



CONFIDENTIAL: DO NOT DISTRIBUTE

## Take-home assignment for Full Stack Software Engineer position

Note: Please do not spend more than a few hours working on the assignment.

**High-level Objective:** Create an app to allow a user to **upload**, **store** and **visualize** **timeseries datafile(s)**.

### Description of the assignment:

Boston Bioprocess regularly runs fermentation tanks (think of a beer brewing process) to produce valuable bioproducts for our clients. Each fermentation run generates a time-series log of multiple process parameters that is analyzed by users. We want you to create a small app that **allows a user to upload such a csv datafile and display the data on a graph**. An example datafile is provided with the assignment (ClientABC\_R001001\_Online\_Report\_BostonBioprocess.csv).

Here is a snapshot of the head of the file:

	A	B	C	D	E	F	G
1		ClientABC_R001001: DataLog Param, PV and Units					
2							
3	Time Stamp		Parameter		Process value		Units
4	0.009		pH		5.99		pH
5	0.009		Pump1		0		%
6	0.009		Pump2		0		%
7	0.009		Temperature		28.01		DegC
8	0.026		pH		5.99		pH
9	0.026		Pump1		0		%
10	0.026		Pump2		0		%
11	0.026		Temperature		28.02		DegC
12	0.043		pH		5.99		pH

The name of the file (also repeated in the first row of the csv) contains 2 key information:

- **ClientABC**: Name of the client
- **R001001**: Unique identifier of the fermentation run (Note: This id will not repeat)

For each **Time Stamp, Process value and Units of 4 parameters** are recorded in this file (**Pump1, Pump2, pH and Temperature**). While pH and Temperature are Parameter names, Pump1 and Pump2 are placeholders that can be replaced by one of 2 options entered by the user:

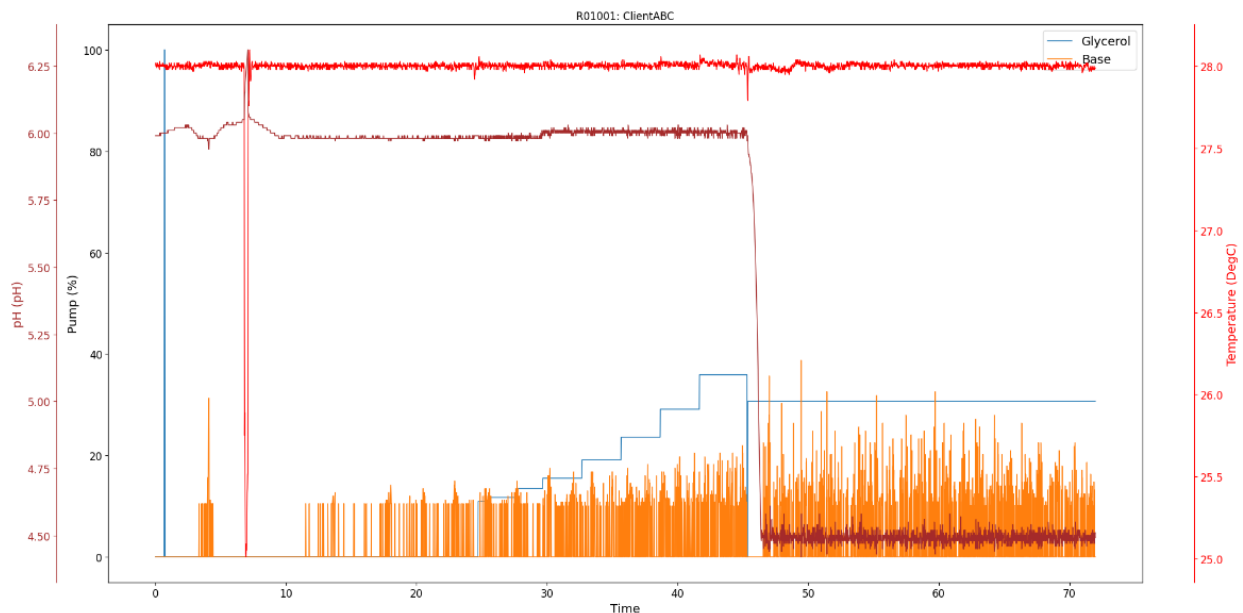
Pump1: Glucose or Glycerol

Pump2: Base or Acid

We want to process the information from the file and the user defined choice for Pump1 and Pump2, and store them in a database. The database would have at least 2 tables:

- **run\_client**: Containing the unique RunID and the Client name
- **run\_time\_series\_data**: Containing the unique RunID, Time, Parameter, Process value and Units. Note that the Parameters Pump1 and Pump2 should be replaced by the user defined choices in the Database entry

Finally, the user wants to see a graphical visualization of the processed data, where the parameter values are plotted against time for the specific RunID. You can use your preferred tool for visualization. Some examples can be python libraries such as [matplotlib](#), [altair](#) and [seaborn](#), to other data plotting frameworks such as [plotly-dash](#) or [streamlit](#), [react](#) etc. A rudimentary example of the visual is shown below, however, you can use your judgement how to implement it:



### Your task:

1. Create a web application that allows a user to upload a fermentation run csvfile, and choose their options for Pump1 and Pump2. You do not need any user authentication step for this assignment, can use your judgement in deciding the frontend framework.
2. The application parses the file and writes into a database. We prefer python as the backend coding language. You can use your judgement to decide the type of database to use.

3. The application provides a single graphical visualization of the parsed data trend of the parameters against time.
4. **Dockerize** your app to allow for ease of deployment and scalability.

**Task submission:**

- Commit your code to your private github repo (name it in the format of <yournameinitials>\_bbp) and send an invite to [achowdhury@bostonbioprocess.com](mailto:achowdhury@bostonbioprocess.com) (github username: anupamc01) to view the results. Alternatively, zip your code and send it back to us.
- Send us a link to your created application.

**Some of the things we will look at in the evaluation:**

- Task execution: How effectively can you translate the assignment's requirements into a working solution?
- Code quality: How you reason about making sure code is readable and maintainable.
- Testing: How you reason about what to test and how to test it.
- Infrastructure and operations: How you would design a scalable system and what are some key criteria to focus on under time constraint.