Research Article
Journal of Research and Review in Science, 100-105
Volume 5, December 2018
DOI:10.36108/jrrslasu/8102/50(0131)

ORIGINAL RESEARCH

Managing Human Induced Crisis with Big Data Infrastructure

Afolabi Ojerinde 1[1] and Philip Adewole 2[2]

[1]Graduate Student University of Lagos, Nigeria.

[2]Department of Computer Science, University of Lagos, Nigeria

Correspondence
Afolabi Ojerinde, [1]Graduate Student University of Lagos, Nigeria.
Email: onafolabi@gmail.com
Philip Adewole, [2]Department of Computer Science, University of Lagos, Nigeria
Email: padewole@unilag.edu.ng

Abstract:
Introduction: This paper presents human induced crisis management system using big data infrastructure. This approach was motivated by the already established fact that human induced crisis are characterized by velocity, variety and volume.
Materials and Methods: This paper therefore employed Hadoop big data stack, web technology to design and implement a crisis management model.
Results: The resulting system comprises analytical engine, custom website and a desktop application called “Channel”. The Hadoop distributed file system was used for data storage in the analytical engine, crisis data were collected via Twitter API and web service generated by the project website using Apache Flume and Channel respectively. Apache Hive was used to analyse the collected data and the analysed result were posted back to custom website using Channel. The system was evaluated using Mean Opinion Score (MOS) to test for its applicability, usability and reliability.
Conclusion: The perceived applicability rating of 74%, usability rating of 73% and reliability rating of 57% were obtained. The resulting system provides insight into crisis; promote rapid situational awareness, aid policy formulation and monitoring.

Keywords: Data Analytics, Big Data, and Data Mining

All co-authors agreed to have their names listed as authors.

JRRS
https:sciencejournal.lasucomputerscience.com

1. INTRODUCTION

(Humanity has dwelt on experience and traditional database systems to manage available crisis data in other to sustain peace and live a life devoid of crisis. However, the proliferation of digital computing devices and explosion of social media website has led to data deluge that hinders the ability to store, process and analyse the massive amount of data available in the bid to curtail crisis. [1] reportedly defines crisis as “a condition characterised by surprise, a high risk of serious values and short reaction time”. Crisis could be categorized into either Natural or Human Induced. Human induced crisis is often accompanied by high velocity of data, high variety and easily accumulate into huge volume – which typify it as big data and makes it amenable to big data infrastructure. Crisis management was defined by [2] as “a systematic process with principal goal to minimize the negative impact or consequences of crises and disasters, thus protecting societal infrastructure”. [2] formulates crisis management cycle to revolve around; Prevention, preparedness, response and recovery. Proper crisis management requires deep analytics of each of the phases, not only to extract information pertinent to the subject under consideration, but also to draw new insight and answer correlation questions. The rest of this paper is presented as follows. Section 2 presents background for crisis management and definition of concepts. Section 3 discusses the model design and implementation. Section 4 present the result and system evaluation. Section 5 concludes the paper and indicates areas of future works.

2. MATERIAL AND METHODS

Management of data generated due to crises occurrence can be overwhelming, as in the case of Haiti [3]. In the bid to resettle Haitian and hasten the crisis recovery process crowdsourced data was translated into actionable information with the use of google crowdsource open map, SMS broadcasting are used in the development of humanitarian applications [4]. Corporate organisation and individuals also used social media and blog to raise fund and provide situation awareness to victims and social workers. None of the technology used was robust enough to warehouse varieties of data being generated and provide insight for better planning. Also, there is no platform in developing nations, Nigeria included where historic crisis information or crisis propensity could be accessed to give insights and adequate situational awareness for citizens and authorities. Thus, a need to model a solution that would accommodate varieties of data moving with high velocity and voluminous to provide online situational awareness of crisis and gives insight.

2.1 Big Data and Infrastructure

[5] defined Big-Data as high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization. Generally, Big Data connote advancement in storage, processing and understanding of all form of data. Data growth challenges has three dimensions; volume, velocity and variety [6]. [7] stated that in addition to volume, variety and velocity; storage, curation, searching, visualization and handling of complications generated from uncertainty are also challenges of big data. The peculiar challenges of Big Data informed its popularity and the development of varieties of tools to manage its.

