The assignment for this Module is a mini project. This is worth 20% of your total course grade.

The objective of this assignment is to incorporate everything you've learned in this course about data warehousing including **scheduling the project, ETL and report development**.

At this stage you will make Project Manager decisions and a certain part you will have hands on tasks. We'll go back to the initial scenario I mentioned, which will serve as the foundation for your project.

The Plot:

You've been assigned to lead a team that will create a data mart that will ultimately become a data warehouse to monitor transportation-related incidents.

The data mart will specifically monitor rail-related incidents throughout the state of Virginia by an organization whose focus is tracking incidents, preventing damage and accidents through regulations information, and education.

General facts are as follows:

* There are five inspectors whose job is to visit rail companies, rail stations rail tracks, et cetera to verify that they follow regulations to follow up on safety violations and to verify that post-incident steps and cleanup are being executed effectively on each trip.
* Inspectors need to hit as many rail locations as possible but there's also a backlog of locations that need inspection visits and an incident with damage can happen at any time requiring a change in the plan.
* In general, inspectors who were all based in Richmond would go on a 2 days inspection tour and then return. Due to the small staff of specialist inspectors, reaching the inspector level requires several years of training and expertise. To establish where the inspectors should travel for each trip, document incidents, track and avoid safety breaches, and make other agency judgments, the agency had to depend extensively on reports.
  + There are 10 major rail cities in Virginia and eight major rail segments.
  + There are 10 passenger lines, and a dozen commercial lines such as CSX, etc.
  + There are ten thousand companies that are directly or indirectly involved with Virginia railroads all over the state.
* There is a plan to double all this in the next few years. So now is the time to establish this data mart to centralize past information and prepare for the future.
* **An incident** is anything that might threaten the safety of people or of the rail system including accidents, collisions, rail cars slipping off tracks, trams going in the wrong direction, chemical events, other chemical spills, oil or chemical-related pollution, and one-time events: terrorist attacks, weather-related incidents.
* **A safety violation** is anything that might lead to any incidents, such as transporting paint or other chemicals in incorrect containers. Incorrect procedures for track maintenance poor electrical maintenance, driving at the wrong speeds approaching curves or tunnels, etc.
* So far, the agency has used three systems for its operational reports.
  1. **The commercial rail system** tracks events, and information about commercial rail companies and companies working with, or servicing the commercial firms, or working with the agency.
  2. **The passenger rail system** tracks everything related to passengers, down to the types of peanuts, and beer sold on the trains.
  3. **The SID (state inspection database)** tracks incidents, safety violations, and agency employee information.

Here are the requirements the client hands you when you start.

**You need to create three reports:**

1. Inspection priority by Inspector
2. Incident by damage costs
3. Incident by rail location

The agency heard that a data mart is the best way to build this, and wants an **incidence data mart**. The agency is depending on you to select the proper tools. But keep in mind that the mart will become a nationwide data warehouse with data that must be shared across similar state agencies.

The Data mart must be up and running in time for the Virginia state agency budget funding conference in two months. You have enough funding to put together a team of four, plus yourself to complete the project.

You have access to whomever you need to help with this project, except that the agency is understaffed and overloaded, and slightly underfunded and people are very busy. More specifically the system will have these users.

Five inspector users and three other staff to support them. Two managers that have a lot of experience. Besides the reports, the users want the ability to generate their own reports in the future. Here's a tiny sampling of the kinds of questions they want to ask of their data.

1. Determine incidents by time of year, at the city and county levels.
2. Conductor information by incidents with certification dates and higher dates.
3. Engineer information by the incident with certification dates, and higher dates.
4. Railroad location mechanical test dates and information.
5. Track equipment misplacement by Rail Road location.
6. Rail Road location personnel, toxicology reports.

Other queries are:

1. Aged trackinspection dates and scores
2. tank material properties per chemical release type
3. Age community exposure levels following chemical releases. Any number of other queries that the staff may want to be run.

Here are a few key tables you encounter in each of the three CIS source systems.

1. In the commercialrail system, you find the organization, equipment, personnel, schedules, freight cargo trips, and general ledger, among other tables.
2. In the passenger rail system, you find the company, passenger trips, trip history, staff, car supplies, cars, and inspection schedules.
3. In the SID database you find employee info, inspections, safety violations, Rail incidents, non-rail incidents, budget, rule book, training schedules, chemical registry, engine registry, equipment registry, and Rail segment info, among other things.
4. Use your experience and imagination to determine what data is in these tables. Visit URL <https://www.ntsb.gov/investigations/AccidentReports/Pages/Reports.aspx?mode=Railroad>

**Project Deliverables**

1. Draw the overall data warehouse diagram including the sources, the data mark and the layer.
2. How would you plan the scope schedule and cost to get the data mart done?
3. What team do you need to get this data mart done?
4. List the positions and high-level tasks such as project management etc.
5. Using the source information available to you, create a data model of your data mart.
   * This should be a Hi-Level model showing the entities and relationships. You don't need to specify any attributes.
6. Select an ETL tool. Why that tool? Create a sample source to target mapping with three rows for every table you will need that will provide data for your data mart’s data model.
7. What types of reports will you create?
8. The result should include a scope statement of Gantt chart and budget for the entire project from requirements to implementation in two months. Note: I have separately uploaded another video on how to make Gantt charts using Excel