

Final Group Project

Instructions

- This is a group assignment worth 20% of the final grade.
- Read each section carefully and provide answers.
- The following are the expected deliverables:
 - **Deliverable 1 - Final Presentation (10%):** This is a technical presentation; Therefore, it must be professionally presented with adequate technical details. You may use a suitable presentation structure to address all FOUR tasks given. Each group must record their presentation using any collaboration tool and submit it along with the PPT file. Both the PPT file and the video must be submitted before the deadline. Please refer to the attached rubrics for the score breakdown.
 - **Deliverable 2 – SQL Script File (7%):** Each group needs to submit an SQL code file. The file must be original and professionally arranged, and each code segment must be clearly explained by using comments. Please refer to the attached rubrics for the score breakdown.
 - **Deliverable 3 – Individual Reflection (3%):** Each “individual” needs to submit their own professional reflection as per the guidelines provided in this project assignment file.
- Time allocated for the presentation: **15-20 minutes** for the recorded group presentation.
- Due date: **February 17, 2022**.
- Toronto School of Management (TSoM) requires students to maintain high standards of academic integrity. Students are responsible for conducting themselves honestly and ethically in all aspects of their academic career and for becoming familiar with this policy and abiding by all aspects of it. To support academic honesty at TSoM, all work submitted by students may be reviewed for authenticity. In submitting their own work to TSoM, students consent to their submissions undergoing such a review and being retained in a database for comparison with other work submitted by fellow students.

Project Requirements (Deliverable 1 & 2)

Goal: Demonstrate proficiency in data handling in SQL (using MySQL).

Synopsis:

VivaK is a successful retail chain in the fashion industry, the head office of which is located in Southlake, Texas, USA. The HR Manager of the VivaK head office has decided to hire a group of analysts (you) to support their HR analytics by devising an Online Analytical Processing (OLAP) Database. Unfortunately, the information management of the HR department has not been robust so far, and it maintains its data in various data/file formats. The data is also known to have anomalies. The HR manager narrates the following story:

We are in a tough situation with this data, and our analysts find it highly challenging to perform their job without a well-designed OLAP database. Allow me to clarify how our organizational structure works. We operate in several regions and countries, but we manage our operations via seven office locations in different parts of the world. We have three offices in the US, one in Canada, two in the UK, and one in Denmark. The President sits at the head office in Southlake, Texas, USA. All the other locations have the same well-structured management team reporting to two Vice Presidents at each location: VP Administration and VP Operations. Each location has 11 departments, 10 of which are managed by a dedicated Manager; the other employees report to these managers. We pay our employees on the USD salary scale, and you may find the monthly salary range per job title and each employee's monthly salary in the sample dataset we have provided to you. We also keep track of each employee's dependents. Our data designer made an error and called the "location_id" column in employees data "department_id".

Your Tasks:

Task 1: Analyse the provided data and design a complete (and effective) MySQL Data Model (File → New Model) call VivaK_Data_Model (This is a self-study task. Consult your professor, should you need any assistance). You must include the diagram in your presentation. You must identify the following:

1. Entities
2. Attributes
3. Constraints (Including Primary Keys, Not Null columns, and Unique Keys)
4. Relationships and Foreign Keys

Task 2: Create a Schema based on VivaK_Data_Model and call it VivaKHR. You must include all the statements in your SQL file and only the important statements/ outputs in your presentation.

1. All salary-related columns must be double. Data must be recorded in 2 decimals.
2. All ID columns and their related foreign keys must be Integer.
3. The employee table MUST contain the report_to column to record the ID of the employee's manager.
4. The data type of All the date columns MUST be DATE.
5. Include a column called experience_at_VivaK in the appropriate table to include the number of months that each employee has worked at VivaK.
6. Include a column called last_performance_rating in the appropriate table to include the performance rating (0-10) of each employee after the annual performance appraisal.
7. Include a column called salary_after_increment in the appropriate table to record the salary anticipated after the annual performance appraisal.
8. Include a column called annual_dependent_benefit in the appropriate table to record the dependent bonus that each employee receives per dependent.

Task 3: Import and clean the data. Import the sample data to test the integrity and efficiency of your OLAP schema. You must include all the statements/outputs in your SQL file and only the important statements in your presentation. Consider the following:

1. **(3 Marks)** Handle Duplicates: All tables must be in the full-normalized form, meaning that there should not be any data redundancy. Check for duplicates in each table using qualifying candidate keys.
2. **(12 Marks)** Format the data according to the designated data types.
 - a. Ensure the floating-point data is represented as double.
 - b. Ensure that the phone numbers are all recorded in the format: '+000-000-000-0000', where the first four characters refer to the country code.
 - c. Ensure the dates are recorded in the format: 'yyyy-mm-dd'.
3. **(10 Marks)** Treat Missing Values:
 - a. Fill up the report_to column by analyzing the available data.
 - b. Devise a strategy to fill in the missing entries in the salary column. Justify your answers and state your assumptions.

Task 4 **[25 Marks]**: Perform the following calculations and updates. You must include all the statements in your SQL file and only the important statements/outputs in your presentation.

