# UCSF JSON data parsing project document

1. Intruduction
2. System requirement

Java: 1.7+

Database: Mysql

Other tools: Maven 3+

1. Configuration
2. Open …/src/main/resources/application.properties and set the database variables in your local environment.
3. Make sure you have a running Mysql database in your local environment.
4. Make sure you have Maven installed in your local environment.
5. Make sure no other processes are using port 8080.
6. Run the program
7. Open command prompt
8. Go to the project root folder
9. Run command >mvn sprint-boot:run
10. Visit <http://localhost:8080/index.html>
11. Directly run it in Heroku

We’ve already deployed a copy of the project to Heroku cloud, it can be directly visited by any browser using this URL: <https://uscf-json-parser.herokuapp.com/index.html>.

1. Run it as AWS service

We deployed another copy to an Amazon EC2 server that allows file upload, download, async processing etc. Open browser and visit this URL: http://52.9.159.178:8080/index.html

1. How it works
2. Upload a JSON file exported from <https://www.clinicaltrialsmatch.org/v5/nonsmallcelllungcancer.html?f=VA>
3. Uploaded file is stored in a “todo” folder on server, click “List all todo files” to list all uploaded todo file names.
4. Click “Process” to import all todo files into database, once completed, the files will be moved to “done” folder.
5. Click “List all done files” to list all done file names.
6. You can download any todo or done files by typing http:// 52.9.159.178:8080/todo/{filename} or http:// 52.9.159.178:8080/done/{filename}
7. Get trials and profile based on profile id, call [http://localhost:8080/profile/{id}/trials](http://localhost:8080/profile/%7bid%7d/trials), replace the path variable {id} with actual profile id.
8. Get all profiles call <http://localhost:8080/profiles>
9. Get all trials, call <http://localhost:8080/trials>
10. Get all accessionNumberBlocks, call <http://localhost:8080/anbs>
11. Database structure

We have 3 tables in the database, t\_accession\_number\_block, t\_profile and t\_trial.

For each uploaded JSON file, the header name is the accession\_block\_id that connects t\_profile table and t\_accession\_number\_block table. In the file content, there’re 3 JSON objects, “profile”, “trials” and “trialsFound”, where “trialsFound” is just a statistics data of “trials” so we don’t need to store it in database. “profile” contains the mutation information and is stored in t\_profile table, “trials” contains a list of trials, each trial is saved as a record in t\_trial table, all the trials in the file is tied to a profile\_id.