**Assignment – Task**

Data Loading with Incremental Processing

1. Table Creation:

• Create 5 tables, where:

• Two tables should have a column with int data type that serves as a unique identifier.

• Three tables should have a datetime column to track the time of data insertion or update.

2. Watermark Table:

• Create a watermark table that will store the last processed value (int or datetime) for each table. This will be used to track the most recent value processed, ensuring that only new data is loaded during subsequent runs.

3. Stored Procedure for Incremental Data Load:

• Write a stored procedure that:

• Takes the table name and the last processed value (from the watermark table) as inputs.

• Fetches the new records based on the last processed value (delta or datetime).

• Updates the watermark table after successfully loading the new records.

4. Dynamic Pipeline Design:

• Design a dynamic data pipeline that performs the following:

• Extracts the latest data incrementally from the source tables.

• Uses the watermark table to track the last processed record for each table.

• Loads the extracted data to the target destination.

• Ensures that only new data is loaded into the target, preventing reprocessing of already loaded data.

5. Screenshots and Documentation:

• Provide screenshots of:

• The SQL script used to create the tables and watermark table.

• The stored procedure that implements the incremental load.

• The pipeline configuration, showing how the dynamic pipeline was built to execute the stored procedure and load data incrementally.

• The execution logs or flow to show that data was loaded successfully and incrementally.

6. Explanation:

• Explain, step-by-step, the approach used to implement the dynamic data loading pipeline. Discuss:

• How the watermark table ensures incremental loading.

• How the stored procedure fetches and processes the new data.

• The role of dynamic pipeline configuration in automating the data load process.