# PROJECT PHASE I-SPECIFICATION

# Nov 22th (Load your report to [DYS](mailto:tsuzek@gmail.com) until midnight AND bring the print-out to class)

The final project in this class will give you the opportunity to design and implement a small database system in a domain of interest to you.

Each group will select a separate task particularly relevant to their outside interests, such as an investment portfolio database, a medical database, an astronomy database, a pharmaceutical database or a small business database.

The purpose of the first phase of the project is for groups to specify which domains and tasks they plan to cover (to confirm that they are suitable and sufficiently challenging) and to begin the database design and implementation process.

**If you have any questions regarding the suitability of your proposed domain, please discuss with Tuba hoca or Erdem during the recitation class.**

In general, your final project will constitute an interface for data input (either from the user entering the data manually using stored procedures, the MySQL WWW interface or Java or Php-based data-extraction from text or on-line data sources), and an interface for data output (through stored procedures and views to provide customized perspectives on the database for different users).

# PHASE I requirements:

In your **written** report, please respond to **all 8** questions:

1) Who are your team members? Each team may consist of up to 3 members. 1 or 2 person teams are not allowed.

**Our team consist of 3 person who names are Adem VAROL (200709078), Mustafa Can İNCE (200709081) and Cihan Sezer ÖZKAMER (200709603).**

2) Explain your data:

* What’s the name/purpose of your project? Why did you pick this topic?

**Project Name: ExamAnalysis\_DB**

**We chose this project in order to analyze the questions of the students who took the exam on a class and student basis.**

**We also plan to have our project available with a UI in the future.**

* How many files exist?

**There are 12 data files in our project.**

* How many rows and columns do they contain?

|  |  |  |
| --- | --- | --- |
| **File** | **Number of rows** | **Number of columns** |
| **students** | **462** | **17** |
| **student\_classes** | **17** | **2** |
| **student\_schools** | **2** | **3** |
| **student\_clubs** | **8** | **2** |
| **student\_projects** | **6** | **2** |
| **cities** | **3** | **3** |
| **states** | **4** | **3** |
| **streets** | **16** | **2** |
| **examsofstudentsdetails** | **625** | **7** |
| **exams** | **8** | **5** |
| **exams\_names** | **4** | **2** |
| **teachers\_names** | **5** | **2** |
| **TOTAL** | **1160** | **50** |

* How many string(non-numeric) columns do you have?

**There are 26 string columns in our project.**

* Provide the links of the **raw data** of your project (e.g. kaggle address, ftp site address etc).

**Our raw data is in the DataSource folder.**

* **Provide the Bitbucket URLs of your ER diagram (.mwb files) in your report AND UPLOAD THE REPORT TO DYS.**

[**https://bitbucket.org/u200709078/dblab2021/src/master/ExamAnalysis\_DB/**](https://bitbucket.org/u200709078/dblab2021/src/master/ExamAnalysis_DB/)

3) Give a reasonably comprehensive and representative list of the English questions you would like your system to be able to answer (find at least 10 questions containing min/max/avg ).

For example,

“Compute the mean expense spent to hobbies by family members, grouped by adult or children"

**OUESTIONS**

**1. What is the total number of female students among the students who took all the exams?**

**2. What is the GPA of male students who take a given exam?**

**3. What is the minimum grade on a given exam?**

**4. What is the interval length (max-min) of a given exam?**

**5. What is the arithmetic mean of a given exam?**

**6. What is the standard deviation of a given test?**

**7. What is the median of a given exam?**

**8. What is the most repeated (mod) grade on a given exam?**

**9. What is the percentage of those who took and did not take the exam in a given group (for example, all 10th graders)?**

**10.What is a list of students in a given exam?**

**(We can't calculate the grades of the students because we can't compare string expressions yet. Therefore, we added 3 new questions and made their solutions.)**

**EXTRA OUESTIONS**

**11. What is the total number of students in the classes?**

**12. Who is the eldest of the students who took the exam?**

**13. What is the list of all students who took the exam on a given date?**

4) Design and show a relational data model (ER diagram) in MySQL Workbench.

In your ER diagram **insert one row of sample values for each relation**. Define all your primary keys and indicate all referential constraints with arrows between foreign keys and the relation/attribute they reference. **Again do not forget to provide one row of sample values from your data!!**

**The requested information is in the ExamAnalysis\_DB\_EER file.**

5) Write at least 5 SQL statements that will implement the English questions from of your target queries (from Question 3). **You will need to provide at least 1 aggregate function with group by.**

**## 1. What is the total number of female students among the students who took all the exams?**

select count(student\_id) as "Number of female students who took all exams:" from students where students.studentgender="F" and students.student\_id in (

select student\_id from examsofstudentsdetails);

**## 10. What is a list of students in a given exam?**

select student\_classes.class\_name, students.studentname, students.studentsurname

from students join student\_classes on students.class\_id=student\_classes.class\_id

where student\_classes.class\_name="9-B" order by students.studentname;

**## 11. What is the total number of students in the classes?**

select student\_classes.class\_name as "Class Name", count(students.class\_id) as "Number of students"

from student\_classes join students on student\_classes.class\_id = students.class\_id

group by student\_classes.class\_name

order by student\_classes.class\_name;

**## 12. Who is the eldest of the students who took the exam?**

select min(students.studentbirthday)

from examsofstudentsdetails join students on examsofstudentsdetails.student\_id=students.student\_id;

**## 13. What is the list of all students who took the exam on a given date?**

select students.student\_id ,students.studentname, students.studentsurname, exams.examdate

from examsofstudentsdetails join exams join students

on examsofstudentsdetails.exam\_id=exams.exam\_id and students.student\_id=examsofstudentsdetails.student\_id

where exams.examdate="2020-4-4"

order by students.studentname;

6) Provide a proposal of how you will load the database with values.

* If you plan to extract/import data from on-line sources, briefly describe what are the sources (e.g. personal data, or WWW URL's) and what any format conversion issues you expect to encounter are.

* If you plan to input your data primarily through a WWW or form-based inter-face, briefly describe this interface and the issues involved.
* **We can convert our data in Excel file to CSV file format using excel vba and transfer them to our database via MySQLWorkBench application. For the conversion process, the "CONVERT ALL DATA TO CSV FILE" button in the ExamAnalysis\_DB\_Data\_v6.xls file should be clicked.**
* **The generated CSV files are in the EXAMANALYSIS\_DB\_CSV folder.**

7) Are you going to implement any front-end user interface? If not, at the minimal you need to indicate here that you will at least implement 1 view and 1 stored procedure with IN, OUT and INOUT.

**We plan to associate an UI with our database in the future, but for now we do not plan to design an UI.**

8) Describe the database platform you plan to use (i.e. MySQL 5.0 on your home computer Pentium II with 26 MB of memory, etc.), including any relevant implementation details or challenging issues.

**Intel Core I3 home computer**

**Minimum 2GB memory**

**MySQL 8.0.xx**

**We think that we may have problems in data processing due to system differences (For example, Turkish character problem in Unix systems).**