International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 7, ISSN No. 2455-2143, Pages 107-110
Published Online November 2022 in IJEAST (http://www.ijeast.com)

AN EFFECTIVE IDENTIFICATION OF TRUTHFUL INFORMATION USING POLYMERIZATION SENTIMENTAL MODLEL

Rajalakshmi R, Nandini Devi T
Department Of CSE,
Thirumalai Engineering College,
Kanchipuram,Tamil Nadu, India.

Abstract— **Despite the increasing use of social media platforms for information and news gathering, its immoderate nature will create peoples like keyboard warriors, rumormongers etc. This often leads to the emergence and spreading of rumors, i.e. pieces of information that are unverified at the time of posting. At the same time, the openness of social media platforms provides opportunities to study how users share and discuss rumors, and to explore how natural language processing and data mining techniques may be used to find ways of determining their veracity. In this survey, we introduce the Big Data technology and discuss two types of rumors that circulate on social media; long-standing rumors that circulate for long periods of time, and newly-emerging rumors spawned during fast-paced events such as breaking news, where reports are released piecemeal and often with an unverified status in their early stages. We provide an overview of datasets into social media rumors with the ultimate goal of developing a rumor classification system that consists of four components: rumor detection, rumor tracking, rumor stance classification and rumor veracity classification. We delve into the approaches presented in the scientific literature for the development of each of these four components. We summarize the efforts and achievements so far towards the development of rumor classification systems and conclude with suggestions for avenues for future research in social media mining for detection and resolution of rumors**

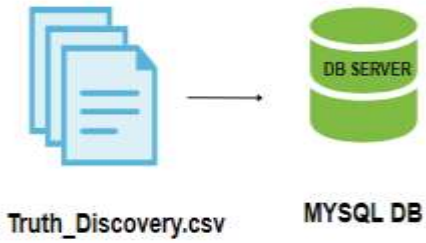
Keywords— **big data, polymerization sentimental analysis, hadoop**

I. INTRODUCTION

The data which is beyond the storage capacity and beyond the processing power such a data is called Big Data. Big data means really a big data; it is a collection of large datasets that cannot be processed using traditional computing techniques. Big data is not merely a data; rather it has become a complete subject, which involves various tools, techniques and frameworks. Data which are very large in size is called Big Data. Normally we work on data of size MB(Wordbook, Excel) or maximum GB(Movies, Codes) but data in Petabytes i.e. 10^15 byte size is called Big Data. It is stated that almost 90% of today's data has been generated in the past 6 years.
In this paper we are analyzing truth data by using Hadoop tool along with some Hadoop ecosystems like hdfs, Map Reduce, sqoop, hive and pig. By using these tools processing of data without any limitation is possible, no data lost problem, we can get high throughput, maintenance cost also very less and it is an open source software, it is compatible on all the platforms since it is Java based. In data is related large volume of storage of research paper publishing website.

II. PROPOSED ALGORITHM

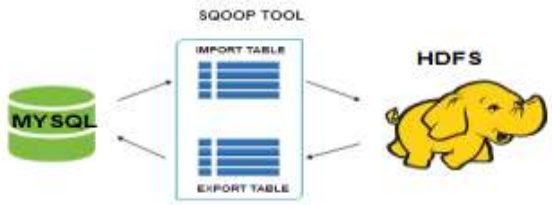
Existing Application (MySQL):
In MySQL is a relational database management system. RDBMS uses relations or tables to store Truth data as a matrix of rows by columns with the primary key. With MySQL language, Textual data in tables can be collected, stored, and processed, retrieved, extracted and manipulated mostly for business purpose. The existing concept deals with providing backend by using MySQL which contains a lot of drawbacks i.e. data limitation is that processing time is high when the data is huge and once data is lost we cannot recover so thus we proposing concept by using Hadoop tool.



