ISSN: 0976-3031

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

CODEN: IJRSFP (USA)

International Journal of Recent Scientific Research
Vol. 8, Issue, 6, pp. 17703-17707, June, 2017

International Journal of Recent Scientific Research

DOI: 10.24327/IJRSR

Research Article

AN IMPROVED LEARNING BASED DISASTER EVENT USING BIG DATA ANALYTICS

Saranya M[1] and Prema A[2]

Department of Computer Science, Raja Doraisingam Govt Arts College, Sivagangai

DOI: http://dx.doi.org/10.24327/ijrsr.2017.0806.0408

ARTICLE INFO

Article History:

Received 06th March, 2017
Received in revised form 14th April, 2017
Accepted 23rd May, 2017
Published online 28th June, 2017

Key Words:

Big Data, Hadoop, ETL, Map Reduce, HDFS

ABSTRACT

Big data is a word for datasets that are so big or multifaceted that traditional data dispensation applications are inadequate to deal with them. It is not merely a data, rather it has become a complete subject, which involves various tools, techniques, and frameworks the face up to of extracting value from big data is parallel in many ways to the age-old problem of distilling business aptitude from transactional data. At the heart of this challenge is the process used to extract data from multiple sources, transform it to fit your analytical needs, and load it into a data warehouse for consequent analysis, a process known as "Extract, Transform & Load" (ETL). The Hadoop Distributed File System (HDFS) is the storage component of Hadoop which is used to implement the disaster management process. Map Reduce method has been calculated in this paper which is required for implement Big Data Analysis using HDFS.

INTRODUCTION

The skill and upward quantity of data(Big Data), necessitate is felt towards implement effective analytics techniques(Big Data analytics) to analyses this big degree of data for unknown and useful facts, pattern, Big data is the term for data sets so large and difficult that it becomes complicated to process using fixed data management tools or processing application. This paper reveals most recent progress on big data networking and big data [1]

Wang, F. Described that Map-reduce is a software framework for distributed processing of large data sets on computer clusters. Map reduce is intended to make possible and simplify the processing of vast amounts of data in parallel on a large cluster of commodity hardware in a reliable, fault-tolerant manner [2]

It is calculated to scale up from single servers to thousands of machines, each contribution local computation and storage Hadoop consists of two components Hadoop distributed file system (HDFS) and Map Reduce framework.[3]

Big Data is a gathering of the large dataset that cannot be processed using conventional Computing technique. Big Data is not merely a data rather it has become a complete subject Which involves various tools techniques as well as framework The necessitate of big data generated from the large companies like Facebook, Yahoo, Google, Youtube etc for the purpose of massive amount of data also Google contain the large amount of in sequence[4]

The ETL method consists of scheming a target, transforms statistics for the target, scheduling and monitor processes. The reason of using ETL tools is to save time and make the whole process more steady and adapted to provide the functionality to meet the venture necessity. Hence many of them choose to make their own data ware house themselves [5, 6, 7]

Disaster management refers to the inclusive approach in all phase of disaster successfully dropping the collision of disaster. Disaster management series consists of the next different phases. Felling, Relief & Rescue, healing and mitigation phases. Over the past decade, innovative use of statement and meteorological means of INSAT system is begin operationally used towards the track, monitor, and forecast of the cyclone. The recent achievement includes deluge mapping of all the major floods in the country in near real time mode, drought severity appraisal using settlement data on journal occasion scales, the landslide zonation pilgrimage route in the Himalayas, monitor of the cyclone and harm estimation. The potential of the geographic positioning system(GPS) to precisely determine the position of a location is being used to measure ground movements associated with plate tectonics ISRO has well-known a decision support center (DSC) at National Remote Sensing Agency, Hyderabad. To provide sensible in order meeting the user needs in provisos of in sequence content, turn-around-time, and set-up

Corresponding author:* **Saranya M
Department of Computer Science, Raja Doraisingam Govt Arts College, Sivagangai

Many regions in India are highly susceptible to usual and another disaster on the explanation of physical Situation. About 60% of the total area of the country is susceptible to seismic injury of buildings in undependable degrees. The most susceptible areas, according to the present seismic zone map of India, area situated in the Himalayan regions. Kutch and the Andaman and Nicobar Islands, which are chiefly earthquake risk prone. Over 8% Indian area of 40 million hectares is flat to floods, and the regular area overstated by floods annually

