Task: Building a Basic Custom Node-Flow System in Google Colab

Description:

In this task, you will demonstrate your ability as a Python developer by designing and implementing a basic custom node-flow system within a Google Colab notebook. The goal is to create a simplified workflow system that handles data extraction, transformation, and loading (ETL) tasks.

Requirements:

1. Data Source Selection: Choose a small public dataset (e.g., a CSV file) relevant to a hypothetical business problem. Ensure that the dataset contains multiple tables or data sources for variety.

2. Google Colab Setup: Set up a Google Colab notebook for this task to leverage Colab's capabilities for running Python code.

3. Custom Node-Flow System: Create a custom node-flow system with the following components:

   a. Node Definition: Define two Python functions or classes, one for data extraction and another for data transformation. These nodes should represent specific data processing tasks.

   b. Node Graph: Design a simplified directed acyclic graph (DAG) that represents the workflow of data processing nodes. The graph should include a clear start and end point.

   c. Node Execution: Implement a mechanism for executing nodes in the correct order based on their dependencies defined in the DAG.

4. Data Processing Steps: Use the custom node-flow system to perform basic data processing tasks:

   a. Data Extraction: Load the dataset from a file into memory.

   b. Data Transformation: Apply a simple data transformation (e.g., filtering, aggregation) to the loaded data.

5. Documentation and Explanation: Provide documentation within the Colab notebook that explains the purpose and logic of each data processing node and the structure of the simplified node-flow system.

Evaluation Criteria:

- Completeness and correctness of the basic custom node-flow system and data processing workflow.
- Simplicity and clarity of the workflow design.
- Clear and concise documentation and explanations within the notebook.
- Quality and maintainability of the Python code.

Submission:

You should submit the Google Colab notebook that contains the basic custom node-flow system, the data processing workflow, and clear instructions on how to set up and run the notebook.