Q) What is zero copy cloning?

A) In Snowflake, zero-copy cloning is a feature that allows users to create fast and efficient copies of data without physically duplicating the underlying data files. Instead of creating full copies, Snowflake creates virtual copies that share the same data blocks, saving storage space and reducing data replication time. This approach significantly improves performance, reduces costs, and simplifies data management tasks, as each clone is independent and can be queried separately without affecting the original data.

Q) What is time trael?

A) Time Travel is a feature that allows users to access historical data at different points in time. It enables querying data as it existed in the past, present, or future, without the need for manual backups or data copies. Snowflake automatically retains historical versions of data for a configurable period, known as the Time Travel retention period, making it easy to analyze data changes and trends over time. Time Travel is a powerful capability that simplifies data auditing, compliance, and historical analysis tasks within the Snowflake data warehouse environment.

Q) Permanent table: It holds persistent data that remains available even after the session ends. It is typically used for long-term storage of important and frequently accessed data.

Transient Table: It is a temporary storage option for intermediate or temporary data. It is used for short-term operations and automatically gets deleted after a certain period of inactivity or at the end of a session.

Temporary Table: Temporary tables are created for specific sessions or tasks and are automatically dropped at the end of that session or transaction. They are useful for storing temporary data during complex queries.

External Table: An external table in Snowflake does not physically store data in the Snowflake database. Instead, it points to data residing in external cloud storage, like Amazon S3 or Azure Blob Storage.

Q) How to do failsafe and retention implementation on all the tables?

A) Failsafe implementation: Regular backups, data replication, disaster recovery plan.

Retention implementation: Data classification, define retention policies, automated data purging, data archiving, monitoring & auditing, security measures.

Q) What will happen in zero copy cloning when you change the data on source?

A) In zero-copy cloning, changes to the data on the source do not affect the cloned data. The clone maintains its own independent state after the initial copy.

Q) What will happen in zero copy cloning when you change the data on target?

A) In zero-copy cloning, changes to the data on the target (cloned data) only affect that specific clone. The source data and other clones remain unchanged and independent from the modifications made to the target.

Q) What is task in snowflake?

A) In Snowflake, a task is a predefined set of SQL statements and operations that can be scheduled to run at specific times or triggered by events. Tasks are used to automate data-related tasks, such as ETL processes, data loading, transformation, and other routine operations. They simplify data management and processing by allowing users to set up recurring jobs or respond to specific events, streamlining data workflows within the Snowflake data warehouse environment.

Q) How to load 10k records from salesforce to snowflake?

A) Export the data from Salesforce using data export or Salesforce API (e.g., REST or Bulk API). And transform the data to match the Snowflake schema if needed. Create a staging area in Snowflake to temporarily hold the data. Load the data from the staging area into the target table using Snowflake's COPY INTO or INSERT INTO commands. Verify the data load for accuracy and completeness.

Q) How do you perform or increase the performance tuning in your project?

A) Performance Analysis, Query Optimization, Data Model Optimization, Indexing and Partitioning, Hardware and Resource Configuration, Caching and In-Memory Processing, Parallel Processing, Compression and Data Pruning, Data Partitioning and Sharding, Query Performance Monitoring, Query Caching and Materialized Views, Optimized ETL/ELT Processes, Regular Performance Testing, Periodic Review and Maintenance.

Q) How do you handle duplicates in Kafka?

A) To handle duplicates in Kafka, utilize producer-side deduplication with unique identifiers, configure the producer to be idempotent, and implement consumer-side deduplication with state tracking or windowing techniques. For critical scenarios, enable Kafka's Exactly Once Semantics (EOS) to ensure strict message ordering and avoid duplicate processing. Thoughtful design and implementation of these strategies will maintain data integrity and optimize data processing in the Kafka ecosystem.

Q) How do you do data frame APIs in ADF?

A) Creating a Data Flow, adding Source and Sink and then adding Data Transformations and configuring Data Frame Transformations, Previewing, Validating and then finally we need to execute the Data Flow.