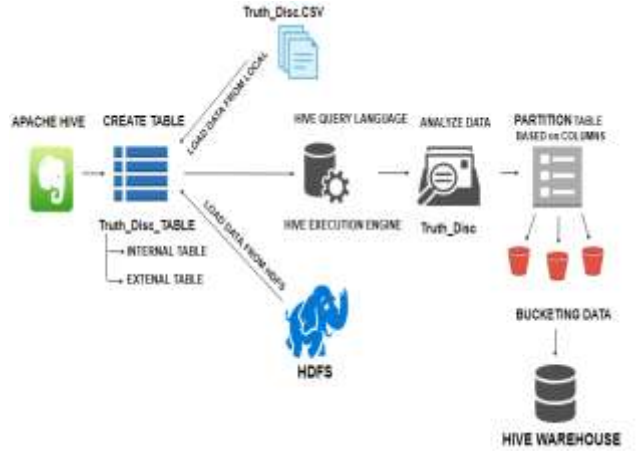
Connector (Sqoop):
Sqoop is a command-line interface application for transferring Truth Discovery data between relational databases (MySQL) and Hadoop. Here in MySQL database having Truth Discovery data have to import it to HDFS using Sqoop. Truth

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 7, ISSN No. 2455-2143, Pages 107-110
Published Online November 2022 in IJEAST (http://www.ijeast.com)

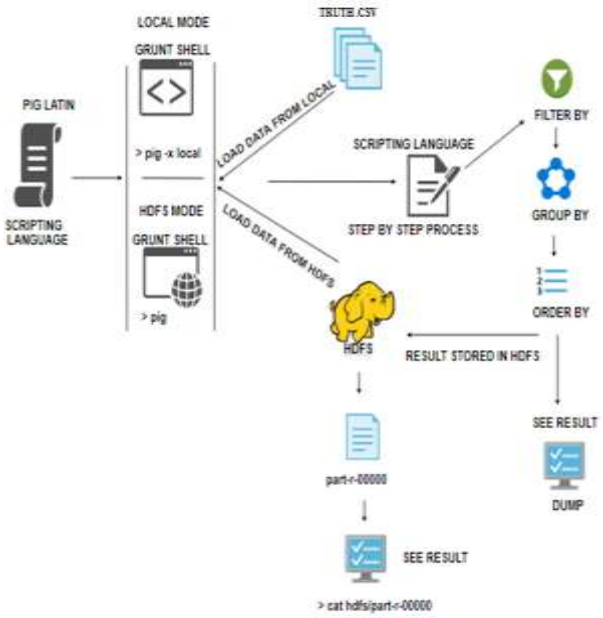
Discovery data can be moved into HDFS/Hive from MySQL and then it will generate the java classes. In previous cases, flow of data was from RDBMs to HDFS. Using "export" tool, we can import data from HDFS to RDBMs. Before performing export, Sqoop fetches table metadata from MySQL database. Thus we first need to create a table with required metadata.



Analysis Query Language (Hive):
Hive is a data warehouse system for Hadoop that runs SQL like queries called HQL (Hive query language) which gets internally converted to map reduce jobs. In Hive, Truth Discovery data tables and databases are created first and then data is loaded into these tables. Hive as data warehouse designed for managing and querying only structured data that is stored in tables. Hive organizes Truth Discovery data tables into partitions. It is a way of dividing a table into related parts based on the values of partitioned columns. Using partition, it is easy to query a portion of the given dataset. Tables or partitions are sub-divided into buckets, to provide extra structure to the Truth Discovery data that may be used for more efficient querying. Bucketing works based on the value of hash function of some column of a table.

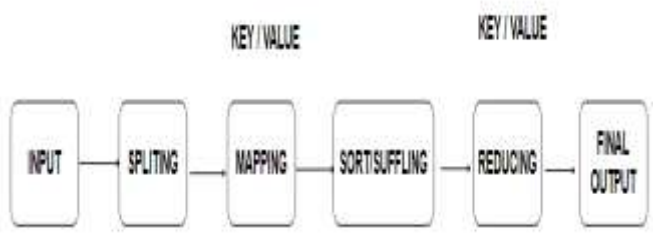


Analysis Latin Script (Pig):
To analyze Truth Discovery data using Pig, programmers need to write scripts using Pig Latin language and execute them in interactive mode using the Grunt shell. All these scripts are internally converted to Map and Reduce tasks. After invoking the Grunt shell, you can run your Pig scripts in the shell. Except LOAD and STORE, while performing all other operations, Pig Latin statements take a relation as input and produce another relation as output. As soon as you enter a Load statement in the Grunt shell, its semantic checking will be carried out. To see the contents of the schema, you need to use the Dump operator. Only after performing the dump operation, the Map Reduce job for loading the data into the file system will be carried out. Pig provides many built-in operators to support data operations like grouping, filters, ordering, etc.