Hadoop, an open source project as major tool for storage and processing of Big Data. Apache Software Foundation [8] describe Hadoop software library as a *“framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models”.* Hadoop as a framework is designed to utilized commodity servers for storage and computation in parallel manner. Hadoop itself have two primary components, the Hadoop distributed filesystem (HDFS) and the MapReduce execution engine. The publication of Google white paper, “**The Google files system” [9]** and the research at Yahoo [10] promoted the development of Hadoop and other tools built around it to form a stable ecosystem.

[11] referred to the Apache Hadoop ecosystem as a Big data stack, it stated that the success of Hadoop has encouraged the building of large number of big data project around it, creating a complete big data project stack in Apache. Yarn, MapReduce, HIVE, and Fume are some of the popular tools in Hadoop ecosystem, reviewed as follows;

i. YARN is a acronym for Yet Another Resource manager. It is incorporated into Hadoop version 2 for a better resource management and utilization. The advent of YARN departs Hadoop from its original monolithic design where there is no separation of resource management function from programming model. YARN delegates many scheduling-related functions to per-job components. The ability of YARN to coordinate intra-application communication and execution flow provides a great deal of flexibility in the choice of programming framework. Apache Mesos [12] also provides similar function like YARN.

ii. MapReduce - is the major data processing framework in Hadoop. [13] describe MapReduce as a programming model for parallel data processing. MapReduce algorithm is often implemented in any of the popular programming languages such as Java, C++, Python, and Ruby. The computation task of Map-Reduce job is shared with aid of HDFS. MapReduce version 2 is an improvement on version 1; it runs on Apache Yarn and used Yarn’s functionalities for the resource allocation and scheduling. The JobTracker and TaskTracker in previous version are replaced with Application master and Node master respectively.

iii. Flume - Apache Software foundation [8] defines Flume as a “distributed a distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of log data”. Flume popularity is informed by its ability to ingest data into Hadoop from multiple source. It uses channel based transaction to ensure secure transmission of data from the source to the sink (Hadoop).

iv. HIVE – [14] defined HIVE as “data warehouse software that facilitates reading, writing and managing large datasets residing in distributed storage using SQL”. Hive has both command line and JDBC driver for user interaction. HIVE leverage on the popularity of SQL as a well-known language. Most users result to the use of HIVE and other high level processing languages like PIG, Shark and MRQL for processing due to their ease of use. They are written once and compiled into MapReduce job provides easy access to data in Hadoop.

3. CRISIS MODEL DESIGN AND IMPLEMENTATION

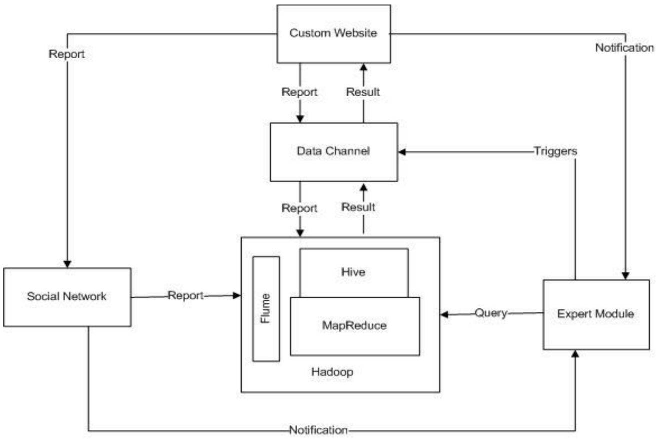


Fig. 1. Crisis Management Model Diagram

Data Channel is the module that pulls data from custom website and also uploads analysis result back to it for user access and APIs consumption. Hadoop is the analysis platform, with tools for storage and processing big data. Crisis data volunteered by users from the data sources are processed inside Hadoop. Flume a component of Hadoop is used to pull data from social media, while Hive is the component that queries the stored data on HDFS. The Expert module is the interface with human expert, the expert initiates ad hoc hive queries (HiveQL) and review analysed result before publication on the custom website.

