

# CA675 Cloud Technologies

## Assignment 2 –Midway Report

### Group 16

#### Project Data source:

This dataset consists of 18 years of user interactions and recipe uploads on Food.com. We are planning on using the raw data on recipes and recipe reviews.

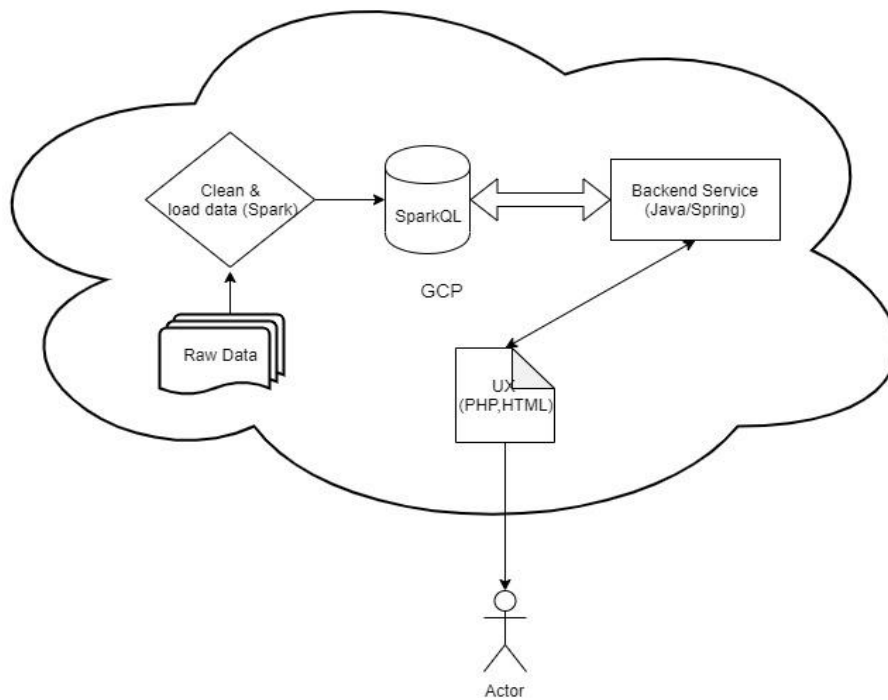
- Recipes
  - Size: 180K+, 287 MB
  - Format: CSV
  - Description: Consist for information like user id, recipe name, preparation description, tags, time required etc.
- Recipe reviews
  - Size: 700K+, 341 MB
  - Format: CSV
  - Description: Consist for information like user id, recipe id, rating given, review comment, time stamp.

More on this data set: <https://www.kaggle.com/shuyangli94/food-com-recipes-and-user-interactions>

#### Project Idea:

- Clean the data using spark, python
- Load the data to SparkQL
- Prepare queries for the data & functionality required by the UI
- Use Spark JDBC to connect to spark from a hosted java-based service
- Using embed prepared quires in java backend service and expose REST API to UI component
- UI component calls the REST API to get the data required.
- UI will have following facility:
  - User can search based on ingredients or recipe or other key words and get matching recipes.
  - User can browse top recipes that are healthy, most rated, commented etc
  - Date visualisation charts and diagrams like the top healthy recipes, the tags of recipes that were trending during a given period of time, and the number of steps/times taken to complete a recipe etc.

## Project High-level Architecture:



## Technologies Used:

1. Data loading and cleaning:
  - Spark
  - SparkQL
2. Connectivity:
  - Java
  - Spring boot
3. Website development:
  - HTML
  - PHP
  - CSS
  - Bootstrap
4. IAAS provider:
  - Google Cloud Platform

## Team Roles and Tasks:

### Roles:

Group Coordinator Name: Akhil Alfons Kodiyan

Student Id	Name	Roles
19210588	Kirthy Francis	Frontend Development
19211063	Barath Raja	Frontend Development
19210412	Tejal Nijai	Backend scripts
19210864	Hithesh Sai K B	Data cleaning and loading
19210912	Akhil Alfons Kodiyan	Components Integration & Backend Development

### Tasks:

#### Front end development:

The task for front end development is to create the website (UI) for the project, where the user enters the queries and the results and visualisations are displayed.

#### Backend SQL:

The task for Back end SQL is to create the queries that are required to search the database for the given input and to test them across all scenarios.

#### Data cleaning and loading:

The task required to perform for this one is to clean the data obtained and to load them in the database.

#### Connectivity and backend development:

The task needed firstly to develop java backend service that connect SparkQL for fetching data to front end, secondly to ensure that various cloud component hosted integrated and able communicate with each other and finally deploy all the services and component to GCP.

### Sample UX Snapshots

