CS6847: Cloud Computing

Assignment 1

Submission via Google Classroom

The objective of the assignment is to use and learn the auto scaling feature of the Cloud platforms such as Amazon Web Services (AWS), Google Cloud Platform (GCP), Microsoft Azure, Digital Ocean, etc using the client-server programming model.

Problem Description

- Create a server program which accepts client requests, process it and responds back to the client. The complexity of the server program may vary in terms of computation, memory usage, I/O operations, etc. For example, a server program which takes input as a string from the client and returns the reversed string.
- Create a client program to vary the rate of the request (requests per second) submitted to the server program. You can choose any programming language for writing a client-server program such as Node.js, Python, Java, etc.

Evaluation

- Evaluate the pattern of the client request rate and the server response time without using the auto-scaling feature.
- Evaluate the behavior of the client-server program with Auto Scaling and load balancing. Test the Auto Scaling feature with different parameters such as CPU utilization, network I/O, memory usage, number of requests, etc. Observe the pattern in the client request rate and the response time, identify the parameters responsible for variation in the response time and provide the reasons for the same.

Submission guidelines

- Submit the source code of client and server program for the assignment. All other supporting files used for generating plots, logs, etc. should also be placed in the zip file (`Roll_number.zip`).
- Submit a `README` file containing the necessary details for running your program on the cloud.
- For different rate of requests (10-1,00,000 request per second), the client program should create a file `rate_n.txt` and write the response times for each request in a new line. Here `'n'` specifies the number of request per second. The client program should run for at least one minute.
- Create a file `Output.txt` containing the average response time for each request rate in a new line. Format: <Request rate,Average Response Time>.
- Create separate folders for output files (With/Without Auto Scaling).
- Prepare a report explaining the plots (request vs response rate) and results in detail. Also, Specify the cloud provider and instance configurations used for the assignment.

Bonus

Compare Auto Scaling feature of AWS and GCP with relevant plots and data using the same client-server program. You are supposed to bring out the relevant details on the differences you observed running the program on these cloud platforms.

Academic Honesty

WARNING ABOUT ACADEMIC DISHONESTY: Do not share your work with anyone else. The work YOU submit SHOULD be the result of YOUR efforts.