**Practice Lab Questions: Data Vault**

**Lab 1: Hub, Link, Satellite Modeling**

* Identify three core business entities from a sample dataset. Define their business keys and design Hub tables for them.
* Determine relationships between these entities and design Link tables capturing these relationships.
* Create Satellites for storing descriptive, time-variant attributes for each Hub and Link.
* Write SQL scripts or use an ETL tool to populate these tables with sample data, generating appropriate hash keys and audit columns.

**Lab 2: Hash Key Generation & Collision Detection**

* Implement a process to generate hash keys for your business keys using MD5 or SHA-256.
* Load a large sample dataset and test for hash key collisions.
* If collisions occur, refine the hashing input (e.g., add delimiters, normalize case) and re-test.
* Document your collision testing methodology and results.

**Lab 3: Build and Query PIT Tables**

* Select a Hub and its related Satellites from your model.
* Create a PIT table that provides the latest satellite data snapshot at specific points in time.
* Write SQL queries that use the PIT table to fetch historical attribute values efficiently.
* Compare query performance with direct joins to satellites.

**Lab 4: Implement Bridge Tables for Many-to-Many Attributes**

* Identify a multi-valued attribute or many-to-many relationship in your dataset.
* Design and implement a Bridge table to capture this relationship.
* Populate the Bridge table and write queries that use it to support drill-across reporting.

**Lab 5: Raw Vault to Business Vault Pipeline**

* Load raw source data into Raw Vault tables (Hubs, Links, Satellites) without transformations.
* Develop and apply automation scripts/macros for hash key generation and audit column insertion.
* Build Business Vault layers by applying business rules and creating PIT and Bridge tables.
* Validate that the Business Vault data supports intended business reporting use cases.

**Lab 6: Develop Automation Macros or Scripts**

* Create a reusable macro/script that generates hash keys from input business keys.
* Develop a script to detect changes in satellite data using hash diffs.
* Write a procedure to insert and manage audit columns automatically during ETL loads.
* Test the macros/scripts on sample data and document the outcomes.

**Lab 7: Create Zero-Copy Star Schema Views**

* Based on your Data Vault model, design star schema views for reporting without physically copying data.
* Create dimension views that join hubs and latest satellites.
* Create fact views from link tables joined with dimension views.
* Test reporting queries against these views and analyze performance.

**Lab 8: Manage and Utilize Audit Columns**

* Modify ETL processes to consistently populate audit columns (load timestamp, record source, load end date, batch ID).
* Write SQL queries to track data lineage and identify the source and timing of data loads.
* Implement soft delete logic using load end date in satellites.
* Demonstrate how audit columns support troubleshooting and data quality validation.