International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 08 | August - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

BIG DATA ANALYTICS IN MOBILE NETWORKING

First author: Gupta Kajal Santosh Sanju
Second author: Mrs. Ritu Sharma
Application ID: 62027, **University Seat No:** 508102

Master of Science in Information Technology Part-II
Bhavna Trust Junior & Degree College of Commerce & Science,
University Of Mumbai

ABSTRACT

As the population of mobile users is increasing day by day, the data generated by the mobile cellular networks increases drastically. These data seems to be high in terms of velocity, variety and value. For efficient use of mobile cellular networks these data need to be analyzed by an effective technique. In this paper we review the various methods of analyzing data generated by mobile cellular networks. We aim to introduce the general background of data generated by mobile cellular networks and review certain technologies related to this. Mobile cellular networks have become both the generators and carriers of massive data. Big data analytics can improve the performance of mobile cellular networks and maximize the revenue of operators. We present an architectural framework for applying the big data analytics in the mobile cellular networks.

INTRODUCTION

Recent decades have witnessed tremendous increase in data in terms of size, speed, variety, value and veracity (called 5 V"s of big data). The term of big data is mainly used to describe this enormous datasets. The big data is comprised of masses of unstructured data which also requires real time analysis. If these datasets are effectively organized and managed many useful and in depth knowledge can be obtained which leads to finding solutions for various unsolved issues.

1.1 Mobile big data

Android Apps has provided more than 2.59 millions applications, covering nearly all categories. Such huge mass of data and numerous applications call for mobile analysis, but also bring about a few challenges. The unique characteristics of mobile data are found to be Mobile sensing, moving flexibility, noise, and a large amount of redundancy. Mobile phones are now useful for building and maintaining communities, and these communities with geographical locations and communities based on different cultural backgrounds and interests with their growing number of users and improved performance. The recent study in wireless sensor networks and mobile phones has lead to various mobile applications like real time health tracking etc., medical data from sensors seem to be of different characteristics in terms of attributes, time and space relations, as well as physiological relations, etc. In addition, such datasets involve privacy and safety protection. Apart from this many real time applications could also be developed which will be helpful to the mobile users. Apart from online health tracking applications many other applications can be developed

from the analysis of big data generated by mobile cellular networks. Many new customers‟ friendly applications can be developed by analyzing this enormous datasets. From these applications the mobile operators would be able to provide better services to their customers thereby improving their revenue. Apart from this many real time applications could also be developed which will be helpful to the mobile users.

1.2 Need for analysis

Wireless cellular networks have witnessed tremendous advances in recent decades. Due to ever increasing mobile applications, mobile cellular networks have become both generators and carriers of massive data. These data are generated while geo-locating mobile devices, recording phone calls, and capturing mobile applications' activities. These enormous data should be paid much attention, for efficient use of the mobile cellular networks and to increase the revenue of the mobile cellular operators. When compared to traditional data analytics Big data analytics would be an efficient method in analyzing such enormous unstructured data. Big data analytics deals with both structured and unstructured data while only structured data is dealt in traditional data analytics. In making real-time decisions traditional data analytics proves to be inadequate. Traditional data analytics fails in such cases such as to improve the performance of mobile cellular networks and to increase the revenue of its operators. In mobile cellular networks the complete data of the customers is scattered in various business department big data analytics is capable of collecting these data and extract useful information from these data while the traditional data analysis concentrates only on specific department.. The big data analytics helps the mobile operators in making dynamic and autonomous decision rather than traditional data analytics.

