International Journal of Information Technology (IJIT) – Volume 6 Issue 6, Nov - Dec 2020

RESEARCH ARTICLE OPEN ACCESS

Influence of CAP Theorem on Big Data Analysis

Dr Anand Kumar Pandey [2], Rashmi Pandey [2]

[1] Computer Science and Application, ITM University

[2] Computer Science, ITM Group of Institute - Gwalior

ABSTRACT

In the current modern world different computing society has developed their innovative solutions to meet the difficult challenges to handle linear expansion in large data collection by different sources. In this paper, we took a research oriented view to achieve more significant ideas in the field of big data world and data science engineering. The CAP Theorem is a commonly cited unfeasibility outcome in distributed computing systems, particularly with NoSQL distributed databases. Cap theorem is very much influenced by cloud providers in the reference of their usability, the latency limit and system requirement. The Cap theorem is also very much influenced by distributer database also. To discover a substitute for CAP with a latency-centric point of view we have to observe how operation latency is exaggerated by network latency at dissimilar levels of consistency.

Keywords:- CAP, Big Data, Analysis, Distributed System, NoSQL.

I. INTRODUCTION

The CAP theorem is also known as Brewer's Theorem, because it was introduced by MIT Professor Eric A. Brewer during 2000 with the concept of distributed computing. Our main purpose in this paper is to consider the influence of CAP Theorem in the broader perspective of big data analysis and distributed computing theory.

Data analysis is usually applicable on current distributed big data centres to accomplish high performance and accessibility. Most of the data science services try to preserve their services stability in all situations. In modern world big data analysis and distributed database system is bound to have partitions in a real-world system due to network failure or some other reason. To describe the practical implementation of CAP theorem we can choose any real big data computing environment for data analysis such as MongoDB, Cassandra and with NoSQL database [2]. CAP theorem describes that before choosing any Database including distributed database, according to your requirement we have to choose only appropriate properties out of three. CAP theorem allows us to find out how we want to operate our distributed database systems when some other database servers decline to communicate with each other due to some imperfection in the system. During data analysis operations we try to retain the originality of actual data received from big data pool and follows all the suitable rules and regulations. Different types of database segments, operators and users are activated during the task of appropriate data analysis. Therefore it is very important to know about that which segment of dataset is consistent and suitable to apply some partition tolerance related operations. Services availability and database partition tolerance are inter-dependent to each other. It seems motivating to investigate which levels of regularity and reliability are strong enough to be straight implicit by the CAP constraints.

II. THE CONCEPTUAL ASPECTS OF CAP THEOREM

The CAP theorem explained the thought that there is a elementary transaction between availability, consistency, and partition tolerance. This transaction, which has become identified as the CAP Theorem, has been commonly discussed ever since. CAP theorem supports non-relational database (NoSQl) are best for distributed network applications and big data analysis [1].

With description of CAP theorem Professor Brewer illustrates that it not possible that a distributed computer system can support the 3 following properties at a time:

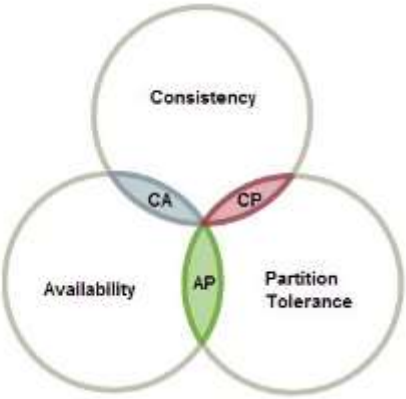

Fig. 1 CAP Theorem and Distributed Database management

CAP Theorem is a considered that a distributed database organization can only consist of 2 of the 3 properties: Consistency, Availability and Partition Tolerance as shown in Figure 1. Some of the well known attention about the CAP Theorem which derived from the fact is as follows:

A. *Safety*

CAP theorem has standard protection property, because every client sent correct request and receives correct response from data centres. These data centre can be personalised or generalized.

