Big data analytics in prevention, preparedness, response and recovery in crisis and disaster management

Dontas Emmanouil [1], Doukas Nikolaos[2]
[1]Hellenic Army, Artillery School, Nea Peramos, Greece
[2]Hellenic Army Academy, Vari, Greece

***Abstract*—** The scientific area of crisis management has been in the center of attention for multiple disciplines especially the computer science. In the information centered and computer driven world, a major aim of computer scientists is to manage and analyze Big Data, extract information from heterogeneous sources and store it in unified structure formats that allow further processing. In this paper, Big Data analytics techniques and tools that are useful in all phases of crisis management are presented. Furthermore, a system-engineering approach of a big data management system will be analyzed that comprises of four phases; data generation, data acquisition, data storage, and data analytics. Benefits of the usage of Big Data for crisis management are analyzed. An innovative view of open problems concerning Big Data in crisis management is introduced.

***Keywords*—**Big data analytics, crisis and disaster management.

I. Introduction

Effective management of crises and disasters, is a global challenge. All communities are vulnerable to crisis, both natural and induced by human activities. A systematic process with principal goal to minimize the negative impact or consequences of crises and disasters, thus protecting societal infrastructure, is called effective crisis and disaster management. It is imperative throughout the world to increase knowledge of crisis and disaster management, for the purpose improving responsiveness. All the above aims may be facilitated by Big Data Analysis.

Big Data and Computer Science

The concept of Big Data project is fundamentally related to computer science since the beginning of computing. The term Big Data describes amounts of data obtained with technological means that are normally unusable by humans due to volume and which with appropriate automated processing will extract actionable information. [1]

Big Data Characteristics

Big Data may be characterized as having four dimensions: Data Volume, measuring the amount of data available, with typical data sets occupying many terabytes. Data velocity is a measure of the rate of data creation, streaming and aggregation. Data variety is a measure of the richness of data representation – text, images, videos etc. Data value, measures the usefulness of data in making decisions. [2]. A further characteristic has recently appeared, namely Variability, which represents the number of changes in the structure of the data their interpretation. Gartner [3] summarizes this in the definition of Big Data as high volume, velocity and variety information assets that demand cost effective processing.

Big Data in Crisis Management- Surveillance

The management of large volumes of data is perhaps one of the biggest challenges to be addressed by computer science. The wide variety of data acquisition sources available in times of crisis creates a need for data integration, aggregation and visualization. Such techniques assist crisis management officials to optimize the decision-making procedure. During the outburst of a crisis, the authorities responsible must quickly make decisions. The quality of these decisions depends on the quality of the information available. A key factor in crisis response is situational awareness. An appropriate, accurate assessment of the situation can empower decision-makers during a crisis to make convenient decisions, take suitable actions for the most affective crisis management [4]. Situational awareness definitions: "*perception*, where elements of the current situation are observed, *comprehension* where information obtained through observation is combined and interpreted and, *projection* where sufficient information and understanding exists to make predictions about impending events" [5]

II. A Big Data Chain

A. *Big Data systems-engineering approach*

A systems-engineering approach of a big data management system operates in four phases: data generation, data acquisition, data storage, and data analytics [6]. A big-data system is complex, providing functions to deal with different phases in the digital data life cycle, ranging from its birth to its destruction. At the same time, the system usually involves multiple distinct phases for different applications. [7]. Raw data can be taken as the raw materials with data generation and data acquisition being the corresponding exploitation process. In the same sense, data storage may be considered as a buffering process and data analysis as the final production process that utilizes the raw material to create new value [8]. The first stage leading to analysis is Data generation. The rate of data generation is increasing due to technological advancements. Indeed, IBM estimated that 90% of the data in the world today has been created in the past two years [9]. The cause of the data explosion has been much debated. A related example is the huge amount of internet data being generated, such as internet forum posts, social media, chatting records. This huge amount of data may be unusable, but via suitable analyses may yield useful information concerning the habits and hobbies of users. Analyzing this information may render possible to predict behaviors, feelings and trends.