3.1 System Design

The system activity diagram is shown in Figure 2. The system requires users to authenticate before performing any activity, using conventional registration on the web or their social media credentials. Users can either make a fresh report or review a previous report which is depicted by the first decision tree. Hadoop pulls reports (tweets) and comment (retweets) from the data source periodically and then replicate the data. The Expert module is the Experts interface with Hadoop console, once the expert receive notification for a new set of report, the expert initiate query and then publish the analytics result back to the custom website, the analytic query could be issue on incidents as they are being reported (stream processing) and it could be run on both fresh and historic data (batch processing). The analytics result is then published in a structured form onto the custom website.

3.2 System Implementation

The model is implemented with Apache Biginsight, desktop application, Custom website, and Twitter API. Apache BigInsight is a stack of Hadoop big data tools, which is used to store and analyse the crisis data collected. Desktop application is developed with Electron, an open-source framework. The custom website is built with Jacascript, PHP and Mysql at the backend. Detail implementation is discussed in the following subsection.

a) Hadoop Cluster Implementation with Apache BigInsight

BigInsights integrates popular stack of big data tools in other to optimized access to all the tools available, such as MapReduce platform, spark, yarn, HDFS and others. Yarn is used to manage the resources on the cluster. HDFS is the storage component of Hadoop. Apache flume is use to stream live tweet into Hadoop for real time analytics. Hive query written in Hive-QL (similar to SQL) is use to query the data in HDFS for the crisis analytics.

b) Custom Website Implementation

The custom website is implemented with PHP, JavaScript and MySQL. The analysis result from the Hadoop platform is published on the custom website. JavaScript and HTML is the scripting language used to implement the front end while the backend is implemented with PHP and MySQL database. User can either register or use their social media (twitter, Facebook, Google++) login credential.

c) The Twitter API

Twitter expose an API where application can be developed, the API is used in pulling data into Hadoop. Consumer Key and Secret key generated by the API are used to authenticate requests on twitter platform. Access level and permission are defined during application creation.

d) Channel Implementation

Channel is the system component that consumes the API from the custom website and also uploads analytics result to the custom website. It pulls newly reported incident from the custom website to a directory on the machine running apache Hadoop in a text file format. The file is then passed on pulled into HDFS for perpetual storage. Electron framework, an open-source framework for creating native application with web technologies like Javascript Html and CSS is used to implement Channel.

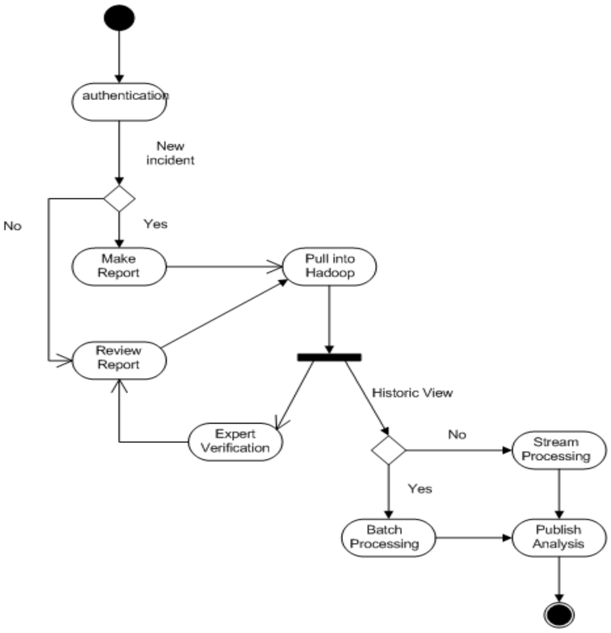
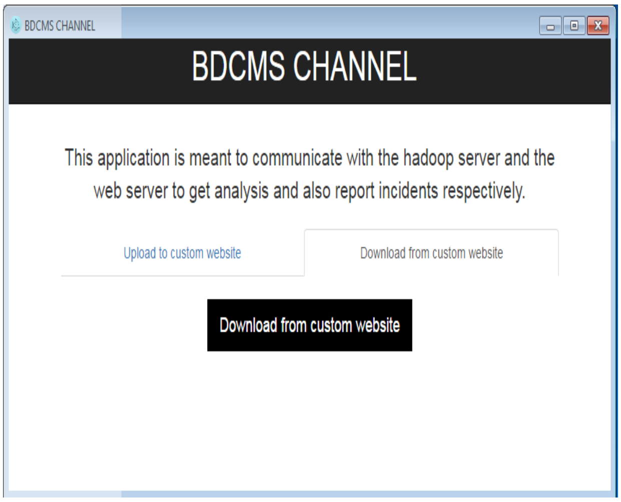


Fig. 2. System Activity Diagram



This application is meant to communicate with the hadoop server and the web server to get analysis and also report incidents respectively.

