

## Задание 3

1. Создайте таблицу с большим количеством данных:

```
CREATE TABLE test_cluster AS
SELECT
    generate_series(1,1000000) as id,
    CASE WHEN random() < 0.5 THEN 'A' ELSE 'B' END as category,
    md5(random())::text as data;
```

2. Создайте индекс:

```
CREATE INDEX test_cluster_cat_idx ON test_cluster(category);
```

3. Измерьте производительность до кластеризации:

```
EXPLAIN ANALYZE
SELECT * FROM test_cluster WHERE category = 'A';
```

*План выполнения:*

| A-Z QUERY PLAN   |
|--|
| Bitmap Heap Scan on test_cluster (cost=5588.98..20238.39 rows=501233 width=39) |
| Recheck Cond: (category = 'A'::text)   |
| Heap Blocks: exact=8334  |
| -> Bitmap Index Scan on test_cluster_cat_idx (cost=0.00..5463.67 rows=501233 w |
| Index Cond: (category = 'A'::text)   |
| Planning Time: 0.479 ms  |
| Execution Time: 96.175 ms  |

*Объясните результат:*

При размере в 1 миллион строк **bitmap scan** часто оказывается быстрее, чем **seq scan**, так как он позволяет избежать полного чтения всех строк и сосредоточиться только на нужных запросу страницах

4. Выполните кластеризацию:

```
CLUSTER test_cluster USING test_cluster_cat_idx;
```

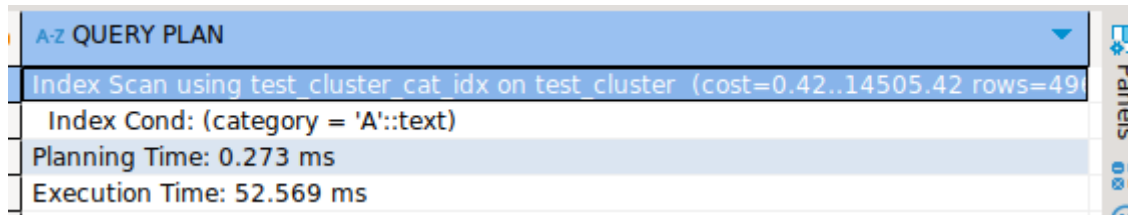
*Результат:*

Не увидел, что нужно скринить этот этап, но выполнение этого запроса у меня выполнялось около 2 сек, тк шло перестраивание по индексу

5. Измерьте производительность после кластеризации:

```
EXPLAIN ANALYZE
SELECT * FROM test_cluster WHERE category = 'A';
```

План выполнения:



| A-Z QUERY PLAN  |
|---|
| Index Scan using test_cluster_cat_idx on test_cluster (cost=0.42..14505.42 rows=49) |
| Index Cond: (category = 'A'::text)  |
| Planning Time: 0.273 ms   |
| Execution Time: 52.569 ms   |

Объясните результат:

Что я и говорил в `task2.md` данных стало побольше, следовательно, началось использование индекса из-за того, что под этот запрос специально индекс подстраивался

6. Сравните производительность до и после кластеризации:

Сравнение:

Видно что время на кластеризации меньше, причем в 2 раза по обоим параметрам `planning time` и `execution time` это свидетельствует о том, что у нас происходит перестройка таблицы под индекс, который мы потом используем для запроса `select * from ... where category = 'A'`