An Epitome of Big Data Pertinence in Cloud Computing Technology

Parimala.S
Dept of Computer Science
SRM Institute of Science and Technology

Abstract- *When the companies need to have an updated technology within a limited budget, for them cloud computing technology becomes a blessing. They can choose what type of resources they want and pay for it. For the management of resources tools made easy in the Big Data environment and they are inexpensive. Big Data has developed in the past few years as a new archetype which affords copious data and chances to recover and/or allow exploration and decision support applications with unparalleled value for digital earth applications which consists of business, sciences and engineering. This paper explicates the applications of Big Data via the cloud computing technology*

Keywords- Big Data, Cloud Computing, applications, issues, challenges

I. INTRODUCTION

Distributed databases had been the boon of vision for research for few decades. But changes in the data patterns and applications has made way for the new type of storage called key value storage which are now being widely used by various enterprises.

For the vision of the researchers distributed databases has been a boon for some decades. Even then some changes in patterns of data and applications has given a way for the innovative method of storage called key value storage which are being extensively used by number of organizations. In the domain of Map reduce [1] and open source implementation of the same known as the Hadoop [2] has been used by majority of the industry and academics. Hadoop increases the usability and performance [3, 4].

II. BIG DATA

Big data is a buzzword practiced for portrayal of colossal amounts of data which are structured, semi structured or unstructured. If the data is cannot able to be handled by the traditional databases and software technologies then we then we classify that data as big data. The term big data [5] is originated from the web companies who used to handle loosely structured or unstructured data. The big data is defined using three v's.

1) Volume: Numerous features subsidize for the growth in capacity like storage of data, live streaming etc.
2) Variety: Several types of data is to be supported.
3) Velocity: The hustle at which the files are formed and processes are carried out mentions to the velocity

III. CLOUD COMPUTING

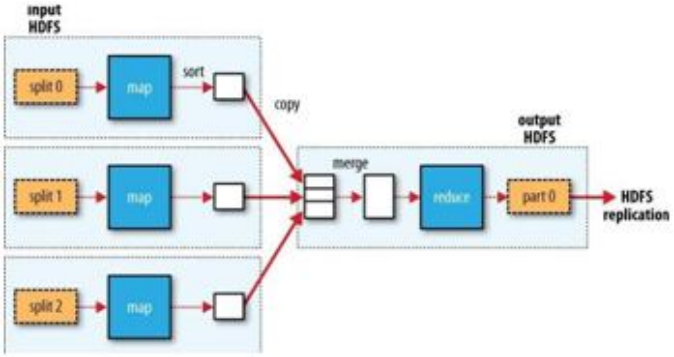
Cloud Computing is a new technology which is determined by on sharing of computing resources than having local servers or personal devices to switch the applications. In Cloud Computing, the word

"Cloud" means "The Internet", so Cloud Computing means a kind of computing in which services are transported through the Internet. local computers no longer have to take the entire burden when it comes to running applications. Cloud computing technology is being used to minimize the usage cost of computing resources [9].

From the user's end the task that should be carried out is to run the cloud interface software to link with the cloud.Cloud Computing comprises the front end and back end. The front end includes the user's computer and software required to access the cloud network. Back end consists of various computers, servers and database systems that create the cloud. The user can access applications in the cloud network from anywhere by connecting to the cloud using the Internet.

IV. HADOOP

This is an easily obtainable java based programming outline for the processing of large sets of facts in a distributed computing environment. Using Hadoop, massive amount of data sets can be managed over cluster of servers and apps may be run on system with thousands of nodes involving terabytes of information. This lowers the risk of system failure even when a huge number of nodes fail.it enables a scalable, flexible, fault tolerant computing solution. HDFS [6], a file system spanning all nodes in a Hadoop cluster for data storage links the file systems on local nodes to make it onto a very large file system thus improving the reliability.



- Task trackers are in charge for running the tasks that the job tracker assigns them
- Job trackers has two primary responsibilities which are managing the cluster resources and scheduling all user jobs
- Data engine entails of all the information about the processing the data
- Fetch manager assists to fetch the data while particular task is running.

V. MAP REDUCE

Map reduce [7] framework is used to write apps that process a large amounts of data in a reliable and fault tolerant way. The application is initially divided into individual chunks which are processed by individual map jobs in parallel. The output of map sorted by a framework and then sent to the reduce tasks. The monitoring is taken care by the framework.

VI. HADOOP DISTRIBUTED FILE SYSTEM (HDFS)

HDFS [8] is a file system that spans all the nodes in a Hadoop cluster for data storage. It links together file systems on local nodes to make it into one large file system. HDFS improves reliability by replicating data across multiple sources to over multiple nodes.

