

Optimization process:

As our optimization process, we first made sure that the queries worked correctly with views, which made our queries easier to read. Next, we checked if there was no unnecessary information or data in the queries or views, and tried to make them faster by changing the way we structured our queries. Then, we checked how often each query would run, and if they would benefit from a materialized view, which was often beneficial especially if views were used in multiple queries and took longer to load. We also tried indices, but these did not always make our queries faster.

By checking each query with and without materialized view and these views with and without indices, we finally came to the conclusion of our chosen optimizations of three materialized views and one index.

Chosen optimizations:

We have chosen to optimize our dataset by using three materialized views and one index. We use views in all our queries, these do not result in any performance optimization, but are useful to simplify our queries.

The materialized view *passed_grades* is used for query 1 and for making other (materialized) views that are connected to the grades. The main benefit of this materialized view is that often grades that are lower than 5 should not be taken into account in the given queries, so a dataset that excludes those values is considerably faster. The total space for this materialized view is 1584 MB. Constructing this materialized view costs 50 seconds. Our performance gain is 1,7 seconds per query run, tested on query 1. Without materialized view this query takes 3,4 seconds, with *passed_grades* it takes 1,7 seconds. So the total performance gain is 170 seconds, since this query runs a hundred times. On top of this, we decided to construct an index on *passed_grades(studentid)*. This index takes 20 seconds to construct, and needs quite a bit of space, namely 803 MB. However, the index brings the time to run query 1 from 170 to almost 0 seconds.

The materialized view *GPA* has a size 245 MB, and consists of the gpa of each studentregistrationid. *GPA* is also needed in two queries, 2 and 7, and was created in 41 seconds. Without this view, our performance for query 2 would have been 37 seconds and with this view the query takes 2,5 seconds. Since this query runs ten times, this is a performance optimization of 345 seconds. The same testing on query 7 gives a performance gain of 35 seconds.

Our last materialized view *degrees_where_all_passed* requires 240 KB of space, and is used in query 2. Due to query 2 running ten times, a materialized view had a huge benefit. The required time to create this view was 1 minute and 14 seconds. Query two with this materialized view and its index runs in 7 seconds. Without the materialized view, the resulting time is 780 seconds. This results in a performance optimization of 773 seconds.

The total gain in our workload is 1493 seconds, which is a little bit more than 23 minutes. The total space used for the index and materialized views is 2632,24 MB. The total time required to construct the materialized views and index is 185 seconds. These values are computed by adding the values of all the materialized views and the index that we created.