

## Сравнение запроса к одной таблице с индексами и без индексов

```
SELECT
    fact_id,
    date_id,
    student_id,
    course_id,
    lesson_type,
    attendance
FROM fact_student_performance
WHERE
    student_id = 20556819
    AND date_id BETWEEN 19725 AND 19997 -- id за 2024/2025 учебный год
    AND attendance = true;
```

	Без индексов	С индексами
План выполнения	Parallel Seq Scan on fact_student_performance (cost=0.00..2043726.67 rows=1 width=28) (actual time=6196.835..11776.523 rows=1 loops=3)  Filter: (attendance AND (date_id >= 19725) AND (date_id <= 19997) AND (student_id = 20556819))  Rows Removed by Filter: 33333333	Index Scan using idx_fact_student_att_date on fact_student_performance (cost=0.57..8.59 rows=1 width=28) (actual time=0.714..0.742 rows=2 loops=1)  Index Cond: ((student_id = 20556819) AND (date_id >= 19725) AND (date_id <= 19997) AND (attendance = true))
Ввод/вывод	shared hit=1057 read=1313503	shared hit=1 read=5
Время	11828.228 ms	0.759 ms

Эффективность: улучшение 99,9%

## Сравнение запроса к нескольким таблицам с индексами и без индексов

```
WITH filtered_dates AS (
    SELECT date_id
    FROM dim_date
    WHERE academic_year = '2024/2025'
),
filtered_students AS (
    SELECT student_id
    FROM dim_student
    WHERE study_status = 'active'
),
filtered_f AS (
    SELECT
        f.course_id,
        f.teacher_id,
        f.attendance,
        f.date_id
    FROM fact_student_performance f
    WHERE
        f.cancelled = FALSE
        AND f.date_id IN (SELECT date_id FROM filtered_dates)
        AND f.student_id IN (SELECT student_id FROM filtered_students)
),
attendance_agg AS (
    SELECT
        course_id,
        teacher_id,
        SUM(CASE WHEN attendance THEN 1 ELSE 0 END)::NUMERIC AS attended_count,
        COUNT(*)::NUMERIC AS total_count,
        MIN(date_id) AS any_date_id
    FROM filtered_f
    GROUP BY course_id, teacher_id
)
SELECT
```

```

c.course_id,
c.course_title,
t.teacher_surname,
t.teacher_name,
t.teacher_patronymic,
ROUND(a.attended_count / NULLIF(a.total_count, 0), 2) AS attendance_rate,
d.academic_year
FROM attendance_agg a
JOIN dim_course c ON c.course_id = a.course_id
JOIN dim_teacher t ON t.teacher_id = a.teacher_id
JOIN dim_date d ON d.date_id = a.any_date_id
ORDER BY attendance_rate DESC
LIMIT 10;

```

	Без индексов	С индексами
План выполнения	<p>-&gt; Parallel Hash Join  (cost=103599.05..1954286.65 rows=739044 width=13)  (actual time=12590.152..12883.707 rows=591257 loops=3)</p> <p>-&gt; Parallel Seq Scan on fact_student_performance f (cost=0.00..1731226.67 rows=40838889 width=17)  (actual time=2.636..9098.062 rows=32666050 loops=3)</p> <p>-&gt; Seq Scan on dim_date (cost=0.00..421.62 rows=366 width=4) (actual time=2.586..2.676 rows=366 loops=3)</p> <p>-&gt; Parallel Seq Scan on dim_student (cost=0.00..96341.33 rows=416361 width=4) (actual time=0.322..356.834 rows=333000 loops=3)</p> <p>-&gt; Seq Scan on dim_course c (cost=0.00..29149.00 rows=1000000 width=24) (actual time=0.048..223.796 rows=1000000 loops=1)</p>	<p>-&gt; Index Only Scan using idx_dim_date_academic_year on dim_date (cost=0.29..14.69 rows=366 width=4) (actual time=0.040..0.335 rows=366 loops=1)</p> <p>-&gt; Index Only Scan using idx_fact_performance_covering on fact_student_performance f (cost=0.57..180.40 rows=4838 width=17) (actual time=0.008..1.194 rows=4851 loops=366)</p> <p>-&gt; Index Only Scan using idx_dim_student_status on dim_student (cost=0.42..30398.63 rows=999267 width=4) (actual time=0.024..91.673 rows=998999 loops=1)</p> <p>-&gt; Seq Scan on dim_course c (cost=0.00..29149.00 rows=1000000 width=24) (actual time=0.041..201.240 rows=1000000 loops=1)</p>