The data generation process is also subject to study, as it comprises of both controlled and unpredictable components. A set of sensors deployed in order to observe a particular situation, is a controllable source of high volume data. On the other hand, there exist in the internet large numbers of users, each one bestirring themselves independently and generating independent data traces. These data traces, when viewed as a total may provide information with serious implications for the economy, the defense and other topics of interest. Hence, the term big data is designated to mean large, diverse, and complex datasets that are generated from diverse data sources, both physically and virtually distributed, that include sensors, video, click streams and many other sources. [10].

B. Data acquisition

Data acquisition refers to the process of obtaining information and is subdivided into data collection, data transmission, and data pre-processing. One of the aims of the data acquisition phase is to aggregate information in a digital form for further storage and analysis. Firstly because data may come from a diverse set of sources, such as websites that host formatted text, images and videos, etc. Data collection refers to dedicated technologies that acquire raw data from specific data production environments. Subsequently, after collecting raw data, high-speed transmission mechanisms are needed, to transfer the data into the proper storage sustaining system for various types of analytical applications. Finally, collected datasets might contain many meaningless data, which unnecessarily increase the amount of storage space required and adversely affect the consequent data analysis [11]. For example, high redundancy is very common among datasets collected by sensors for environment monitoring. Data compression technology can be applied to reduce the redundancy. Therefore, data pre-processing operations are indispensable to ensure efficient data storage and exploitation [12].

Special data collection techniques are utilized in order to acquire raw data from specific data generation environments. This statement refers to the process of retrieving raw data from real-world objects. The process needs to be well designed [13]. Otherwise, inaccurate data collection would impact the subsequent data analysis procedure and ultimately lead to invalid results. At the same time, data collection methods not only depend on the physical characteristics of the data sources, but also on the objectives of data analysis. As a result, there are many kinds of data collection methods. In the following sections, three common methods for big data collection will be explained, while some related methods will be outlined [14].

Data Collection methods

1. **Log files**: As one widely used data collection method, log files are record files automatically generated by the data source system, so as to record activities in designated file formats for subsequent analysis. Log files are typically used in nearly all digital devices. For example, web servers record in log files number of clicks, click rates, visits, and other property records of web users [15]. To capture activities of users at the web sites, web servers mainly include the following three log file formats: public log file format (NCSA), expanded log format (W3C), and IIS log format (Microsoft). All the three types of log files are in the ASCII text format. Databases other than text files may sometimes be used to store log information to improve the query efficiency of the massive log store [16, 17]. There are also some other log files based on data collection, including stock indicators in financial applications and determination of operating states in network monitoring and traffic management.
2. **Web Crawlers**: A crawler [18] is a program that downloads and stores webpages for a search engine. Roughly, a crawler starts with an initial set of URLs to visit in a queue. All the URLs to be retrieved are kept and prioritized. From this queue, the crawler gets a URL that has a certain priority, downloads the page, identifies all the URLs in the downloaded page, and adds the new URLs to the queue. This process is repeated until the crawler decides to stop. Web crawlers are general data collection applications for website-based applications, such as web search engines and web caches. The crawling process is determined by several policies, including the selection policy, re-visit policy, politeness policy, and parallelization policy [19]. Traditional web application crawling is a well-researched field with multiple efficient solutions. With the emergence of richer and more advanced web applications, some crawling strategies [20] have been proposed to crawl rich Internet applications. Currently, there are plenty of general-purpose crawlers available as enumerated in the list [21].
3. **Other methods**: In addition to the methods discussed above, there are many data collection methods or systems that pertain to specific domain applications. For example, in certain government sectors, human biometrics [22], such as fingerprints and signatures, are captured and stored for identity authentication and to track criminals.

Data Collection Tools

The role of technology could easily be integrated into various subtopics on crisis and disaster management. The advantages in sensing, networking and communication produce improvements in crisis management from both the research and practice perspectives. Technological advances are necessary to promote the effectiveness of crisis management systems. Reference must be made to the role Geographical Information Systems (GIS), the Global Positioning System (GPS) and Remote Sensing Technologies have in the context of data acquisition [23].

