Contents

0.0 Description

0.1 Group Assignments and Scoring

This document outlines the requirements for the final project. As a class, we will split up into teams. Teams develop a project that incorporates the spectrum of content covered in the class. The project will provide students with the opportunity to solve a real-world problem using a data warehouse and business intelligence solution. We will check-in on the final project midway throughout the semester. **The project is due at the end of the last week of class via email submission.**

- **Groups: TBD**

The project will have several components:

1. Peer Score
2. Design and Planning
3. Development of the Warehouse Solution
4. Visualization
5. Executive Presentation

The last two weeks of class are reserved for in class finalization of the project and presentations to the group. Here is a summary of how you will be graded:

COMPONENT	DESCRIPTION	POINTS
PEER SCORE	You will fill out a survey about your peers in your Group Assignments	5 pts
DESIGN AND PLANNING	Several deliverables will be submitted as part of the assignment	5 pts
DEVELOPMENT OF THE WAREHOUSE SOLUTION	You are required to develop a working solution of the design and development of what you designed	5 pts
VISUALIZATION	Some visualization of the data flowing through your solution is important and must be a component to the project	5 pts
EXECUTIVE PRESENTATION	You need to prepare a summary of the work you developed for executives, which includes a demo	5 pts

	of the end-to-end solution. This is to be presented during class.
TOTAL	25 pts

0.1 Choosing a topic

It is important that you choose a topic that is simple enough to develop in the period of the class. Choose a topic that you know or is relevant to you. It can be a problem at work, a capstone to the program, or you can consult me on the topic and I can help you with this. This is important, as your ability to develop the solution is critical to the success of the project. You should test and validate the concept by prototyping it in the platform as this will help you gauge the level of effort required for the topic you selected. Please reach out to me if you have questions on the topic selection. You should be able to develop the solution in the AWS Educate instance or an AWS Free Tier instance that you created in class. Thus, a low cost topic is ideal.

0.2 Core Requirements

The basic requirement is to design, development and construct an end-to-end data warehouse solution using AWS as the platform. The core components of the solution outside of the components to the project are summarized below:

Requirement	Description
Data Types	• Structured dataset: This can be any available structured dataset • Unstructured dataset: this can be documents with metadata, twitter APIs, information that can be streamed into the platform.
Integrations	• Batch/Migration – flat file interface to your system • Real-time – integrated to your solution
Platform	• Amazon Web Services • Any visualization software
Documentation	• You are expected to develop the necessary documentation outlined in each of the subsections
Architectural Components	• You should be using fundamental architectural components to the solution: source, ingestion, staging, processing, visualization. • You have full license to be creative in designing your solution • You should think about core network and data privacy considerations as you develop your solution • You should have a development environment and a production environment to manage code and testing

1.0 Peer Score

A survey will be provided to each student at the end of class to score your group's participation. Each student will be evaluated by their peers based on their contributions and effort provided to the project. The combined score will be averaged and that will comprise 5 points of the total 25 points of the final project.

2.0 Design and Planning

As a practitioner, you will to work with the business representatives to come up with an appropriate solution. Ideally, you would work with these representatives throughout the design process, but here you are both the designer, planner, and a business representative. You will need to work autonomously and develop and articulate these requirements on your

own. This requires you to think about what is important to the ultimate consumer of your project. Be sure to think first about the result and work backwards using the techniques taught in class.

You are required to develop the following design and planning deliverables as part of your project submission:

- Bus matrix
- Basic use cases/user stories for analytics
- Dimensional Model Matrix with ETL instructions
- Conceptual, Logical, and Physical Data Model
- Architecture diagram with network information
- Data Dictionary
- High level waterfall project plan

It is strongly recommended that you review these deliverables with me prior to beginning the development of the project.

3.0 Development of the warehouse solution

Once you have designed the solution, you should begin developing the solution. It is entirely up to you on how to manage the code and split up the work, but it is highly recommended that you use code commit and some code repository to manage. You should manage your code using the best practices for deployment (including commenting, version and branch managing, etc.) from a development environment to a production environment and incorporate a plan for testing your code into your overall project.

As part of the submission, you will submit links and provide access to:

- Any ETL or ELT orchestration service you used (Matillion) if applicable
- Any code or scripts written (including any files executed in Lamba)
- Access to your repository with the latest, production ready code version for your solution
- Access to the databases and with example queries to test that access. For example, if you developed an OLAP database as part of the solution, you should add me to a relevant security group and provide me with the endpoint, port, and access credentials to run basic queries.

4.0 Visualization

Here is where you have the most creative license. The analytics you defined as part of your design document should be visualized accordingly. Ideally you would connect your AWS Instance to the Public Version of Tableau or use AWS QuickSight as a visualization service.

As part of the submission, you will provide links to your visualizations. They will also be incorporated as a demo in the presentation (no. 5.0).

5.0 Executive Presentation

Finally, you are required to present on the topic as if your final solution were presented to a team of Executives. The presentation should be 20 minutes, 15 minutes for the presentation and demo and about 5 minutes of questions and answers. You should use demos and Microsoft PowerPoint to deliver the presentation. Here is the summary of the format of the presentation below. You should have a minimum of these components to the presentation, but should feel free to add sections should you feel it necessary to make the presentation more accessible to the audience:

1. Executive Summary – Introduce yourselves, and summarize the problem and then your solution at a high level
2. Timeline and assumptions --- summarize the project milestones and the assumptions you made in solving the solution

3. Design summary--- describe the design of the solution using architecture diagrams, and data modeling description. This should be targeted toward someone who may not necessarily be familiar with data modeling
4. Demo of the solution – Showcase the analytics and the development of the solution. You can log-into an environment to show the solution, you can use screenshots. Be creative.
5. Challenges – Highlight the challenges you faced
6. Lessons learned and improvement– describe what you would do differently and what would be good opportunities for enhancements to the solution in the future