## **Activity 4: Serverless Architectures**

1. Run siege using the same number of clients (supposedly large) as Activity 2 and 3. Observe the response times, and the variations in response times across all requests. Do all requests see the same response times? Are there variations? Explain what you observe and why (same response time or variation)?

Using this command to run siege

```
siege -c200 -d1 -r10 --content-type "application/json" 'https://nwvomsw9i4.execute-api.us-east-2.amazonaws.com/default/activity4 POST { "a": "9", "b": "6", "op": "+"}'
```

```
Transactions:
                                  406 hits
Availability:
                                28.25 %
Elapsed time:
                                 4.71 secs
Data transferred:
                                 0.04 MB
Response time:
                                 0.09 secs
Transaction rate:
                                86.20 trans/sec
                                 0.01 MB/sec
Throughput:
Concurrency:
                                 8.13
Successful transactions:
                                  406
Failed transactions:
                                 1031
Longest transaction:
                                 0.19
Shortest transaction:
                                 0.01
```

The response time is generally very fast, typically ranging between 0.01 to 0.03 seconds. However, it varies considerably within the range of 0.01 to 0.16 seconds, from the fastest to the slowest response times. Not all requests will experience the same response time. Additionally, apart from the expected status code 200, clients may also receive a response status code 503, indicating 'Service Unavailable.' This occurs when many clients send requests simultaneously.

2. Wait for at least 20 minutes. Rerun siege. Observe the response times, and the variations in response times across all requests. Are the results the same as question 1? Do you observe any start up delays? Are they shorter or longer than before?

Initially, the response time is longer than in question 1, but after many requests have passed, the response time returns to the same as question 1. And the average response time is 0.13 seconds, which is longer than in question 1 (0.09 seconds) due to startup delay.

```
408 hits
Transactions:
Availability:
                                28.37 %
Elapsed time:
                                 4.83 secs
Data transferred:
                                 0.04 MB
Response time:
                                 0.11 secs
Transaction rate:
                                84.47 trans/sec
Throughput:
                                 0.01 MB/sec
Concurrency:
                                 9.33
Successful transactions:
                                  408
Failed transactions:
                                 1030
Longest transaction:
                                 0.25
                                 0.01
Shortest transaction:
```

3. How much did this experiment cost you? How does this compare to if you were to use EC2 to run your application? How does this compare to if you were to use Elastic Beanstalk to run your application?

From https://dashbird.io/lambda-cost-calculator/\_show that I don't pay anything for Lambda from this experiment. Because Lambda is the most cost-effective choice low traffic with noncomplicated application, benefitting from its pay-per-request and pay-per-use billing model. On the other hand, EC2 instances is more economical for applications with consistent and predictable traffic, as charges are based on instance uptime. Elastic Beanstalk is combining the convenience of managed services (which not convenience that much) with cost control similar to EC2.

Execution costs:	\$0/month
Request costs:	\$0/month
Results	
Yes No	
nclude free tier?	
300	0
Estimated average duration (ms)	
128 MB	~
Memory allocation	
10000	0
Number of executions (month)	

4. Apparently, with AWS Lambda, auto-scaling happens without user configuration. Compare the pro's and con's with configuring auto-scaling for laaS and PaaS in Activity 3.

	Pros	Cons
Lambda	- Simplicity with automatically configuration adjustment	- Limited control
	- Reduced operational overhead	- Cold start issues
laaS and PaaS	- Fine-grained customize option	- Manual configuration adjustment
	- Customize scale up policy	- Complexity

5. Compare the ease of development and ease of deployment between getting an app ready to work on AWS Lambda vs. if you were to get it to work on EC2.

Developing and deploying on Lambda is generally easier than on EC2, especially for simpler functions. However, for more complex functions that involve file management, EC2 is often better suited, allowing local management before uploading remotely.