> edit	d0975b90-660a-11ee-91c6-0242ac130003	Далида	Демидова	Demidova.Dalida1643@gmail.com	1991-08-01	Et ut distinctio pariatu ex tenetur. Soluta est molestiae quia con
<input type="checkbox"/> edit	d09764aa-660a-11ee-87ef-0242ac130003	Руслана	Чернова	Černova.Ruslana1661@gmail.com	2009-08-29	Qui blanditiis sed modi fugit laboriosam. Maxime et consequatu
<input type="checkbox"/> edit	d0976cd4-660a-11ee-9665-0242ac130003	Иннокентий	Нефедьев	Nefed'ev.Innokentij1660@gmail.com	1993-01-07	Ut delectus non pariatu commodi amet.
<input type="checkbox"/> edit	d09777e2-660a-11ee-8b8d-0242ac130003	Юрий	Прашкевич	Praškevič.Jurij1662@gmail.com	1997-03-11	Molestias voluptatum qui non accusamus et. Qui dolorem lauda
<input type="checkbox"/> edit	d09782b4-660a-11ee-87ef-0242ac130003	Филат	Филиппов	Filippov.Filat1668@gmail.com	2010-11-13	Id et facere repellendus tenetur id. Eos natus excepturi ut a maç
<input type="checkbox"/> edit	d0978f2a-660a-11ee-8f76-0242ac130003	Родослав	Котов	Kotov.Rodoslav1646@gmail.com	1971-05-19	Omnis odit ut eligendi iure ea. Voluptas eveniet expedita et vitae
<input type="checkbox"/> edit	d097998e-660a-11ee-8b8d-0242ac130003	Гульнара	Амарантова	Amarantova.Gul'nara1667@gmail.com	2008-07-20	Unde dicta harum beatae dolores at. At dolor ullam possimus qu
<input type="checkbox"/> edit	d0996d4a-660a-11ee-a88a-0242ac130003	Стефания	Игнатьева	Ignat'eva.Stefanija1723@gmail.com	2009-05-10	Sunt consectetur aut rem voluptatem harum. Et quia atque prov

Load more data

Page 1 2 3 4 5 ... last Whole result ~ 1,274,140 rows Modify Save Selected (0) Edit Clone Delete Export (~ 1,274,140)

3. Для тестирования поиска используется endpoint метод принимает параметры **name** и **surname**. Условие хотя бы одно поле должно содержать значения. возвращает не более 100.

