**15619 Project Phase 1 Report**

**Performance Data and Configurations**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Front end | Web service with HBase | Web service with MySQL |
| Query | q1 | q2 (small dataset) | q2 (small dataset) |
| Scoreboard request ID | 517 |  | 1931 |
| Instance type | m1.small |  | m1.large |
| Number of instances | 3 |  | 3 |
| Queries Per Second (QPS) | 6388.9 |  | 1423.1 |
| Error rate | 0 |  | 0 |
| Correctness | 100 |  | 100 |
| Cost per hour | 0.043 |  | 0.72 |

**Task 1: Front end**

**Questions**

1. Which front end system solution did you use? Explain why did you decide to use this solution.

2. Explain your choice of instance type and numbers for your front end system.

3. Did you do any special configurations on your front end system? Explain your design decisions and provide details here.

4. What is the cost to develop the front end system.

**Task 2: Back end (database)**

**MySQL:**

-Design:

Run MySql server at backend

-Table Structure:

The structure of each row is:

|  |  |  |  |
| --- | --- | --- | --- |
| idx(primary key) | userid | timestamp | twitterid |

Created index on attributes userid and timestamp

-Instance Number and Type:

Used 3 m1.large instances for backend

Only in this way could we satisify the test requirement

-Cost:

For large type instance, we used spot price at $0.04. We ran both backend and frontend on same instance simultaneously, so the price should be 3\*0.04/2=$0.06/h

**Hbase:**

**Task 3: ETL**

**MySQL:**

-Programming Model:

We used sequential programming model for MySQL ETL job. In extracting and transforming parts, we use python script, and in loading part, we use java application to add data into MySQL database twitter by twitter.

The reason for using this model is that the dataset used in this phase is not very large and the whole procedure did not take very long time.

-Instance Type:

We used one m1.medium type instance for whole ETL procedure. With medium type we could finish this procedure within two hours.

-Spot Price:

$0.02/h

-Execution Time:

1.5 hours

-Overall Cost:

$0.07

-The Number of Incomplete ETL Runs:

0

-Difficulties:

Gets access to MySQL server using java

-Backup:

**HBase:**

-Programming Model:

-Instance Type:

-Spot Price:

-Execution Time:

-Overall Cost:

-The Number of Incomplete ETL Runs:

-Difficulties:

-Backup:

**Questions**

1. Describe a MySQL database and typical use cases.
2. Describe an HBase database and typical use cases.
3. What are the advantages and disadvantages of MySQL?
4. What are the advantages and disadvantages of HBase?