

Final Project

Due Dec 1, 2021 by 11:59pm **Points** 100 **Submitting** a file upload
File Types sql, txt, docx, pdf, ppt, and pptx **Available** until Dec 2, 2021 at 11:59pm

This assignment was locked Dec 2, 2021 at 11:59pm.

Instructions:

- The open-ended project provides you an opportunity to design and implement an integrated data management architecture for a real-world scenario.
- For this final project, you will work with your groups and present in-class.
- The presentation should take 10-minutes.
- Important note: peer evaluations will be applied to this project since this is a group project

Tasks:

1. Please select a real-world goal that includes database management component:
 - A. This can be a company from an industry where you have experience and/or an industry that you are passionate about (Such as: Fidelity, United Health, Walmart)
 - B. Or, alternatively, this can be a technical idea you would like to start and turn into a project but never found the opportunity (Such as: A new app or a system you want to develop)
 - C. Review publicly available information and then articulate what you believe should be the data management strategy, including the benefits that will come from this strategy
 - D. For this part of the project, you should be guided by the discussions we had in class to motivate the value of this project
 - E. You will want to be able to justify this decision and briefly share your reasoning during your presentation
 - F. This is a suitable step to prepare one (or more) motivation slide for your presentation
2. For the selected company (or idea), hypothesize 3 transaction management applications that you believe this company needs to run their business.

- A. Examples of transaction management application include software solutions to support business processes such as order processing, accounts receivables, accounts payables, fixed asset management, inventory management, and call center operations
 - B. Figure out information requirements (Entities and Attributes)
 - C. For each of these applications, articulate the data model of the relational database that supports the application using LucidChart (or any other ER diagram tool we used in HW1)
 - D. Please keep in mind that the workload supported by a transactional database involves, for the most part, single row inserts and updates and simple read queries that do not include many joins
 - E. Write the DDL for this data model and implement this model in Oracle
 - F. Next, either using publicly available data or manually created data, seed the tables with data. You will not be evaluated on the volume of seeded data
 - G. This is a suitable step to prepare an ERD slide for your presentation
3. Use the 3 transactional databases created in the previous step, to design the data model for an enterprise data warehouse for this company using LucidChart.
- A. If you need more information on DW Design and ETL, read [the file I posted on data warehousing](https://utexas.instructure.com/courses/1318318/files/63379513/download?download_frd=1) ↓ (https://utexas.instructure.com/courses/1318318/files/63379513/download?download_frd=1)
- You can also find the file under Canvas/Files/[381 Class 23 Reading - DW Design and ETL.pdf](https://utexas.instructure.com/courses/1318318/files/63379513/download?download_frd=1) ↓ (https://utexas.instructure.com/courses/1318318/files/63379513/download?download_frd=1)
- B. Please keep in mind that the workload supported by an enterprise data warehouse involves, for the most part, high volume batch inserts and complex queries that involve multiple joins
 - C. Write the DDL for the enterprise data warehouse data model and implement it in Oracle
 - D. Develop the ETL that will source the data from the 3 transactional databases and populate the data in the data warehouse in Oracle
 - E. Articulate and document the ETL decisions you are making in this process
 - F. This is a suitable step to start preparing the code and technical slides on the database design
4. Starting with the enterprise data warehouse, select a subset of data to create a data lake
- A. This section might use a NoSQL database of your choice (you don't actually implement

anything but if your information requirements include non-relational data include this in your slides)

- B. Articulate the rationale for the subset of data that you have selected
 - C. Augment the data lake with additional data elements that are required for the specific purpose of the data lake (For example, if the data lake is for marketing what other data elements or denormalizations will you need to develop to make the data lake easy to use for marketing)
 - D. This is a suitable step to prepare a "Beyond Relational" slide for your presentation
5. Connect this project to the other courses
- A. Implement at least two learnings from your other courses to this project (For example, you can include analysis patterns against this data lake from your machine learning or other courses)
 - B. Articulate the importance and value from the analysis and relate it back to the benefits we were expecting from the data strategy
6. Critically reflect on your learning from executing this final project
- A. Answer the following questions:
 - B. What did you learn?
 - C. What was most valuable?
 - D. How can you use this learning going forward?
 - E. What are additional opportunities for learning that this project did not capture?
 - F. How can we change this project to capture these opportunities?

Come to class and present. More importantly, have fun

Deliverables (You will present this project)

- **You will present this project, in-class, with your team.** There is no chance to change the dates after we set them because this is the end of the semester
- **You will turn in (submit) the presentation slides along with any code you wrote**
- **Obviously, you will prepare presentation slides (typically MS PowerPoint but any other common presentation tool is acceptable)**
 - You will use these slides to present your final project in class, and then you will submit them under the Canvas assignment

under the Canvas assignment.

- Presentation should be 10 minutes maximum
- **Script in a (.sql) file format.** If you have issues doing this, at least save it as a .txt file. Do not submit in any other format other than .sql and .txt.
 - Save your code into one SQL file with the naming format: Group_Number.SQL. Please make sure the team number you use matches what is in Canvas/People/Groups. For example:
Team_1.SQL
 - Save your file either as a .sql file or as a .txt file
- I recommend submitting both files before you present. I would be fine with a little delay such as a few hours later the same day, but please do not delay Canvas submission. If I don't see any files during grading, your entire team can get a zero.