Description:

Client1:

- Client class: The Client class is responsible for creating a new client instance and sending a POST request to an EC2 instance.

- MultiThread class: This is the main class that uses an executor service to send 500k requests.

- Metrics class: This class is used to stored important information such as the number of requests and the latency of each request.

Workflow

The MultiThread class is where we start by specifying the total number of requests that we wat to send to the EC2 instance, as well as the maximum number of threads that we allow in the executor service. Then, we create client instances iteratively and add them to the executor for execution.

In each client instance, we first set up the base URL, which is the IP address of the EX2 instance, followed by the API. After that, we use the Metrics class to keep track of important timestamps such as the start time, latency, and wall time. Finally, in the MultiThread class, we print out all the relevant information that was recorded by the Metrics class.

Client 2:

- CSVWriter class: In order to plot a diagram of the throughput per second, we have a class specifically designed to store all the information of each request in a CSV file. This allows us to easily collect and analyze the data needed to create an accurate representation of the performance of our system.
- Metrics class: The only difference is that we collect and store more information in the Metrics class in order to perform the calculations need for part 2.
- Result class: This POJO class is responsible for storing information about each request, which will help us plot the throughput per second diagram.

The expected throughput using Little Law's prediction: 64.9 * 50 = 3245

```
Wall Time: 154644 ms
The number of threads: 1
The number of successful requests: 10000
The number of unsuccessful requests: 0
The average latency: 15
The total throughput: 64.93506493506493(requests/second)
```

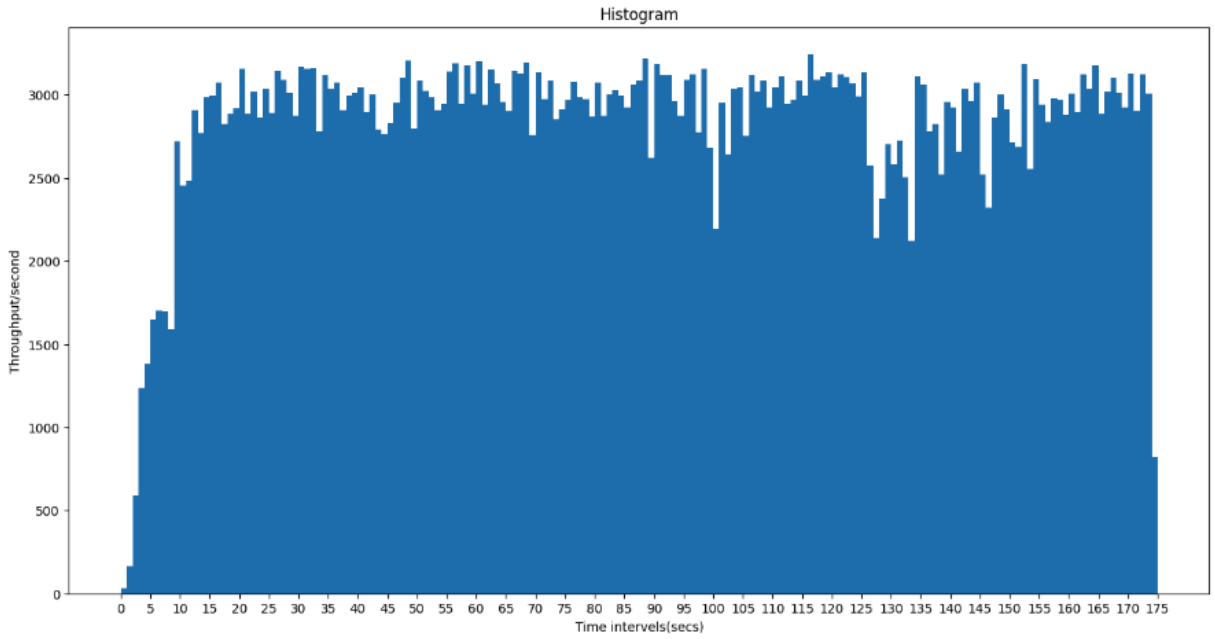
Part1

```
Wall Time: 169258 ms
The number of threads: 50
The number of successful requests: 500000
The number of unsuccessful requests: 0
The average latency: 16
The total throughput: 2958.579881656805(requests/second)
```

Part2

```
The mean response time: 17.3565112192294 milliseconds
The median response time: 16 milliseconds
The total throughput: 2873.5632183908046(requests/second)
The p99 response time: 43 milliseconds
The max response time: 1580 milliseconds
The min response time: 10 milliseconds
```

Task 4


Histogram
Throughput/second
0
500
1000
1500
2000
2500
3000
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175
Time intervels(secs)