Github link: https://github.com/LiiaDulher/docker_project
My project repository contains:

1. Cassandra database files
    a. run-cassandra-cluster.sh
    b. shutdown-cassandra.sh
    c. DDl.cql
    d. Dockerfile1
2. App files
    a. cassandra_api.py
    b. Dockerfile2
    c. cassandra-app.sh
    d. shutdown-app.sh
3. Data processing files
    a. read-from-stream-write-to-cassandra.py
4. Client files
    a. client.py
5. Documentation
    a. documents
    b. example_queries
    c. Readme.md

As a database for my project I used Cassandra, because it is easy to write to, I do not need to update my data. Also it is the database I am the most aware of. I have six tables in my keyspace. Their diagrams can be found in *"Cassandra database diagrams.pdf"*(There are fields, their types and keys, where K - primary key, C-clustering key). There are three statistics tables for A type queries and 3 for B type. Also page_creation is used for collecting statistics.

For processing data I use a python program: it has 2 threads: one is processing a stream and writes data to the database all the time, another one wakes up at X:01:00 and collects statistics for the last hour.

App is running in a docker container and maps 8080 port to communicate. It receives GET requests and sends responses with the json body.

You can run my client for communication with the app. It will give you all information about the query, say in which format data should be given, print the query and response bodies.

It also gives you **UTC+0 time**, because it is the timezone in all Wikipedia data. It is used in all my programs and all responses are with this timezone. It is the time required in 5th type B query.