Fig. 3. Channel for Upload/ Download on custom Website

4. RESULTS AND EVALUATION

The screenshots of some of the interfaces of the crisis management system are presented and the system evaluation is presented using mean opinion score (MOS) approach.

4.1 Results of Implementation

The desktop application called "Channel" is shown in Figure 3; it is used for download of report and upload of analysed report onto the custom website. An analysed incident is displayed on the custom website report page as shown in Figure 4. The information dsipaly on the report page are:

i. Report – displayed the detail incident reported.

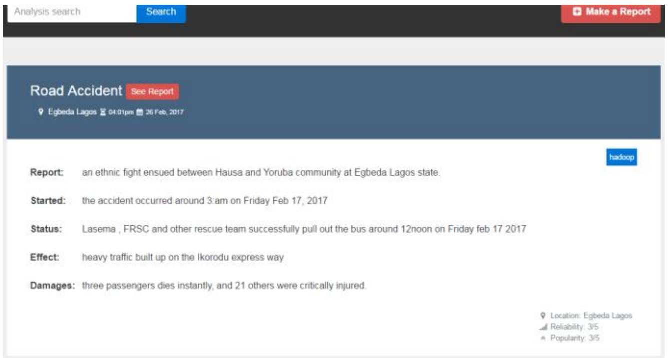
ii. Started - is the time the incident begins.

iii. Status - display information such as whether the incident is about to happen, is ongoing or is over.

iv. Effect - display the impact of the incident at the immediate environment.

v. Damages - display information available about the casuality recorded at the Location of the reported incident.

vi. Location, Reliability and Polpularity is also dispalyed at the bottom right of the analysis screen. The "location" is the place where the incident occurred. The "reliability" is measure of how reliability the analysis is. Reliability is measured on the scale of 5, that is based on the number users that retweet or quote the reported incident. The "popularity" is the measure of the number of user that like or replied to a reopted incident, a single user like/replied is scored 1, two users is scored 2. To have a score of 5 on either reliabilty or popularity at least 5 users must comment/retweet or like as the case might be.



Road Accident

Egbeda Lagos 04:01pm 26 Feb, 2017

Report: an ethnic fight ensued between Hausa and Yoruba community at Egbeda Lagos state.

Started: the accident occurred around 3:am on Friday Feb 17, 2017

Status: Lasema , FRSC and other rescue team successfully pull out the bus around 12noon on Friday feb 17 2017

Effect: heavy traffic built up on the Ikorodu express way

Damages: three passengers dies instantly, and 21 others were critically injured.

Location: Egbeda Lagos
Reliability: 3/5
Popularity: 3/5

Fig. 4. Incident Analysis Report

4.2 System Evaluation

The The implemented model is evaluated using Mean Opinion Score (MOS), a numerical method of evaluation. MOS is use to test for the reliability, usability and applicability of the implemented model. Twenty users are selected users (Ten with computing knowledge and Ten without knowledge of computing) evaluate the implemented system and grade the system on the 5-point MOS scale (excellent, good, fair, poor, bad). The implemented model achieved a perceptual rating of 74% on applicability, 73% on usability, and 57% on reliability. The MOS range and score for each of the tested feature is display in Table 1. Column one is the feature tested, column two to six is the grade point for instance one person grade the system as excellent (5-point) on usability, eleven person graded the good(4-point) on usability while no one graded the system as poor (2-point) on usability. Column seven is the sum of users that evaluated the system.

Table 1. The MOS range and score for each of the tested features

Feature	Excellent 5	Good 4	Fair 3	Poor 2	Bad 1	Sum of Tester	Sum of Point	Average	MOS %
Usability	1	11	8	-	-	20	73	3.65	73
Reliability	-	4	9	7	-	20	57	2.85	54
Applicability	3	7	9	2	1	20	74	3.7	74

Column eight depicted the sum of grade point for each criteria, the total point for Usability, reliability and applicability are 73, 57 and 74 respectively. The average score (column nine) is the 'sum of point' divided by 'number of tester'. The mean opinion score (MOS) is calculated as the percentage of average score (column nine) divided by grade point (5-point likert scale).