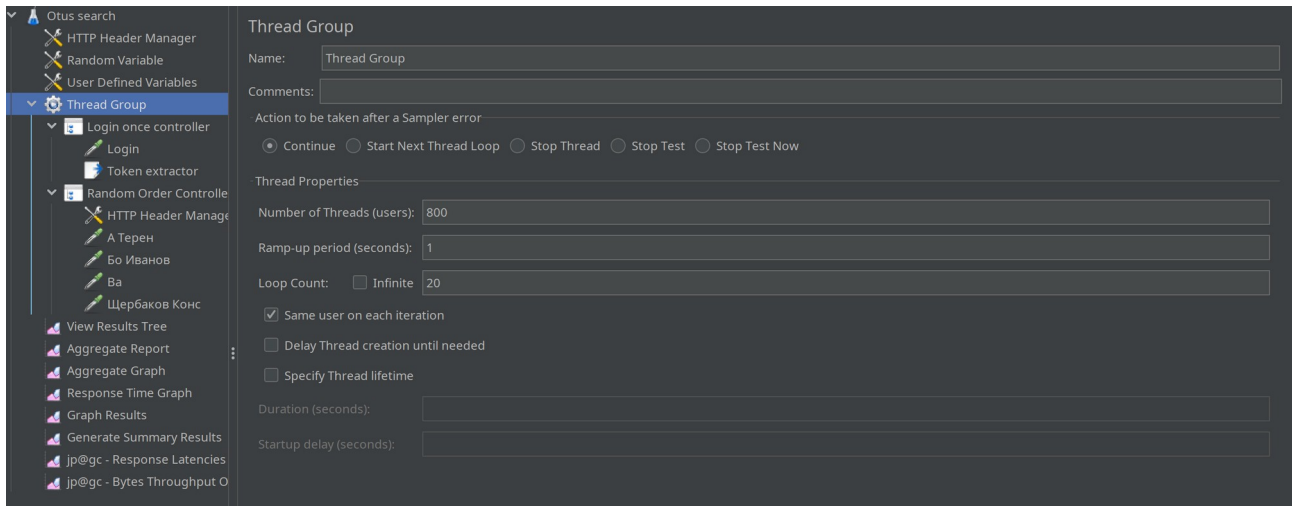
http://localhost:8080/users/search

Чтобы разнообразить выборку и кеш используются следующие параметры для запросов:

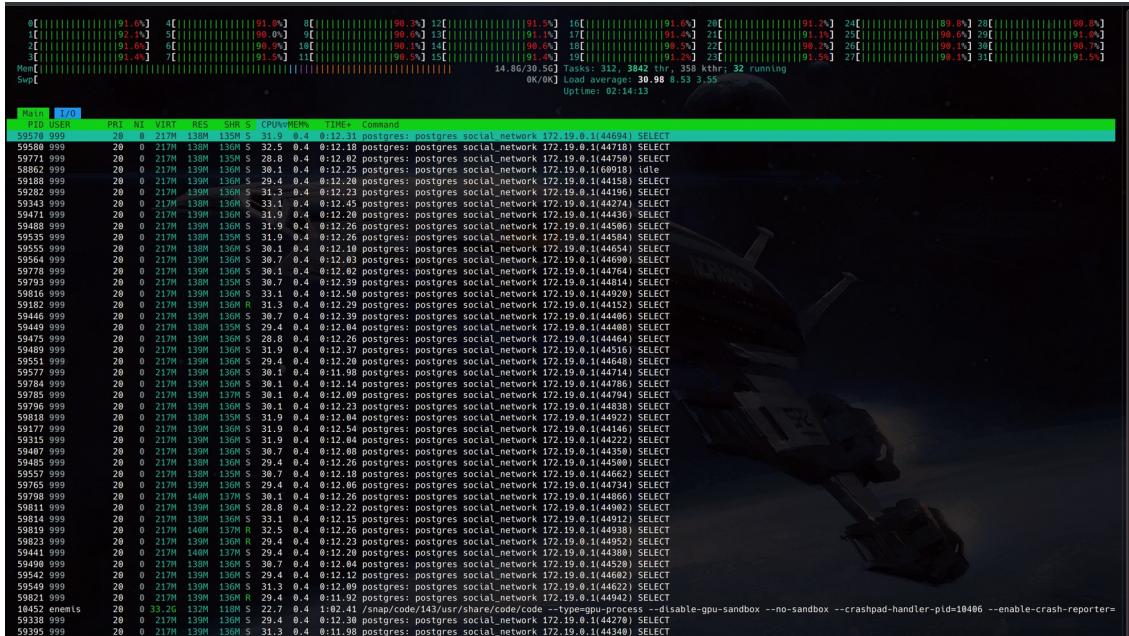
Параметры	Запрос который выполнится	Количество результатов
Name: A	"SELECT * FROM users WHERE (lower (name) LIKE 'A%' and lower (surname) LIKE 'Терен%') OR (lower (surname) LIKE 'A%' and lower (name) LIKE 'Терен%') ORDER BY id LIMIT 100;"	100+ пользователей
Surname: Терен	"SELECT * FROM users WHERE ((lower (name) LIKE 'Бо%' and (lower (surname) LIKE 'Иванов%') OR ((lower (surname) LIKE 'Бо%' and (lower (name) LIKE 'Иванов%') ORDER BY id LIMIT 100;"	5 пользователей
Name: Бо	"SELECT * FROM users WHERE (lower (name) LIKE 'Ba%' ORDER BY id LIMIT 100;"	100+ пользователей
Surname: Иванов		
Name: Ba		
Name: Щербаков	"SELECT * FROM users WHERE ((lower (surname) LIKE 'Конс%' and (lower (surname) LIKE 'Щербаков%') OR ((lower (surname) LIKE 'Конс%' and (lower (name) LIKE 'Щербаков%') ORDER BY id LIMIT 100;"	3 пользователя
Surname: Конс		

4. Проведение тестирования.

