Group - 5 (Monday Batch)

Question:

A bundle of questions was asked concerning the performance testing performed by Group-5.
Question included:

1. How were the performance stats verified?
2. Are the numbers shown, an average of multiple tests performed, or a single query?
3. Are there any factors that you believe affected performance results?
4. Were you able to find any scenarios where indexing degraded performance?

Answer provided:

1. Performance testing was performed using simple queries to the database and measuring time.
2. Single query
3. Memory and internet
4. A scenario was mentioned in detail.

Your feedback on the answer and their project work:

Vishwak Tateni shared a satisfactory answer for part 4 by mentioning the scenario where they faced performance degradation. For the first three parts, the answers that a single query executed once was used to declare the performance stats, showed a lack of knowledge and experience among the team members in the area of performance analysis. Considering that last term performance analysis was one of the core subjects, measuring performance by checking the average of results captured multiple times was an expected answer. Additionally, answer 3 was incomplete. Performance factors include not only available memory and an internet connection, but also server configuration, load on the server, test cases, and verification methodology.

As a project, group-5 picked a very interesting topic and did well in researching and learning the concept of sharding. My only concern here was that the team pronounced the topic as performance evaluation of queries using concepts of indexing and sharding but missed a lot of research on the performance evaluation part. When we talk about performance evaluation, it shall be evaluated with different factors, environments and also the cases shall be executed multiple times to reach a stable performance number. Given the time constraint and scope of the project, even defining 1 env and its factors clearly would have made the project more presentable and appreciable.

To conclude, Team-5 presented a different idea and shared a ton of useful information with the classmates while explaining the concept of sharding very clearly.

Group - 1 (Wednesday Batch)

Question:

How is the idea focused on databases?

Answer provided:

Answer-1 Project focused on machine learning stream of databases.
Answer-2 Focus was on database as they were modifying, analyzing, and processing the database.

Your feedback on the answer and their project work:

The question was answered by 2 team members, Sanju Atwal and Prabjot Singh respectively. Sanju's answer was completely wrong as machine learning is not a advance database topic. On the other hand, Prabjot's answer was an unsuccessful attempt at evading the question.

While the topic of "Analysis of Suicidal and Depressing Posts based on Reddit" is very much interesting and unique, it lacks the major requirement of an advanced database topic project. The project focused on analyzing and querying database only. It does not involve the application of any advanced topic thereby making it not a suitable project choice. While it may have been an excellent machine learning project idea, it majorly skipped the database unit of the project.

Another reason for the topic not seeming suitable was the presentation of the project. The team's focus was on the machine learning and analysis part of the project instead of making it seem database-oriented. It should have been about query performance, data storage in the database, challenges in storing the amount of data, and challenges in performing operations when millions of rows of data are available.

In conclusion, Group-1 chose a very interesting topic which could have been one of the best topics if they had focused on the database part more.