B. *Liveness*

The theorem is well defined but overall, it is quite lively. Availability is a classic liveness property because, every request receives a response.

C. *Unreliable*

There are so many different ways in which a system seems unreliable. Since every request at last receives a response so that structure can be *unreliable.* There may be partitions, as is discussed in the CAP Theorem.

III. THE ROLE OF CAP THEOREM IN BIG DATA ANALYSIS

During the review of CAP theorem as a consider its role and its influence for Big data analysis to achieve the distributed solutions. May be the first time solution is not perfect, so it needs to repeat the process of data analysis and again identify the best appropriate solution. In this case we are trying to get solutions for such kind of big data distributed system it is impossible to assurance all three features of Cap theorem like Consistency, availability and partition tolerance all at the equivalent time. Here we try to analyses the applicability and relationship of the CAP theorem with Big Data and distributed systems [1]. The CAP theorem is also relate with Hadoop, Big data Analytics, DBMS, network communication system and with advanced data structure.

During data analysis at first we have to partition the data in to appropriate segments. At the time of partitioning the user interacts with operator and operator interacts with database. Lets discuss the relationship of CAP theorem with big data centres and their nodes associated with wireless area network. As shown in figure 2 we have two data centres with their individual single server nodes and interconnected with dedicated network [3]. Here we try to propose a framework of interconnected big data centres for specifying a huge set of distributed data consistency model. The relationship of distributed big data described the correlated aspects of CAP theorem as per their requirements and their needs in this real world also.

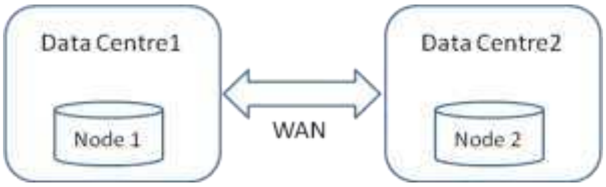

Fig. 2 Relationship of database

In this framework the CAP theorem involves as a software service consist in distributed system which makes wisdom to believe those reliability models in this extent. In above relational database, we have achieved Availability of the both data centres with respect to their concerned nodes as well as Partition Tolerance in both data centres even if they cannot correspond [5]. A communicated network divider is a particular type of communication defect that divides the network into subsets of nodes such that nodes in one subset cannot communicate with other nodes.

IV. FUTURE OF CAP THEOREM

The concept of CAP theorem is best possible futuristic approach to discuss basic trade off for available big data analytic solutions and distributed systems. In future it will easily scrutinize the inbuilt trade off some insights into that how system can be considered to gather an application's needs, in spite of unpredictable networks [6]. With the reference of CAP theorem we must know all theoretical aspects to achieve these challenges, and some modern techniques for supporting with the problem in real big data world system.

In this artificial intelligent and Big data world, the networked world has altered significantly in the last two decades, creating new challenges for system designers, and new areas in which these same inherent trade-offs can be explored. We need new theoretical insights to address these challenge, and new techniques for coping with the problem in real-world systems.

A. *Mobile Wireless Network*

The CAP Theorem primarily determined on modern wireless network services. Now days, we illustrate that its considerable growth proportion of network communication going to initiated by advanced mobile devices. A big data distributed database system is dedicated to comprise partitions in a real-world system due to network failure or some other reason.

Especially, wireless network communication is particularly untrustworthy. The main problem that the frequency is going to change quickly, so it is not easy to motivated the CAP Theorem for stable partitions. In every wireless networks, partitions are less common. After re-evaluating the CAP Theorem in the framework of wireless networks, we expect to better recognize the best appropriate solutions that take place in these types of scenarios [4]. By re-examining the CAP Theorem in the situation of wireless networks, we might hope to enhanced understand the unique trade-offs that occur in these types of scenarios.

B. *Scalability*

The CAP Theorem describes that in the current scenario of a network partition, the administrator *has* to decide between consistency and availability. Gradually, we involve that our systems be designed, scalable not just for today's consumers but also for future growth. Spontaneously, we believe that structure as scalable if it can mature resourcefully, using innovative resources capably to handle extra load.

