For the Final Project, you will be work in 4 to 6-person teams.

The project provides you an opportunity to design and implement an integrated data management architecture for a real-world company.

**Project description**

1. Please select a real-world company (e.g. Johnson & Johnson, United Health, Walmart) from an industry where you have experience and/or an industry that you are passionate about. Review publicly available information about this company and then articulate what you believe should be the data management strategy for this company, including the benefits that will accrue to the company from this strategy. For this part of the project, you should be guided by the case discussions in class. You will want to be able to justify this decision and briefly share your reasoning during your presentation.
2. For the selected company, hypothesize 3 transaction management applications that you believe this company needs to run their business.  
     
   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.   
     
   For each of these applications, articulate the data model of the relational database that supports the application using LucidChart.   
     
   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.   
     
   Write the DDL for this data model and implement this model in Oracle.   
   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.
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.   
     
   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.   
     
   Write the DDL for the enterprise data warehouse data model and implement it in Oracle.   
     
   Develop the ETL that will source the data from the 3 transactional databases and populate the data in the date warehouse in Oracle.   
     
   Articulate and document the ETL decisions you are making in this process.
4. Starting with the enterprise data warehouse, select a subset of data to create a data lake using HBase, MongoDB or another NoSQL database of your choice. Articulate the rationale for the subset of data that you have selected. 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.
5. Implement at least 2 analysis patterns against this data lake. 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.
   1. What did you learn?
   2. What was most valuable?
   3. How can you use this learning going forward?
   4. What are additional opportunities for learning that this project did not capture?
   5. How can we change this project to capture these opportunities?

**Deliverables**

1. Create a 20-minute PowerPoint presentation (you may use a presentation tool of your choice). All team members should have a role during the presentation.  
     
   The presentation should logically describe the work you have done for all the six steps outlined above including the results you have obtained.  
     
   The final slide should share your thoughts on what you learned, the value, and how you can use this learning going forward.
2. Submit a word document that includes the following:
   1. 4 LucidChart data models (3 transactional applications + 1 data warehouse)
   2. DDL for the four data models
   3. ETL code to populate the data models
   4. The model for your data lake
   5. The code for your analysis patterns

**Grading Rubric**

1. 8 points for each of the 6 deliverables as described in the PowerPoint, total of 48 points.
2. 32 points for the deliverables in the Word document.
3. 10 points – faculty discretion regarding presentation and team dynamics
4. 10 points – peer grading.