

Rate my Beer

A Distributed Data Management System of Beer Data

Team Details

- Haoran Wang
 - USC ID: 3125586391
 - Email: hwang822@usc.edu
- Xintong Jiang
 - USC ID: 9689084620
 - Email: xintongj@usc.edu
- Rahul Aggarwal
 - USC ID: 9285519625
 - Email: rahulagg@usc.edu

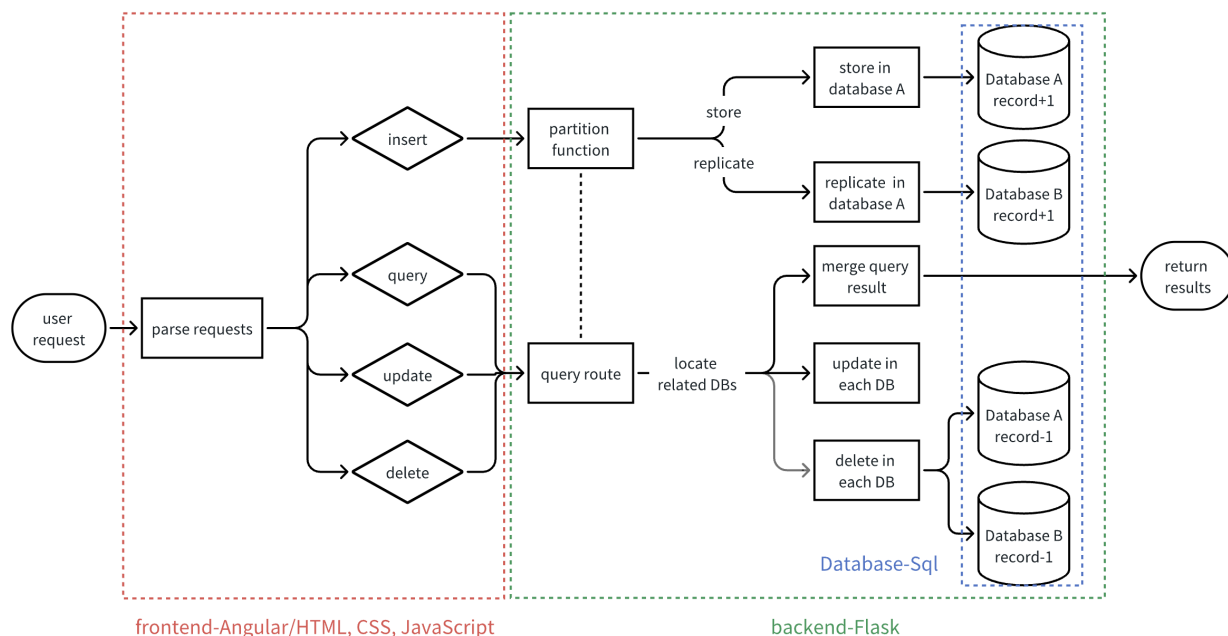
Team members background and skills

- Haoran Wang: Computer Science, Python backend, Data analysis, SQL.
- Xintong Jiang: Applied Data Science, Python, Data analysis, SQL, NoSQL.
- Rahul Aggarwal: Computer Science, Web Technologies, SQL.

Project requirements

Our project a distributed, robust, fast data management system of beer rating data. It allows users to manage their beer data on the website. The workflow and key steps are as follows.

Figure 1. Workflow of out project



1) Dataset Preparation

Our data comes from a Kaggle dataset *Beer Data Analytics*¹. This dataset contains different types of beers and various aspects of them, with a total of 13 attributes and 528870 pieces of data. Some main attributes are as below.

Table 1. Main Attributes of Beer Dataset

¹ <https://www.kaggle.com/datasets/gauravharamkar/beer-data-analytics?select=BeerProject.csv>

Attribute name	Description	Data type
beer_name	name of the beer	string
beer_style	style of the beer	string
review_appearance	appearance score	numeric
review_taste	taste score	numeric
review_overall	overall score	numeric
review_text	text comment	string

Due to the fact that the original data lacks a unique identifier, we will generate a unique identifier ***review_id*** for each piece of data and use it as the primary key in the database.

2) Partition and Storage Strategy

In order to store a large amount of data with high safety, consistency, fault tolerance ability, we design a partition and storage strategy for beer data.

Partition is a technique for dividing data, including horizontal partitioning and vertical partitioning. Considering the length of attributes and dataset, we will use horizontal partitioning method based on the range of the primary key ***review_id***.

Based on the partition result, our system will achieve distributed storage. According to the load of the database, it will allocate a storage location for each block of data. In the meanwhile, it will replicate data in other databases.

3) User interface and applications

Users can access our system through web pages.

Our system provides two kinds of applications. One is to allow users to manage the data through CRUD. Users can insert, query, update and delete reviews through buttons, text boxes

and menus. The other one is to provide some analysis results of current beer reviews. For example, it will show “Top rated beers” “Keywords for Blue Whale Ale”, etc.

Planned Implementation

- Frontend: Angular/HTML, CSS, JavaScript
- Backend: Flask + SQL database

Team responsibilities

- Haoran Wang: backend development, data processing
- Xintong Jiang: backend development, data processing
- Rahul Aggarwal: frontend development

Timeline

Time	Milestones and checkpoints
Jan	-- Project Kickoff -- Define project scope and objectives -- Conduct initial research and feasibility analysis -- Assemble project team and allocate responsibilities -- Develop a detailed project plan and timeline
Feb	- Complete detailed system requirements - Design system architecture and database schema - Set up development and testing environments - Begin initial coding and development - Regular progress check-ins and team meetings
Mar	- Complete initial implementation of core features - Conduct first round of internal testing - Address and fix any identified issues - Begin integration testing and debugging - Documentation of code and processes

Apr	<ul style="list-style-type: none">- Feature freeze for first release- Conduct thorough testing of all functionalities- Prepare for beta testing and gather user feedback- Revise and enhance based on beta testing feedback- Finalize documentation and user guides
-----	---