Lewis University  
CPSC 50900: Database Systems  
Term Project

TITLE OF YOUR PROJECT:

**Employee Management System**

Proposed and submitted by:

Syed Aziz Ullah Hussaini, azizullahhussainis@lewisu.edu

Work products stored in the GitHub repository:

Table of Contents

[Initial Proposal 2](#_Toc85814569)

[Data Sources 3](#_Toc85814570)

[Data Storage Alternatives 5](#_Toc85814571)

[Relational Database Design Process 7](#_Toc85814572)

[Relational Database Design 7](#_Toc85814573)

[Data Definition Language (DDL) Scripts 8](#_Toc85814574)

[Data Manipulation Language Scripts 8](#_Toc85814575)

[Indexes 9](#_Toc85814576)

[Views 9](#_Toc85814577)

[Triggers 10](#_Toc85814578)

[Transactions 10](#_Toc85814579)

[Database Security 10](#_Toc85814580)

[Locking and Concurrent Access 10](#_Toc85814581)

[Backing Up Your Database 11](#_Toc85814582)

[Python Programming 11](#_Toc85814583)

[PHP Programming 12](#_Toc85814584)

[Suggested Future Work 12](#_Toc85814585)

[Activity Log 13](#_Toc85814586)

# Initial Proposal

*Description: You will describe the data you aim to store. What data will be storing? Why are you interested in this data? Why is it important? Where will the data come from? Who will use this data? What kind of application do you plan to build with it?*

*Rubric: Your response to each of these six questions will be graded out of 3 points.*

* *3 points: clear, complete descriptions that convey the importance and meaning of your data*
* *2 points: mostly clear descriptions, although some additional data would have helped in some sections*
* *1 point: necessary details are lacking in many of your responses.*

*You will also earn 2 additional points for coming up with a descriptive title for your project.*

*Total points possible: 20*

**Employee Management System**

***Employee Management System*** is the idea behind implementation of the project which is used to keep all details of the employees of an Organization including all the personal, official, banking, experience, qualification, and identification details. All this data is very necessary for any organization to get the information of an employee at any moment of time for example if a manager wants to see the current salary of his reporting employees then he can just execute a query based on the requirement and can fetch the details required and if at all he wants to give hike to them then he can easily give it to them just by running a query instead of changing the salary for each person. My personal interest in this data is because if at all in future I got a chance to lead any team then it’ll be very helpful for me, also as I have already been a part of Project Management Team in my previous company (DXC Technology Ltd.) which led me having a good exposure to this kind of data almost for a year. Generally this data will be taken directly from the employee when he joins the organization and will be stored in the respective tables or data sources which will be managed by individual leads of each team considering an example let us say that John is going to join XYZ company then as a part of Onboarding process all the necessary details including personal, experience, address, passport/driving related, and qualification will be submitted by John whereas the other details like Branch, Team, Location will be given by manager and Salary, banking details will be given by HR, Finance Team respectively. This data will be mostly used by HR and Managers to give salary hike, if at all the manager is getting replaced then to update the same which includes team details as well. What location an employee is placed on etc will also has to be updated by respective Manager, also this will be acting as a directory to all the employees who work in the organization in which only the basic details like Name, Mobile Number, Location, Email ID, manager name, Employee ID, Address will be displayed when querying with the employee id.

The idea of the application is not yet finalized by me as there are a lot of applications that I can build using this database or data sources. Most probably I’ll be concentrating to build a website named as ‘Employee Portal’ that will act as directory to all employees and coming to HR and manager’s they’ll be updating the database based on the employee’s performance like giving hike, assigning the projects, giving promotions and all which will be reflected in the directory post the changes are done by HR and manager respectively. Depending on the tasks which we’ll be performing going further there might be some addition of data sources as well depending on the requirement which I’ll be keeping you updated with on a regular basis.

