

CSC 400, Final Project Directions, Spring 2024

The final project for this course will be in place of a final exam and cover the last third of the course. The focus of the material is databases, data warehouses (and data lakes), big data, OLAP, parallel processing, etc.

Deliverables: The main deliverable is a report of 10-12 pages in length (not counting references and title page), with a minimum of 5 cited sources. The structure of the project should follow this general scope:

- Describe the differences between databases, data warehouses and data lakes. What is the purpose and scope of each type of data storage architecture? How do they differ from each other in terms of size, data format, and query languages.
- Describe the data management lifecycle and how each one is impacted by the architecture of the data storage device.
- Describe parallel processing, the major architectures (such as Hadoop, Hive, etc.), and the types of processing that work well in such an environment. Be sure to note any types of processing/analysis that is not well-suited to such an environment.
- How does Online Analytic Processing (OLAP) work? Provide some examples of services. You should discuss at least one platform in detail (such as AWS, Apache Kylin, Azure Analysis Services, or others). Describe what services they offer, their cost, etc.
- Use the information you gained from the Census assignment as it relates to the above elements.
- Provide examples of the kinds of analysis that would be ideal for OLAP. Include size and context elements (you should describe situation, but you do not need to provide data).

All sources should be properly cited (you may use whatever standard you prefer, but APA is fairly common).

Your sources should be high quality sources. They should **not** include Wikipedia or ChatGPT. The report should be original and not a summary of other articles. If you use images to illustrate key details, the diagrams should be original, or properly cited. Each image included should be numbered and have a description underneath.

The structure of the report should have an abstract at the start. The main report should include an introduction, body and conclusion. The body should address each element described above. An estimated word count is around 4000 words.

The assignment is due the day of the scheduled final exam for this course (there is no in-class final exam). The project is worth **300 points**.