Processing (Map Reduce):
Map Reduce is a framework using which we can write applications to process huge amounts of Truth Discovery data, in parallel, on large clusters of commodity hardware in a reliable manner. Map Reduce is a processing technique and a program model for distributed computing based on java. The Map Reduce algorithm contains two important tasks, namely Map and Reduce. Map Reduce program executes in three stages, namely map stage, shuffle stage, and reduce stage**.** The map or mapper's job is to process the input data. Generally the input data is in the form of file or directory and is stored in the Hadoop file system (HDFS). The input file is passed to the mapper function line by line. The mapper processes the data and creates several small chunks of data. This stage is the combination of the Shuffle stage and the Reduce stage. The Reducer's job is to process the data that comes from the mapper. After processing, it produces a new set of output, which will be stored in the HDFS.

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 7, ISSN No. 2455-2143, Pages 107-110
Published Online November 2022 in IJEAST (http://www.ijeast.com)



III. EXPERIMENT AND RESULT

Apache Spark is an open source processing engine built around speed, case of use, and analytics. If you have large amounts of data that requires low latency processing that a typical Map Reduce program cannot provide, Spark is the alternative. Spark provides in-memory cluster computing for lightning fast speed and supports Java, Scala, and Python APIs for ease of development**.**

SOFTWARE REQUIREMENTS:
INTRODUCTION:
The following describes the requirement specification for An efficient and scalable detection of truth in social media using big data.
Purpose

The purpose of this document is to outline required hardware and software specifications that are needed to big data platform.

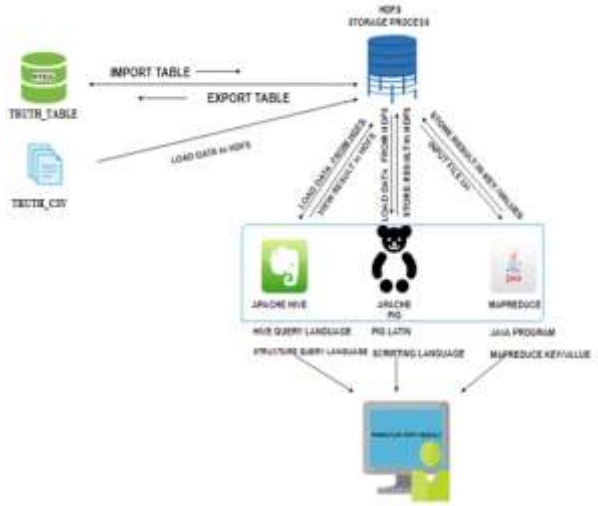
SOFTWARE DESCRIPTION:

Framework : Hadoop
Operating System : cent OS
IDE : Eclipse
Database : MySQL

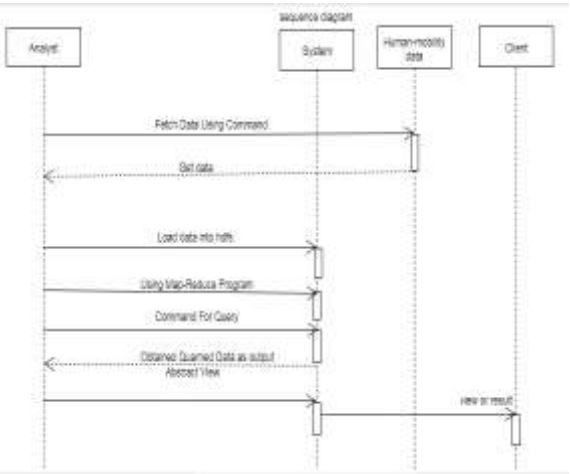
HARDWARE REQUIREMENTS:

Processor : Pentium IV 2.6 GHz, Intel Core 2 Duo.
Ram : 4GB DD Ram
Monitor : 15" color
Hard disk : 80 GB