Postgres ограничена 80 одновременными соединениями. Экспериментальным путем обнаружено что при запуске 800 потоков начинает сильно расти latency. Поэтому выбраны такие настройки



заставляет компьютер немного поработать, видно как создаются процессы базы.



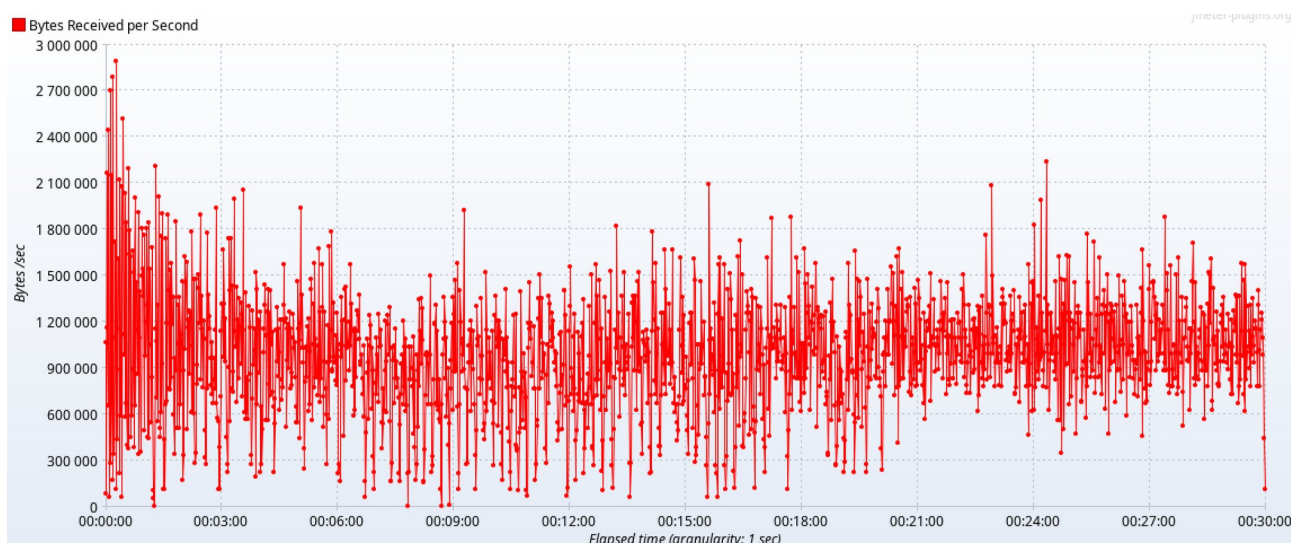
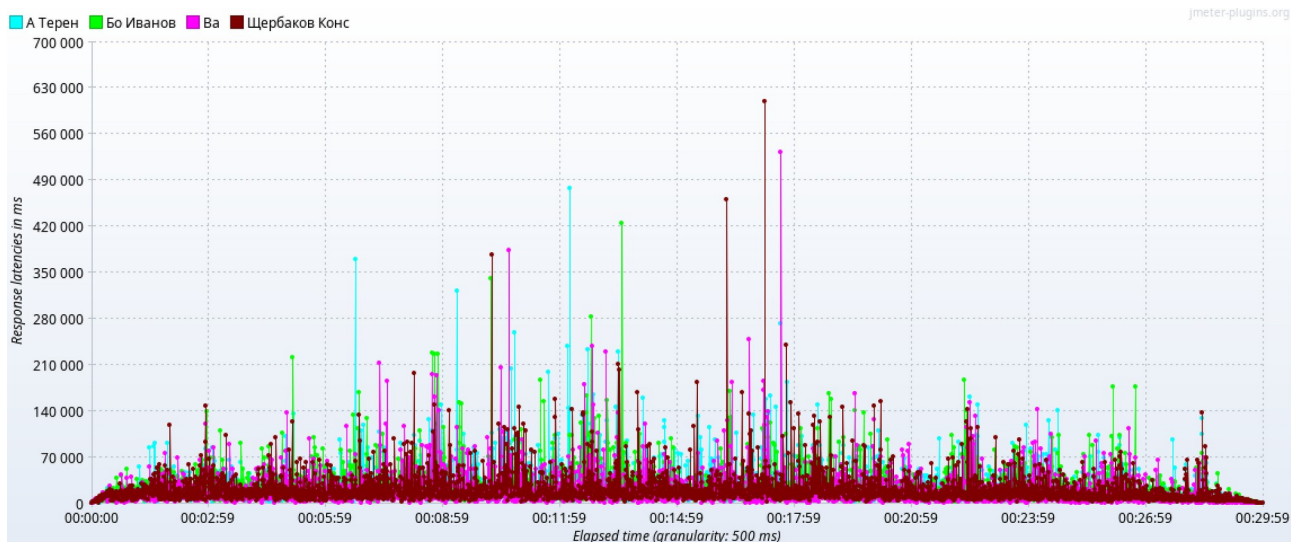
5. Отчет до добавления индекса