Related Works

The design and assessment of a data conscious cache framework that requires a minimum change to the original map reduce encoding model for provisioning incremental elsewhere for Big data application using the Map Reduce mock-up [9]. Kulkarni *et al* stated the meaning of the technology that grip data like Hadoop, HDFS, and map reduce. And not compulsory about various schedulers used in Hadoop and about the scientific aspects of Hadoop. He also focused on the weight of YARN which overcomes the boundaries of Map Reduce [10]

The Keane *et al* has given some vital emerging frame for big data Analytics and a three-tier structural design model for Big Data in Data Mining. In the future, the three-tire design model is more scalable in operational with different situation and also profit to trounce with the main matter in Big Data Analytics for the store, analyze, and hallucination. The framework model given for Hadoop HDFS spread data storage, real-time NoSQL databases, and Map Reduce distributed data giving out over a cluster of product servers [11]

Inmon *et al* conquered the week points of established Extract, Transform and Load tool's architecture and proposed a three layers architecture based on metadata. That built ETL process more bendable, multipurpose and efficient and finally they designed and implemented a new ETL tool for drill data ware house. A systematic review method was projected to identify, extract and analyze the main proposal was branded and compared based on the skin, activities, and document of ETL processes and finished the study by reflecting on the approach being considering and providing an update frame for prospect study[12] every year, an amount of natural disasters hit across the sphere, killing hundreds and causing billions of dollars in property and infrastructure damage. Minimizing the impact of disasters is very important in today's society. As the capabilities of software and hardware evolve, so does the role of information and communication technology in disaster improvement, grounding, answer, and revival. A large measure of disaster-related data is accessible, together with reply plans, records of earlier incident, imitation data, social media data, and websites. Though existing data management solutions offer little or no mixing capabilities. Moreover, the recent advance in cloud compute, big data, and NoSQL opens the door for the new solution in disaster data management. In this paper, a knowledge as a service (KaaS) skeleton is future for disaster cloud data management (Disaster-CDM), with the objectives of

1. Storing large amount of disaster-related data from various source
2. Facilitate search, and
3. Behind their interoperability and addition.

Data are stored in a cloud setting using a mixture of relational and NoSQL databases. The case study accessible in this paper illustrates the use of disaster CDM on a case in point of imitation models [13].

Hristidisr *et al*. surveyed data management and examination in the disaster area. The main focus of their survey was on data therapy techniques lacking the storage facet. In disparity, in disaster-CDM, storage and therapy are considered as vital parts. Hristidis *et al* described in order extraction, information retrieval, information filtering, data mining, and conclusion support. Similarly, disaster-CDM uses an amount of technologies as of information extraction and retrieval. The survey reveals that the greater part of the research has listened carefully on a very narrow area of disaster management, for example, a precise disaster event such as an earthquake or flood, or specific disaster-related behavior such as statement among actors, estimating disaster damage, and use of mobile devices. They also predict the need for elastic and customizable disaster-management solution that could be practical in a different disaster situation. Disaster-CDM aims to supply such solution using cloud and NoSQL approach. [14]

Halevy *et al* has given the design and estimate of a data alert cache framework that require lowest amount change to the original map reduce encoding model for provisioning incremental giving out for Big Data application using the Map Reduce mock-up[15]

Kulkarni *et al* declared the meaning of some of the technologies that switch Big Data Like Hadoop, HDFS, and Map Reduce. The author-optional about a variety of schedules used in Hadoop and About the technical aspect of Hadoop the author also focus on the importance of YARN which overcomes the confines of map reduce[10]

S.Vikram Phaneendra *et al* illustrated that in historic days the data was less and simply Handled by RDBMS but in recent times it is difficult to handle vast data through RDBMS tools, Which is favored as "Big Data "in this they tell that big data differ from other data in five size such as volume, velocity, variety, value, and complexity. They also described the Hadoop structural design consisting of name node, data node, edge node, HDFS to handle big data systems. S.Vikram Phaneendra *et al* also listening carefully on the challenge that needs to be faced to enterprises while handling Big Data: -Data privacy, search analysis, etc [16]

ETL with Hadoop

An ETL workflow with Hadoop process comprises various tasks as follows:

