PDF 1: Data Storage Design

Subjective Case Study Questions and Answers

Q1. Azure Synapse Storage Architecture

Scenario: You are designing an analytical solution using Azure Synapse Analytics for a retail company. The solution should handle large historical datasets stored in Azure Data Lake and provide fast query performance.
Answer: The Dedicated SQL Pool is suitable for predictable, high-performance workloads where data is pre-loaded and optimized for analysis using distributed storage and compute. It supports partitioning, indexing, and materialized views for faster queries. Serverless SQL Pool is suitable for ad-hoc analysis on unstructured or external data without ETL overhead. For repeated analytical queries on large structured data, a Dedicated SQL Pool is ideal.

Q2. Choosing Between Azure SQL Database and Cosmos DB

Scenario: A global e-commerce company needs a low-latency, globally distributed database to handle product catalog and user session data.
Answer: Azure Cosmos DB provides global distribution, multi-region writes, and tunable consistency. Using the Session consistency model ensures each user sees their own updates. Replication across regions reduces latency for international users.

Q3. Optimizing Storage Costs

Scenario: Your client stores 500 TB of raw IoT telemetry data. They want to reduce costs without losing analytics ability.
Answer: Use Azure Data Lake Storage Gen2 lifecycle management to move old data automatically from Hot to Cool or Archive tiers. Infrequently accessed data can be moved to Cool after 30 days and Archive after 90 days, lowering costs while retaining accessibility.

Q4. Designing for High Availability

Scenario: Designing a mission-critical banking application requiring 99.99% availability.
Answer: Use Azure SQL Database with Active Geo-Replication across regions. Readable secondary databases and automatic failover ensure high availability. Zone-redundant storage ensures data durability across availability zones.

MCQs

1. Which Azure storage solution is most suitable for unstructured big data analytics?
   Answer: B. Azure Data Lake Storage Gen2 — supports hierarchical namespace and analytics workloads.
2. Which Synapse feature reduces query latency on frequently accessed data?
   Answer: A. Materialized Views — pre-compute and store results for faster queries.
3. Main advantage of Cosmos DB over Azure SQL for global apps?
   Answer: C. Multi-region writes with low latency — active-active replication.
4. Best tier for rarely accessed data retained for compliance?
   Answer: C. Archive — lowest-cost storage for long-term retention.
5. Which ensures automatic failover in Azure SQL Database?
   Answer: B. Geo-Replication — enables failover across regions.
6. Serverless SQL Pool is ideal when:
   Answer: B. On-demand query execution over external data.

7. Service with automated data tiering?
   Answer: C. Azure Data Lake Storage Gen2 — supports lifecycle policies.
8. To achieve 99.99% database availability?
   Answer: C. SQL Database with Zone Redundant Configuration — replication across zones.