Note:

Personal, Qualification, Experience, Identification 🡪 This will be coming from the Employee that joins Organization

Banking, Official, Branch, Salary 🡪 This data will be coming from the Human Resource Manager and Technical Manager based on the project/resource requirement

**====================================================================================**

# **Data Sources**

*Description:* *Gather your data in text files. The text files may be csv, tab-delimited, xml, json, or some other custom format. Not all the files need be of the same type. Identify what each file contains by indicating where it came from, explaining in detail how it structured, and describing how you will reorganize the data into a relational database. Post your data files to your GitHub repository, and provide samples of the data in your Word doc.*

*Rubric: Your work will be graded as follows:*

* *5 points: you gathered multiple data files that contain the data that will populate your databases. If you do not use multiple data files, you will not receive credit.*
* *5 points: you described the contents of the data files in detail, including referencing their origin and explaining how they were structured.*
* *3 points: you identify which fields you plan to include in your database, including their data types and any constraints you expect to impose on the data or steps you'll have to take to clean up the data.*
* *2 points: you post the data files to your GitHub account and make it possible for me to see them.*

*Total points possible: 15*

There are various data sources we are using here in different formats like .txt, .json, .csv and .xlsx formats which we’ll need to insert into the respective tables into database the data sources and their relative information are as given below

1. ***address.json*** 🡪 This will contain the address of the employee and has the primary key as the zipcode, this data will be provided by the employee while onboarding process. It is basically needed to keep the information of employee i.e., where he currently lives in.
2. ***bankdetails.json*** 🡪 This data we are taking as the json format and will be provided by the Finance team when an employee joins basically this will contain the details of bank where the account is created for the respective employee
3. ***branchdetails.txt*** 🡪 These are the details of the work location of the employee that will be which will be shared by the manager once the projects get assigned to him. It purely depends on the client or the team that a location will be assigned to him
4. ***Employee.csv*** 🡪 This is our main file/data source which contains all the necessary information of the employee in the form of CSV, most of the fields are provided by the Employee but there are a couple of fields that’ll be determined by HR and respective project Manager for example Salary
5. ***identificationdetails.xlsx*** 🡪 These are more likely the details which are collected as the address proof from the employee which consists of a couple of Government recognized IDs that’ll be helpful in identifying the employee during background verification. We have taken Driving License and the Passport details of employees as the source of information for this purpose.
6. ***officialdetails.json*** 🡪 These details are allotted by the HR after the employee onboards this will mostly consists of the manager details the bank details, HR details etc. This will be directly managed by the HR and Manager depending on the changes what an employee goes through with the project perspective
7. ***Qualification.csv*** 🡪 These are the details of the qualifications of the employees, these will be updated by the employee itself.
8. ***WorkExperience.xlsx*** 🡪 This table/data source depicts the employee experience and address of the company where the employee is currently working and its joining date to the company.

# **Data Storage Alternatives**

*Description: We will study alternatives to storing data in a relational database. Some of the alternatives come from several decades ago, including the hierarchical and network models. Some are newer options, such as NoSQL databases that use JSON or some other encoding. Describe in detail how to store the data using two alternatives to relational databases. Be sure to describe how you would implement the alternatives and the advantages and disadvantages of each.*

*Rubric: Your work will be graded as follows*

* *5 points for clearly describing how your data could be stored using one alternative to relational databases and what the advantages and disadvantages of that approach would be.*
* *5 points for clearly describing how your data could be stored using another alternative to relational databases and what the advantages and disadvantages of that approach would be.*

*Total points possible: 10*

**Alternate Ways**

The primary database that we are using in the project is a relational database(SQL), we do have N number of tables one for storing Personal Information, one for Storing Contact Information, and so on, these tables are used for storing the data and performing CRUD operations on it. All the tables are linked to each other in one or another way.

As an alternate, we can also make use of MongoDB and Redis

1. **MongoDB**

MongoDB is a document-oriented database that stores data in JSON-like documents with dynamic schema. which makes it easier for you to store the data without thinking about the data structure to be used such as the number of fields or types of fields to store values. MongoDB documents are like JSON objects.

In MongoDB instead of creating tables we’ll be creating the collections with below syntax and will be performing similar operations like select using its related commands

* For creating collection: db.createCollection("mycol ")
* For querying collection: db.mycol.find()
* For Inserting the documents: db.mycol.insert([records we want to insert])
* For dropping the collection: db. mycol.drop()

**Advantages of using MongoDB**

* Handles large volume of data
* Developer Friendly
* Easy to deal with Cloud with zero Downtime.

**Disadvantages of using MongoDB**

* It doesn’t support transaction.
* It uses high memory for data storage.

