

# Disaster Management Analysis using Hadoop

Vivek Mittal

Dept. of Computer Science and  
Engineering  
SRM Institute of Science and  
Technology  
Chennai, Tamil Nadu

Vrishank Vajpayee

Dept. of Computer Science and  
Engineering  
SRM Institute of Science and  
Technology  
Chennai, Tamil Nadu

D. Viji

(Asst. Prof. (OG))  
Dept. of Computer Science and  
Engineering  
SRM Institute of Science and  
Technology  
Chennai, Tamil Nadu

**Abstract:-** Of late, different sorts of standard and man-made disasters like shivers, typhoons, pandemics, fear-based oppressors' assaults, twisters, and floods causing epic authentic beating and the disastrous death toll and properties any spot in the world.

The advancement of the calamities incited improvement in their information other than utilizes the computational force for the examination. To manage these giant and complex datasets, gigantic information is utilized. It manages the evaluation which prompts dynamic with different accommodating ways to deal with figure out how to coordinate generous easing off, joint undertakings, gigantic length recuperation, post-fiasco, and Overall catastrophe the heads. It begins from get-together data from online media, sensors, individual's data, and so forth, managing the accumulated data into information, investigating the information, and using it into settling on a choice.

We will present an examination on Hotspot data assessment, will explore the Disaster data in the Hadoop organic framework and will try to improve the zone by controlling the catastrophe slanted and occasion in a locale.

**Keywords:-** Big Data, Emergency Management, Information Centric Disaster Management, Analysis.

## I. INTRODUCTION

The rise in the occurrences of disasters ranging from natural and manmade create a huge impact on the society, with affecting the economy to taking lives and impacting the governance. The information which is passed beyond what many would consider possible and past the dealing with force such information is called Big Data. Huge data proposes immense information; it's an assortment of mammoth datasets that can't be ready utilizing customary enlisting systems.

Monstrous information isn't simply information; rather it's gotten an outright subject, that fuses completely different devices, procedures, and structures. info that's huge in size is termed Enormous Data. constantly we tend to chip away at data of size MB (Wordbook, Excel) or most cutoff GB at any rate data in Petabytes i.e.,  $10^{15}$ -byte size is called Big Data. It's sent that essentially 89% of the current data has been made inside the past six years.

In our paper we discuss about the amalgamation of this technology with disaster management that could help to reduce the impact of the manmade as well as natural disasters.

Moreover, recently there has been a humongous rise in the number of natural disasters occurring around the world which worsens the situation for the government. We will focus on merging the big data and emergency management which has not been fully utilized in the past.

## II. LITERATURE SURVEY

In [1] the author discussed about China which was affected by the string of awful events having different sorts of calamities that have a wide degree, high rehash, ordinary solid irregularity, and present more tangled fiascos. This paper introduced structure portions of the Integrated Disaster Reduction and Quick Service Platform. The results of the seismic shiver by using Integrated Disaster Reduction and Quick Service Platform were in the like way pronounced in this appraisal.

[2] focused on major horrible experiences since the 1990s. Elective risk move approaches are projected to assist the insistence business in covering and insurability issues, among that capital business areas. During this paper, the manufacturers cite the piece of the general public professional in supporting the market-based upheaval risk funding approach. this can be done employing a one-decent, two-country, and two-period-covering age model.

[4] It is an encounter report that follows the checking of a cell phone-based improvement method for the disaster of the centre's space. The framework is rather than growing new advances were made by the application structure that disrupting impact data revealing and status transmission framework dependent on a telephone. These applications are made out of four confines that (1) bafflement uncovering application, (2) catastrophe data amassing and planning organized gifted, (3) unsettling influence enlightening mix, (4) disappointment data the board web point, ace. The advancement system has been perhaps applied for these perspectives making the disaster data exposure and status transmission structure.

[5] The premise of this paper is on contradictory climate varieties and distinctive earth science disasters within the Gansu district, with the institution of natural

modification flooded by warming, the difficulties accomplished by meteorological catastrophes are developing. The enthusiasm for catastrophe hazard qualities and its reaction to regular change must be constrained to be dealt with direly. Considering this, this appraisal reliant upon the data of the debacle situation of major meteorological stations in Gansu area since 1961, made the calamity hazard record expansive misfortune rate, and intentionally slowed down the space-time continuum combination attributes of the dry season, downpour and flood, wind, and hail and low-temperature catastrophes in Gansu zone, and referenced the effect of ecological adjustments.

Paper [7] is mainly on the Japanese National Police Agency report which attested 15,883 passing's, by the 2011 Tohoku seismic quake wave. Dismissing the way that Japan had a strong and consistent tempest rebuking structure, the offbeat size of waves and short takeoff time (the principal wave appearance time was only 25 minutes) brought various difficulties. As such, information system for emergency response has been a noteworthy concern of policymakers to reduce hurts from the wave. By at that point, a phone is basic for fiasco response because of its adaptability and information clearness. Evacuees can get unsurprising disaster information which helps with making a clearing judgment. Especially, in excess of 30 million are using cells in South Korea (The origin of the paper), so it is significant to use wireless as debacle information passing on the device. In this appraisal, a cell (application) is made to decrease the mischief to human life from the storm.

