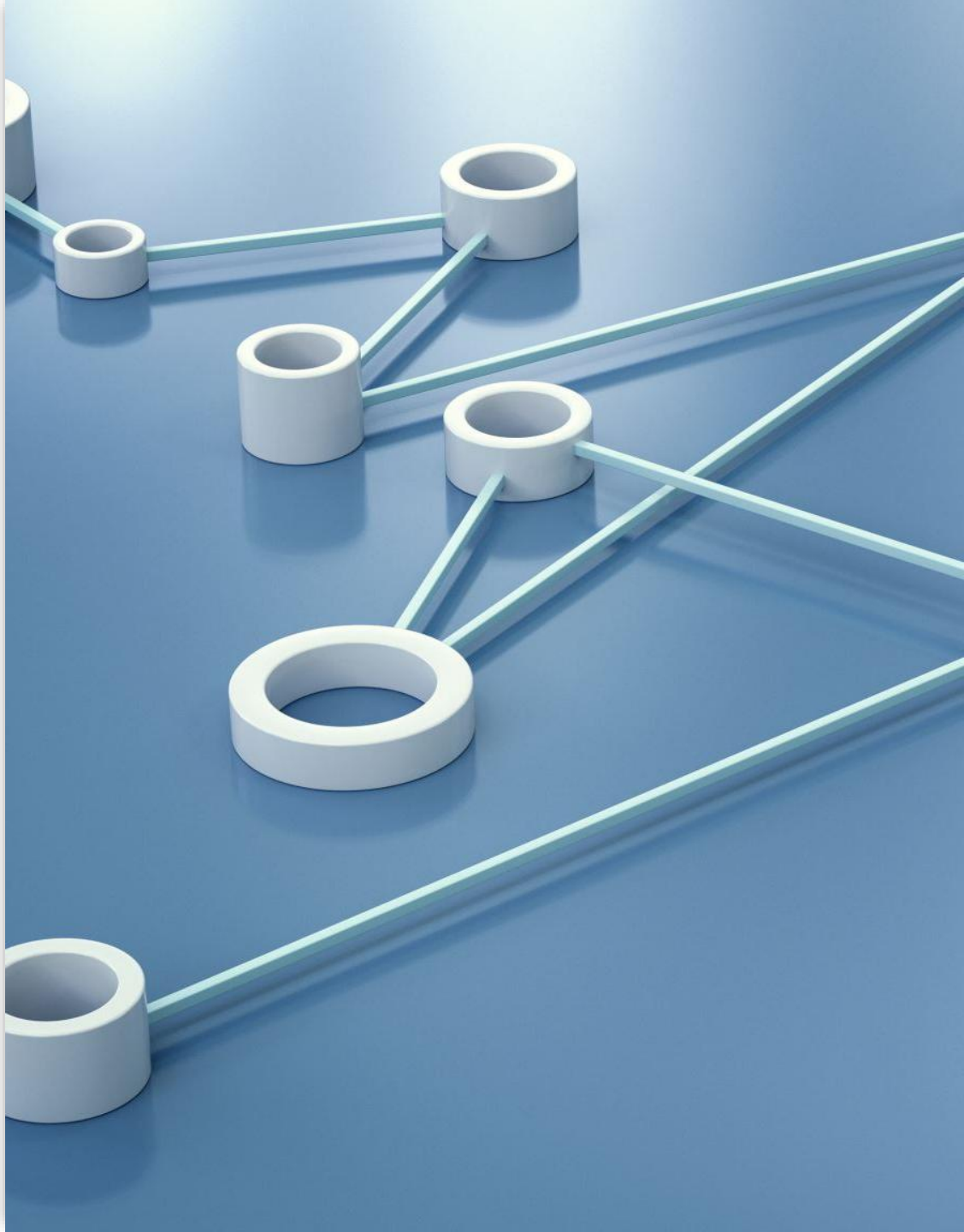
Data Architect Case Study – Data & Business Technology Executive