Example showing how data is stored in SQL and Mongo DB

A record in SQL:

|  |  |  |  |
| --- | --- | --- | --- |
| IFSCCode | BankName | Location | BranchManager |
| FR0PRS | Bank of France | Paris | John Walker |

A document in Mongo DB:

{

"IFSCCode": "FR0PRS",

"BankName": "Bank of France",

"Location": "Paris",

"BranchManager": "John Walker"

}

1. **Redis**

As an Alternate, we can also make use of PostgreSQL which is open-source database also used as message broker or caching system. It has built-in replication, scripting, transactions, and different levels of on-disk persistence, and provides high availability via Redis Sentinel and automatic partitioning with Redis Cluster. Depending on our requirement or use case we can persist data either by periodically dumping the dataset to disk or by appending the command to disk based log

In Redis we have to setup a cluster and then it’ll be direct for us to create a database using + symbol and to load the data we’ll using the import. We’ll having to do most of the things from console in Redis.

**Advantages of using Redis**

* We can use Redis from most of the programming languages
* Redis supports almost all data structures.
* Redis uses its own hashing mechanism called Redis Hashing

**Disadvantages**

* It requires a huge ram because it is in-memory so are not supposed to use it on ram servers.
* Clients connecting to the Redis cluster should be aware of the cluster topology, causing overhead configuration on Clients.

Redis stores the data in following format

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rediskey | IFSCCode | BankName | Location | BranchManager |
| bank:1000 | FR0PRS | Bank of France | Paris | John Walker |

SQL stores as below in contrast

|  |  |  |  |
| --- | --- | --- | --- |
| IFSCCode | BankName | Location | BranchManager |
| FR0PRS | Bank of France | Paris | John Walker |

But in our case, we are using SQL database as

SQL databases provide great benefits for the data whose structure doesn’t change frequently and where data integrity is paramount. It’s also best for fast analytical queries. And in our employee management system, we have no plans for changing Database Schema, and as well data Integrity is the priority.

**====================================================================================**

# **Relational Database Design Process**

*Description: Consider the list of fields you identified in part c. Identify functional dependencies that exist among them. For each functional dependency, identify the determinants and the fields they determine. This becomes the basis for identifying your entity sets, which then become your tables. Give each entity set or table you identify in this way a unique and clear name, making sure that the names you use are singular nouns. Then list the relationships that exist among the various entity sets. For each relationship, identify its connectivity (one-to-one, one-to-many, many-to-many) and participation (optional or mandatory). Finally, make sure that none of the attributes you've assigned to each entity set are multi-valued. If they are, take the steps needed to break them down.*

*Rubric: Your work will be graded as follows:*

* *8 points for identifying all the functional dependencies, including determinants and the columns whose values they determine.*
* *2 points for naming the entity sets that make up your data with clear, easy-to-understand names.*
* *6 points for identify the relationships among the entity sets and identifying connectivity and participation for each.*
* *2 points for breaking down multi-valued attributes.*

*Total points possible: 18*

ENTER YOUR RELATIONAL DATABASE DESIGN DESCRIPTION HERE. INCLUDE SOURCE CODE AND SCREEN SHOTS.

# Relational Database Design

*Description: This is where you will complete your database design. For each of the entity sets you identified in the preceding section, analyze them to make sure they pass 2nd, 3rd, 4th, and Boyce-Codd Normal Form. If they do not, introduce additional entity sets or key changes to make sure that they do. Then, add foreign keys to connect entity sets that are related. For many-to-many relationships, introduce bridge entity sets to convert them into two one-to-many relationships. Also, consider whether you should introduce surrogate keys to create a more efficient primary key for some of your entity sets. Finally, diagram your design in Vertabello. Make sure your ER diagram correctly shows all entity sets, their primary and foreign keys, the data types for each attribute, and the connectivity and participation characteristics of each entity set. Your final Vertabello design should be something you could actually implement in a relational database management system.*

*Rubric: Your work will be graded as follows:*

* *4 points for the normalization analysis of your entity sets.*
* *3 points for introducing bridge entity sets.*
* *3 points for choosing foreign keys and perhaps more efficient surrogate keys*
* *10 points for correctly depicting your physical database model in Vertabello*

*Total points possible: 20*

ENTER YOUR RELATIONAL DATABASE DESIGN HERE

# Data Definition Language (DDL) Scripts

*Description: Use Vertabello to generate a script of SQL commands that build the database and its table structures. Write scripts or build Excel spreadsheets that take your data files and generate scripts of SQL insert statements from them. Use the MySQL source command to run the various scripts needed to build and populate the database in MySQL. Include the source code and / or Excel spreadsheets you use to manipulate and populate the data. Make sure all your tables have at least three records in them and that you've linked the tables through their foreign keys.*