### III. METHODOLOGY

The proposed thought figures out some approach to give an enlightening arrangement by using Hadoop contraption we can segregate no premise of data and fundamentally add the number of machines to the gathering and we get results with less time, high throughput and upkeep cost are less and we are using joins, segments, bucketing structures in Hadoop.

Fig. 1. depicts the architecture diagram for our application where firstly the data is pre-processed by combining the datasets from verified sources such as Kaggle, official governments records publicly available. The data is pruned and processed for the accurate information available to the users, the data is cleansed with the removal of noisy data, data filling with the exact value from the trusted source and removal of the missing value data and column if it couldn't be replaced by the trusted source.

The dataset is then loaded to the SQL database after which it is further loaded into HDFS using SQOOP commands for query analysis. After loading the dataset into HDFS, with the help of HIVE we are able to perform the query analysis on the dataset and store that obtained result into HDFS, which can later be downloaded in the form of .txt file.

We also created a Web Application for our project where our aim is to connect the government, and common people to a common web framework, which gives the precautions, safety measurements and helpline number in case of a disaster. The user dashboard has additional features which offers users to select the country from the data which gives the visual representation about that particular nation, about the disaster, the fatalities it created, the diseases which were spread due to disaster, etc.

### ARCHITECTURE DIAGRAM

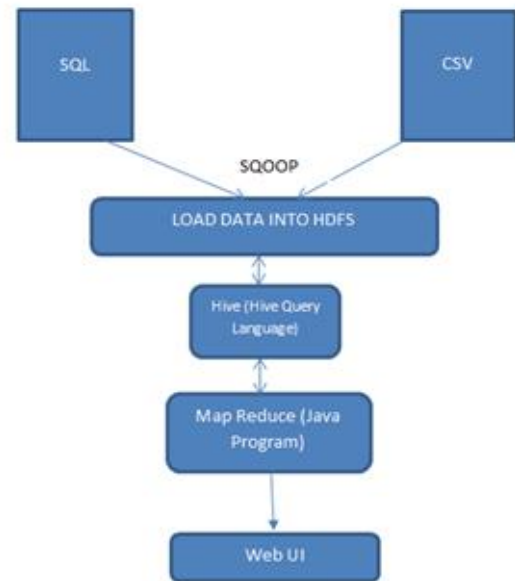


Fig. 1. The figure depicts the overview of our system

#### A. PREPROCESSING DATABASE

In this module, we are collecting the data acquiring to disasters around the world from genuine sources such as the official reported numbers to the government, this covers the calamities all around the world from the 17<sup>th</sup> century A.D. to the 21<sup>st</sup> century of the humankind. The data comprises the havoc created naturally like the seismic, tectonic movements, droughts, floods to the disasters caused due to human error and activities with the likes of the Bhopal Carbide gas tragedy to the most devastating Chernobyl nuclear reactor blasts. Data (with more than 40,000 entries) with different field types in Microsoft Excel was then changed to a comma-separated mapping, which should be a comma-separator (CSV) record, and moved to MySQL support via the database. Prior to this the data is pruned and processed for the accurate information available to the users, the data is cleansed with the removal of noisy data, data filling with the exact value from the trusted source and removal of the missing value data and column if it couldn't be replaced by the trusted source. The cleansed data comprising the spatial location (comprising the latitudes and the longitudes), the number of casualties, the record of the gases and the hazardous gases that were recorded at the site after the havoc, the problems it caused to the people situated in the epicenter of the disaster and the nation which encountered the dreaded situation is passed on the MySQL.

We are getting all the help information that we have dealt with in MySQL and securing each and every one of that information by use of Sqoop orders to HDFS (Hadoop Distributed File System). As of now, all the information is dealt with in HDFS where it is configured to be managed by using Hive.

## B. ANALYZE QUERY

In this module, all data is transferred from HDFS (Hadoop Distributed File System) to HIVE using the Sqoop import command, with Hive ready to be parsed. Here in HIVE, we can process only structured data to analyze by extracting only the meaningful data and neglecting unclenched data we can analyze the data in more effective manner by use of hive. Hive functionalities lie in the part that it provides the necessary SQL abstraction to integrate with the SQL like queries (Hive QL) into the preempted Java without the necessity of the implementation of the low-level Java API. The query data obtained from the Hive becomes the hand which provides which provides the input to the mouth of the MapReduce's mapper stage, where eventually the data will be processed parallelly in the shorter batches for the quick end result.

## C. PROCESSING (MAPREDUCE)

MapReduce algorithm else the paradigm is based on the sending the workstation to where the existence of data resides. It executes the task into 3 subtasks mainly pronounced as (1) Map Stage (2) Shuffle stage (3) Reduce stage. MapReduce will divide try into little parts and cycle each part uninhibitedly by giving out them to various constructions. After the entirety of the parts are dealt with and examined, the yield of every PC is amassed in one single area, and from that point, a yield instructive assortment is ready for the given issue.