1. **(4 Marks)** experience_at_VivaK: calculate the time difference (in months) between the hire date and the current date for each employee and update the column.
2. **(4 Marks)** last_performance_rating: to test the system, generate a random performance rating figure (a decimal number with two decimal points between 0 and 10) for each employee and update the column.
3. **(6 Marks)** salary_after_increment: calculate the salary after the performance appraisal and update the column by using the following formulas:

salary_after_increment = salary * increment

increment = 1 + (0.01 * experience_at_VivaK) + rating_increment

last_performance_rating	rating_increment
>= 0.9	0.15
>= 0.8	0.12
>= 0.7	0.10
>= 0.6	0.08
>= 0.5	0.05
< 0.5	0.02

* Salary caps at max_salary.

4. **(6 Marks)** annual_dependent_benefit: Calculate the annual dependent benefit per dependent (in USD) and update the column as per the table below:

Title	Dependent benefit per dependent
Executives	0.2 * annual salary
Managers	0.15 * annual salary
Other Employees	0.05 * annual salary

5. **(5 Marks)** email: Until recently, the employees were using their private email addresses, and the company has recently bought the domain VivaK.com. Replace employee email addressed to '<emailID>@vivaK.com'. emailID is the part of the current employee email before the @ sign.

Deliverable 3:

Individual Reflection and Peer Review. *[3% of the final score / 30 points]*

1. Each student on your team must write their independent reflection on the project. Your reflection article should be no more than 1 page (excluding the peer evaluation) and must be uploaded separately. Your task:
 - a. *[10 Marks]* Choose a topic from the SQL series that you found most interesting and explain it.
 - b. *[10 Marks]* Explain how well the chosen topic was applied in the project.
 - c. *[5 Marks]* Explain the strengths the team demonstrated when executing the project.
 - d. *[5 Marks]* Discuss the areas to improve when working in this group.
2. Complete and include the peer evaluation.

Marking Rubric: Deliverable 1 – PowerPoint File and Live Presentation

Question		0-20%	20-50%	50-80%	80-100%	Final Marks
Sections		The answer does not meet the question's requirements. Invalid or incorrect answer.	A poorly structured and presented answer that demonstrates the lack of understanding of the subject matter. The section addresses some of the topic's objectives. The answer has major grammatical errors.	A well-structured and presented answer. The answer addresses most of the underlying subject matters of the question. The section addresses most of the topic's objectives. The answer has limited grammatical errors.	An exceptional answer that covers all the underlying subject matters of the question. The answer exerts critical thinking and includes examples. The section addresses all of the topic's objectives. The answer has no grammatical errors.	Marks Out of 100
Introduction		0-2	2-3	3-4	4-5	
Task 1: Data Model		0-5	4-10	10-16	16-20	
Task 2: Schema		0-4	4-7	7-11	11-15	
Task 3	1: Duplicates	0	1	2	3	
	2: Formats	0-2	2-5	5-8	8-10	
	3: Missing Values	0-2	2-5	5-7	7-9	
Task 4	1: experience_at_VivaK	0	1	2	3	
	2: last_performance_rating	0	1	2	3	
	3: salary_after_increment	0-2	2-3	3-4	4-6	
	4: annual_dependent_benefit	0-2	2-3	3-4	4-6	
	5: email	0-2	2-3	3-4	4-5	
PPT Presentation File		0-2: Poor quality	2-3: Structure is maintained. Include Notes.	3-4: Structure is maintained. Descriptive Notes. Accurate referencing.	4-5: Structure is maintained. Accurate referencing. Well-written Notes. High-quality presentation.	
Group Presentation (Audio/Visual)		0-2: Poor Quality Presentation	2-5: Satisfactory Presentation	5-8: Very Good Presentation	8-10: Exceptional Presentation.	
Total						

Marking Rubric: Deliverable 2 – SQL Code

Question		0-20%	20-50%	50-80%	80-100%	Final Marks
Sections		Invalid or incorrect answer. The code doesn't run, and the steps are incorrect.	The code works but the intermediate steps are not compatible with the answer. A correct answer with a poor approach. Some of the steps are correct, although the code doesn't work.	The code works and most of the steps meet the question's expectations. Some comments are available. Most of the steps are correct, although the code doesn't work.	The code works and all the steps meet the question's expectations. The code is well commented and well-structured.	Marks Out of 70
Task 2: Schema		0-5	4-10	10-16	16-20	
Task 3	1: Duplicates	0	1	2	3	
	2: Formats	0-2	2-5	5-7	7-9	
	3: Missing Values	0-2	2-5	4-7	7-8	
Task 4	1: experience_at_VivaK	0	1	2	3	
	2: last_performance_rating	0	1	2	3	
	3: salary_after_increment	0-2	2-3	3-4	4-5	
	4: annual_dependent_benefit	0-2	2-3	3-4	4-5	
	5: email	0-1	1-2	2-3	3-4	
SQL code file quality		0-2: Poor quality	2-5: Structure is maintained. Some code is explained using comments.	5-8: Structure is maintained. Most of the important code sections are explained using comments.	8-10: Structure is maintained. All important code sections are explained using comments.	
Total						