C. *Tolerating Attacks*

Partition tolerance describes those clusters that must continue to work even though any number of communication breakdowns between nodes in the system. The CAP Theorem focuses on network partitions: occasionally, a number of servers did not communicate consistently [5]. Progressively, however, we felt that more rigorous attacks on networks.

Tolerating these extra problematic forms of interruption requires a somewhat different understanding of the fundamental consistency/ availability trade-offs. A rejection of service attack, however, cannot basically be modeled as a network partition.

V. SEGMENTING TASK OF DATA ANALYSIS

Those systems who are using aspects of CAP theorem some of them many systems do not include a single uniform requirement. Some aspects of the system require strong consistency, and some require high availability. In this section of the paper, we describe few of the dimensions along which a system might be partitioned. It is not always clear the specific guarantees that such segmentation provides, as it tends to be specific to the given application and the particular partitioning.

A. *Data Partitioning*

In this big data world different types of data analysis may require different levels of consistency and availability. For example, an on-line shopping cart may be highly available, responding rapidly to user requests; yet it may be occasionally inconsistent, losing a recent update in anomalous circumstances. The on-line product information for an e-commerce site may be somewhat inconsistent: users will tolerate somewhat out-of-date inventory information. The check-out/billing/shipping records, however, have to be strongly consistent.

B. *Functional Partitioning*

Many services can be divided into different subservices which have different requirements. For example, an application might use a service for coarse-grained locks and distributed coordination. Whatever service or function we need to be use can be partition as per condition and requirement.

C. *Operation Partitioning*

Different operations may require different levels of consistency and availability. Moreover, different types of updates might provide different levels of consistency. The CAP theorem and its data analysis provide with differing trade-offs for different types of read and write operations.

D. *User Partitioning*

Network partitions, and unfortunate network performance in general, normally correlate with real geographic distance: users that are far away are more likely to see poor performance. Usually, one could imagine that a social networking site might try to partition its users, ensuring high availability among groups of friends.

VI. CONCLUSIONS

In this paper we discussed several aspects of the CAP theorem: the definitions, the conceptual aspects of Cap theorem, the role of cap theorem in Big data analysis, future of CAP theorem in the literature are fairly paradoxical and counter- intuitive. The Cap theorem is also very much influenced by distributer database also and it is not possible to make available reliable data on both the nodes and accessibility of complete data. The CAP theorem describes that proposed distributed database system has to compose a transaction between Consistency and Availability when a Partition occurs.

REFERENCES

[1] Brewer EA (2012) CAP twelve years later: How the "rules" have changed. IEEE Computer 45(2):23–29, DOI 10.1109/MC.2012.37.

[2] Daniel J Abadi. Consistency tradeoffs in modern distributed database system design. IEEE Computer Magazine, 45(2):37– 42, February 2012. doi:10.1109/MC.2012.33.

[3] Dobre D, Viotti P, Vukolic M (2014) Hybris: Robust hybrid cloud storage. In: ACM Symposium on Cloud ´ Computing (SoCC), Seattle, WA, USA, pp 12:1–12:14, DOI 10.1145/2670979.2670991.

[4] Fekete A, Gupta D, Luchangco V, Lynch NA, Shvartsman AA (1996) Eventually-serializable data services. In: 15th ACM Symposium on Principles of Distributed Computing (PODC), Philadelphia, PA, USA, pp 300–309.

[5] Francesc D. Munoz-Esco, Ruben de Juan-Martin, J. R. Gonzalez de, , Jose M. Bernabeu, CAP Theorm: Revision of its Related Consistency Models, Technical Report TR-IUMTI-SIDI-2017/002, Universitat Politecnica de Valencia, 46022 Valencia (Spain).

[6] Martin Kleppmann, A critique of the CAP Theorem, article published in researchgate on Sept 2015.