

Assignment 3: HBase Database for Food Nutrition Facts

see *Seven Databases in Seven Weeks* by Luc Perkins, 2nd Edition, p.82

The **objective** of this assignment is to provide you with experience in processing big data in a column-oriented database and distributed mode.

The **tasks** are the following:

1. Download the MyPyramid Raw Food Data set from <https://catalog.data.gov/dataset/mypyramid-food-raw-data> and extract the zipped contents to find the Food_Display_Table.

This data consists of many pairs of <Food_Display_Row> tags. Inside these, each row has a <Food_Code> (integer value), <Display_Name> (string), and other facts about the food in appropriately named tags.
2. Create a new HBase table called `foods` with a single column family to store the facts. Argue your choice of a row key. Suggest some alternative column family options that would make sense for this data.
3. Create code for importing the food data into the new table, use programming language of your choice. Pipe the food data into your import script on the command line to populate the table.
4. Using either a client application or the `HBase shell`, query the `foods` table for providing information about your favorite foods.

This is a group project.

Your solution and a Peergrade review of three other solutions bring 20 credit point in total to your accumulated score.