ISSN: 0976-3031

Available Online at http://www.recentscientific.com

CODEN: IJRSFP (USA)

International Journal of Recent Scientific Research
Vol. 9, Issue, 1(D), pp. 23132-23135, January, 2018

International Journal of Recent Scientific Research

DOI: 10.24327/IJRSR

Research Article

STUDY OF MASSIVE SCALE COMPUTING IN IOT

Rahul Vasant.Nale[1]., Shubham Srivastava[2] and Aditya Kumar Sinha[3]

[1] Vel Tech University, Chennai
[2,3]CDAC-ACTS INDIA

DOI: http://dx.doi.org/10.24327/ijrsr.2018.0901.1402

ARTICLE INFO

Article History:

Received 17th October, 2017
Received in revised form 21st November, 2017
Accepted 05th December, 2017
Published online 28th January, 2018

Key Words:

Internet of Things, Big Data, Cloud Computing, Data Analytic Platforms.

ABSTRACT

Nowadays because of availability of smart devices, programming techniques, sensors, radio frequency technologies as well as wireless sensors networks, the internet of things (IoT) is giving lives smart and easy as compared to the previous decade. Due to multiple connected sensors, generation of data is in massive scale which refers to "Big data". Collecting, storing, processing, analyzing this data it is a hard job. The solution for this problem needs some data storing and analytic platform. Huge, fastest and secure cloud is one of the best solutions, which gives perfect scope for IoT applications. The paper focuses on various big data aspects and big data analytics platforms for cloud computing area.

INTRODUCTION

The internet of things is the collection of some important things physical and Virtual. This is transforming normal human life into smart human life. In software and embedded industry 'IoT' is popular but the reason behind this Popularity is cost-effective, tiny in size, low powered sensors. They are playing a crucial role in home automation, industrial automation, smart parking, wearable devices, health monitoring, etc. The technology of wireless sensors plays backbone role to give seamless connection to a real world (physical) with help of virtual world getting physical things sense by environment when supports with the inputs and responses given by sensors changing data. As per the static report of international data corporation (IDC),(R. Villars, C. Olofson and M. Eastwood,2011), every two years data size is double, and up to 2020 the data generate and copy annually will reach 44 zettabytes.

Motivation

Big data becomes an insignificant trend and it will change the whole science scope. In figure .1, big data characterized by four terms (zhang yaoxue1, ren ju1, liujiagang):

1. Volume- continuously growing volume of data from different sensors and sources, which is unable to handle traditional database.
2. Variety- various data collected via sensors, smart phones and social networks. The data collected in Structured data: Relational data, Semi-Structured data: XML data, Unstructured data: Word, PDF, Text, and Media Logs.
3. Velocity- the acquisition speed of the data and additionally the speed.4) value-Big Data implies the extraction of knowledge or patterns out of the raw data. Characteristics give great challenges in front of the Traditional computing in terms of massive-scale data and complex data collecting, storage, staging, processing.

Figure 1 The "4V" characteristics of big data

Worldwide popularity in terms of big data. Besides, some mature distributed data storage and processing technologies, like (HDFS) Hadoop distributed file system, MapReduce, NoSQL database, and Hive, can be implemented in a cloud

Corresponding author:* **Rahul Vasant.Nale
Veltech University, Chennai

environment and integrated with the cloud computing framework to address the challenges of big data storage, processing and analytics. Amazon Elastic MapReduce (EMR) is one of the delegates that utilize a hosted Hadoop framework running on (EC2) cloud to provide cost-effective big data services (l. chang, r. ranjan, z. xuyun, *et al*.,2013).cluster computing is one of the best solutions for the big data analysis, the experimental results show that the search time and response time is improved [4].data analytic frameworks like MongoDB Repository, IMDB, CiDAP, IoTDSF it will give different characteristics, used technology and its limitations(zhiming ding · jiajiexu · qi yang "seaclouddm).

Internet within one minute

When thinking about the internet connectivity, how much data will collect within one minute from social media form different social applications those most peoples are using likeWhatsApp, Facebook, E-Mail, Google, Instagram and so on. In the year of 2016 -17 analysis of these social data Shown in figure2. (Twitter "collection of data from social media in 2016 and 2017 online)

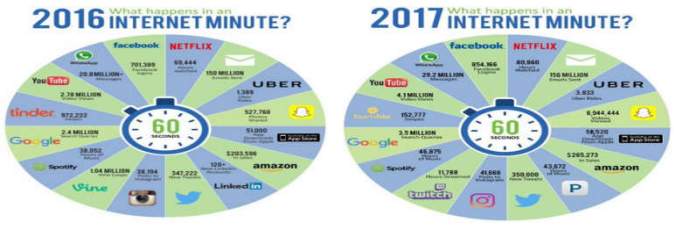

Figure 2 Collection of data from social media.