Label	# Samples	Average	Median	90% Line	95% Line	99% Line	Min	Max	Error %	Throughput	Received KB/sec	Sent KB/sec
Login	800	3448	1753	8810	13965	27491	0	57718	0,000%	13,75658	4,68	3,39
Ба	16000	17726	7354	35380	69202	194327	2	643160	0,000%	8,89181	447,69	3,58
Бо Иванов	16000	21044	10426	39575	72427	193940	173	633332	0,000%	8,89178	22,96	3,97
А Терен	16000	21860	10503	40886	79367	202918	176	649357	0,000%	8,89087	466,53	3,86
Щербаков Конс	16000	20553	10256	38092	70536	185340	165	609333	0,000%	8,89005	16,15	4,18
TOTAL	64800	20088	9621	38009	71867	194499	0	649357	0,000%	36,00466	953,34	15,69

explain

```
SELECT * FROM users WHERE (lower(name) LIKE 'Конс%' and lower(surname) LIKE 'Щербаков%') OR (lower(surname) LIKE 'Конс%' and lower(name) LIKE 'Щербаков%') ORDER BY id LIMIT 100
```

QUERY PLAN
Limit (cost=84084.45..84090.51 rows=52 width=392)
-> Gather Merge (cost=84084.45..84090.51 rows=52 width=392)
Workers Planned: 2
-> Sort (cost=83084.42..83084.49 rows=26 width=392)
Sort Key: id
-> Parallel Seq Scan on users (cost=0.00..83083.81 rows=26 width=392)
Filter: (((lower((name)::text) ~ 'Конс%::text) AND (lower((surname)::text) ~ 'Щербаков%::text)) OR (lower((surname)::text) ~ 'Конс%::text) AND (lower((name)::text) ~ 'Щербаков%::text)))



6. Сделать подходящий индекс.

```
CREATE INDEX CONCURRENTLY IF NOT EXISTS name_index ON users (lower((name)::text) varchar_pattern_ops);
CREATE INDEX CONCURRENTLY IF NOT EXISTS surname_index ON users (lower((surname)::text) varchar_pattern_ops);
```

