PART I: Concepts and Theory, Algorithms [60 points]

Please provide accurate, concise (maximum 10 sentence) answers.

*1. Consider equation (12.1, pg 643) from the textbook for the response time of an app. [15 points]*

**a.** What is the typical response time of a web application (research using the web). What is an acceptable range for it, and what problems occur if it exceeds that range.

What is the typical response time of a web application?

For a web application, response time is the time interval between the moment when the user submits a request to the web application and the moment when the user receives a response.

As the equation from the textbook shows:

D: the amount of data transfer required to serve the user request

B: the minimum bandwidth across all links in the network from the user to the application deployment

: the round trip time for user-application interactions that are needed to generate the response

: the total processing time required by all the tiers of the application deployment

: the total processing time required by the user's device.

What is an acceptable range for it, and what problems occur if it exceeds that range?

A one-second response time is generally the maximum acceptable limit, as users still likely won't notice a delay. Anything more than one second is problematic, and with a delay of around five or six seconds a user will typically leave the website or application entirely.

*b. Consider 3 different types of workloads from this overview:*

*CPU Intensive, Memory Intensive and Storage Intensive. Provide real world examples of each of these. Comment on how the different Parameters on the right-hand side of the equation (12.1) will be different for each of these workload type and why. Please explain.*

**Provide real world examples:**

CPU Intensive: application of tasks of sorting, search, graph traversal, matrix multiply, Video editing and processing · Software compiling.

Memory Intensive: web browsers, 3D graphics, IDEA/Eclipse/PyCharm

Storage Intensive: database applications (such as MySQL, IBM DB2, AWS DBS)

For the parameters on the right-hand of the equation:

D: the amount of data transfer required to serve the user request

B: the minimum bandwidth across all links in the network from the user to the application deployment

: the round trip time for user-application interactions that are needed to generate the response

: the total processing time required by all the tiers of the application deployment

: the total processing time required by the user's device.

**CPU Intensive**:

the amount of data transfer required to serve the user request would be higher;

the total processing time required by all the tiers of the application deployment would be lower;

**Memory Intensive**

the round trip time for user-application interactions that are needed to generate the response would be lower;

**Storage Intensive**

D/B could be lower as storage intensive application could handle large amount of data transfer required to serve the user request quicker.

*2. Write a 1-2-page Design Note on how you would design a Performance Testing project for your eCommerce SaaS on AWS Cloud. Please refer to the paper Mukherjee et al. Performance Testing Web Applications on the Cloud, and provide similar approach, with diagrams. [15 points]*

**TEST PROCEDURE**

For my eCommerce SaaS on AWS Cloud, I would use three types of AmazonEC2 instances: small, medium and large. The characteristics of these instances are as follows:

Table

Description automatically generated

All these instances are located in the same region (us-west1) and their IP addresses belong to the same network subnet.

Ubuntu server version 12.04 (kernel version 3.2.0-40-virtual) is installed as the Operating System in all three instances.

Apache 2 (version 2.2) is installed as the Web server.

PHP with FastCGI (version 5.3.6) as the application tier.

MySQL (version 5.1.49-1) as the database server in all instances.

Collectd is used in each instance to monitor the CPU utilization.

The httperf Web workload generator tool [7] is used to generate synthetic requests to the Web Server.

Server response time is the only focused metrics in my eCommerce SaaS.

I will conduct a subset of experiments over a week to look for performance variability in the EC2 platform. In the performance Testing project, httperf is run at the beginning of each hour for 24 hours in a day over the course of a week. Each httperf test to an EC2 instance runs for 100 seconds. Performance data from the instances was captured only once every hour of the day for a week in order to minimize the cost of running the EC2 instances continuously.

**Steps:**

**First step**: run experiments to rule out bottlenecks in the test setup that can prevent us from inferring the true impact of the EC2 platform on Web server response time. (Specifically, I will conduct experiments for determining whether the workload generator should be located inside the cloud or in a dedicated physical host in a remote location) Based on these results, it was clear that the performance impact of the EC2 platform is better evaluated when the workload generator resides inside the cloud.

**Second step**: conduct experiments to determine the right type of EC2 instance to serve as the workload generator. The experimental results show how choosing a wrong instance as the workload generator can badly affect our estimates of server performance.

**Third step**: I will continuously monitor the network bandwidth for a week between the workload generator and the Web server instances under test using the Iperf tool.