*Rubric: Your work will be grades as follows:*

* *Database and table creation statements from Vertabello saved as an sql script file: 3 points*
* *Scripts you write or Excel spreadsheets you create to generate SQL commands for populating the tables, uploaded to GitHub: 8 points*
* *Descriptions of the scripts and Excel spreadsheets you wrote along with code excerpts included in the Word document: 5 points*
* *Screenshots of your successful attempts to use the MySQL source command to populate each table with at least three records: 4 points*

*Total points possible: 20*

ENTER YOUR DDL WORK HERE

# Data Manipulation Language Scripts

*Description: Write the SQL commands for twelve queries. Two queries should be insert statements, two should update statements, one should be a delete statement, one should be a simple select statement that selects a subset of the rows and columns from one table, two should be a select statements that select data from a joining of two tables, two should use summary functions to generate statistics about the data, one should be a multi-table query, and one should be another query of your choice. Show the queries and screenshots of the results in your Word document, and save your queries in a commented sql script to GitHub.*

*Rubric: Your work will be graded as follows:*

* *1 point each for the two insert statements*
* *1 point each for the two update statements*
* *1 point for the delete statement*
* *1 point for the simple select statement*
* *2 points each for the 2 join statements*
* *2 points each for the two that use summary statements*
* *2 points for the multi-table query*
* *2 points for the query of your choice.*
* *12 points for showing the query and a screenshot of the corresponding result set back-to-back for each of these queries in your Word document.*

*Total points possible: 30*

ENTER DML WORK HERE

# Indexes

*Description: Improve the performance of your design by adding indexes to various tables. Show the SQL needed to add the indexes. Explain why you chose the ones you added. Explain how you would demonstrate the impact the indexes had on the performance of various queries.*

*Rubric: Your work will be graded as follows:*

* *6 points for clearly defining at least three indexes and explaining why you chose them.*
* *3 points for showing the sql needed to generate the indexes*
* *3 points for explaining how you would demonstrate the performance improvement afforded by the indexes.*

*Total points possible: 12*

ENTER YOUR INDEX WORK HERE

# Views

*Description: Add two views to your database to provide easy access to combinations of data from multiple tables.*

*Rubric: Your work will be graded as follows:*

* *2 points for including the SQL for generating the two views in your Word document*
* *2 points for including screenshots for the data contained in each view in your Word document*
* *2 points for explaining why each view is a valuable addition to your database*
* *2 points for explaining who might benefit most from having access to each view.*

*Total points possible: 8*

ENTER YOUR WORK WITH VIEWS HERE

# Triggers

*Description: Add a trigger to a table so that data will be updated when a certain event occurs*

*Rubric: Your work will be graded as follows:*

* *2 points for including the SQL for the trigger in your Word document*
* *2 points for clearly explaining the purpose of the trigger*
* *2 points for a screenshot and explanation that shows the trigger in action.*

*Total points possible: 6*

ENTER YOUR WORK WITH TRIGGERS HERE

# Transactions

*Description: Demonstrate that you know how to define and use a transaction. Why are transactions important for ensuring ACID behavior?*

*Rubric: Your work will be graded as follows:*

* *3 points for clearly explaining the importance of transactions to ensuring ACID behavior*
* *3 points for including a screenshot and accompanying explanation of a MySQL transaction.*

*Total points possible: 6*

ENTER YOUR WORK WITH TRANSACTIONS HERE

# Database Security

*Description: Identify the different kinds of users who will use your database. Write GRANT statements to define the privileges for these different kinds of users.*

*Rubric: Your work will be graded as follows:*

* *6 points for clearly identifying and describing the various kinds of users who will use the databases and identifying and justifying what privileges each should have.*
* *4 points for writing GRANT statements that assign privileges to these different kinds of users.*
* *4 points for demonstrating with screenshots that your GRANT statements do distinguish among different kinds of users in regard to what they can do with the database.*

*Total points possible: 14*

ENTER YOUR WORK WITH DATABASE SECURITY HERE

# Locking and Concurrent Access

*Description: Explain the purpose of locking tables and show how to do that to prevent inconsistencies that may arise in your data when concurrent transactions take place.*

