Unique Research Paper Based Questions

GFS (Google File System)

1. Specify three most important considerations/rationale behind Google to design GFS. Also specify what design choices were made to address these in GFS.
2. Illustrate through examples what happens if a chunk server goes down in GFS, i.e., how the self-healing process takes place.
3. Explain the key steps involved when a file is submitted to a GFS to be stored including a diagram that illustrates the interaction among the key components.
4. Suppose you run out of space in a GFS system. You decided to add a new server with a set of new disks. Illustrate how rebalancing takes place to take advantage of the new added space in GFS.
5. Using Fig 1, describe how a chuck server failure is handled in GFS.
6. Why single GFS master is used? How does GFS recover from GFS master failure?
7. If the master node in GFS faces a sudden increase in workload due to high-frequency read requests, what steps could be taken to prevent bottlenecks?

Dremel

1. Explain briefly but precisely how Dremel architecture is designed both based on concepts from Web Search and Parallel DBMS.
2. What are the two key reasons Dremel is so fast and why?
3. Discuss the specific challenges and trade-offs involved in implementing Dremel's nested columnar storage format compared to traditional flat relational column

stores.

4. What is the key problem that Dremel handles? How is it different or compared to BigQuery?

BigQuery

1. What are key differences between BigQuery and MapReduce? Give examples of use cases where BigQuery is preferred over MapReduce and vice versa.
2. Explain the diffeernce between BigQuery and Dremel in tabular format.

Borg

1. In the Borg system, what role does the priority system play in ensuring resource allocation for high-priority tasks, especially during resource contention
2. Howdoes Borg prioritize between the need for rapid task startup and efficient resource utilization, and what trade-offs does this prioritization entail
3. Borg employs a priority and quota system for job scheduling. Describe how this system works and explain how it contributes to balancing resource allocation across production and non-production tasks.

DynamoDB

1. How does DynamoDB handle consistency? Why is it different from typical transactional system?
2. Describe the concept of eventual consistency in DynamoDB and how it supports high availability.
3. How does DynamoDB use data partitioning to handle scaling?