- In the beginning, input data from various data source, which contains data in different formats to HDFS
- Map the consolidated data into a table to make it query table
- The target data is distorted within a finalized format and is mapped to destination source.
- Convert all input data sources information into aim format and make it available at Central.
- Use the finalized data available at Central for reporting analytics. [17]

T.K.Das *et al* have given that Hadoop can be reasonably considered as the evolution of next-generation Data Warehousing systems, with particular regards to the ETL phase

of such systems, And this paper, explanation points of Hadoop-based skill have been mentioned. According to it, In Hadoop file system, once data has been loaded, no alteration can be made on it. It is just like once-write-read- many. It also mentions that Hadoop is not an ETL tool. It supports the ETL environment. Once data has been loaded into HDFS; it is required to write transformation code. But we focus on natural disaster events and that data sets are given below.

Proposed Work

Very big large data value to prepare the new methodology with combine the big data, Hadoop, Map-Reduces used to take the right decision in Disaster Management. It is observed that the Map Reduce framework generates a large amount of in-between data. And such abundant information is thrown away after the task finish because Map reduce is unable to utilize them. Therefore, we propose ditched, a data-aware cache framework for big data applications then its tasks submits their in-between results to the cache manager. The undertaking queries the cache manager before executing the tangible computing work. An original cache explanation scheme and a cache request and reply protocols are designed.

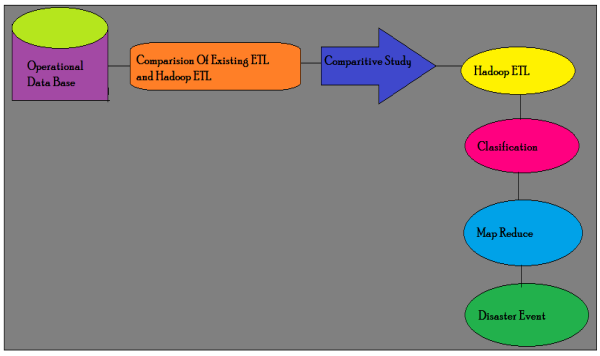

Fig 1 Methodology for proposed work

SVM Classification

Support Vector Machine" (SVM) is a supervised mechanism knowledge algorithm which can be used for in cooperation categorization and regression challenges. It is mostly used in classification problems. We plot each data item as a point in n-dimensional space with the value of each feature being the value of a particular coordinate. It is easy to have a linear hyper-plane between these two classes. We need to add this feature manually to have a hyper plane. SVM has a technique called the kernel trick. These are functions which take low-dimensional input space and transform it into a higher dimensional space i.e. it converts not divisible problem to divisible problem, these functions are called kernels. It is most valuable in the non-linear division problem. Simply locate it does some particularly complex data transformations, and then find out the process to separate the data based on the labels or outputs you've defined.

We access Hadoop ETL with Support Vector Machine methodology in disaster data then evaluate ETL solution which is used to the take right decision. After this process, we tested this result with traditional ETL through some disaster record, Load the data into the databases and do the extraction and preprocessing using this dataset and compare the answer to which the existing ETL and Hadoop ETL responded to the Map Reduce and I would apply to disaster events. We have taken disaster event as our application area to implement the proposed concept.

The United Nations defines a disaster as a grave disruption of the implementation of a community or a society. Disasters engage extensive human, material, monetary or environmental impacts, which exceed the ability of the affected community or society to cope using its own resources. There is no country that is protected from disaster, though susceptibility to disaster varies. There are four main types of Disaster namely Natural Disasters, Environmental emergencies, Complex emergencies, and Pandemic emergencies. We give attention to on the national disaster which includes floods, hurricanes, earthquakes and volcano eruptions that have instant impacts on human health and secondary impacts causing further death and suffering from (for example) floods, landslides, fires, tsunamis. [19]. But we focus on natural disaster events and that data sets are given below.

Data Set for Fatalities is shown in table 1.

