CSP554—Big Data Technologies

Assignment #8

Worth: 10 points (5 points for each problem)

Exercise 1) Read and provide a half page summary and analysis of this article available on the blackboard in the 'Articles' section: Dynamo: Amazon's Highly Available Key-value Store.

Dynamo, as the title of the paper suggests, is Amazon's highly available, scalable distributed data (key-value) storage system. Dynamo only supports primary-key access to data, which is useful for services such as provide best seller lists, shopping carts, customer preferences, session management, sales rank, and product catalog etc. Storage nodes can be added and removed from Dynamo without requiring any manual partitioning or redistribution.

Dynamo is only good and feasible for for certain applications. Reason is that it stores simple key value pairs. There is nothing to do with ACID properties. There might be many applications that wouldn't work on dynamo efficiently because they require the values of multiple keys to be updated simultaneously. Although, perhaps it would be okay because Dynamo is "always writable". Hence, all the keys would most likely get written to the storage, as requests are rarely rejected.

Why Amazon is using? Most of the services on Amazon only store and retrieve data by primary key and do not require the complex querying and management functionality offered by an RDBMS. Dynamo is incrementally scalable. It also allows the service owners to customize their storage system to meet their desired performance, durability and consistency

Conclusion: Amazon DynamoDB is a key-value and document database that delivers single-digit millisecond performance at any scale. It supports some of the world's largest scale applications by providing consistent, single-digit millisecond response times at any scale

Exercise 2) Read and provide a half page summary and analysis of this article available on the blackboard in the 'Articles' section: Data management in cloud environments: NoSQL and NewSQL data stores.

The article depicts on the storage aspects of cloud computing and the storage management in cloud computing. Nowadays most of the application is touched by the concept of cloud computing. Cloud computing has emerged as a computational paradigm. It is used in all kind of applications such as Web technologies, mobile devices, sensory data, IoT etc.

Conventional social databases were outlined in an alternate equipment and programming period and are confronting challenges in meeting the execution and scale prerequisites of Big Data. NoSQL and NewSQL information stores present themselves as choices that can deal with gigantic volume of information. Traditional relational databases were designed in a

different hardware and software era and are facing challenges in meeting the performance and scale requirements of Big Data.

The paper reviewed NoSQL and NewSQL data stores with the objective's objective of:

(1) giving direction to experts and specialists to pick the proper information store, (2) providing a perspective in the field (3) identifying challenges and opportunities in the field.

Besides, utilize cases and situations in which NoSQL and NewSQL information stores have been utilized are talked about and the reasonableness of different answers for various arrangements of uses is inspected.

Conclusion: This article focuses on aspects of cloud computing and its data management techniques. Apart from this it focuses on NOSQL concepts and tried to compare the concepts of the NoSQL and NewSQL data stores, together with the comparison of data stores on several dimensions such as Data Models, Querying, Scaling, and Security.