Part 1

1 Database tables definition

Notes:

The definition of tables in sql: (in the beginning, all of the follow tables are empty)

CREATE TABLE `tables` (

`table\_name` varchar(32) NOT NULL,

`file\_location` varchar(256) NOT NULL,

PRIMARY KEY (`table\_name`),

UNIQUE KEY `file\_location\_UNIQUE` (`file\_location`)

) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4\_0900\_ai\_ci;

CREATE TABLE `columns` (

`table\_name` varchar(32) NOT NULL,

`column\_name` varchar(32) NOT NULL,

`type` enum('text','number') NOT NULL,

`nullable` tinyint(4) NOT NULL,

PRIMARY KEY (`table\_name`,`column\_name`),

CONSTRAINT `c\_to\_t` FOREIGN KEY (`table\_name`) REFERENCES `tables` (`table\_name`)

) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4\_0900\_ai\_ci;

CREATE TABLE `index\_info` (

`table\_name` varchar(32) NOT NULL,

`column\_name` varchar(32) NOT NULL,

`kind` enum('PRIMARY','UNIQUE','INDEX') NOT NULL,

`position` int(11) NOT NULL,

`index\_name` varchar(32) NOT NULL,

PRIMARY KEY (`table\_name`,`column\_name`,`index\_name`),

CONSTRAINT `i\_to\_c` FOREIGN KEY (`table\_name`, `column\_name`) REFERENCES `columns` (`table\_name`, `column\_name`)

) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4\_0900\_ai\_ci;

2 Files structures

Code:

CSVCatalog.py: simulating catalog function similar to information\_schema

CSVTable.py: perform dumb\_join and smart join

**Test files 1:** Unit\_test\_catalog.txt: Test files for CSVCatalog.py

function1: def test\_build\_CSVCatalog(): test for building a CSVCatalog

function2: def test\_create\_table(): test for creating an table with complete definition

this also could be used to test the following function since these function were called by this function:

TableDefinition()

CSVCatalog.ColumnDefinition()

CSVCatalog.IndexDefinition()

save\_core\_definition()

add\_column\_definition()

define\_index()

save\_index\_function() which called by define\_index()

function 3: test\_get\_table(): test for get\_table()

notes: first run test\_create\_table(), do not run two times, otherwise duplicate key issues will arise

By this test function, the following function also being tested:

load\_core\_definition()

load\_columns()

load\_indexes()

function4: def test\_drop\_table(): test for drop\_table()

by this function, the following test functions could be tested:

drop\_column\_definition()

drop\_index()

function 5: def test\_column\_definition(): test for CSVCatalog.ColumnDefinition()

function 6: test\_index\_definition(): test for CSVCatalog.IndexDefinition

**Test files 2:** Unit\_test\_csv\_table.txt: Test files for CSVTable.py

Test files for CSVTable.py

1 test\_dumb\_join():test for dumb\_join

2 test\_execute\_smart\_join(): test for execute\_smart\_join