

Problem Statement

You will create a stream processing pipeline for an IOT use case using Kafka (<https://kafka.apache.org/>). Sensors send data every minute in the following JSON format to keep the network utilization at minimum.

```
{
  "id": int,
  "ts": unix_timestamp,
  "t": int,
  "v": [int, int, int],
  "i": [int, int, int]
}
```

This data is then processed to following format:

```
{
  "id": int,
  "time": DateTime String,
  "temperature": float,
  "voltage": [int, int, int],
  "current": [float, float, float],
  "power": [float, float, float]
}
```

Example:

<i>Data from Sensor</i>	<i>Processed Data</i>
<pre>{ "id": 110, "ts": 1483262700, "t": 2750, "v": [230, 232, 230], "i": [1504, 1508, 1505] }</pre>	<pre>{ "id": 110, "time": "1/1/2017, 9:25:00", "temperature": 27.50, "voltage": [230, 232, 230], "current": [15.04, 15.08, 15.05], "power": [3.4592, 3.49856, 3.4615] }</pre> <p>// All float values are divided by 100 // Power(kWh) = (V*I)/1000</p>

The details of required tasks are mentioned on the next page.

The following tasks are required:

Task 1:

Generate 50k sample data from sensors corresponding to the JSON schema above. You can use any online tool e.g. <https://www.mockaroo.com/>

Task 2:

Ingest the json data into Kafka.

Task 3:

Using Kafka Streams (<https://kafka.apache.org/documentation/streams>) transform the data to processed JSON format:

- a) You must read data from Kafka topic created in *Task #2*.
- b) Republish processed data back to Kafka.
- c) Save processed data to file.

Task 4:

Using Kafka Streams further process sensor data to output format:

- a) Running Cost per sensor per hour
 - i) $\text{Running Cost} = (\text{Sum (Total power for one hour)} / 60) * 15$
- b) Out of order data points should be handled
- c) Save output to file.