*Rubric: Your work will be graded as follows:*

* *3 points for clearly explaining an example that shows why you should lock tables to prevent inconsistencies.*
* *3 points for providing a screenshot and accompanying explanation of locking tables.*

*Total points possible: 6*

ENTER YOUR WORK WITH LOCKING AND CONCURRENT ACCESS HERE

# Backing Up Your Database

*Description: How you will back up your database. What commands will you issue? How frequently will the commands run? How can they be automated? Where will the backups be stored?*

*Rubric: Your work will be graded as follows:*

* *12 points for clearly explaining and justifying your database backup strategy, including the frequency with which you will back up the database, how you will automate backups, where you will store them, and how you will secure them. You will earn three points for addressing each factor (frequency, location, automation, and security)*
* *3 points for providing a screenshot of the command you would issue to back up the database and for including a portion of the resulting file.*

*Total points possible: 15*

ENTER YOUR WORK ON DATABASE BACKUPS HERE

# Python Programming

*Description: Write a Python program that generates a report that contains a subset of the data from your database. Include the code for your Python program in your Word document, and also post the program to your GitHub repository.*

*Rubric: Your work will be graded as follows:*

* *12 points for writing a Python script (and including its code in the Word doc) that will pull data from a database and store it to a text file and present it to the screen. Your code must have comments in it that explain how it works. You will be awarded 3 points for successfully connecting to the database, 3 points for successfully querying it, and 4 points for presenting the data to the screen and to a file. Internal comments count for 2 points.*
* *2 points for posting the code to GitHub*
* *4 points for showing a screenshot of your running the script and showing the results it produces on the screen.*

*Total points possible: 18*

ENTER YOUR PYTHON DATABASE PROGRAMMING WORK HERE

# PHP Programming

*Description: Build an HTML form that enables the user to specify criteria to search by. Use PHP to show the results of the query on a resulting web page. Make sure you include protections against an SQL injection attack. Include your HTML and PHP code in your Word document, and also post the files to your GitHub repository.*

*Rubric: Your work will be graded as follows:*

* *4 points for writing an HTML form the user will use to enter search criteria*
* *8 points for a PHP script that uses the search criteria and returns results*
* *4 points for an HTML page that shows the results*
* *4 points for explaining what SQL injection might be run on your website and explaining how you prevented it.*
* *4 points for providing screen shots of your PHP website in action.*
* *2 points for posting your code to GitHub*

*Total points possible: 26*

ENTER YOUR PHP DATABASE APP PROGRAMMING WORK HERE

# Suggested Future Work

*Description: Describe the limitations of your current database and explain how you or someone else could improve the design to address these shortcomings. Also describe how you might take advantage of leverage cloud services to increase the performance and availability of your database. Finally, explain the advantages and disadvantages of storing your data in a NoSQL format instead.*

*Rubric: Your work will be graded as follows:*

* *3 points for clearly describing the limitations of your databases*
* *3 points for explaining how you would address these shortcomings*
* *3 points for explaining how you might migrate the database to the cloud and describing what advantages you might gain from doing that.*
* *3 points for explaining the advantages and disadvantages of storing your data in a document-based NoSQL format instead.*

*Total points possible: 12*

ENTER YOUR SUGGESTED FUTURE WORK IDEAS HERE

# Activity Log

*Description: As an appendix, the team will keep a daily diary or log of their activity. What did you or your team study in this class each day? What did you learn? What did you accomplish or build or design? You don't have to enter something every day, but there should be at least three entries each week. Since we have eight weeks, that means you should make 3 posts to the Activity Log each week, for a total of at least 24 posts. Each post will be worth 1 point.*

*If you are working as part of a team, make sure you clearly identify which team member worked on which tasks. The Activity Log should help me figure out how each team member contributed to the project. If I cannot discern who worked on what aspects of the project from the activity log, no points will be awarded for it.*

*Total points possible: 24*

|  |  |  |
| --- | --- | --- |
| S. No | Activity | Week |
| 1 | Idea Evaluation | Week 1 and 2 |
| 2 | Initial Proposal |
| 3 | Data source identification |
| 4 | data gathering with the perspective of employee |
| 5 | data gathering with perspective of HR and manager |
| 6 | Identification of alternate ways to store data |