1. What is Zero copy cloning in snowflake?

   Creating a virtual clone of the existing data without consuming any additional storage, it shares the data with the original object. So any changes made to the clone will not affect the original object.

2. What is time travel?

   This feature allows us to go back to historical data points and query them as they existed at that specific timestamp. Retention period will be 1-90 days depending on the type of account. It is also helpful in recovering columns or values that are accidentally deleted. So querying the historical data creates a time travel clone, note that time travel consumes additional storage, and can be costly.

3. Transient vs temporary vs persistent vs external tables in snowflake
   - Temporary: tables that are created and that exist in a particular session with a long life span or until the user drops it. It is useful when computing intermediate results across multiple queries.
   - Transient: similar to temporary tables but shorter life span and it exists in that session for a particular query. It is lost automatically when query is executed or session is ended
   - Persistent: these are standard tables in snowflake and typically stored for long term or until the retention period ends, they can be shared across sessions, users and queries.
   - External: these tables can be created to query data from external sources directly, like S3, without loading the data into snowflake thus saving cost as well.

4. Failsafe and retention implications on types of tables in snowflake?
   - Temporary: not fail safe, can not be recovered after session ends. No retention, as they are deleted when session ends or dropped in the session.
   - Transient: not fail safe, since they are intended for short duration, they are not backed up and cannot be recovered after query ends. No retention, table gets deleted after query is executed automatically.
   - Persistent: failsafe, automatically backed up by snowflake, snapshots are maintained continuously to recover in case of accidental deletions. Default retention is 1 day (can be extended based on requirement)
   - External: snowflake cannot perform backups on external data and it should be managed by the platform. Same for retention policies as well.
5. What happens to zero copy cloning when you change the data on Source?

   the clone will only maintain snapshots of the data that was available at time of cloning, so it cannot track changes like updates or insertions on the source. The source and clone are isolated and do not get affected by changes.

6. What happens to the source when you change the data on a clone?

   When data in the clone is changed, it does not affect the source data, it only makes the modifications to the clone, any deletion or updates remain unchanged to the source. Changes made to clone do not get propagated to the source in any way.

7. What is a task in snowflake?

   Feature in snowflake that automates execution of SQL queries. They can be scheduled based on time or frequency, and can be integrated with external notification services to provide alerts, it maintains logs of the tasks executed as well.

8. How to load 10k records to snowflake?

   For larger datasets, the COPY command and crating a stage to s3

   ```
   CREATE OR REPLACE STAGE your_stage_name
   URL = 's3://your_s3_bucket/path'
   CREDENTIALS = (AWS_KEY_ID = 'your_aws_access_key' AWS_SECRET_KEY =
   'your_aws_secret_key');
   ```

9. How do you increase the performance tuning in a project?
   - Proper Sizing and Scaling of warehouses so that we dont pay extra for unused resources like in redshift or snowflake
   - Ensure Proper data partitioning and minimized data movement techniques for query performance
   - In snowflake we can use data compression to reduce storage costs and in S3 we can archive or store data in glacier for historical data which is not used frequently.
   - Separating services for storage and compute (external tables in snowflake)

10. How to handle duplicates in kafka

    Using Kafka Message Keys, if the message has unique keys, we can ensure that message with same key go to the same partition, then set a retention policy
    Or set up alerts to the admin if the duplicate rate exceeds a certain amount.

11. Read data from external API in json format and send notifications to certain teams/units how do you do it in AWS?
    - Created required IAM roles and permissions
    - Create an AWS Lambda function to fetch data from the external API.
    - API gateway endpoint is created to trigger lambda function.
    - HTTP request is stored in JSON format
    - Create an SNS topic to send each team the notification. (emails/sms)