Table 1. Shows the difference between traditional data & IoT related big data. The traditional data supports Atomicity, Consistency, Isolation, and Durability (ACID) model whereas IoT related data support Consistency, Availability, Partition-tolerance (CAP) theorem & Basic Availability, the traditional data supports data in Gigabytes. The IoT related data supports in petabyte and Exabyte and database in NoSQL, Hadoop and important thing is data source in distributed format (A. Silberschatz, H. F. Korth, S. Sudarshan,2001)(Salomé Simon. (2012)). Shifting India towards IoT platform, The Indian Government proclaimed that $15 billion would be invested for the development of IoT marketplace up to the year 2020 and it will.Inintegration (IoT&Big data) to this, they have superimposition multiplier effect as shown in figure 3(DeitY.online)

Table 1 Traditional Data versus IoT

Parameter	Traditional data	IoT related data
Unit Of Data	Gigabytes (10^10)	Petabyte(10^15)/Exabyte(10^18)
Database	Relational	NoSQL database, Hadoop
Analytics	Batch/interactive/stored	Real time
Model	ACID	CAP theorem, BASE
Data Source	Centralized	Distributed

CAP theorem states that for a latest distributed system it's not possible to fulfill three requirements of consistency, availability and partition-tolerance at the same time. It is not probable to process IoT related big data in traditional ways. So cloud computing platform the IoT applications with suitable and enough resources in terms of storage, processing power and communication. The Concourse of IoT, Big Data and Cloud Computing in figure 4.

Figure 3 Value-based Multiplier Effect of the confluence of IoT and Big Data

Challenges of Big Data

Due to the different characteristics of big data, already implemented technologies and its theories of data management cannot fulfill the needs of big data (C. Chen and C. Zhang, 2014).

1. **Variety of data-** generated from heterogeneous devices makes discovering the related knowledge and information patterns hidden in data become challenging.
2. **Traditional data management-** technologies are not able to satisfy the requirements of big data storage (P. Russom, 2011)(S. Chaudhuri, 2012). Thus, how too easily and effectively store a large volume of information in order to efficiently process the data and it will give better recognize related knowledge hidden in information.
3. **Heterogeneous data-** like multimedia streaming data is continuously created at a high rate and sometimes has to be processed in a real-time manner. (S. Meng and L. Liu, 2013)

As per CAP a distributed system it's not possible to fulfill requirements. Different data cloud platform, security and privacy of important data. (A. R. Pathak and B. Padmavathi, 2014).

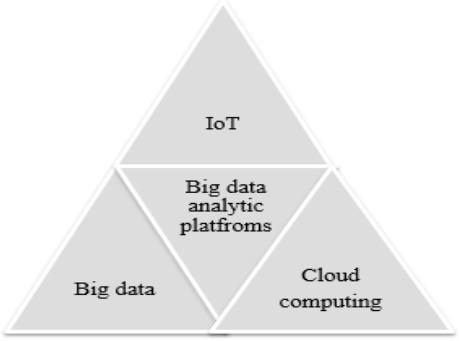

Figure 4 Concourse of emerging technologies

Cloud Computing and Its Related Technologies For Big Data

The Cloud computing is a network computing example, which supports on-demand access to a huge number of and any time for organizations and individuals(zhang yaoxue1, ren ju1,2017) Service provides via a network at any place significant changes in modern communications technology.

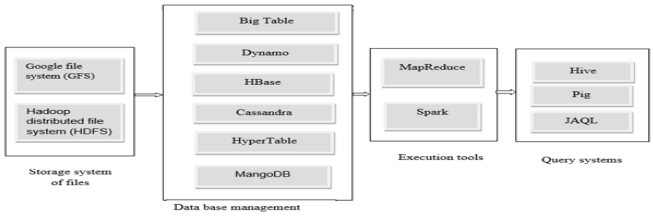

Figure 5 Framework for big data in cloud computing

Storage system of files

Cloud computing platform supports as per demand network access to computing sources. Files systems are the basic building of the architecture of cloud computing. It is consist of many computer nodes with an availability of fault-tolerance and good scalability. From the all the available files Google file system (GFS) (S. Ghemawat, H. Gobioff and S.T. Leung,2003)and Hadoop distributed file systems (HDFS) (T. White, Hadoop,2009). Because of the larger volume and continues generation, big data requires a larger amount of data storage space. The file systems GFS and HDFS is the implementation a common interface include applications which are in IaaS platform, they are handling big data with good effect.