7. Повторить нагрузочное тестирование

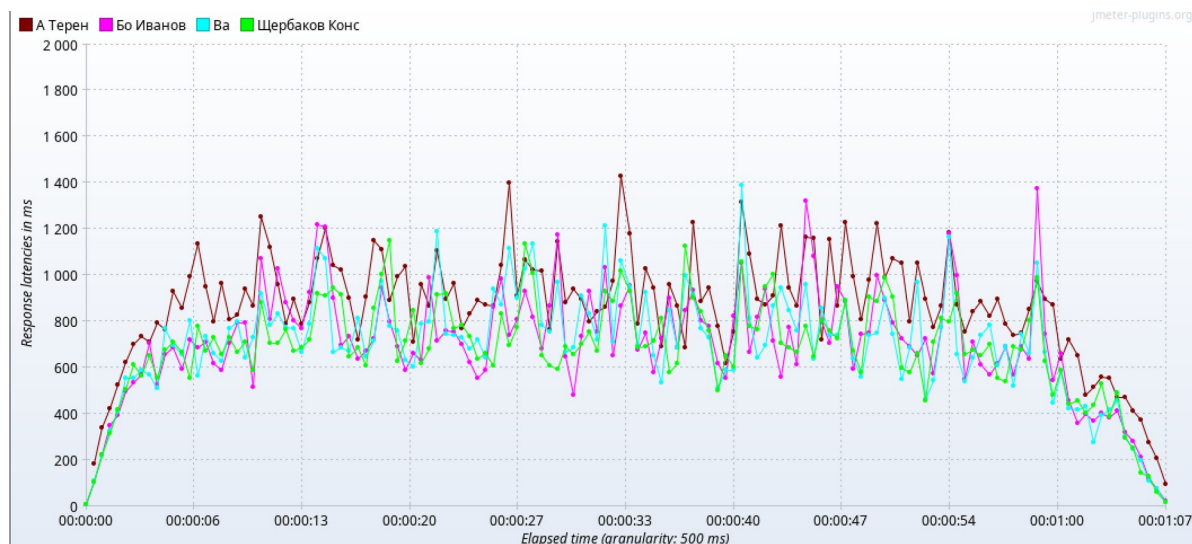
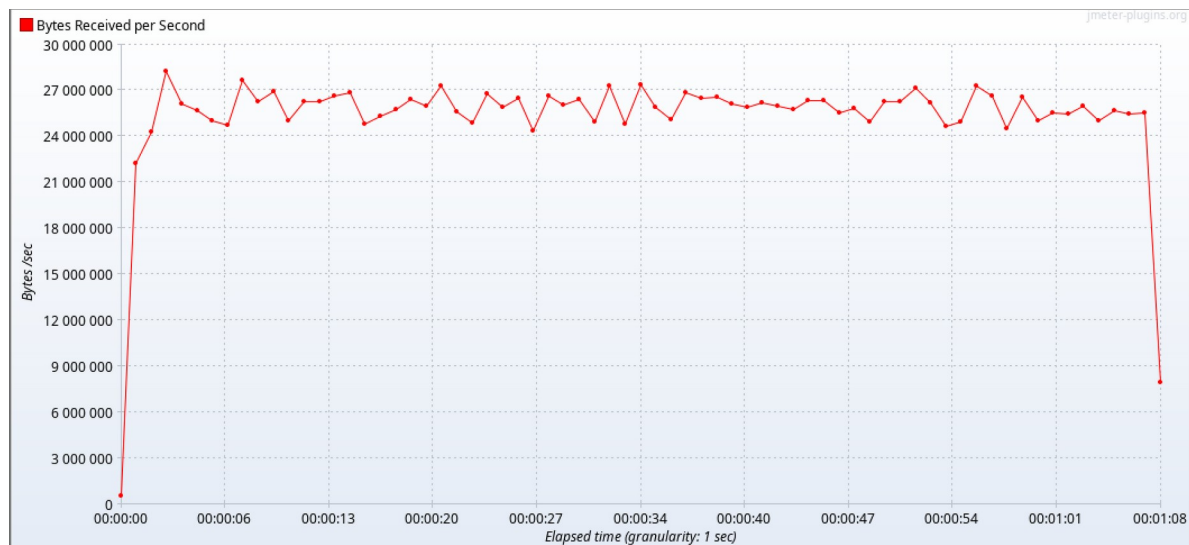
Label	# Samples	Average	Median	90% Line	95% Line	99% Line	Min	Max	Error %	Throughput	Received KB/sec	Sent KB/sec
Login	800	239	88	444	812	3235	0	6057	0,000%	103,41262	35,14	25,45
Бо Иванов	16000	710	350	1446	2592	6588	2	15499	0,000%	237,38168	612,93	105,94
Щербаков Конс	16000	693	345	1408	2537	6321	0	17630	0,000%	237,45566	431,32	111,54
Ва	16000	708	347	1437	2660	6764	1	15490	0,000%	237,45566	11955,57	95,54
А Терен	16000	875	504	1603	2974	6975	15	15832	0,000%	237,39225	12456,6	103,16
TOTAL	64800	740	388	1468	2673	6643	0	17630	0,000%	961,36728	25455,28	419,02


```

explain
SELECT * FROM users WHERE ((lower(name) LIKE 'Конс%' and lower(surname) LIKE 'Щербakov%') OR (lower(surname) LIKE 'Конс%' and lower(name) LIKE 'Щербakov%')) ORDER BY id LIMIT 100

