ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology
(IJARTET) Vol. 5, Special Issue 14, April 2018

EMPIRICAL APPROACH FOR BIG DATA PROCESSING AND ANALYSIS IN CLOUD

1. **S.MD.MUJEEB,** Research Scholar, Jawaharlal Nehru Technological University Anantapur
2. **Dr. R.PRAVEEN SAM,** Professor, Department of CSE ,G.Pulla Reddy Engineering College
3. **Dr. K.MADHAVI,** Associate Professor & Additional Controller of Examinations, Department of CSE, JNTUA

ABSTRACT

Addressing big data has been a challenging task and time-demanding task that requires a large computational infrastructure to ensure successful data analysis and data processing.The continuous increase in the volume of data captured by organizations, such as the popularity of social media, Internet of Things (IoT), and multimedia, has produced avery great amount of data in either structured or unstructured format. Data creation is occurring at a record rate, referred to herein as Big data.Big Data has gain attention from the academia, government and industry, and has emerged as a widely recognized trend. Big Data are characterized by 3 aspects: (a) Data cannot be categorized into relational databases, (b) Data are great in number, and (c) Data are generated, captured, and processed very quickly. Moreover, big data is transforming healthcare, science, engineering, finance, business, and eventually, the society.

Big data and Cloud computing have a strong relation. Big data provides users the ability to use computing to process distributed queries across multiple datasets and return resultant sets in a timely manner. On other sideCloud computing is a powerful technology to perform complex and massivescale computing. It eliminates the need to maintain expensive computing hardware and dedicated space.

Cloud computing provides the underlying engine through the use of Hadoop, a class of distributed data-processing platforms. The size of data at present is huge and continues to increase every day.The velocity of data generation and growth is increasing because of the mobile devices and other device sensors connected to the Internet. The use of cloud services to store, process and analyze the data has been available frompast few years. Reducing the cost of data analytics in the cloud thus remains a main challenge and there is an increasing interest in complex analytics to derive value out of this big data.

RESEARCH PROBLEM

AIM

The aim of my proposal is to develop a better and faster processing and analysis methods for Big data.

OBJECTIVES

As using extent amounts of data is not an easy task. Only through their size alone, getting insight from Big data and ensuring quality can be difficult. This work will be consider successful if it results in strong, well-justified methods for Processing and Analyzing different types of Big Data, along with a feasibility of establishing a shared approach for using different Big Data sources.

PROPOSED RESEARCH WORK

Numbers of studies have addressed a few significant problems and issues related to the storage and processing of big data in cloud environment. The amount of data continues to increase at a rapid rate, but the improvement in the processing mechanisms is relatively slow. Only a few tools are available to address the issues of big data processing in cloud environments. The newest ideas, techniques and technologies in many

ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology
(IJARTET) Vol. 5, Special Issue 14, April 2018

important big data applications cannot solve the existing problems of storing and querying big data. For instance, Hadoop lack query processing strategies and have low-level infrastructures with respect to data management and processing. Despite the excessive amount of work performed to address the problem of storing and processing big data in cloud computing environments, certain important aspects of storing and processing big data in cloud computing are yet to be solved. Research issues that require substantial research efforts are summarized below:

- DATA PROCESSING

The most important issue is regarding big data processing is related to the diverse nature of data. Data gathered from different sources do not have a unique format. For instance, mobile-cloudbased applications, social networking sites and blogs are unsatisfactorily structured similar to pieces of text messages, videos, and images. Transforming and cleaning such unstructured data before loading those into the warehouse for analysis are time taking and challenging tasks. Efforts have been made to simplify the transformation process by adopting technologies such as Hadoop to support the distributed processing of unstructured formatted data. However, understanding the circumstances of unstructured data is necessary, particularly when meaningful information is to be mined.

- DATA ANALYSIS

The selection of an appropriate model for large-scale data analysis is very important. However, current algorithms are failing in terms of big data analysis. Therefore, efficient data analysis technologies and tools are required to process such huge data. Each algorithm performance decreases with increasing computational resources. As researchers continue to explore the issues of big data in cloud computing, new problems in big data processing emerges from the transitional data analysis techniques. The speed of stream data arriving from different data sources must be processed and compared with past information within a specific period of time. Such data sources may contain different kind of formats, which makes the integration of multiple sources for analysis a difficult task.

- BIG DATA CHALLENGES

Handling large data sets is a major task. Challenges are integrating complex and large datasets, getting started with the right big data project, developing and implementing infrastructure for managing and processing. Data Challenges include Volume, Veracity, Variability, Visualization and Value.Process Challenges include Data Acquisition and Warehousing,Data Mining and Cleansing,Data Aggregation and Integration,Data Analysis and Modeling,Data Interpretation. Management Challenges include privacy, Data Governance, Security, Operational expenditure, Data ownership etc.,

- **Big Data Analytical Methods**:

Big data can be analyzed in three ways namely descriptive, Predictive and prescriptive.

