Documentation Milestone 1

All the running times of this document are calculated by using our own computers. So in the cluster it will be much faster.

Optimization process:

1. First we made different unlogged tables so that we could load the data in these tables
2. Then we loaded the data in these tables and immediately added primary keys
3. We join the tables CourseOffer_tmp and Courses as CourseOffers on several variables
4. We dropped CourseOffer_tmp, because all variables are also in CourseOffers, for space
5. We join the tables CourseOffer, StudentRegistrationsToDegrees, and CourseRegistrations as CourseOfferRegistrations on several variables
6. We drop the table CourseRegistrations, because all variables are also in CourseOfferRegistrations, for space
7. Then we create the index student_idx on CourseOfferRegistrations(studentId)
8. Then we create the materialized view ects_per_degree so we know easily what the current ECTS and GPA is per student per degree
9. Then we create the materialized view CompletedDegrees so we know easily who passed their degree.
10. Then we create the index offer_idx on CourseOfferRegistration(CourseOfferID)

Chosen optimizations:

1. Empty tables were made so we could load the data in the tables
2. Primary keys were added because primary keys are used as indexes. We added primary keys: CourseId for the table Courses, StudentRegistrationId for StudentRegistrationstoDegrees, DegreeId for Degrees, StudentId for Students, TeacherId for Teachers.
3. We joined these tables so we needed to do less joinings during the queries. This saved us a lot of time.
4. Because we were already short on space, we choose to delete this unnecessary table.
5. We joined these tables because we needed to join them a lot, and joining them permanently saves us a lot of time. The creation of this table takes 12 minutes. And we use this table for all the queries. This means that if we didn't create this table, all the eight queries need to join a lot of tables with each others. This saved us hours of waiting.
6. Because we were already short on space, we choose to delete this unnecessary table.
7. Indexes because the first query takes before the index is made 1:07.471 minutes and after the index is made it only costs 45.498 ms. So this query is much faster and because of the fact that this query needs to be runned a 100 times this is very positive. And the creation of this index takes only 3:14.057 minutes. This index has a size of 1714 MB
8. This materialized view is used two times. To create the materialized view CompletedDegrees and for query 7. If we didn't create this view we did lost 3:43.943 minutes extra. This view has a size of 333 MB.
9. This materialized view is used 11 times: 10 times for query 2 and 1 time for query 7. The time we need to create this materialized view is 0:06.689 minutes. If we did not create this materialized view, we did lose 1:06.989 minutes for these queries in total. This view has a size of 7904 kB.
10. Indexes because the first query takes before the index was made 2:16.677 minutes and after the index was made it only costs 0:47.806 minutes. This is a lot faster and especially because this query needs to be runned three times. It takes 1:45.930 minutes to create this index and it has a size of 1714 MB.
11. The data is loaded in 16 minutes and 10 seconds
12. The confirming queries run in 22 seconds
13. The creation of auxiliaries takes 8 minutes and 50.616 seconds
14. The queries run in: Q1: 45.498ms * 100, Q2: 33.702s * 10, Q3: 25.948s, Q4: 2.563s * 10, Q5: 54.653s * 5, Q6: 29.147s * 3, Q7: 49.179s, Q8: 216.053s . This is a total of 16:59.086 min.
15. Our database is 10 GB: CourseOfferRegistrations is 9155 MB, StudentRegistrationsToDegrees 509 MB, Students is 472 MB, StudentAssistants is 28 MB, CourseOffers is 23 MB, TeacherAssignmentsToCourses is 14 MB, Courses is 5192 kB, Teachers is 4880 kB and Degrees is 1016 kB