**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 |  |  |  |
| Instance type |  |  |  |
| Number of instances |  |  |  |
| Queries Per Second (QPS) |  |  |  |
| Error rate |  |  |  |
| Correctness |  |  |  |
| Cost per hour |  |  |  |

**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)**

**Questions**

1. Describe your table design for both HBase and MySQL. Explain your design decision.

2. What is the cost to develop your back end system.

**Task 3: ETL**

Since ETL was performed for both HBase and MySQL, you will be required to submit information for each type of database.

MySQL:

1. The code for the ETL job
2. The programming model used for the ETL job and justification
3. The type of instances used and justification
4. The number of instances used and justification
5. The spot cost for all instances used
6. The execution time for the entire ETL process
7. The overall cost of the ETL process
8. The number of incomplete ETL runs before your final run
9. Discuss difficulties encountered
10. The size of the resulting database and reasoning
11. The time required to backup the database on S3
12. The size of S3 backup

HBase:

1. The code for the ETL job
2. The programming model used for the ETL job and justification
3. The type of instances used and justification
4. The number of instances used and justification
5. The spot cost for all instances used
6. The execution time for the entire ETL process
7. The overall cost of the ETL process
8. The number of incomplete ETL runs before your final run
9. Discuss difficulties encountered
10. The size of the resulting database and reasoning
11. The time required to backup the database on S3
12. The size of S3 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?