An SQL Query engine, which functions on top of Hadoop's MapReduce. Data stored on HDFS is used to perform queries provided to the engine.

Queries supported:

- PUT: Transfer data from local filesystem to HDFS
- SHOW TABLES: Show all database/tables loaded
- DESCRIBE: Show schema of database/table specified
- LOAD: Load the schema of respective data
    - Specify database/table.csv and schema
    - Optionally specify separator of csv file (default : ',')
- SELECT:
    - Select multiple columns
    - Aggregations supported: Max, Min, Count

ALGORITHM/DESIGN

Control flow:

- Get query; parse query
- Select type of query
- Check match between structure of data and submitted schema (LOAD)
- Store schema (LOAD)
- Change mapper and reducer files accordingly
- Perform Hadoop MapReduce job
- Get output of MapReduce job; Display the same

Structure of schema:

{column_name1 : (column_index1, datatype1),

column_name2 : (column_index2, datatype2),

"separator" : 'your_seperator'}

Schema storage: Using Python's **Pickle** module(object serialization to store as a dictionary) **on HDFS**

Order of operations: SELECT>>PROJECT >> AGGREGATE

EXPERIMENTAL RESULTS