Project Proposal

Group Members: Xuesong Shen (xuesongs@usc.edu), Haoze Zhu (haozezhu@usc.edu)
Class Session: Tuesday Afternoon

The basic design of the system: We will use Python to design and implement our systems. For Storage, we will use CSV files for the relational model and JSON files for NoSQL data.

Data model: For a relational model, data will be represented in tables with rows and columns. Each table will be stored in a CSV file. For a NoSQL model, data will be represented as Collections and Documents. Each Collection will be stored in a JSON file.

Query language:

We will design our query language.
For example(temporary):
"from table 'STUDENT' find 'names' where 'age' > 25 or 'wages' <100"
"from table 'STUDENT' find 'name' and 'age' and 'wage' "
"join table ''STUDENT' and table 'Classes' on 'Student ID' "

Data modification:

We will design our own command in our user interface for CRUD operations.
For example:
" create table 'Course' column 'ID' and 'Semester' insert (DSCI551, Fall 2023)"
" from table 'STUDENT' insert (19, 100 , …)"
" from table 'STUDENT' update (19, 100 , …) to (20, 100 , …)"
"from table 'STUDENT' delete (19, 100 , …) "

Ideas on how to implement them:

At first, with different real-world data sets, we will create a relational database and a non-relational database. Then, we will create tables with CSV files and collections with JSON files.

Specifically, to avoid reading the entire dataset into the main memory, we will use loops and indexes to read the data in smaller chunks, execute our query to each chunk, and combine our query to a new table or a new collection.

We will implement commands for projection, filtering, joining, grouping, aggregation, and ordering, as well as inserting, deleting, and updating the data in our query language through

our interactive command line interface. For example, we will store a set of reserved words such as “create”, “find”, “table”, etc. If a reserved word is detected, the systems will read the data, which will be marked by marks such as single quotes or parentheses. After that, the Python program will do the operation on the corresponding data.

Background and Responsibilities:

Xuesong Shen
Graduate Major: Computer Science
Undergraduate Major: Computer Science
Skills: Python, SQL, NoSQL
Responsibilities: Design and implement the relational database system

Haoze Zhu
Graduate Major: Applied Data Science
Undergraduate Major: Statistics and Data Science
Skills: Python, R
Responsibilities: Design and implement the NoSQL database system