**Forth step**: analyze the results to see if there is a wide variation in available bandwidth and whether the available bandwidth is sufficient to sustain our test workloads.

**Finally step**: run experiments to analyze the performance of the three types of EC2 Web server instances

**VALIDATING THE TEST PROCEDURE**

A. Using a Remote Workload Generator

Internet round trip latencies dominate the response times measured by httperf.

Table

Description automatically generated

B. Using In-Cloud Workload Generator

The performance of each EC2 instance is measured in terms of response time (in milliseconds).

Table

Description automatically generated

C. Measuring Network bandwidth for Tests

The available bandwidth between the workload generator and the small instance shows some variations for different hours of the day.

Graphical user interface, chart

Description automatically generated

3.

*a. Make 2 or 3 basic tables (only few example rows) for your eCommerce App (SaaS) following these schemata.* *What are the advantages of using postgreSQL for this data?*

|  |
| --- |
| user |
| user\_id |
| username |
| password |
| name |
| phoneNum |
| createdTime |
| address |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| user\_id | username | password | name | phoneNum | createdTime | address |
| 0 | Hello | 123 | Jason | 6693084725 | 2021-02-17 | A St. |
| 1 | World | 258 | Edward | 6673089753 | 2021-05-06 | 32 way |
| 2 | Cloud | 1314 | Sarah | 6587222587 | 2020-11-23 | B St. |

|  |  |  |  |
| --- | --- | --- | --- |
| user\_id | session\_id | product\_id | price |
| 0 | 258 | 69 | 29.99 |
| 1 | 467 | 945 | 3.65 |
| 2 | 32 | 364 | 12.8 |

|  |
| --- |
| cart\_item |
| user\_id |
| session\_id |
| product\_id |
| price |

*What are the advantages of using postgreSQL for this data?*

Postgres is an object-relational database, while MySQL is a purely relational database. This means that Postgres includes features like table inheritance and function overloading, which can be important to certain applications. Postgres also adheres more closely to SQL standards. Postgres handles concurrency better than MySQL for multiple reasons: Postgres implements Multiversion Concurrency Control (MVCC) without read locks Postgres supports parallel query plans that can use multiple CPUs/cores Postgres can create indexes in a non-blocking way (through the CREATE INDEX CONCURRENTLY syntax), and it can create partial indexes (for example, if you have a model with soft deletes, you can create an index that ignores records marked as deleted) Postgres is known for protecting data integrity at the transaction level. This makes it less vulnerable to data corruption.

*b. How do you map (translate) this schema to a Graph Db schema? Answer this question using the reference resources.*

**STEP ONE**:

The first step is to identify the data entity types. In my schema there are two types of data entities: user and cart items

**STEP TWO**:

In the second step, I need to find the semantic relations between those entities.

We find five relations between them which are:

Relation: Cart items – CreatedBy -> User

Logical Link: User\_id

**STEP THREE**:

|  |
| --- |
| user |
| user\_id |
| username |
| password |
| name |
| phoneNum |
| createdTime |
| address |

When data entities and relations are correctly identified, we can create nodes for each data entity and edges for each relation we just discovered. Now we’ve created a graph from the relational data model based on the predefined relation.

|  |
| --- |
| cart\_item |
| user\_id |
| session\_id |
| product\_id |
| price |

Created By

**STEP FOUR**:

Promoting those common attributes to nodes with proper relations to the entities could add value to my graph.

|  |
| --- |
| user |
| user\_id |
| username |
| password |
| name |
| phoneNum |
| createdTime |
| address |

When data entities and relations are correctly identified, we can create nodes for each data entity and edges for each relation we just discovered. Now we’ve created a graph from the relational data model based on the predefined relation.

|  |
| --- |
| cart\_item |
| user\_id |
| session\_id |
| product\_id |
| price |

Created By

Related to

Related to

|  |
| --- |
| User |
| User\_id |

PART II: LAB [30 points]

Collectd

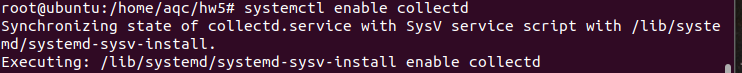
Text

Description automatically generated



Text

Description automatically generated





Text, chat or text message

Description automatically generated



Text

Description automatically generated

Httpf





Text

Description automatically generated



Text

Description automatically generated

Final result

Text

Description automatically generated