Table 1 Disaster Fatalities

YEAR	EVENT	COUNTRY	FATALITIES
2005	Hurricane	USA	16514
2005	Floods	Maharashtra	7201
2005	Earthquake	Kashmir	20615
2008	Earthquake	China	17755
2008	Cyclone	Myanmar	17332
2009	Earthquake	Andaman	15734
2010	Earthquake	Haiti	15660
2011	Earthquake	Sikkim	15065
2012	Earthquake	New delhi	10760
2013	Cyclone	Somalia	19785
2014	Earthquake	Andaman	15617
2014	Flash floods	Congo	18737
2015	Floods	Gujarat	16423
2015	Earthquake	Pakistan	9391
2016	Floods	Assam	21888
2016	Earthquake	North east India	17203

Diagrammatic representation of the above table is shown in figure: 2 Disaster Events

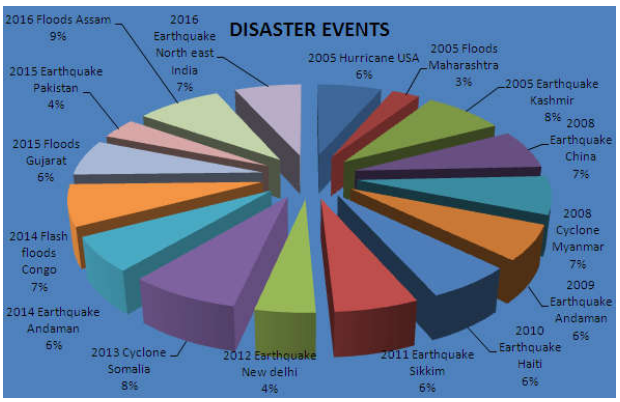

Fig 2 Fatalities Analysis

Data Set for cyclone is shown in table 2.

Table 2 Data Set for Cyclone

Year	Event type	Country	Fatalities
2005	Cyclone	TamilNadu	25000
2008	Cyclone	TamilNadu	24189
2010	Cyclone	TamilNadu	70054
2012	Cyclone	TamilNadu	150200
2013	Cyclone	Somalia	19785
2013	Cyclone	Somalia	19785
2013	Cyclone	Somalia	19785
2013	Cyclone	TamilNadu	18449
2016	cyclone	TamilNadu	15738

Diagrammatic representation of the above table is shown in figure: 3 cyclones

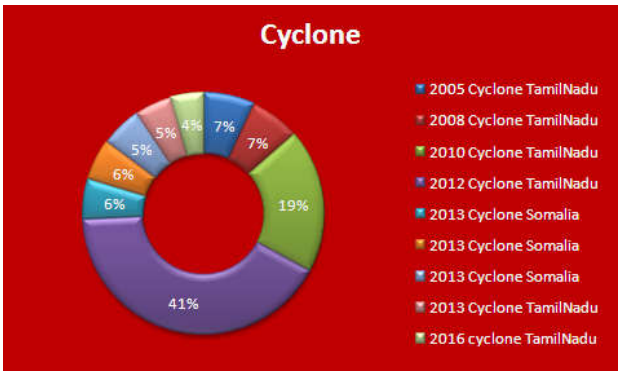

Fig 3 Cyclone Analysis

Data Set for Earthquake is given in table 3.

Table 3 Data Set for Earthquake

Year	Country	Earthquake
2008	China	17755
2008	China	17755
2009	Andaman	15734
2010	Haiti	15660
2010	Haiti	15660
2011	Sikkim	15065
2012	Andaman	19755
2012	New delhi	10760
2014	Andaman	15541
2015	Pakistan	9391
2015	Assam	15617
2015	North east I	24131
2015	Pakistan	9391
2015	Assam	15617
2016	North east I	17203
2016	North east I	17203

Diagrammatic representation of the above table is shown in figure: 4 Earthquake

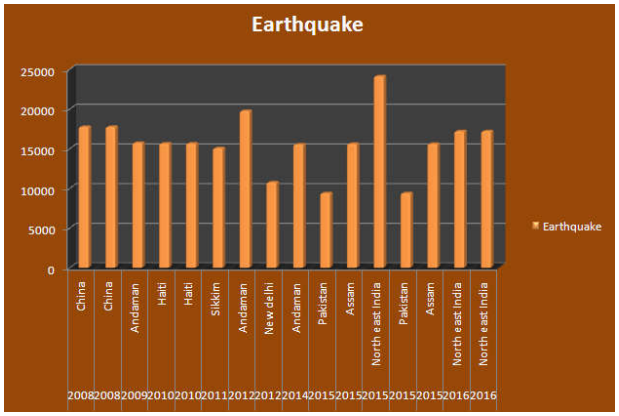

Fig 4 earthquake Analysis

The result of the comparative study between Existing ETL and Hadoop ETL is shown in figure 5.

