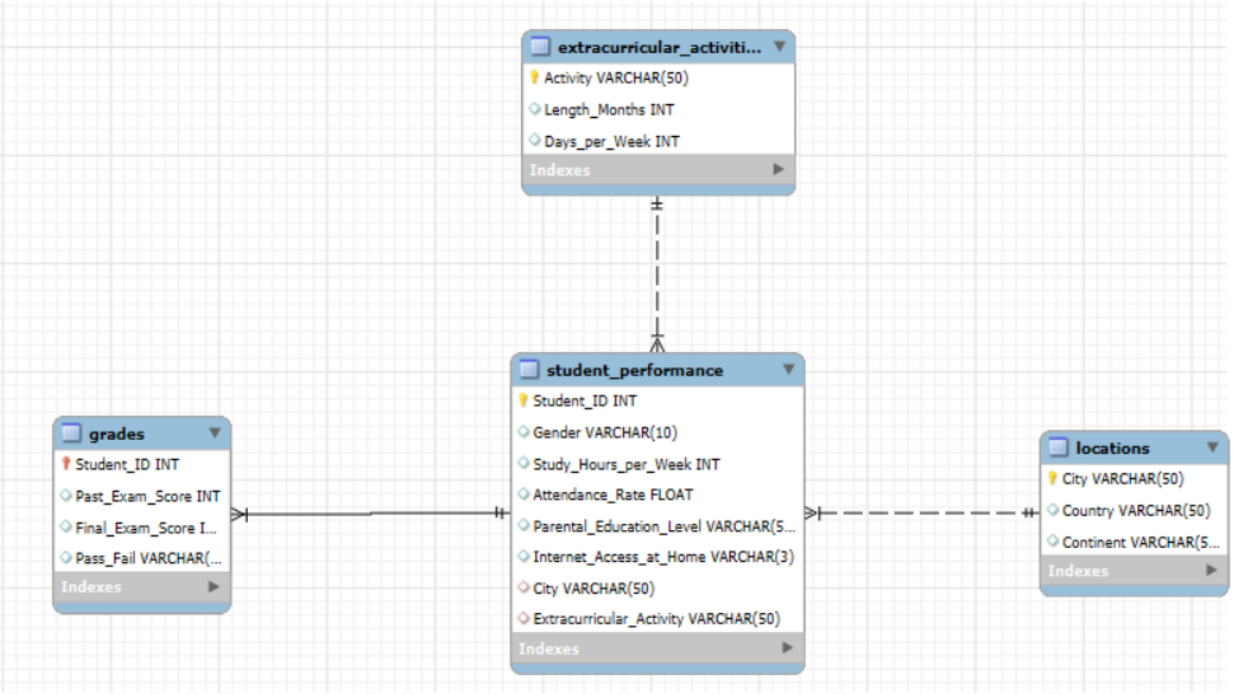
Jack Woodland

Jc253700@dal.ca | B00976388

CSCI 2141 – Database Systems

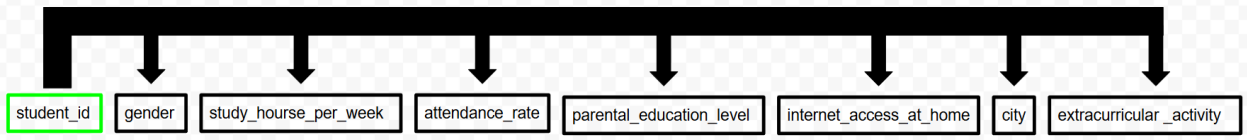
Course Project Part 3a – Explaining the Student Database

I have obtained a dataset from kaggle it consists of statistics believed to have an impact on student performance. The table "student_performance" is the only table taken from this dataset, and has been modified to follow the 3nf requirements. The other three tables are grades, locations, and extracurricular_activities.



