HBase

Hrishikesh Borkar

October 18, 2021

1 Data

The dataset from MyPyramid Food Raw Data has been used for this project. Since the "Display_Name" column in the file 'Food_Display_Table' has commas within it's value, it was not possible to save the file in csv format with commas as the delimiter. Therefore, semi colon was used as a delimiter and the file was generated through the below steps:

- Open 'Food_Display_Table.xlsx' and 'Foods_Needing_Condiments_Table.xlsx' in Excel
- Go to File → Options → Advanced → Editing Section
- Uncheck the "Use system separators" setting and put a comma in the "Decimal Separator" field.
- Save them in 'CSV UTF-8' format in the /data directory. Resulting files will be semi-colon separated.

Schema

Table: foods

Row key: Food_Code

Column families:

- identifiers: Food_Code
- portions: Portion_Default, Portion_Amount, Portion_Display_Name, Factor,Increment, Multiplier
- contents: Grains, Whole_Grains, Vegetables, Orange_Vegetables, Drkgreen_Vegetables, Starchy_vegetables, Other_Vegetables, Fruits, Milk, Meats, Soy, Drybeans_Peas, Oils
- macros: Solid_Fats, Added_Sugars, Alcohol, Calories, Saturated_Fats

Table: condiments

Row key: Survey_Food_Code

Column families:

- identifiers: display_name
- condiment: cond_1_code, cond_1_name, cond_2_code, cond_2_name, cond_3_code, cond_3_name, cond_4_code, cond_4_name, cond_5_code, cond_5_name

2 Question

Part 1 scans the table: "foods" for the food that has been input and returns the details regarding the portions, contents and macros. This data can be useful for a chef when he decides to cook a dish.

Part 2 scans the table: "condiments" and returns the macros of the food that has been input plus the macros of the condiments that go along with it. This could be useful for the health conscious to help decide which condiment would be the right choice for them if any.

3 Run

The repo is saved at HBase Assignment

Build hbase image:

```
docker build -t adv-db/hbase .
```

Run container using:

```
cd csci-540-hw02
docker run –rm -d -p 8080:8080 –volume "$PWD:/mnt" –name hw2 adv-db/hbase
```

Run bash:

```
docker exec -it hw2 bash
```

Run create, load and query files for part 1:

```
hbase shell /mnt/part1/create.rb
hbase shell /mnt/part1/load.rb
hbase shell /mnt/part1/query.rb
```

Run create, load and query files for part 2:

```
hbase shell /mnt/part2/create.rb
hbase shell /mnt/part2/load.rb
hbase shell /mnt/part2/query.rb
```