Database management

Different database systems are developed to process various datasets at a different scale for supporting various applications .the traditional database (relational database) unable to handle large scale of data. NoSQL is playing the very important role promoting the solution of big data storage. NoSQL database is given in table 2.

Table 2 The classic DBMS based on NoSQL DBMS based

DBMS based on NoSQL	Authorizer	Description
Big Table	Google	database engine based on GFS, which includes set of key-value pairs that are of sparsity, distribution, durability and multi dimension
Dynamo	Amazon	Provides a tightly handle over tradeoffs between consistency, availability and extendibility and the technology of consistent hashing
HBase	Open source	Column-oriented database is built on HDFS, which supports executing of MapReduce tasks and Java API
Cassandra	Open source	hybrid-type NoSQL database which has a hierarchical structure based on columns
Hypertable	Open source	database provides Storage of data represented a multidimensional table and interfaces to a low-level API
MongoDB	Open source	NoSQL-type database supports a flexible processing of JSON type documents.

Execution tools

Big data processing modes are classified in two ways that is stream processing and batch processing (R.M. Chang, R.J. Kauffman and Y.O. Kwon,2014).in the stream processing value of data reduction as times goes, while in batch processing firstly data is stored and then can be processed off-line or online.MapReduce is one of the popular tools for batch processing modes.it is programming model developed for processing large scale of data in parallel by separating work in the group of independent tasks. MapReduce is having two important phase in data processing that is map phase and reduce phase. . The Framework of MapReduce is shown in figure 6. Spark (M. Zaharia, M. Chowdhury, M.J. Franklin, *et al*., 2010)is a framework in-memory data-processing that is compatible with Hadoop data sources but runs faster than MapReduce. For machine learning task spark is particularly suitable and interactive query base as well as it is very easy for a developer because it includes API in Python, Java, and Scala. Spark streaming is an extenuation of core Spark API that able to do real-time stream data processing. "DStream" stands for a sequence of Resilient distributed datasets (RDDs).it is integration into Spark framework and it can handle graph processing algorithm and spark learning But, Spark cannot conduct the group reduction operation and the only uses one task (at a time) to collect the results. It may seriously affect the scalability. The framework of spark is shown in fig.

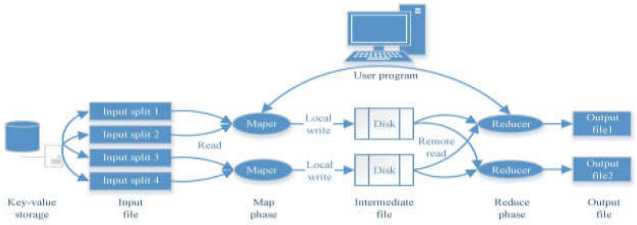

Figure 6 Map Reduce framework

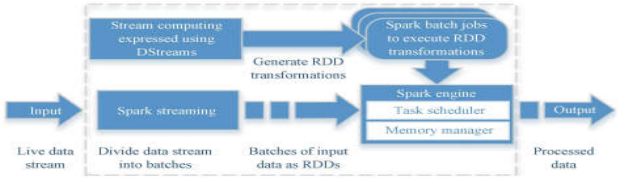

Figure 7 The Spark framework

CONCLUSION

The number of physical devices being connected to the Internet is increasing at a high rate driving the world to the area of Internet-of-things. Increasing volume, velocity in big data making big influences in IoT era. Different kinds of data processing and platform services can be seamlessly attended in the cloud environment. The data storage and analysis is the core and most important part of any system. Some important terms like scalability, availability, consistency and interoperability should be considered while designing the analytic platform.