Geographical Information Systems are informative systems capable of storing, analyzing, sharing, and displaying geographically referenced information data. With the usage of GIS crisis administrators are in position to collect spatial information over a wide geographic area, to analyze and collect up to date information. In addition, given the information from GIS can be easily tabulated, providing a pictorial overview of what happening in area

ISBN: 978-1-61804-320-7

was hit by the crisis. GIS applications can be useful in the following activities:

- To promote situational awareness. Situational awareness is a prerequisite in any
- To create hazard inventory maps. At this level GIS can be used for the pre-feasibility study of developmental projects, at all inter-municipal or district level.
- Locate critical facilities. The GIS system is quite useful in providing information on the physical location of shelters, drains and other physical facilities. The use of GIS for disaster management is intended for planners in the early phase of regional development projects or large engineering projects.
- Create and manage associated databases. The use of GIS at this level is intended for planners to formulate projects at feasibility levels, but it is also used to generate hazard and risk maps for existing settlements and cities, and in the planning of disaster preparedness and disaster relief activities.
- Vulnerability assessment. GIS can provide useful information to boost disaster awareness with government and the public, so that (on a national level) decisions can be taken to establish or expand disaster management organizations. At such a general level, the objective is to give an inventory of disasters and simultaneously identify “high-risk” or vulnerable areas within the country.

GIS technology can provide the user with accurate information on the exact location of an emergency situation. This would prove useful as less time is spent trying to determine where the trouble areas are. Ideally, GIS technology can help to provide quick response to an affected area once issues are known. Mapping and geo-spatial data will provide a comprehensive display on the level of damage or disruption that was sustained as a result of the emergency. GIS can provide a synopsis of what has been damaged, where, and the number of persons or institutions that were affected. This kind of information is quite useful to the recovery process. [24]. An indispensable tool provided by GIS technologies is the GRP, that facilitates real time tracking of the accurate position of perties of interest. By the use of suitable hardware, GPS can be used for a variety of activities from navigation to observing volcanic activity [25].

Remote Sensing

Remote sensing refers to sensors that are attached to aircrafts or satellites. Robotic vision systems the use of remote sensing shows the following features: Data acquisition far away from the emergency area, regular renewal of the data and also provides big image data of very large areas. [26] Sensors also are used commonly to measure a physical quantity and convert it into a readable digital signal for processing (and possibly storing). Sensor types include acoustic, sound, vibration, automotive, chemical, electric current, weather, pressure, thermal, and proximity. Through wired or wireless networks, this information can be transferred to a data collection point. Wired sensor networks leverage wired networks to connect a collection of sensors and transmit the collected information. This scenario is suitable for applications in which sensors can easily be deployed and managed. For example, many video surveillance systems in industry are currently built using a single Ethernet unshielded twisted pair per digital camera wired to a central location. [27]

Social Media

Big Data analytics provides a great opportunity to reveal many sources of data. Exploring social media represents a significant challenge for big data analytics in crisis and disaster management. Research has emerged that deals with monitoring the trends of social media like Facebook, twitter, etc, before or during times of crisis. Thus, social media represent another big data source of interest. [28]

C. Data storage

The explosive growth of data imposes strict requirements on storage and management. Big data storage refers to the storage and management of large-scale datasets while achieving speed, reliability and availability of data access. It is necessary to review important issues including massive storage systems, distributed storage systems, and big data storage mechanisms. On one hand, the storage infrastructure needs to provide information storage service with reliable storage space; on the other hand, it must provide a powerful access interface for query and analysis of a large amount of data.[29] The data storage subsystem in a big data platform organizes the collected information in a convenient format for analysis and value extraction. For this purpose, the data storage subsystem should provide two sets of features:

1. The storage infrastructure must accommodate information persistently and reliably.
2. The data storage subsystem must provide a scalable access interface to query and analyze a vast quantity of data.

This functional decomposition shows that the data storage subsystem can be divided into hardware infrastructure and data management tools. Hardware infrastructure is responsible for physically storing the collected information. The storage infrastructure can be understood from different perspectives. Typical storage technologies include RAM and cache memory, hard disk drives and disk arrays.

Storage infrastructure can be classified from a networking architecture perspective [30]. In this category, the storage subsystem can be organized in different ways, including, but not limited to the following.

