CSCI 5408
DATA MANAGEMENT AND WAREHOUSING

Group-9
Project: Tiny DB

Group Members:
Naqib Ali(B00981531)
Shrey Nimeshkumar Patel(B00960433)
Syeda Misbah Hussain(B00982800)

Gitlab: https://git.cs.dal.ca/naqib/csci_5408_s24_9

Data Structure Selection

Chosen Data Structures: HashMap and Array

1. **HashMap (Non-linear)**:
    - **Purpose**: Used to store table metadata, such as column names and their types, and row data during operations.
    - **Justification**: The HashMap provides constant time complexity (O(1)) for insertion, deletion, and lookup operations. This efficiency is crucial for managing column details and row data. Its key-value pairing nature is ideal for associating column names with their respective values or types.
2. **Array (Linear)**:
    - **Purpose**: Used for handling ordered data and maintaining sequences of elements, such as rows in the table.
    - **Justification**: Arrays allow for efficient data access with O(1) time complexity for accessing elements by index. They are particularly useful when the size of the data is known beforehand or does not change frequently. Arrays are simple and provide fast access, making them suitable for operations that require traversal of rows, such as selecting data and iterating over row values.

Custom File Format

Custom Format: Pipe-separated Text File (.txt)

Format Details:

1. **Structure**: Each table is stored in a text file where:
    - The first line contains the column headers, separated by a pipe (|) symbol.
    - Subsequent lines represent rows of the table, with each value corresponding to the headers and separated by pipes.

Workflow Diagram:

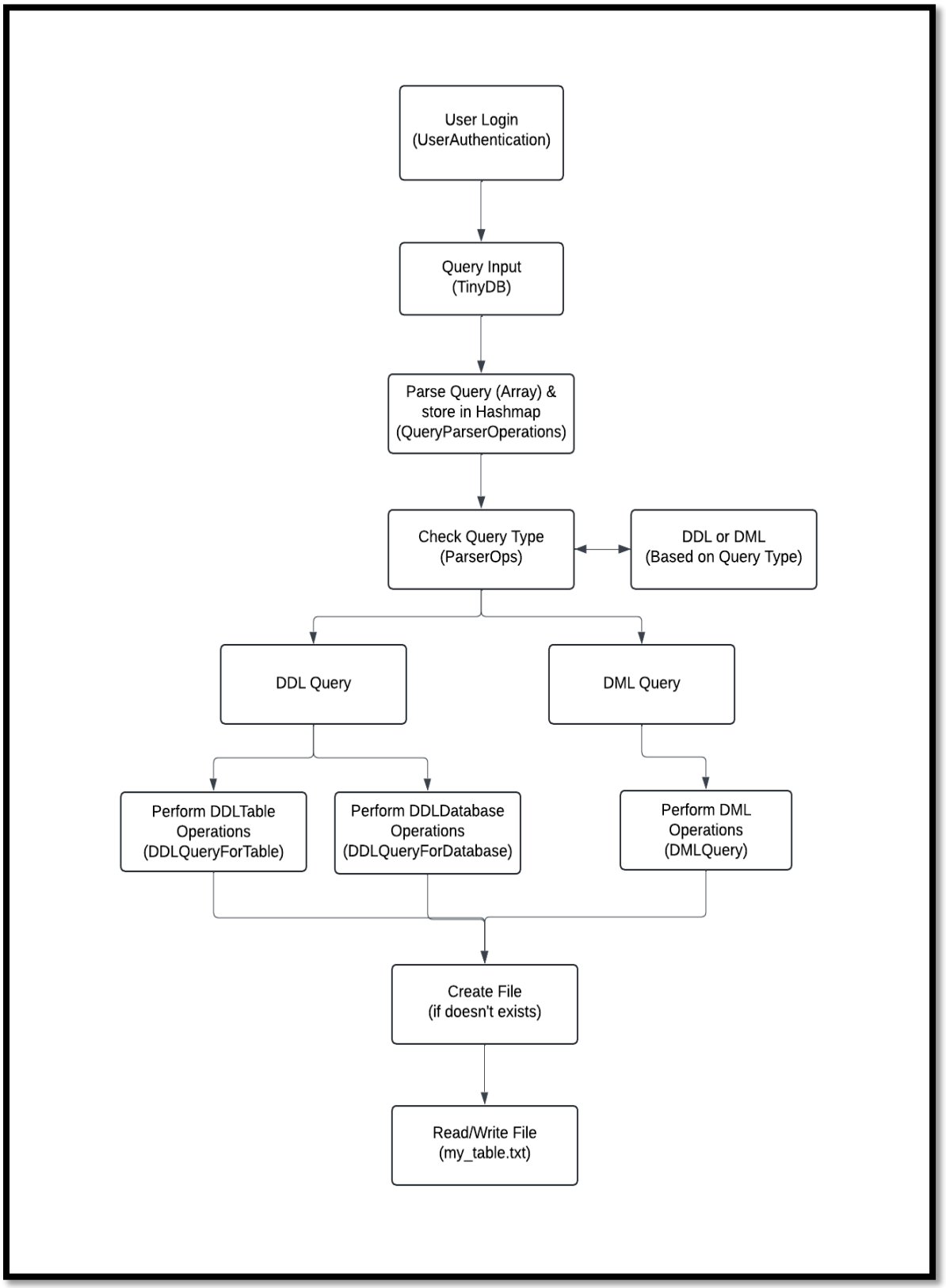

Figure 1: Workflow diagram of Tiny DB