5. CONCLUSION

A model that captured the process and components of a crisis management system has been designed. The model is implemented using big data tools and approach, such that it would be scalable, robooust and be able to be able process large set of unstructured data quickly. The implemented system offers quick response in times of a cisis, promotes situational awareness, useful for crises management and policy development to forstall or minimse the consequencies of crisis incident within the monitored polpulation.

The system is applicable in area of security managrmrnt, governace and serves as citizen guide. The security operatives can explore the system tract crisis prevalance per location, and and use to monitered the effectiveness of security impact in a minitored population. The government can use the system in making policy, for instance if crisis often reported based on the activities of motorcycle riders, a policy could be make to ban or curb their activities. Citizen could also utilze the system in getting both historical and comtemporaray information on crisis incident.The field of computation can also advance on the project to develop analytical tools specific for crisis management.

A further research should done on area of data collection such that data would be collected through a uniform platform. The process of analysis should be improve on to increase relaibilty of analysis report. The system should enforce capturing of user location and identity to avoid mischievous reporting.Work should be done in area of sensitive information management, such that the system would not be exposing sensitive information to wrong people.Also futher work should be done in the area of processing voice and video in crisis situation so that corresponding media would analysed along text

REFERENCES

1 Pu, C., & Kitsuregawa, M. (2013). Big data and disaster management: a report from the JST/NSF joint workshop. Georgia Institute of Technology, CERCS.

2 Emmanouil, D., and Nikolaos, D.(2015). " Big data analytics in prevention, preparedness, response and recovery in crisis and disaster management". In The 18th International Conference on Circuits, Systems, Communications and Computers (CSCC 2015), Recent Advances in Computer Engineering Series (Vol. 32, pp. 476-482).

3 Nelson, A, Sigal I, and Zambrano D (2010). Media, Information systems and communities: lessons from Haiti. John S. and James L. Knight Foundation.

4 Reiersgord, B. (2011). Technology and disaster: The case of Haiti and the rise of text message relief donations. Case Specific Briefing Paper. Humanitarian Assistance in Complex Emergencies. University of Denver.

5 Genovese Y., Prentice S., (2011), "Pattern-Based Strategy: Getting Value from Big Data. Gartner Group press Release". Retrieved from: http://www.gartner.com/newsroom/id/1731916.

6 Laney D. (2001), "3D Data Management: Controlling Data Volume, Velocity and Variety". Application Delivery Strategies, Vol 949.

7 Schroeck M., Shockley R., Smart J., Romero-Morales D., and Tufano P. (2012), “Analytics: the real-world use of big data: how innovative enterprises extract value from uncertain data”, Executive Report, IBM Institute for Business Value and Said Business School at the University of Oxford.

8 Apache Software Foundation (2016). Welcome to Apache Flume. Available: https://flume.apache.org/Retreived February 8, 2017.

9 Ghemawat Sanjay, GobioffHoward, and Leung Shun-Tak (2003), “The Google File System”, ACM SIGOPS Operating Systems Review.

10 Shvachko K., Kuang H., Radia S., Chansler R., (2010), “The Hadoop Distributed File System”, 2010 IEEE 26th Symposium on Mass Storage System and Technologies, MSST2010.

11 Kamburugamuve, S., Fox, G., Leake, D. and Qiu, J., (2013). Survey of Apache Big Data Stack (Doctoral dissertation, Ph.D. Qualifying Exam, Dept. Inf. Comput., Indiana Univ., Bloomington, IN).

12 Hindman, B., Konwinski, A., Zaharia, M., Ghodsi, A., Joseph, A.D., Katz, R.H., Shenker, S. and Stoica, I., (2011). Mesos: A Platform for Fine-Grained Resource Sharing in the Data Center. In NSDI (Vol. 11, No. 2011, pp. 22-22).

13 White, T., 2012.Hadoop: The definitive guide. “O’Reilly Media, Inc."

14 Thusoo, A., Sarma, J.S., Jain, N., Shao, Z., Chakka, P., Zhang, N., Antony, S., Liu, H. and Murthy, R., (2010). Hive-a petabyte scale data warehouse using hadoop. In Data Engineering (ICDE), 2010 IEEE 26th International Conference on (pp. 996-1005). IEEE.