PDF 10: Data Migration Strategies

Subjective Case Study Questions and Answers

Q1. Migrating On-Prem SQL Server to Synapse
Scenario: A company wants to move historical sales data from SQL Server to Azure Synapse.
Answer: Use **Azure Data Factory** to copy data from on-prem SQL to Synapse. Enable **staging in Blob storage** for large datasets. Consider **schema mapping**, data validation, and incremental loads to minimize downtime.

Q2. Lift-and-Shift vs Refactoring
Scenario: Legacy ETL processes are on-prem.
Answer: Lift-and-shift moves workloads with minimal changes (e.g., VM-hosted SQL). Refactoring redesigns pipelines for cloud-native services like ADF and Synapse. Refactoring improves scalability, cost-efficiency, and cloud integration.

Q3. Handling Large Data Volumes
Scenario: 50 TB of historical logs must be migrated.
Answer: Use **staged bulk copy**, compress data, and perform migration during off-peak hours. Use **PolyBase** in Synapse for fast ingestion and consider partitioning to improve load performance.

Q4. Validating Migration Success
Scenario: Data must be 100% consistent post-migration.
Answer: Use row counts, checksums, and sample queries to validate. Implement reconciliation reports and confirm data integrity before switching production workloads.

MCQs

1. Fastest method for bulk data migration?
   Answer: C. PolyBase — parallel ingestion into Synapse.
2. When to refactor ETL?
   Answer: B. For cloud-native optimization — scalability and maintainability.
3. Ensuring post-migration data integrity?
   Answer: A. Checksums + reconciliation — detect discrepancies.
4. Minimizing downtime during migration?
   Answer: B. Incremental loads — only move changed data.
5. Compressing data helps?
   Answer: C. Reduce transfer time and storage costs.
6. Lift-and-shift approach?
   Answer: B. Minimal changes to migrate as-is workloads.
7. Staging data in Blob storage is useful when?
   Answer: A. Handling very large datasets — intermediate storage.
8. Validating schema consistency?
   Answer: C. Compare source and destination columns and data types.