```

QUERY PLAN
Limit (cost=600.40..600.55 rows=63 width=392)
-> Sort (cost=600.40..600.55 rows=63 width=392)
Sort Key: id
-> Bitmap Heap Scan on users (cost=350.40..598.51 rows=63 width=392)
Recheck Cond: (((lower((surname)::text) ~~ 'Щербakov%::text' AND (lower((name)::text) ~~ 'Конс%::text')) OR ((lower((surname)::text) ~~ 'Конс%::text' AND (lower((name)::text) ~~ 'Щербakov%::text'))))
Filter: (((lower((name)::text) ~~ 'Конс%::text' AND (lower((surname)::text) ~~ 'Щербakov%::text')) OR ((lower((surname)::text) ~~ 'Конс%::text' AND (lower((name)::text) ~~ 'Щербakov%::text'))))
-> BitmapOr (cost=350.40..350.40 rows=63 width=0)
-> BitmapAnd (cost=175.08..175.08 rows=31 width=0)
-> Bitmap Index Scan on surname_index (cost=0.00..87.40 rows=6297 width=0)
Index Cond: ((lower((surname)::text) >= 'Щербakov%::text' AND (lower((surname)::text) <= 'Щербakov%::text'))
-> Bitmap Index Scan on name_index (cost=0.00..87.40 rows=6297 width=0)
Index Cond: ((lower((name)::text) >= 'Конс%::text' AND (lower((name)::text) <= 'Конс%::text'))
-> BitmapAnd (cost=175.08..175.08 rows=31 width=0)
-> Bitmap Index Scan on surname_index (cost=0.00..87.40 rows=6297 width=0)
Index Cond: ((lower((surname)::text) >= 'Конс%::text' AND (lower((surname)::text) <= 'Конс%::text'))
-> Bitmap Index Scan on name_index (cost=0.00..87.40 rows=6297 width=0)
Index Cond: ((lower((name)::text) >= 'Щербakov%::text' AND (lower((name)::text) <= 'Щербakov%::text'))



В результате добавления индекса время тестирования уменьшилось с 30 минут до 1 минуты. А производительность увеличилась на порядок. Был выбран индекс btree с оператором varchar_pattern_ops, функция lower() от полей name, surname. Индексы отдельные так как есть возможность искать отдельно по имени и отдельно по фамилии. varchar_pattern_ops — оператор нужен для того чтобы искать в локе отличной от C. Функция понадобилась чтобы иметь возможность искать как с заглавной так и строчной букв. Добавление в индекс результата вычисления lower() позволило так значительно ускорить работу.