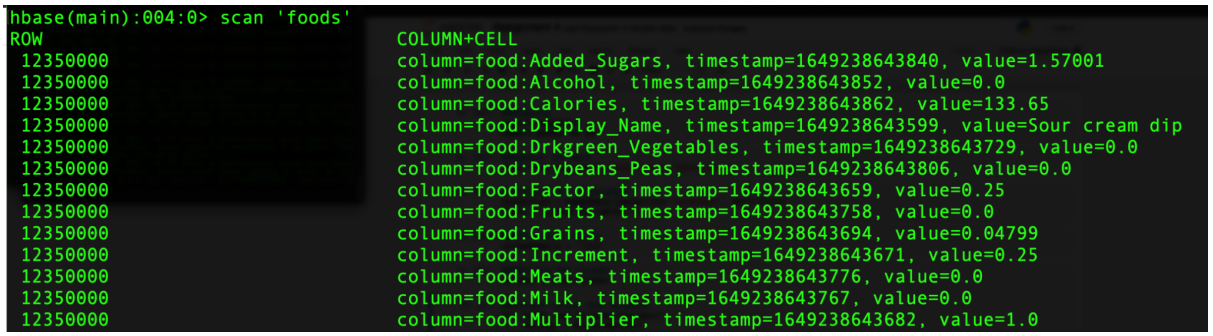
# Assignment 4: HBase Database for Food Nutrition Facts

### Create Table

We created a table named foods and a single column family named food. In food we inserted all columns from the excel table except the first which is the row key. This can be seen in this screenshot:




The reason we chose Food_Code to be the row key, because it was unique for search if you wanted to find a specific Display_Name. But the key itself wasn't unique. There was the same key for the same Display_Name's, but that is okay here because it won't affect the searches we can do.

The single column family we had to create could be distributed to several column families like this instead:

**Column family 1 - Name:** Display_name

**Column family 2 - Portion:** Portion_Default, Portion_Amount, Portion_Display_Name

**Column family 3 - Factors:** Factor, Increment, Multiplier

**Column family 4 - Ingredients:** Grains, Whole_Grains Vegetables, Orange_Vegetables, Drkgreen_Vegetables, Starchy_Vegetables, Other_Vegetables, Fruits, Milk, Meats, Soy, Drybeans_Peas,Oils

**Column family 5 - Product_Info:** Solid_Fats, Added_Sugars, Alcohol, Calories, Saturated_Fats

This would be a fine way to group the data together.

### Import Code

The code for piping down the food data can be seen here:

```
In [52]: ###Imports
         import happybase
         import pandas as pd

In [53]: ###Import the data
         food = pd.read_excel('Food_Display_Table.xlsx')

In [54]: ###Connect to the database
         connection = happybase.Connection('localhost', port=16010)
         table = connection.table('foods')

In [55]: ###Set column Family
         columnFamily = 'food'

In [56]: ###Getting the columns
         columns = food.columns

In [58]: ###insert values into database
         for index, row in food.iterrows():
             row_key = row[columns[0]]
             for col in columns[1:]:
                 table.put(str(row_key), {columnFamily + ':' + col: str(row[col])})
```

We use happybase to connect to the hbase table ‘foods’ that we created in hbase shell. We declare our column family to be ‘food’ and then we loop over the columns and insert the data from the excel table.

### Code Querying:

We used the row key ‘94210100’ to get information on Fitness Water. We would like to see how many calories it contained, and the result is 35,1. This can be seen in the following:

```
hbase(main):025:0> get 'foods', '94210100', {COLUMN => ['food:Calories']}
COLUMN                                CELL
 food:Calories                        timestamp=1649239237459, value=35.1
1 row(s) in 0.1120 seconds
```

The rest of the information can be found like this:

```
hbase(main):024:0> get 'foods', '94210100'
COLUMN                                CELL
 food:Added_Sugars                    timestamp=1649239237453, value=35.1
 food:Alcohol                         timestamp=1649239237456, value=0.0
 food:Calories                        timestamp=1649239237459, value=35.1
 food:Display_Name                    timestamp=1649239237280, value=Fitness Water (Propel)
 food:Drkgreen_Vegetables             timestamp=1649239237316, value=0.0
 food:Drybeans_Peas                   timestamp=1649239237443, value=0.0
 food:Factor                          timestamp=1649239237294, value=1.0
 food:Fruits                          timestamp=1649239237426, value=0.0
 food:Grains                          timestamp=1649239237303, value=0.0
 food:Increment                       timestamp=1649239237297, value=0.25
 food:Meats                           timestamp=1649239237435, value=0.0
 food:Milk                            timestamp=1649239237432, value=0.0
 food:Multiplier                      timestamp=1649239237300, value=0.25
 food:Oils                            timestamp=1649239237447, value=0.0
 food:Orange_Vegetables               timestamp=1649239237313, value=0.0
 food:Other_Vegetables                timestamp=1649239237418, value=0.0
 food:Portion_Amount                  timestamp=1649239237286, value=1.0
 food:Portion_Default                 timestamp=1649239237283, value=2
 food:Portion_Display_Name            timestamp=1649239237290, value=sports bottle (23.7 fl oz)
 food:Saturated_Fats                  timestamp=1649239237464, value=0.0
 food:Solid_Fats                      timestamp=1649239237450, value=0.0
 food:Soy                             timestamp=1649239237439, value=0.0
 food:Starchy_vegetables              timestamp=1649239237369, value=0.0
 food:Vegetables                      timestamp=1649239237310, value=0.0
 food:Whole_Grains                    timestamp=1649239237306, value=0.0
25 row(s) in 0.1210 seconds
```