LITERATURE SURVEY

Wireless network virtualization and information-centric networking (ICN) are two promising techniques in software-defined 5G mobile wireless networks. Traditionally, these two technologies have been addressed separately. In this paper we show that integrating wireless network virtualization with ICN techniques can significantly improve the end-to-end network performance. In particular, we propose information-centric wireless network virtualization architecture for integrating wireless network virtualization with ICN. We develop the key components of this architecture: radio spectrum resource, wireless network infrastructure, virtual resources (including content-level slicing, network-level slicing, and flow-level slicing), and information-centric wireless virtualization controller. Then we formulate the virtual resource allocation and in-network caching strategy as an optimization problem, considering the gain of not only virtualization but also in-network caching in our proposed information-centric wireless network virtualization architecture. The obtained simulation results show that our proposed information-centric wireless network virtualization architecture and the related schemes significantly outperform the other existing schemes. Another new technology, called information-centric networking (ICN), has attracted great interests from both academia and industry. The basic principle behind ICN is to promote the content to a first-class citizen in the network. A significant advantage of ICN is to provide native support for scalable and highly efficient content retrieval while enabling the enhanced capability for mobility and security. ICN can realize in-network caching to reduce the duplicate content transmission in networks. The ICN-based air caching technique has been recognized as one of the promising-candidate techniques to efficiently implement the SDN-based 5G wireless networks. A number of research efforts have been dedicated to ICN, including the EU funded project Publish- Subscribe Internet Technology (PURSUIT) and the US funded project Named Data Networking (NDN). Although some excellent works have been done on wireless network virtualization and ICN, these two important areas have traditionally been addressed separately in the literature.

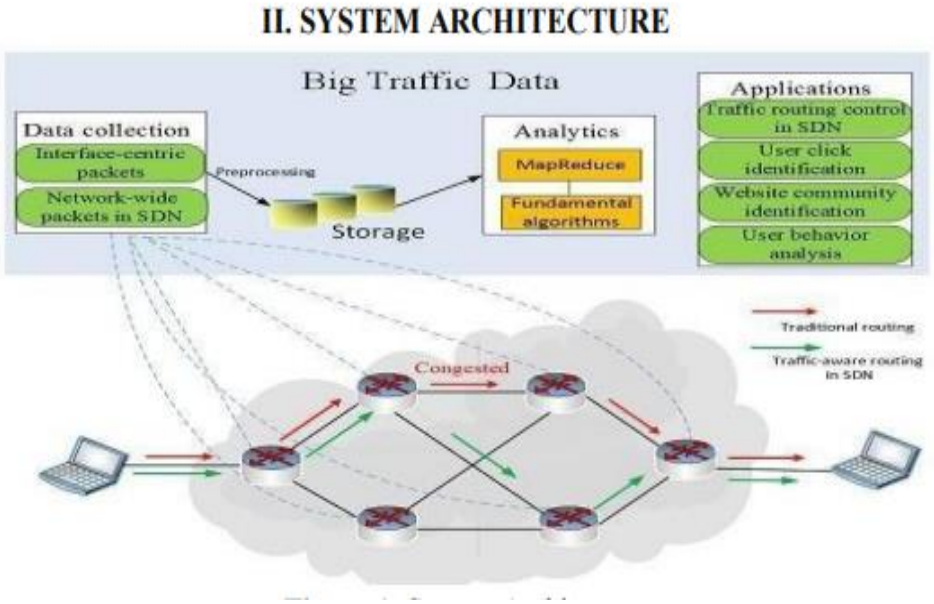
ADVANTAGES

- Effective large-scale analysis requires all of this to be done totally automatically.
- To compensate for lost data, redundancy might be investigated.
- We find Associations, patterns and to analyze the large data sets. Different methodologies associated with different algorithms used to handle such large data sets Of the data can be used to classify it as high.

