

Banking Data Engineering – Interview Scenario Questions & Answers

Scenario 1: Pipeline ran twice for the same day

Q: Your daily pipeline ran twice. How do you avoid duplicates?

A: In our Microsoft Fabric project, the Gold fact table is loaded by business date. Before inserting data, we delete existing records for the processing date partition and then append refreshed data. This ensures idempotent behavior and prevents duplicates even if the pipeline reruns.

Scenario 2: Late-arriving transactions

Q: What if a transaction arrives days later than expected?

A: Our pipeline processes data based on transaction date, not ingestion date. Late-arriving records are handled through controlled backfill runs by rerunning the pipeline with the affected process date. The Gold layer safely reprocesses that date without impacting other data.

Scenario 3: Dimension updated after fact load

Q: What happens when customer attributes change after transactions are loaded?

A: We use SCD Type 2 for customer dimensions. During enrichment, transactions join to the customer record valid at the transaction date. This preserves historical accuracy and ensures past reports do not change when attributes update later.

Scenario 4: Business reports mismatched numbers

Q: Dashboard numbers do not match expectations. What do you do?

A: I start by validating audit and reconciliation metrics between Silver and Gold layers. If those match, the issue is likely reporting filters or joins. If not, I trace the problem through enrichment logic and date filters before changing any code.

Scenario 5: Pipeline failure at night

Q: The pipeline failed overnight. How do you recover?

A: Fabric pipelines stop downstream execution on failure, so Gold data is not partially updated. Business users continue seeing the last successful data. After fixing the issue, we rerun the pipeline safely using idempotent logic.

Summary

These scenarios demonstrate production-ready thinking using Microsoft Fabric, Lakehouse architecture, SCD Type 2 modeling, idempotent pipelines, and audit-driven validation aligned with enterprise banking standards.