1. You need to create a table to store sales data for a retail company. The table should be created such that querying on monthly sales is efficient. Write the SQL statement to create this partitioned table.
2. The `sales_data` table now needs a partition for the year 2024 and the month of August. Write the SQL statement to add this partition to the existing table.
3. You have sales data located at `/path/to/data` that needs to be loaded into the `sales_data` table for the year 2024 and the month of August. Write the SQL statement to load this data.
4. You are working on a data pipeline that ingests sensor data from IoT devices. The data has a simple schema but includes a mix of structured and semi-structured data, such as numerical readings and JSON metadata. Data is ingested continuously and then queried in Hive for periodic reporting and analysis. The schema is likely to evolve over time, and the file format should support schema evolution. Which file format would you choose in Hive to store this data, considering both flexibility and efficient query performance?
5. Write a complex INSERT statement to load data into a partitioned table `sales_partitioned` with `year` and `month` partitions. Assume `sales_data` is a temporary table with columns `product_id`, `product_name`, `sale_date`, and `sale_amount`.
6. You need to create a bucketed table `user_logs` that distributes data into 25 buckets based on `user_id`. Write the SQL statement to create this bucketed table.