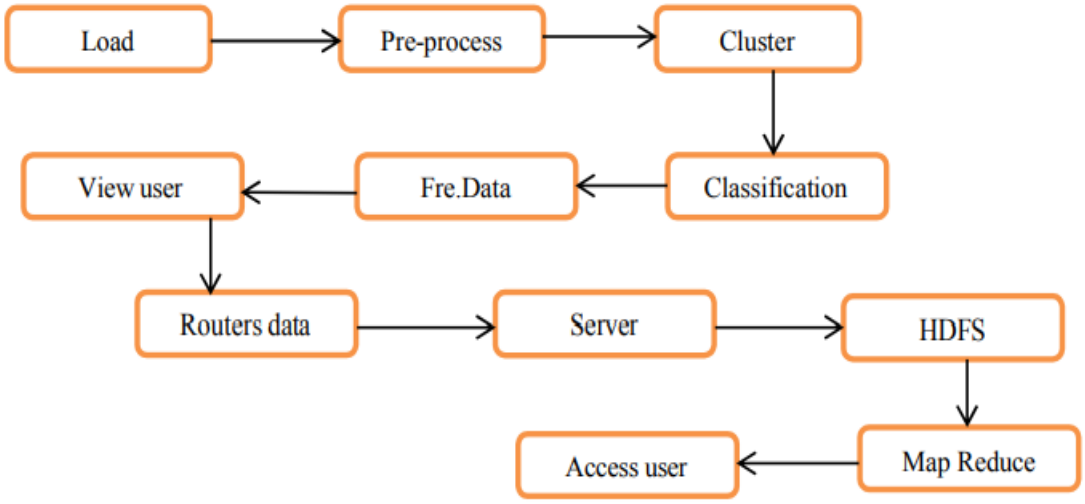
DISADVANTAGES

- All data which is not structured and is in free format is unstructured. In fact, most individuals and organizations achieve their lives around free data.
- Machine analyses algorithms expect homogeneous understand guidance.
- It also describes about the various security issues, application and trends followed by a large data set.

SYSTEM ARCHITECTURE



FLOW CHART

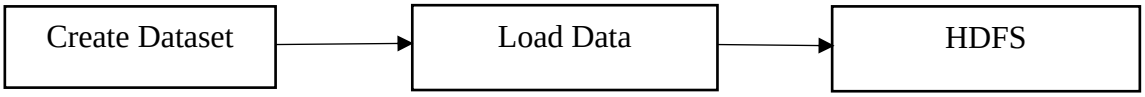


MODULES

1. Load Dataset
2. Data Clustering
3. Node Selection
4. Hadoop performance

A. Load Dataset

A data set or dataset, although this spelling is not present in many contemporary dictionaries is a collection of data. Most commonly a data set corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in question. Here Multiple attributes based create dataset and respective to the real world data transmissions with Processing data.



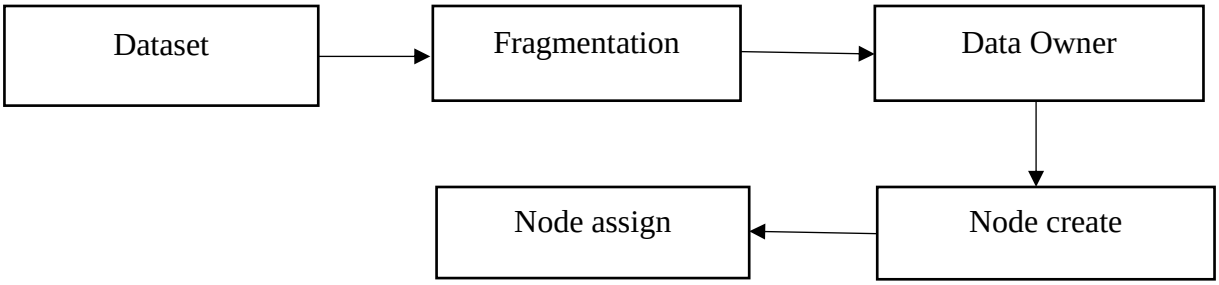
B. Data Clustering

It is a primary objective of exploratory data mining and a typical approach for statistical data analysis that is utilised in a variety of domains such as machine learning, pattern recognition, image analysis, information retrieval, bioinformatics, data compression, and computer science. graphics. As a result, clustering may be stated as a multi- objective optimization

problem. The best clustering technique and parameter settings (such as the distance function to employ, a density threshold, or the number of predicted clusters) are determined by the unique data set and intended application of the findings.

