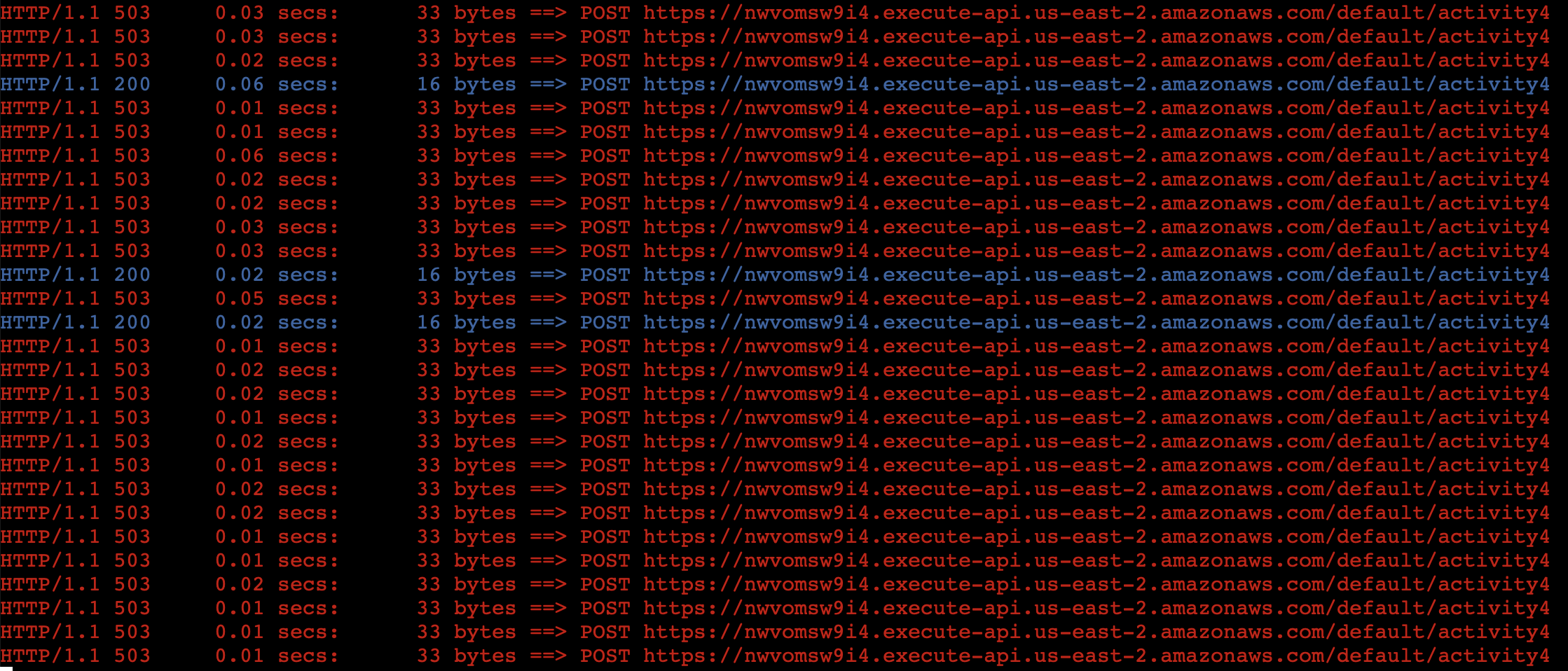
**6432021021 Jiratchaya Prajonkitchai**

**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": "+"}'





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?**

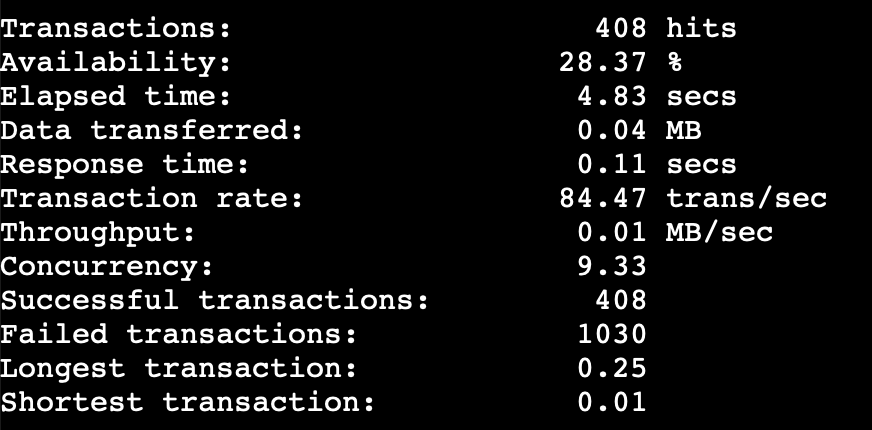
Initially, the response time is longer than in question 1, but after many requests have passed, the response time returns to the same as question 1. And the average response time is 0.13 seconds, which is longer than in question 1 (0.09 seconds) due to startup delay.

**A screen shot of a computer screen

Description automatically generated**

**A red and blue lines on a black background

Description automatically generated**

****

**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. |  |

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

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