## D. WEB INTERFACE

Java is a programming language and a phase. Java is an obvious level, solid, object-arranged, and secure programming language. We'll be using Java for the customer-savvy web UI as the yield. With the use of Java, we're able to display the unarranged data in a form which is easily understood by lame person. It is being used to visually present the analyzed data in a more structured and planned way.

## IV. RESULT AND CONCLUSION

In this paper, we presented an examination on Hotspot data assessment. We took our dataset from Kaggle (Natural Disaster 1900-2021). With the use of Hadoop ecosystem, we were able to analyse the dataset and reduce the analysis time by using the map-reduce algorithm by processing the queries in parallel. The map-reduce algorithm also helped us achieve the scalability. To explore the Disaster data in the Hadoop organic framework and to improve the zone by controlling the catastrophe slanted and occasion in a locale. Hadoop climate is using hive, pig, map-reduce devices for planning whether the yield will put to the side less exertion to gauge and the outcome will be uncommonly brisk. Along these lines in this endeavor, Disaster data which is by and

large going to store in RDBMS going to less execution therefore by using the Hadoop contraction, it will be snappier and gainfully setting up the data.

## FUTURE WORK

Apache Spark is an open-source managing an engine worked around speed, a representation of use, and evaluation. Just in case you have got a vast heap of information that needs low dormancy fixing that a customary Map scale back program can't give, Spark is that the alternative. The radiance provides in-memory cluster computation to lightning-fast speed and supports Java, Scala, and Python API's for simplicity of progress.

## REFERENCES

- [1]. Yan Cui, Suju Li, Liying Wang, MoQuan Sha, Yang Shu., "Disaster event management based on Integrated Disaster Reduction and rapid Service Platform", IEEE International Geoscience and Remote Sensing Symposium (IGARSS) ., 2016
- [2]. Ye, T., Yokomatsu, M., & Okada, N. "Combined international and intergenerational disaster risk diversification: An innovative instrument for government intervention into the private disaster insurance program", IEEE International Conference on Systems, Man and Cybernetics, 2016
- [3]. Y. Yuan, "Advance in the assessment of natural disaster situation", Advanced in Earth Science., vol. 25, no. 1, pp. 22-32, 2019.
- [4]. Seop, S. H., Young, M. G., & Hoon, J. D., "A study on the development of disaster information reporting and status transmission system based on smart phone"., 2017
- [5]. Han, L., & Zhang, Q. "The Characteristics of Agricultural Meteorological Disasters and Responses to Climate Change in Gansu Province", 7th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)., 2018
- [6]. P.J. Shi, T. Ye and J.A. Wang, "Integrated Governance of Natural Disaster Risk", Journal of Beijing Normal University(Social Sciences)., no. 5, pp. 130-136, 2016.
- [7]. Hyoungseong, P., Si-bum, C., Dongseag, K., & Sungjin-hong. "Development of a smartphone application for disaster response"., MTS/IEEE Washington, 2015
- [8]. X.R. Guo, P. Huang and W.Y. Zhang, "3D visualization management system of remote sensing satellite data", Procedia Environmental Sciences., no. 10, pp. 1059-1064, 2018.
- [9]. Xuan Song, Haoran Zhang, Rajendra A. Akerkar, Huawei Huang, Song Guo, Lei Zhong, Yusheng Ji, Andreas lothe Opdhal, Hemant Purohit, Big Data and Emergency Management: Concepts, Methodologies and Application., pp. 1-1, 2020
- [10]. Vagelis Hristidis, Shu-Ching Chen, Tao Li, Steven Luis, Yi Deng, Survey of data management and analysis in disaster situations., no. 5, pp. 178-163, 2018.

- [11]. S. Rao, M. Ali, M.M. Naqvi et al., "Satellite remote sensing(SRS)and geographical information system(GIS) for earthquake damage assessment and disaster management", IEEE Geoscience & Remote Sensing Letters., no. 25, pp. 190-195, 2016.
- [12]. Y.F. Liao, F. Zhao, Z.Q Wang et al., "Spatial Pattern Analysis of Natural Disaster in China from 2000 to 2011", Journal of Catastrophology., vol. 28, no. 4, pp. 55-60, 2018.
- [13]. P.J. Xu and L. Deng, "Application of remote sensing in the field of disaster reduction and response", Remote Sensing Technology and application., vol. 26, no. 4, pp. 512-519, 2017.
- [14]. Y.D. Fan, "Disaster management supported by remote sensing technology", Disaster Reduction in China., vol. 12, no. 2, pp. 46-48, 2015.
- [15]. Y.D. Fan, L. Wang and S.J. Li, "A study on the construction of a spatial information cooperative service mechanism for emergency disaster applications", Journal of Remote Sensing., vol. 15, no. 5, pp. 1053-1064, 2019.