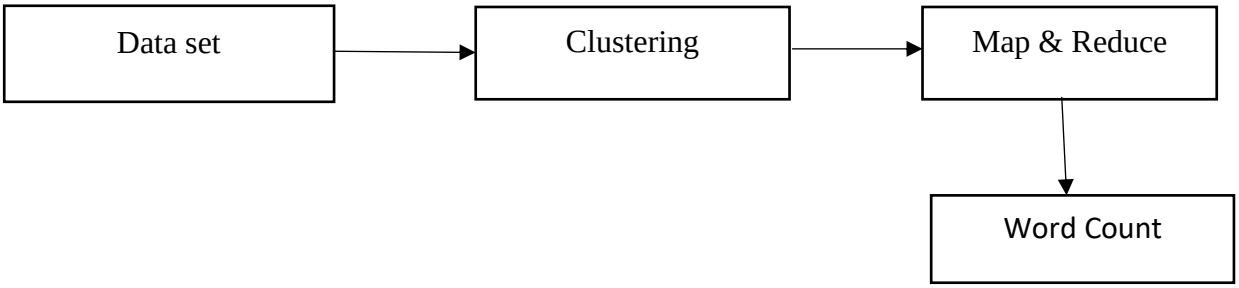
C. Node Selection

In the second iteration, the node with the lowest RC in combination with the previously selected node is chosen. The procedure is performed for each file fragment. The centrality metric is the same for both every every node. As a result, the same node is chosen to store the file fragment. As a result, the performance was the same, and all three lines intersected at the same position sign.



D. Hadoop Performance

Hadoop is a framework that enables Petabytes of data may be processed by programmes running on massive clusters with thousands of hardware-based nodes. Having a Hadoop cluster in production is only half the fight. It is critical for a Hadoop administrator. Optimize the Hadoop cluster configuration for best performance. During Hadoop installation, the cluster is setup using default configuration options that correspond to the minimum hardware
configuration. Hadoop with specific data as big data, with data set development based on performance analysis.



SYSTEM IMPLEMENTATION

- The ultimate installation of the package in its real environment, the satisfaction of the active user must be aware of the benefits of using the system
- Their confidence in the software built up
- Proper guidance is impaired to the user so that he is comfortable in using the application
- Intended users, and the operation of the system are all referred to as software implementation.
- People are unsure whether the software is intended to make their jobs simpler.

Before going ahead and viewing the system, the user must know that for viewing the result, the server program should be running in the server. If the server object is not running on the server, the actual processes will not take place.

1. User Training

To accomplish the goals and advantages anticipated from the proposed system, the individuals who will be participating must be confidence in their position in the new system. As the system gets more complicated, the importance of education and training grows. Education is a supplement to training. It brings formal training to life by describing the context of the resources available to them. Education entails providing the correct environment and inspiring user employees. Training may be made more entertaining and clear by providing educational content.

2. Training on the Application Software

Users will need to be educated on the new application software after receiving the appropriate basic computer awareness training. This will provide the basic philosophy of the new system's usage, such as the screen flow, screen design, kind of help on the screen, type of mistakes while entering data, the associated validation check at each input, and methods for correcting the data entered. This training may change amongst user groups and at different levels of organisation.

CONCLUSION

This study has been done to explore various methods of big data analytics in mobile cellular networks. Big data analytics will be an indispensable part of the mobile cellular operators' consideration of network operation, business deployment, and even the design of the next-generation mobile cellular network architectures. In this paper, the connection between big data analystics and mobile cellular net-works has been systematically explored.

FUTURE SCOPE

Finally, we reviewed several research problems and the opportunities for big data analytics in next-generation cellular networks. Future effort is being done to solve these issues. We blended several methodologies and domains to improve speed and security.

REFERENCES

- C. Liang, F. R. Yu, and X. Zhang, “Information-centric network function virtualization over 5G mobile wireless networks,” *IEEE Netw.*, vol. 29,no. 3, pp. 68_74, May/Jun. 2015.
- L. Cui, F. R. Yu, and Q. Yan, ``When big data meets software-defined networking: SDN for big data and big data for SDN," *IEEE Netw.*, vol. 30, no. 1, pp. 58_65, Jan./Feb. 2016.
- R. C. Qiu, ``Large random matrices and big data analytics," in Big Data of Complex Networks. Boca Raton, FL, USA: CRC Press, 2016.
- R. C. Qiu and P. Antonik,Smart Grid and Big Data. New York, NY, USA: Wiley, May 2016.