Descriptive analytics:

It is the simple analytical method which includes description of knowledge patterns and required statistical measures such as mean, median, mode, variance and standard deviation.

Predictive analytics:

It includes statistical modeling on supervised, unsupervised and semi supervised learning models. Predictive analytics aims to predict the future by analyzing current and historical data. For example, determination of customers' propensity to churn, by correlating behavior over a period of time with network event data

Prescriptive analytics:

It identifies cause-effect relationship among analytic results. Here organizations optimize their business process models based on the feedback provided by predictive analytic models.

RESEARCH METHODOLOGY

ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 14, April 2018

When huge amount of data is being analysed, traditional methods, developed for analysis of small samples, run into inconvenience or problems. There comes the need fornew tools and/or methods:

a. Methods to quickly discover information from massive amounts of data available,such as visualisation methods that are able to 'make Bigdata small'. Increasing computer power is a first way to assist this step.
b. Methods capable of integrating the information discovered in the statistical process, such as linking at massive scale, macrointegration and statistical methods specifically suited for large datasets. Methods need to be developed that rapidly produce consisten results when applied to very large datasets.

ETHICAL CONSIDERATIONS

"Less is more," stands true. In today's business world, many are wrapped up in the thought of, "the more data the better." But in actuality, to gain the most return on your investment, the key is to have just the "right" amount of data to solve your problem. Interesting many businesses actually need only 1%-10% of the amount of data they are currently collecting. Maybe this is something that businesses need to start taking a closer look at. The technology doesn't matter. It's about how to use that data to get some kind of return (more profit, savings, better data, reports, charts, etc) for an organization or business.

IMPACT OF RESEARCH STUDY

The importance of Big Data to the industry has been raised during recent years. Recommending appropriate environments and methods for analyzing and processing different types of Big Data to delivers results that are of important to the community can be done by improving data analytic techniquesby gathering all data and filtering them out on certain restrictions and use them to take beneficial decisions.

REFERENCES

[1] Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, Samee Ullah Khan,The rise of "big data" on cloud computing: Review and open research issues, Information Systems, ScienceDirect, Volume 47, January 2015, Pages 98–115
[2] J.J. Berman, Introduction, in: Principles of Big Data, Morgan Kaufmann, Boston, 2013, xix–xxvi (pp).
[3] Marcos D. Assuncao, Rodrigo N. Calheiros, Silvia Bianchi, Marco A.S. Netto, Rajkumar Buyya. Big Data computing and clouds: Trends and future directions, Journal of Parallel and Distributed Computing, ScienceDirect, Volumes 79–80, May 2015, Pages 3–15
[4] Marcos D. Assuncao, Rodrigo N. Calheiros, Silvia Bianchi, Marco A.S. Netto, Rajkumar Buyya. Big Data computing and clouds: Trends and future directions, Journal of Parallel and Distributed Computing, August 25, 2014
[5] Changqing Ji, Yu Li, Wenming Qiu, Uchechukwu Awada, Keqiu Li. Big Data Processing in Cloud Computing Environments, IEEE 2012, International Symposium on Pervasive Systems, Algorithms and Networks.
[6] Satoshi Tsuchiya, Yoshinori Sakamoto, Yuichi Tsuchimoto, Vivian Lee. Big Data Processing in Cloud Environments, FUJITSU Sci. Tech. Journal, Vol. 48, No. 2,pp. 159-168 (April 2012)
[7] Easwar Krishna Iyer, Sachin Sood, Neha Gupta & Tapan Panda. What Drives Big Data Analytics To Cloud? International Journal of Consumer & Business Analytics, 2014
[8] Divyakant Agrawal, Sudipto Das, Amr El Abbadi. Big Data and Cloud Computing: Current State and Future Opportunities, *EDBT 2011*, March 22–24, 2011, Uppsala, Sweden.
[9] P.Zikopoulos, K. Parasuraman, T. Deutsch, J. Giles, D. Corrigan, Harness the Power of Big Data The IBM Big Data Platform, McGraw Hill Professional, 2012.
[10] D.Loshin, Chapter5 – data governance for bigdata analytics: considerations for data policies and processes, in: D.Loshin(Ed.), Big DataAnalytics, MorganKaufmann, Boston,2013, pp.39–48.
[11] A paper on automatic Computing using Clouds by Giridhar et al., in International Journal of Advanced Trends in Computer Science and Engineering, Vol 2, No 1.
[12] Zhou, Z-H, Chawla, NV, Jin, Y and Williams, GJ (2014) Big Data Opportunities and Challenges: Discussions from Data Analytics Perspectives IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE, 9 (4). 62-74. ISSN 1556-603X
[13] Xindong Wu , Gong Quing Wu and Wei Ding " Data Mining with Big data ", IEEE Transactions on Knowledge and Data Engineering Vol. 26, No.1, Jan 2014
[14] H.Demirkan, D.Delen, Leveraging the capabilities of service- oriented decision support systems: putting analytics and bigdata in cloud, Decis.SupportSyst.55 (2013)412–421.