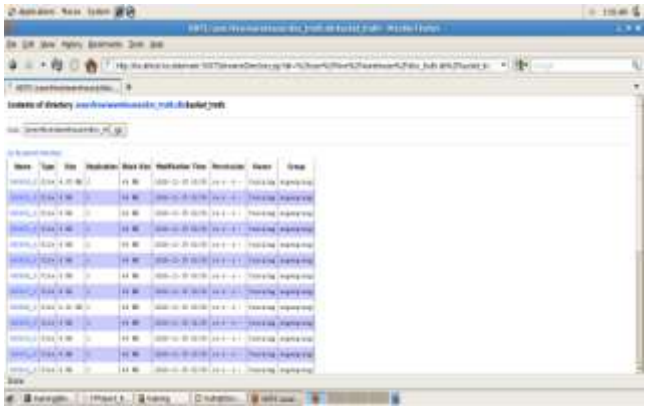
SYSTEM ARCHITECTURE:



WORKING OUTLINE:



EXPERIMENT SAMPLE RESULT:



IV. CONCLUSION

In this paper, we presented a study Hadoop ecosystem is having hive, pig, map reduce tools for processing whether output will take less time to process and result will be very fast. Hence in this project already textual data which is traditionally going to store in RDBMS which has less

performance hence by using Hadoop tool we can be faster and efficiently processing the data.

V. FUTURE WORK:

Apache Spark is an open source processing engine built around speed, case of use, and analytics. If you have large amounts of data that requires low latency processing that a typical Map Reduce program cannot provide, Spark is the alternative. Spark provides in-memory cluster computing for lightning fast speed and supports Java, Scala, and Python APIs for ease of development.

VI. REFERENCE

[1]. M. Litzkow, "Remote Unix," Proceedings of 1987 Summer Usenix Conferences, Phoenix, Arizona, (June, 1987).

[2]. D. Chavey, Private Correspondence, University of Wisconsin, Madison, Wisconsin, (December, 1986).

[3]. P. Sandon, "Learning Object-Centered Representations," Ph. D. Thesis, University of Wisconsin, Madison, Wisconsin, (August, 1987).

[4]. Bo Pang and Lillian Lee., "Opinion Mining and Sentiment Analysis", Foundations of Information Retrieval, Vol. 2, Nos. 12,1-135,2008.

[5]. Alistair Kennedy , Diana Inkpen, "Sentiment Classification of Movie Reviews Using Contextual Valence Shifters", Computational Intelligence, Volume 22, 2006.

[6]. Andrew McCallum , Kamal Nigam, "A comparison of Event models for Naive Bayes text classification", AAAI-98 workshop on learning for text categorization, 1998.

[7]. J. J. McAuley, C. Targett, Q. Shi, and A. van den Hengel, "Image based recommendations on styles and substitutes," in SIGIR, 2015, pp. 43–52.

[8]. E. M.Rogers, Diffusion of Innovations. New York: The Rise of High- Technology Culture, 1983.

[9]. K. Sarkar and H. Sundaram, "How do we find early adopters who will guide a resource constrained network towards a desired distribution of behaviors?" in CoRR, 2013, p. 1303.

[10]. D. Imamori and K. Tajima, "Predicting popularity of twitter accounts through the discovery of link-propagating early adopters," in CoRR, 2015, p. 1512.

[11]. D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on. IEEE, 2000, pp. 44–55.

[12]. D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Eurocrypt, vol. 3027. Springer, 2004, pp. 506–522.

[13]. P. Xu, H. Jin, Q. Wu, and W. Wang, "Public-key encryption with fuzzy keyword search: A provably secure scheme under keyword guessing attack," IEEE Transactions on computers, vol. 62, no. 11, pp. 2266–2277, 2013.

[14]. R. Chen, Y. Mu, G. Yang, F. Guo, and X. Wang, "A new general framework for secure public key encryption with keyword search," in Australasian Conference on Information Security and Privacy. Springer, 2015, pp. 59–76.

[15]. Q. Huang and H. Li, "An efficient public-key searchable encryption scheme secure against inside keyword guessing attacks," Information Sciences, vol. 403, pp. 1–14, 2017.