INCREMENTAL LOGIC EXPLANATION

Objective: The objective of incremental processing in this task is to process only new or changed data instead of reprocessing the entire dataset during every pipeline run. This approach improves scalability, reduces compute cost, and ensures efficient pipeline execution for large-scale data.

- The pipeline implements timestamp-based incremental ingestion using a watermarking approach.
- Only records with a transaction_timestamp greater than the last successfully processed timestamp are considered during each pipeline run.
- The Silver layer reads the last(max) transaction timestamp from the existing Silver table:
    - If the Silver table exists the watermark is retrieved
    - If it is the first run then full load is performed
- During subsequent runs, Bronze data is filtered using the watermark:
    - Records older than or equal to the watermark are ignored
    - Only newly arrived records flow through the Silver pipeline

- last_processed_ts = None

  ```
  if spark.catalog.tableExists("main.default.silver_sales_s3"):
      last_processed_ts = spark.sql("""
          SELECT max(transaction_timestamp)
          FROM main.default.silver_sales_s3
      """).collect()[0][0]
  ```

  This is the piece of code that deals with incremental logic explanation.