Day 25: Basics of Azure Cosmos DB

Welcome to Day 25 of our Azure Data Engineer interview questions and answers series! Today, we will delve into Azure Cosmos DB, a globally distributed, multi-model database service.

1. What is Azure Cosmos DB and what are its key features?

Answer: Azure Cosmos DB is a globally distributed, multi-model database service provided by Microsoft Azure. Key features include:

- Global distribution
- Multi-model support (e.g., document, key-value, graph, and column-family)
- Horizontal scaling
- Low latency with guaranteed speed at the 99th percentile
- Automatic indexing
- Multi-region writes
- Comprehensive SLA guarantees

2. What are the different consistency levels offered by Azure Cosmos DB?

Answer: Azure Cosmos DB offers five consistency levels:

- Strong: Guarantees linearizability.
- Bounded Staleness: Guarantees reads lag behind writes by at most k prefixes or t time.
- Session: Guarantees within a single client session.
- Consistent Prefix: Guarantees reads never see out-of-order writes.
- Eventual: Guarantees eventual consistency without ordering guarantees.

3. How does Azure Cosmos DB ensure high availability and low latency?

Answer: Azure Cosmos DB ensures high availability and low latency through:

- Multi-region replication: Distributes data across multiple regions.
- Partitioning: Distributes data across multiple partitions for scalability and low latency.
- SLA guarantees: Provides SLAs for availability, latency, throughput, and consistency.

4. Explain the concept of partitioning in Azure Cosmos DB.

Answer: Partitioning in Azure Cosmos DB involves distributing data across multiple partitions to achieve scalability and manage large datasets efficiently. Each partition is a distributed subset of the data and is managed by a partition key that determines how data is divided among partitions. This ensures efficient query processing and load balancing.

5. What are the different APIs supported by Azure Cosmos DB, and how do they differ?

Answer: Azure Cosmos DB supports multiple APIs, including:

- SQL API: For document-oriented data using SQL-like syntax.
- MongoDB API: For compatibility with MongoDB applications.
- Gremlin API: For graph databases and traversals.
- Cassandra API: For column-family data models compatible with Cassandra.
- Table API: For key-value data models compatible with Azure Table Storage. Each API is designed to interact with Cosmos DB using the respective data model and query language.

6. How does Azure Cosmos DB handle indexing, and what are the benefits?

Answer: Azure Cosmos DB automatically indexes all data without requiring schema or secondary indexes. This ensures high performance and query efficiency. Benefits include:

- Simplified development: No need to manage indexes manually.
- Improved query performance: Fast and efficient query execution.
- Flexibility: Indexes adapt automatically to data model changes.

7. What is the Request Units (RUs) model in Azure Cosmos DB?

Answer: The Request Units (RUs) model in Azure Cosmos DB is a performance currency that abstracts the cost of database operations (reads, writes, queries). RUs provide a uniform measure of throughput, allowing users to provision and scale their database's performance according to the workload's needs.

8. Describe the process of migrating data to Azure Cosmos DB.

Answer: Migrating data to Azure Cosmos DB involves:

- Assessment: Evaluate the source data and determine the target Cosmos DB model.
- Data Preparation: Transform and prepare data for compatibility with Cosmos DB.
- Tools: Use migration tools like Azure Data Factory, Cosmos DB Data Migration Tool, or custom scripts.
- Validation: Verify data integrity and consistency post-migration.
- Optimization: Optimize partitioning, indexing, and performance settings.

9. How does Azure Cosmos DB handle security and compliance?

Answer: Azure Cosmos DB handles security and compliance through:

- Encryption: Data is encrypted at rest and in transit.
- Access Control: Fine-grained access control using Azure Active Directory and role-based access control (RBAC).
- Compliance: Adheres to various industry standards and compliance certifications (e.g., ISO, HIPAA, GDPR).

10. Explain the pricing model of Azure Cosmos DB.

Answer: Azure Cosmos DB pricing is based on two primary factors:

- Provisioned Throughput: Measured in Request Units (RUs) per second, provisioned based on the required performance.
- Storage: Charged based on the amount of data stored, including indexes. Additional costs may include multi-region replication, backup, and network egress.