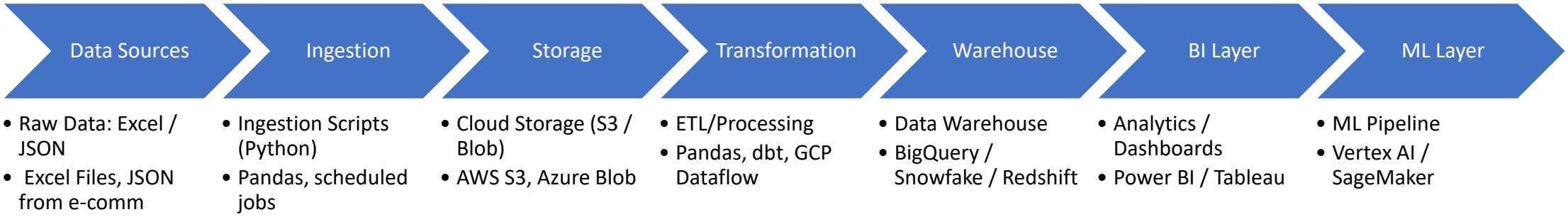
Emmanuel Ramon M. Cerrer

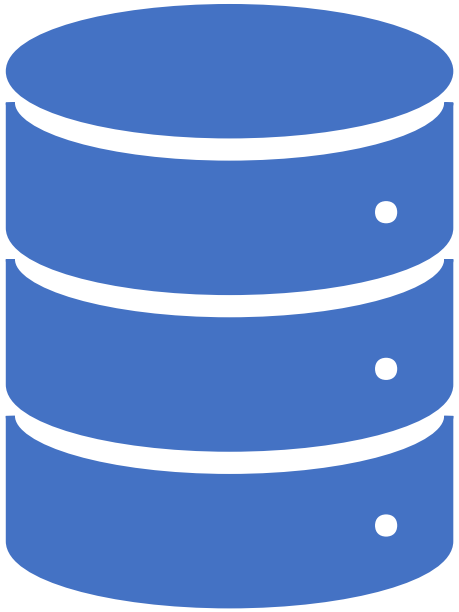
Business Case Summary

- The client is a mid-sized retail company with data coming from multiple sources:
  - B2B and B2C transactions
  - Logistics records
  - Real-time customer activity from their e-commerce platform
- Data is currently spread across Excel files and database systems.
- The client wants to:
  - Consolidate all data in the cloud
  - Enable weekly business reporting
  - Prepare data for advanced analytics and ML use cases
- Solution must be scalable, cost-effective, and able to handle data inconsistencies.



Data Architecture – Process Diagram Flow

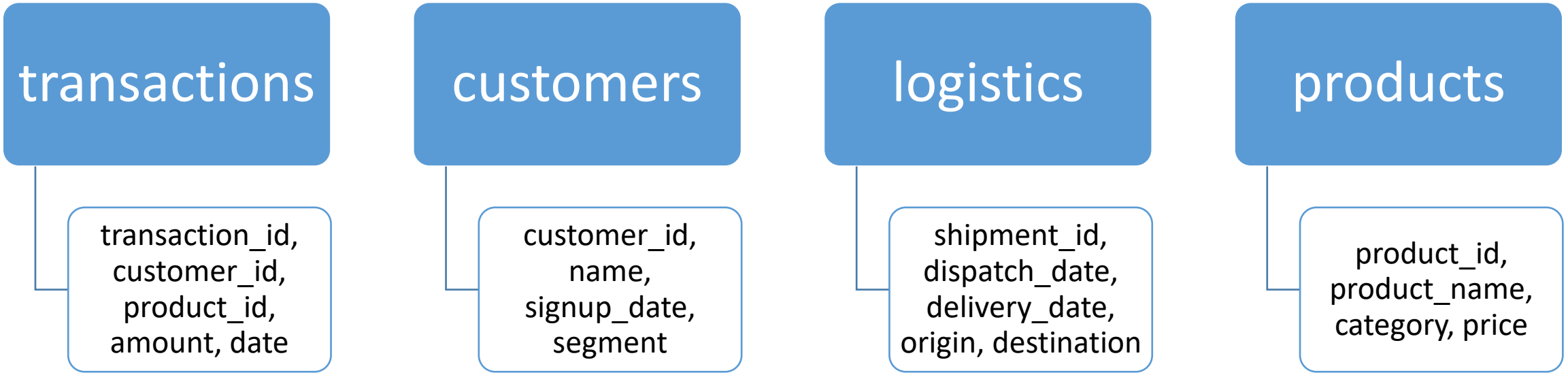




Data Flow Explanation

- This architecture supports the ingestion of raw B2B/B2C transactions (Excel), logistics data (Excel), and customer data (JSON). Using Python scripts, these are ingested into cloud object storage, transformed with pandas/dbt, stored in a cloud warehouse (e.g., BigQuery), and made available for business dashboards and ML pipelines.

Ideal Database Schema



Cost Estimate

Storage

- 50GB/day transactions: ~$25/month
- 10GB/month logistics: ~$2/month
- 100GB/day customer JSON: ~$50/month

Compute

- ETL/cleaning: ~$30/month

Total: ~$100–120/month

Known Issues & Mitigation

Inconsistent Excel headers

- schema mapping

Missing values

- fillna() or dropna()

Delayed JSON ingestion

- retry + logging