Internal Model for StudentDB

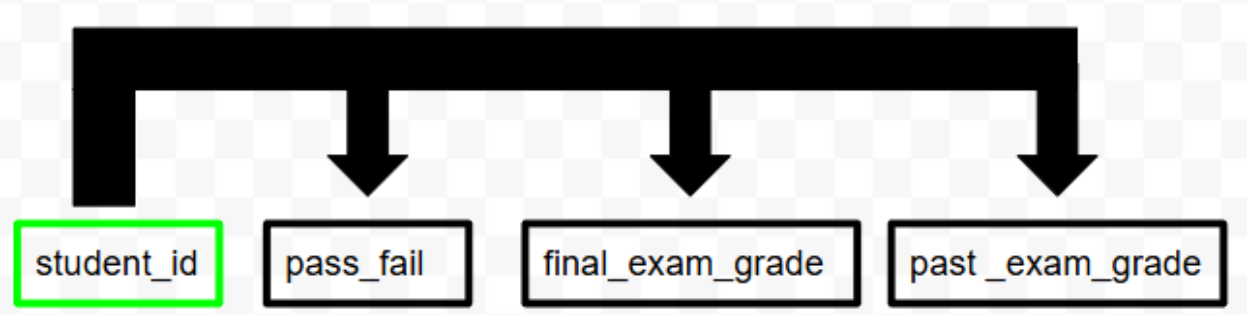
"student_performance" (8x708) -> the parent table of the database, found on Kaggle and edited to fit the 3nf guidelines by removing some aspects like the columns in the grades table to remove unwanted dependencies



Student_performance Dependency Diagram

The student_performance table has primary key student_id and is the main table of this database. It contains information on the students gender, hours working per week, attendance rate, parents education level, access to internet while at home, city of origin, and extracurricular activities they participate in. the student_performance table is 8 columns by 708 rows.

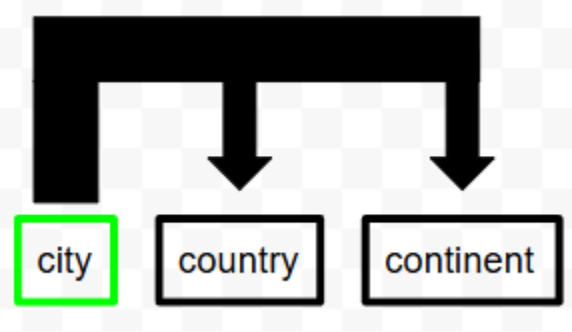
“grades” (4x708) -> a table created using columns extracted from the “student_performance” table



grades Dependency Diagram

The grades table consists of primary key student_id which connects it to student_performance. This table is used to ensure that student_performance is in 3nf and was a large change since the last submission for this project. Grades is 4 columns by 708 rows.

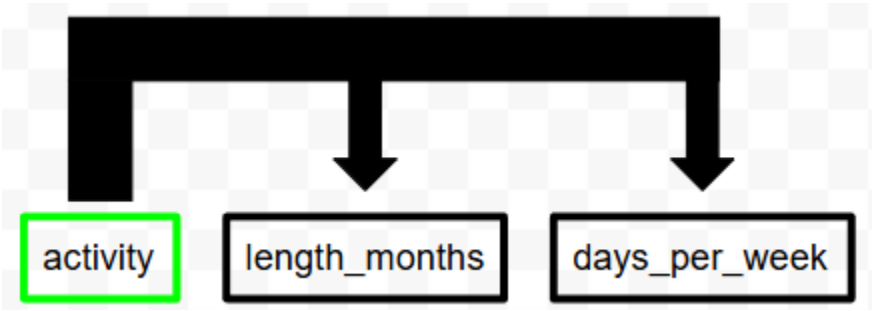
“loacations” (3x92) -> a table generated by chatgpt that links Cities in “student_performace” to Countries and Continents



Locations Dependency Diagram

The locations table consists of primary key city which is also used to connect the table to other tables. The connection is formed between city in student_performance and the primary key for locations. The locations table is 3 columns by 92 rows.

“extracurricular_activities” (3x10) -> a table generated by chatGPT to link activities information to “student_performance”



extracurricular_activities Dependency Diagram

The extracurricular_activities table has primary key activity which connects it to student_performance. It gives information on the different activities students can participate in when needed. Extracurricular_activities is 3 columns by 10 rows.

I would like to use this database to figure out what factors affect the student’s final exam grade outcome the most whether that is parents’ educations level, extracurricular activity involvement or even if certain cities or countries perform better than others.

Some of these questions will be approached as queries listed below:

1. Show students who study more than 10 hours per week and calculate their estimated weekly study time in minutes.
2. Get student final scores and their attendance rate.
3. Average final exam score by parental education level.
4. Show average past exam score for students with internet access at home.

Before this database was altered it was not in 3nf form because of the dependency that past_exam_grade, final_exam_grade, and pass_fail had on each other. I fixed this by removing these rows from the parent table while also extracting the student_id column so that I could create a new table that can link each student to their grades. The database is now 3nf because no table has internal complex dependencies that could cause disruption when inputting data in the future.