MSc in CSTE, Cloud Computing Assignment 2023-24

Collecting, Processing and Distributing IoT Data to Clients

Stuart Barnes, Cranfield University

November 17th, 2023
Hand-in date 2/1/2024 (FT), 15/1/2024 (PT), 9.30am

1. Introduction

The goal of the assignment is to store the latest air quality data captured from a network of small IoT environmental sensors and make the data available to clients on-demand. The data can be accessed in JSON format from:

- Averaged data from last 24 hours for each sensor: https://data.sensor.community/static/v2/data.24h.json
- Averaged data from last 5 minutes for each sensor (for testing): https://data.sensor.community/static/v2/data.json

An example of the data format is given below:


```
{
     "id":10995737600,
     "sampling_rate":null,
     "timestamp":"2022-06-21 12:56:19",
     "location":
         {
                "id":16232,
                "latitude":"48.8",
                "longitude":"9.002",
                "altitude":"365.6",
                "country":"DE",
                "exact_location":0,
                "indoor":0
         },
     "sensor":
         {
                "id":92,
                "pin":"1",
                "sensor_type":
                     {
                          "id":14,
                          "name":"SDS011",
                          "manufacturer":"Nova Fitness"
                      }
          },
     "sensordatavalues": [
```

```
            {
                "id":24428549678,
                "value":"3.68",
                "value_type":"P1"
             },
             {
                "id":24428549711,
                "value":"2.80",
                "value_type":"P2"
             }
        ]
}
```


The data set consists of sensor location, sensor type and sensor values. Note that different sensors may give different types of data, specifically PM_{10} and $PM_{2.5}$ values corresponding to the two main sizes of airborne particulate pollution, as well as temperature, pressure and humidity. For this exercise, the particulate data can be used to calculate the air quality index (AQI).

To calculate the AQI you will need to use the following table to identify the index from values of the P_{10} and $P_{2.5}$ data. For example, if $PM_{2.5}$=39 and PM_{10}=36, the AQI will be 4. Note, as in this example, that if the P_{10} and $P_{2.5}$ measurements provide different AQI values, choose the largest value.

Range	**Air Quality Index**	**$PM_{2.5}$ Particles, 24 hour mean (µg/m^3)**	**PM_{10} Particles, 24 hour mean (µg/m^3)**
Low	1	0-11	0-16
	2	12-23	17-33
	3	24-35	34-50
Medium	4	36-41	51-58
	5	42-47	59-66
	6	48-53	67-75
High	7	54-58	76-83
	8	59-64	84-91
	9	65-70	92-100
Very High	10	>70	>100

Table 1. UK air quality index, "*Review of the UK Air Quality Index",* 2011

2. Tasks

- Collect environmental sensor data via the on-line sources and store in an appropriate Cloud-located database.
- Provide a means for updating the database with new data as it becomes available.
- Calculate the Air Quality Index (AQI) for each entry in the database, according to the information provided in Table 1. Decide whether this calculation will take place at the point of bulk sensor data download or on each client request, according to what you think will be the most efficient and/or most cost-effective solution.

- Provide a front-end implementation that distributes data to clients on demand.
    - Clients should be able to download data according to user-specified criteria, such as geographical location, time frame, etc.
    - Simulate increasing workloads on the system by increasing the number of clients and therefore the number of requests being made.
    - Measure the current system workload to determine whether a new worker node is required. You can decide whether the system load is measured directly from the amount data being downloaded or the number of currently connected clients, or some other criteria.
    - Implement a solution that demonstrates elasticity, such that the number of worker nodes increases as the work-load increases.
    - Think about the cost implications of instantiating new worker nodes and find the optimal point at which new nodes are required.
- Discuss the implications of data security and sovereignty with respect to the environmental sensor data and comment on any specific fields that might trigger additional consideration, if they were present in the data.

The clients can be served via html requests (ie a Web server) or can simply receive a csv file containing the requested data. Whichever approach you take, think about how you will demonstrate the elasticity of your solution.

The 5-minute averaged data stream can be used to test your algorithms over a shorter time scale.

3. Source Code and Report requirements

Write a report to present and discuss your findings. The report should be no less than 1500 words and must not exceed 3000 words. The report can contain any number of figures/tables, however all figures/tables should be numbered and discussed. All code used in the analysis should be included in an Appendix along with appropriate documentation.

4. Assignment Submission

The source code and documentation should be submitted electronically via the technical work submission point by 9.30am on the 2nd Jan (FT) and 15th Jan (PT). The report should be submitted electronically via the TurnItIn submission point by the prescribed deadline, for the assignment submission to be considered complete.

This is an individual assessment. Respect the University regulations on plagiarism.

5. Marking

The assignment will be assessed based on the following marking scheme:

- 20% Introduction, methodology, conclusions

- 40% Software: system architecture, code organization and efficiency
- 30% Discussion and analysis of the results
- 10% Report structure, presentation, clarity, references