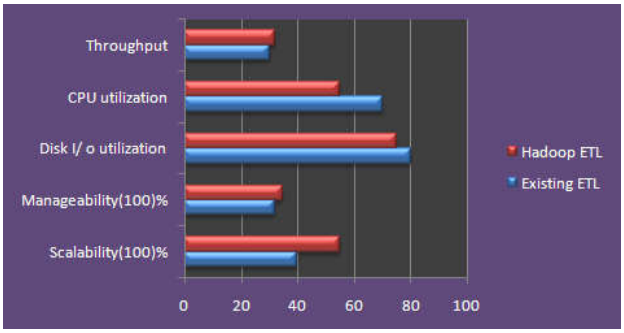

Fig 5 Comparison between Existing ETL and Hadoop ETL

CONCLUSION

I have entered a period of big data. The paper describes the concept of big data along with operational Vs Analytical system of big data. We compared Existing and ETL and Hadoop ETL and also proved that Hadoop ETL is better than an existing ETL and describes the advent of Big Data, Architecture, and characteristics with an integration of Hadoop ETL, Map Reduce and SVM classification in disaster Events for making a better decision in the natural hazard. Our future work focuses on the analysis part of the big data classification by implementing a different data mining techniques in it.

References

1. "Survey of Recent Research Progress and Issues in Big Data" www.cse.wustl.edu/~jain/cse570- 13/ftp/bigdata2 /index.html 1/13
2. Yongqiang He *et al* RCFile: "A Fast and Space-efficient Data Placement Structure in Map Reduce-based Warehouse Systems," ICDE, 2011
3. Varsha B.Bobade" Survey Paper on Big Data and Hadoop" *International Research Journal of Engineering and Technology*
4. Varsha B. Bobad, "*International Research Journal of Engineering and Technology (IRJET)*", Volume: 03 Issue: 01
5. Inmon, William "Data Mart Does Not Equal Data Warehouse". DM Review.com. (2000-07-18).
6. Jeffrey R. Bocarsly, "The Data Warehouse Toolkit." Complex ETL Testing-A Strategic Approach
7. R. Kimball and M. Ross. Wiley Publishing, Inc., 2002.

8. Ms. Vibhavari Chavan, Prof. Rajesh. N. Phursule,-"Survey Paper on Big Data"‖ *International Journal of Computer Science and Information Technologies*, Vol. 5 (6), 2014.
9. Amogh Pramod Kulkarni, Mahesh Khandewal, -"Survey on Hadoop and Introduction to YARN"‖, *International Journal of Emerging Technology and Advanced Engineering* Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 5, May 2014)
10. Tekiner F. and Keane J.A., Systems, Man and Cybernetics (SMC),-"Big Data Framework‖" 2013 IEEE International Conference on 13-16 Oct. 2013, 1494-1499
11. Inmon, William (2000-07-18). "Data Mart Does Not Equal Data Warehouse". DMReview.com
12. Katarina Grolinger, Miriam A.M. Capretz." Knowledge as a Service Framework for Disaster Data Management"
13. V. Hristidis, S. Chen, T. Li, S. Luis, and Y. Deng, "Survey of Data Management and Analysis in Disaster Situations," *Journal of Systems and Software*, vol. 83, no. 10, pp. 1701-1714, 2010.
14. Franklin, M., Halevy, A., Maier, D., 2005. "From databases to data spaces: A new Abstraction for Information Management"
15. S.Vikram Phaneendra & E.Madhusudhan Reddy "Big Data- solutions for RDBMS problems- A survey" In 12th IEEE/IFIP Network Operations & Management Symposium (NOMS 2010) (Osaka, Japan, Apr 19{23 2013).
16. Kuldeep Deshpande, and dr. Bhimappa desai limitations of data warehouse platforms and Assessment of Hadoop as an Alternative, Volume 5, Issue 2, pp. 51-58, IJITMIS, 2014
17. T.K.Das and Arati Mohapatro, A Study on Big Data Integration with Data Warehouse, *International Journal of Computer Trends and Technology* (IJCTT)-volume 9 number 4-Mar 2014
18. http://www.wcpt.org/disaster-management/what-is-disaster-management

How to cite this article:

Saranya M and Prema A.2017, An Improved Learning Based Disaster Event Using Big Data Analytics. *Int J Recent Sci Res.* 8(6), pp. 17703-17707. DOI: http://dx.doi.org/10.24327/ijrsr.2017.0806.0408

****