You work as an analyst for MSBA Financial Group, a specialized investment firm. MSBA holds information on publicly traded companies from three different sources, saved in the following tables:

- A vendor, DataCorp, provides a collection of data from company balance sheets and income statements into the table, **datacorp_financials.csv.**
- Analyst A is responsible for calculating certain important accounting ratios, using data from DataCorp. These ratios are stored in a table called **msba_fg_ratios.csv**.
- Finally, Analyst B is responsible for tracking publicly traded companies that have filed for bankruptcy. They store this data in a table called **msba_fg_bankruptcy.txt**.
- A **data dictionary** including data field definition and explanations can be found at MSBA Financial Group Data Dictionary.

The new Chief Information Office (CIO) is frustrated with the decentralized nature of MSBA's data architecture. She has tasked you with prototyping a cloud-native data architecture in AWS that will achieve the following objectives:

1. Provide a landing place for incoming data of various data types, sizes, and sources.
2. Provide a data warehouse that can facilitate a more centralized "single source of truth" of data for the company's analyst teams to use regularly. This prototype warehouse should house the data from DataCorp, Accounting Ratios from Analyst A, and Bankruptcy Data from Analyst B.
    a. For example, this may look like a table with balance sheet financials and income statement financials consolidated for easier downstream consumption, along with bankruptcy data gathered by Analyst B in a separate table. Use your critical thinking to design what you believe makes sense for the organization.
3. Optional: Make future data ingestion easy to replicate or automate. (You are not tasked with automating data ingestion, only creating a system that could be easily upscaled when the need arises.)
4. Connect to Machine Learning tools that can:
    a. perform cursory Exploratory Data Analysis, and
    b. create a Machine Learning model that can predict a company's likelihood of bankruptcy in the next fiscal year.
5. Use the model to make a prediction. MSBA is considering 10 companies for a new portfolio. Their information is saved in **company_profiles_to_predict_unlabeled.csv.** Use the model to determine each of the 10 companies' likelihood of filing for bankruptcy.

Finally, you have been invited to make a 3-5 minute presentation on your project to the CIO. She is familiar with the cloud and with AWS in general, but is less familiar with its data-specific tools. She has asked that your presentation include:

- a data flow diagram outlining the architecture of the project;
- a demonstration (live or screen shot) of each step in the data flow (including a very brief explanation of the tool(s) used for that step, as well as its function or purpose);
- at least two findings from Task 4a that will be interesting to the portfolio strategy team;
- your recommendation regarding which, if any, of the 10 companies from Task 5 that MSBA Financial Group should include in their new portfolio.

SUBMISSION:

Record your video presentation and submit a URL link to your video presentation via Canvas. ScreenPal is preferred, but you may use Loom, Zoom, or another free recording software.

Note: If Senior Leadership gives you 3-5 minutes for a presentation, you should ALWAYS stay within that time constraint. Just as your CIO will stop listening to you after 5 minutes, I will end the video and will not watch anything after 5 minutes. **You will not receive credit for any requirements met after the 5:00 point of the video.**

RUBRIC:

Requirement		Points Possible
Data diagram clearly outlines the architecture of the project		4
Demonstration and explanation of each step in the project's data flow		
	Data Lake	1
	ETL or ELT	1
	Data Warehouse	1
	Machine Learning	1
2 Exploratory Data Analysis Findings		4
Bankruptcy prediction and subsequent recommendation		4
Professionalism and clarity in presentation		4
Total		**20**