

01 Azure Cloud Overview and Azure Storage

Azure Cloud Overview and Azure Storage

Basic Level Questions (1-8)

1. What is Microsoft Azure and what are its main benefits for data engineering projects?

What to look for: Understanding of cloud computing basics, cost efficiency, scalability, global reach, and managed services.

2. What is the difference between Azure Blob Storage, Azure Files, and Azure Queue Storage?

What to look for: Clear distinction between unstructured data storage (Blob), file shares (Files), and message queuing (Queue).

3. Explain the different Azure Blob Storage access tiers and when you would use each.

What to look for: Hot, Cool, and Archive tiers with understanding of cost vs. access frequency trade-offs.

4. What is an Azure Storage Account and what authentication methods can be used to access it?

What to look for: Container concept, access keys, SAS tokens, Azure AD authentication, and managed identities.

5. What is Azure Data Lake Storage Gen2 and how does it differ from regular Blob Storage?

What to look for: Hierarchical namespace, POSIX permissions, better performance for analytics workloads.

6. Explain the concept of Resource Groups in Azure and their importance in data projects.

What to look for: Logical grouping, lifecycle management, access control, and billing organization.

7. What are the main Azure regions and availability zones, and why do they matter for data storage?

What to look for: Geographic distribution, disaster recovery, compliance, and latency considerations.

8. What is Azure Resource Manager (ARM) and how does it help in managing cloud resources?

What to look for: Declarative templates, consistent management layer, and infrastructure as code concepts.

Intermediate Level Questions (9-17)

9. How would you implement data lifecycle management in Azure Blob Storage?

What to look for: Lifecycle policies, automated tier transitions, deletion rules, and cost optimization strategies.

10. Explain the difference between locally redundant storage (LRS), zone-redundant storage (ZRS), and geo-redundant storage (GRS).

What to look for: Durability levels, cost implications, and appropriate use cases for each redundancy option.

11. How would you securely transfer large datasets (multiple TBs) to Azure Storage?

What to look for: Azure Data Box, AzCopy, Azure Storage Explorer, network optimization, and security considerations.

12. What is Azure Private Link and how would you use it to secure access to storage accounts?

What to look for: Private endpoints, network isolation, VNet integration, and security best practices.

13. How would you monitor and troubleshoot performance issues in Azure Storage?

What to look for: Azure Monitor, Storage Analytics, metrics, diagnostic logs, and performance optimization techniques.

14. Explain how you would implement data partitioning strategies in Azure Data Lake Storage Gen2.

What to look for: Folder structure design, date-based partitioning, performance optimization, and query efficiency.

15. What are the different ways to encrypt data in Azure Storage?

What to look for: Encryption at rest, encryption in transit, customer-managed keys, Azure Key Vault integration.

16. How would you implement a data archival strategy using Azure Storage?

What to look for: Archive tier usage, retrieval strategies, cost analysis, and compliance requirements.

17. Describe how you would set up cross-region replication for critical data in Azure.

What to look for: Read-access geo-redundant storage (RA-GRS), failover procedures, and RTO/RPO considerations.

Advanced/Difficult Level Questions (18-25)

18. Design a multi-tenant data lake architecture in Azure with proper isolation and security.

What to look for: Container/folder strategies, RBAC implementation, data governance, and compliance considerations.

19. How would you optimize Azure Storage performance for a high-throughput data ingestion pipeline processing 100GB/hour?

What to look for: Parallel uploads, optimal blob sizes, hot partitions avoidance, and network optimization.

20. Explain how you would implement a disaster recovery strategy for a data lake spanning multiple Azure regions.

What to look for: RTO/RPO requirements, automated failover, data consistency, and testing procedures.

21. How would you handle schema evolution in a data lake while maintaining backward compatibility?

What to look for: Versioning strategies, metadata management, delta lake concepts, and migration approaches.

22. Design a cost optimization strategy for a data lake with 500TB of data across different access patterns.

What to look for: Intelligent tiering, lifecycle policies, compression strategies, and cost monitoring.

23. How would you implement data lineage and governance in an Azure-based data lake environment?

What to look for: Azure Purview, metadata management, data cataloging, and compliance tracking.

24. Explain how you would handle concurrent access and consistency issues in a large-scale data lake with multiple writers.

What to look for: Delta Lake, optimistic concurrency, transaction handling, and conflict resolution strategies.

25. Design a real-time data processing architecture that combines Azure Storage with stream processing services.

What to look for: Event-driven architecture, Azure Event Hubs, Stream Analytics, change data capture, and lambda architecture patterns.