References

1. R. Villars, C. Olofson and M. Eastwood, "Big data: What it is and why you should care", White Paper, IDC, MA, USA, 2011.
2. ZHANG Yaoxue1, REN Ju1, LIU Jiagang1, XU Chugui1, GUO Hui1 and LIU Yaping2 "A Survey on Emerging Computing Paradigms for Big Data" Chinese *Journal of Electronics* Vol.26, No.1, Jan. 2017
3. L. Chang, R. Ranjan, Z. Xuyun, *et al*., "Public auditing for big data storage in cloud computing: A survey", Proc. of IEEE16th International Conference on Computational Science and Engineering (CSE), pp.1128-1135, 2013.
4. Nandhakumar Ramachandran, Varalakshmi Perumal, Sakithya Gopinath and Monika Jothi "Sensor Search Using Clustering Technique in a Massive IoT

Environment" Springer Nature Singapore Pte Ltd.2017 DOI 10.1007/978-981-10-3953-9_27
5. Zhiming Ding · Jiajie Xu · Qi Yang "SeaCloudDM: a database cluster framework for managing and querying massive heterogeneous sensor sampling data"Published online: 18 April 2012 Springer Science+Business Media, LLC 2012
6. A. Silberschatz, H. F. Korth, S. Sudarshan, "Transactions" in Database System Concepts, McGraw-Hill, 4th ed. 2001, section 5, ch. 15, pp. 564- 568.
7. Salomé Simon. (2012). "Brewer's CAP Theorem" [Online]. Available:https://fenix.tecnico.ulisboa.pt/downloadFile/1126518382178117/10.eCAP-3.pdf
8. C. Chen and C. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on big data", Information Sciences, Vol.275, No.8, pp.314-347, 2014
9. P. Russom, "Big data analytics", TDWI Best Practices Report, pp.1–35, 2011.
10. S. Chaudhuri, "What next?: A half-dozen data management research goals for big data and the cloud", Proc. of the 31st ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, Scottsdale, AZ, USA, pp.1-4, 2012.
11. S. Meng and L. Liu, "Enhanced monitoring-as-a-service for effective cloud management", IEEE Transactions on Computers, Vol.62, No.9, pp.1705–1720, 2013.
12. DeitY. "Draft Policy on IoT by Government of India", [Online]. Available:http://deity.gov.in/sites/upload_files/dit/files/Revised-DraftIoT-Policy_0.pdf
13. A. R. Pathak and B. Padmavathi, "Analysis of Security Techniques Applied in Database Outsourcing", *Proc. Int. J. Computer Science and Information Technologies*, 2014, vol. 5, pp. 665-670.
14. twitter "collection of data from social media in 2016 and 2017", [Online] Available: https://twitter.com/cNews_iWitness/status/851837386870988801
15. S. Ghemawat, H. Gobioff and S.T. Leung, "The Google file system", Proc. of ACM SIGOPS Operating Systems Review, NewYork, USA, Vol.37, No.5, pp.29-43, 2003
16. T. White, Hadoop: The Definitive Guide, O'Reilly Media, Sebastapol, CA, 2009.
17. R.M. Chang, R.J. Kauffman and Y.O. Kwon, "Understanding the paradigm shift to computational social science in the presence of big data", Decision Support Systems, Vol.63, No.7, pp.67-80, 2014.
18. M. Zaharia, M. Chowdhury, M.J. Franklin, et al., "Spark: Cluster computing with working sets", Proc.of Hot Cloud, Boston,MA, pp.1–7, 2010.
19. Yong-Shin Kang, Il-Ha Park, Jongtae Rhee, and Yong-Han Lee, "MongoDB-Based Repository Design for IoT-Generated RFID/Sensor Big Data", Proc. *IEEE Sensors J.,* Vol. 16, 2016, pp. 485-497.
20. Tingli Li *et al*, "Storage Solution for Massive IoT Data Based on NoSQL" Proc. IEEE. Int. Conf. Green Computing and Communications, 2012, pp. 50-57.
21. Bin Cheng, Salvatore Longo, Flavio Cirillo, Martin Bauer, Ernoe Kovacs, "Building a Big Data Platform for Smart Cities: Experience and Lessons from Santander" Proc. IEEE. Int. Congr. On Big Data, 2015. pp. 592-599.
22. Lihong Jiang *et al*, "An IoT-Oriented Data Storage Framework in Cloud Computing Platform" Proc. IEEE Trans. Industrial Informatics, 2014, vol. 10, pp. 1443-141.

How to cite this article:

Rahul Vasant.Nale *et al*.2018, Study of Massive Scale Computing In IOT. *Int J Recent Sci Res.* 9(1), pp. 23132-23135.
DOI: http://dx.doi.org/10.24327/ijrsr.2018.0901.1402

****