# 🎯 Challenge: InsightHub – Resilient Sales Aggregator

## 🔍 Assumptions & Thought Process

- **Network instability** is realistic for distributed pipelines; we simulate retry with exponential backoff.

- **Data may be malformed or inconsistent**, requiring robust validation.

- **Duplicates are defined by same store_id + product_id + timestamp ±2s.** I'll sort and deduplicate accordingly.

- **Output must include product-level aggregation** (top 5 products) and additional stats.

- **Code is modular and readable**, leveraging OOP for Sales and Batch logic.

## 🧠 Candidate Explanation Notes

- **OOP Use**: Used `Sale` as a typed structure for clarity and `SaleBatch` to encapsulate all logic for separation of concerns.

- **Edge Case Handling**: All invalid data (missing fields, type errors, suspicious totals, dupes) are safely skipped with metrics tracked.

- **De-Duplication**: Based on combination key with ±2s comparison using `DATE_DIFF`.

- **Retry Logic**: Implemented via exponential backoff using `SLEEP` and structured `TRY/CATCH`.

- **Sorting & Aggregation**: Output shows top 5 products by total revenue.