	<pre> -&gt; Seq Scan on dim_teacher t (cost=0.00..104271.00 rows=1000000 width=31) (actual time=0.082..715.408 rows=1000000 loops=1)  -&gt; Seq Scan on dim_date d  (cost=0.00..371.10 rows=20210 width=14) (actual time=0.011..2.109 rows=20210 loops=1) </pre>	<pre> -&gt; Seq Scan on dim_teacher t (cost=0.00..104271.00 rows=1000000 width=31) (actual time=0.081..719.011 rows=1000000 loops=1)  -&gt; Seq Scan on dim_date d  (cost=0.00..371.10 rows=20210 width=14) (actual time=0.018..1.955 rows=20210 loops=1) </pre>
Ввод/вывод	Buffers: shared hit=3739 read=1516144, temp read=49160 written=49271	Buffers: shared hit=1748556 read=110581, temp read=51909 written=59528
Время	Execution Time: 18522.210 ms	Execution Time: 9746.583 ms

Эффективность: улучшение 47,4%

## Полнотекстовый поиск, фильтрации с использованием массива и json-формата

```
SELECT course_id, course_title, course_description
FROM dim_course
WHERE to_tsvector('english', course_description) @@ to_tsquery('english', 'systems | computational');
```

	Без индексов	С индексами
План выполнения	<pre>-&gt; Parallel Seq Scan on dim_course (cost=0.00..128524.00 rows=4156 width=113) (actual time=8.964..4106.021 rows=74466 loops=3)  Filter: (to_tsvector('english')::regconfig, course_description) @@ '''system''   'comput'''::tsquery)  Rows Removed by Filter: 258867</pre>	<pre>-&gt; Bitmap Index Scan on idx_dim_course_desc (cost=0.00..146.81 rows=9975 width=0) (actual time=25.561..25.562 rows=223399 loops=1)  Index Cond: (to_tsvector('english')::regconfig, course_description) @@ '''system''   'comput'''::tsquery)</pre>
Ввод/вывод	Buffers: shared hit=16281 read=3136	Buffers: shared hit=36 read=19153 written=3
Время	Execution Time: 4209.063 ms	Execution Time: 225.861 ms

```
SELECT
    teacher_surname,
    teacher_name,
    teacher_patronymic,
    subjects_taught,
    teacher_metadata
FROM dim_teacher
WHERE subjects_taught @> ARRAY['Practice']
AND teacher_metadata->'projects' @> '["Development of Computer Algebra Systems"]';
```

	Без индексов	С индексами
План выполнения	<pre>-&gt; Parallel Seq Scan on dim_teacher (cost=0.00..101562.67 rows=635 width=654) (actual time=1.155..428.784 rows=2536 loops=3)</pre>	<pre>Bitmap Heap Scan on dim_teacher (cost=159.63..5663.29 rows=1525 width=654) (actual time=6.678..16.082 rows=7608 loops=1)</pre>

	<pre>Filter: ((subjects_taught @&gt; '{Practice} '::text[]) AND ((teacher_metadata -&gt; 'projects' ::text) @&gt; '["Development of Computer Algebra Systems"]' ::jsonb))  Rows Removed by Filter: 330797</pre>	<pre>Recheck Cond: ((subjects_taught @&gt; '{Practice} '::text[]) AND ((teacher_metadata -&gt; 'projects' ::text) @&gt; '["Development of Computer Algebra Systems"]' ::jsonb))  -&gt; Bitmap Index Scan on idx_dim_teacher_sub_proj (cost=0.00..159.25 rows=1525 width=0) (actual time=5.654..5.654 rows=7608 loops=1)  Index Cond: ((subjects_taught @&gt; '{Practice} '::text[]) AND ((teacher_metadata -&gt; 'projects' ::text) @&gt; '["Development of Computer Algebra Systems"]' ::jsonb))</pre>
Ввод/вывод	Buffers: shared hit=964 read=93599	Buffers: shared hit=7413
Время	Execution Time: 515.359 ms	Execution Time: 16.531 ms

## **Оптимизация для таблицы фактов**

Для таблицы объемом больше 100 млн. записей произвести оптимизацию, позволяющую быстро удалять старые данные, ускорить вставку и чтение данных.

Надо произвести секционирование таблицы.