CS5200 DBMS Project Guidelines

Overview

We will approach this project like we are building a new product. Collaborate in a team of three-to-four members. Think about a product/service offering revolving around a strong relational model and how it can benefit a set of users. By the end of the semester, you will deliver a data-driven web application that is built on top of a relational model that you have defined. The project will focus on the design of your relational model, the insights you derive from your data, and how the web application and users interact with the relational model.

Course Syllabus

Link

Milestone 1: Definition

Assigned: 05/17. Due: 05/31.

Deliverables:

- One submission per team. On no more than 1 page, saved as a PDF.
- Team name, team members, project name.
- A one-sentence value proposition using the format:
  (Product name) Is A *(description of market and/or offering)*
  That Provides *(key benefits, state around three)*
  For *(target customers and/or segment)*
  Who Are Unhappy With *(current barriers, alternatives, and/or competitors).*
- What do you plan to deliver this semester, and what could you add in the future?
- Data sources and how you plan to extract the data. Include links to the data sources.
- Each member should be prepared to present the project definition in class on the due date.

Milestone 2: Relational Model and Data

Assigned: 05/31. Due: 06/14.

Deliverables:

- One submission per team, as a single PDF.
- UML for your relational model. At least 10 classes must be defined.
- CREATE TABLE statements for all tables. Ensure the CREATE TABLE statements are in an order that respect referential integrity. Also include DROP TABLE IF EXISTS statements at the top so the relational model can be easily recreated. I should be able to copy-paste and run your statements.
- Row counts for each table. You can submit a screenshot of the row counts in MySQL Workbench. At least 100K rows must be loaded (in aggregate).
- Each member should be prepared to demo the relational model implementation in Workbench in class on the due date.

Milestone 3: Business Insights

Assigned: 06/07. Due: 06/28.

Deliverables.

- One submission per team, as a single PDF file.
- Identify 10 questions that can be answered with your data. Make sure your questions are insightful and contribute to your value proposition.
- Provide a single SELECT statement to answer each question above. Make sure the SELECT statements are well formatted for ease of reading. Also include the result output. At least 5 SQL statements should use advanced features, such as functions, joins, aggregation, filtering after aggregation, ordering, etc.
- Does your relational model design suffice, or do you need to make changes to answer the questions? If so, then describe your changes and include an updated UML. Also include your new CREATE TABLE statements. I should be able to copy-paste to create your tables and then run your SELECT statements.
- Each member should be prepared to demo the SQL statements in class on the due date.

Milestone 4: Hello World

Assigned: 06/28. Due: 07/19.

Deliverables:

- One submission per team.
- Use JDBC to build the data access layer for all classes. Use JSP to build the web application.
- At least 4 screenshots that demonstrate the ability to perform the following operations on your database: create, read, update, delete (CRUD). Put the screenshots in a single PDF file. For each screenshot, provide a quick narrative of the workflow.
- Zip'ed Java code (no .rar files).
- Each member should be prepared to demo CRUD operations in class on the due date.

Milestone 5: Data Warehousing

Assigned: 07/12. Due: 07/26.

Deliverables:

- One submission per team.
- Combine your existing data with at least 2 external data sources that can provide interesting insights. Describe the external data sources used (including links, what information the data contains, how you plan to use the data). Include this in your final report.
- Define at least 2 ETL workflows that combine and/or transform the external data sources. Provide screenshots of the ETL workflows, and describe the ETL workflows (components used and what they do). Include this in your final report.
- Using Excel and/or Google Sheets, create at least 5 charts from your data warehouse (so the charts should reflect the results of ETLs and should utilize the external data). Include this in your final report for each chart:
  - Your hypothesis for combining the data.
  - The results of combining the data, and if it validates or invalidates your hypothesis.
  - Briefly describe the chart's significance for your application and the action you could take (if any) given the new information.
- Export the entire report as a single PDF: include the external data source descriptions, ETL descriptions, and chart descriptions/conclusions.
- Zip'ed CloverETL workspace (exclude large data files) and report (no .rar files).
- Each member should be prepared to demo the ETL workflows and present the visualizations in class on the due date.

Milestone 6: Conclusion

Assigned: 07/26. Due: 08/01.

Deliverables:

- One submission per team.
- Zip'ed application code (no .rar files).
- On no more than 2 pages, saved as a PDF:
  - Did you fulfill your value proposition? Explain.
  - List what you initially planned to deliver (as stated in Milestone 1), and compare to what you actually delivered.
  - Your final UML. Include a description of what changed in your UML (including which milestone did you make changes and why).
  - What went well?
  - What would you do differently?
  - What do you plan to do next?
- One submission per individual, no more than 1 page saved as PDF:
  - Did your team work well together? Why or why not?

- ○ Specifically, what did you contribute (implemented part of UML, loaded data source, wrote SELECT statements, wrote JDBC/JSP code, create ETL, etc.)?
- ○ Feedback for each of your team members.

Final Presentation
08/02

Deliverables:
- 30-minute presentation:
  - ○ Demo your final product/service.
  - ○ Summarize all the Milestone 6 conclusions into a slide deck.
- Presentation schedule: http://goo.gl/ENQiuN.