VII. BIG DATA APPLICATIONS

The big data application refers to the large scale distributed applications which usually work with large data sets. Data exploration and analysis turned into a difficult problem in many sectors in the span of big data. With large and complex data, computation becomes difficult to be handled by the traditional data processing applications which triggers the development of big data applications [10]. Google's map reduce framework and apache Hadoop are the defacto software systems [11] for big data applications, in which these applications generates a huge amount of intermediate data. Manufacturing and Bioinformatics are the two major areas of big data applications.

VIII. THE CLOUD AS AN ENABLER FOR BIG DATA ANALYTICS

How are big data and cloud technologies converging to make big data analytics in clouds a reasonable option? For big data analytics:

Vertical scaling attains elasticity by accumulation of supplementary illustrations with each of them portion a part of the demand. Software like Hadoop are precisely intended as distributed systems to take benefit of vertical scaling. They process small self-governing tasks in enormous corresponding scale. Distributed systems can also serve as data stores like NoSQL databases, e.g. Cassandra or HBase, or file systems like Hadoop's HDFS. Replacements like Storm provide synchronized stream data processes in near real-time through a cluster of machines with multifaceted workflows.

Archetypal cloud big data projects focus on scaling or accepting Hadoop for data processing. MapReduce has

become a practicing standard for large scale data processing. Tools like Hive and Pig have emerged on top of Hadoop which make it feasible to process huge data sets easily. Hive for instance converts SQL like queries to MapReduce jobs. It reveals data set of all sizes for data and business analysts for reporting and Greenfield analytics projects.

IX. ISSUES AND CHALLENGES

Cloud computing comes with abundant security concerns since it incorporates many skills which includes networks, operating systems, virtualization databases, transaction management, load balancing, resource scheduling ,concurrency control and memory management. Hence, security issues of these systems and tools are applied to cloud computing. For instance, it is very significant for the network which interrelates the systems in a cloud to be more secure. Also, virtualization paradigm in cloud computing results in several security concerns. For example virtual machines mapping to the physical machines has to be accomplished very confidentially.

X. CONCLUSION

This paper gave a description of an organized flow of about big data in the environment of cloud computing. We also discussed about some applications, issues and challenges of Big Data Technology via Cloud computing technology. In future, the challenges are need to be overcome and make way for the even more efficient use of the big data by the user on a cloud computing environment. It is very much needed that the computer scholars and IT professionals to cooperate and make a successful and long term use of cloud computing and explore new ideas for the practice of the big data over cloud environment.

REFERENCES

[1] D. Borthakur, “The hadoop distributed file system: Architecture and design,” Hadoop Project Website, vol. 11, 2007.

[2] The Apache Hadoop Project. http://hadoop.apache.org/core/, 2009.

[3] A. Abouzeid, K. B. Pawlikowski, D. J. Abadi, A. Rasin, and A.Silberschatz. HadoopDB: An Architectural Hybrid of Map Reduce and DBMS Technologies for Analytical Workloads.
PVLDB, 2(1):922–933, 2009.

[4] A. Thusoo, J. S. Sarma, N. Jain, Z. Shao, P. Chakka, S. Anthony,H.Liu, P. Wyckoff, and R. Murthy. Hive - A Warehousing Solution Over a Map-Reduce Framework. PVLDB, 2(2):1626–1629, 2009

[5] A, Katal, Wazid M, and Goudar R.H. "Big data: Issues, challenges, tools and Good practices.” Noida: 2013, pp. 404 – 409, 8-10 Aug.2013.

[6] K, Chitharanjan, and Kala Karun A. "A review on hadoop — HDFS infrastructure extensions.".JeJu Island: 2013, pp. 132-137, 11-12 Apr. 2013

[7] Wie, Jiang , Ravi V.T, and Agrawal G. "A Map-Reduce System with an Alternate API for Multi-core Environments.". Melbourne, VIC: 2010, pp. 84-93, 17-20 May. 2010.

[8] K, Chitharanjan, and Kala Karun A. "A review on hadoop — HDFS infrastructure extensions.".JeJu Island: 2013, pp. 132-137, 11-12 Apr. 2013.

[9] Y, Amanatullah, Ipung H.P., Juliandri A, and Lim C. "Toward cloud computing referencearchitecture: Cloud service management perspective.". Jakarta: 2013, pp. 1-4, 13-14 Jun. 2013.

[10] F.C.P, Muhtaroglu, Demir S, Obali M, and Girgin C. "Business model canvas perspective on big dataapplications." Big Data, 2013 IEEE International Conference, Silicon Valley, CA, Oct 6-9, 2013, pp.32 - 37.

[11] Zhao, Yaxiong , and Jie Wu. "Dache: A data aware caching for big-data applications using the MapReduce framework." INFOCOM, 2013 Proceedings IEEE, Turin, Apr 14-19, 2013, pp. 35 - 39.