Direct Attached Storage (DAS): DAS is a storage system that consists of a collection of data storage devices. These devices are connected directly to a computer through a host bus adapter (HBS) with no storage network between them and the computer. DAS is a simple storage extension to an existing server.

ISBN: 978-1-61804-320-7

Storage Area Network (SAN): SANs are dedicated networks that provide block-level storage to a group of computers.
SANs can consolidate several storage devices, such as disks and disk arrays, and make them accessible to computers such that the storage devices appear to be locally attached devices.[31]
Network Attached Storage (NAS): NAS is file-level storage that contains many hard drives arranged into logical, redundant storage containers. Compared with SAN, NAS provides both storage and a _le system, and can be considered as a file server, whereas SAN is volume management utilities, through which a computer can acquire disk storage space.

Crisis management data storage tools

Storage mechanisms for big data may be classified into three bottom-up levels: file systems, databases, and programming models. File systems are the foundation of the applications at upper levels. Google's GFS is an expandable distributed file system to support large-scale, distributed, data-intensive applications [32]. GFS uses cheap commodity servers to achieve fault-tolerance and provides customers with high performance services. GFS supports large-scale file applications with more frequent reading than writing. However, GFS also has some limitations, such as a single point of failure and poor performances for small files. Such limitations have been overcome by Colossus [33], the successor of GFS. In addition, other companies and researchers also have their solutions to meet the different demands for storage of big data. For example, HDFS and Kosmosfs are derivatives of open source codes of GFS. Microsoft developed Cosmos [34] to support its search and advertisement business .Facebook utilizes Haystack [35] to store the large amount of small-sized photos. Taobao also developed TFS and FastDFS. In conclusion, distributed file systems have been relatively mature after years of development and business operation. Some of the available tools to facilitate big data storage are:

1. *Google BigTable*: a distributed, structured data storage system, which is designed to process the large-scale (PB class) data among thousands commercial servers [36]. The basic data structure of BigTable is a multi-dimension sequenced mapping with sparse, distributed, and persistent storage. Indexes of mapping are row key, column key, and timestamps, and every value in mapping is an unanalyzed byte array. BigTable is based on many fundamental components of Google, including GFS [37], cluster management system, SSTable file format, and Chubby [38]. GFS is used to store data and log files.

2. *Cassandra:* a distributed storage system to manage the huge amount of structured data distributed among multiple commercial servers [39]. The system was developed by Facebook and became an open source tool in 2008. It adopts the ideas and concepts of both Amazon Dynamo and Google BigTable, especially integrating the distributed system technology of Dynamo with the BigTable data model. Tables in Cassandra are in the form of distributed four-dimensional structured mapping, where the four dimensions including row, column, column family, and super column. The partition and copy mechanisms of Cassandra are very similar to those of Dynamo, so as to achieve consistency.[40]

3. *Hadoop:* a top level Apache project that started in 2006. Hadoop can process extremely large volume of data with different structures. Is used commonly in industrial applications, analyzes big data with specific functions such as spam filtering, network click stream analysis and social recommendations. [41, 42]. In fact, Hadoop has long been the mainstay of the big data movement, Instead of relying on expensive, proprietary hardware to store and process data, Hadoop enables distributed processing of large amounts of data on large clusters of commodity servers. Hadoop offers scalability, cost efficiency, flexibility and fault tolerance. Hadoop can recover the data and computation failures caused by node breakdown or network congestion. The Apache Hadoop software library is a massive computing framework consisting of several modules, including HDFS, Hadoop MapReduce, HBase, and Chukwa. [43]

4. *MapReduce*: a software framework for easily writing applications which process vast amounts of data (multi-terabyte data-sets) in-parallel on large clusters (thousands of nodes) of commodity hardware in a reliable, fault-tolerant manner. The computational model consists of two user defined functions, called Map and Reduce. The framework takes care of scheduling tasks, monitoring them and re-executes the failed tasks. [44] The concise MapReduce framework only provides two opaque functions, without some of the most common operations (e.g. Projection and filtering). [45]

5. *Dryad*: a general-purpose distributed execution engine for processing parallel applications of coarse-grained data. The operational structure of Dryad is a directed acyclic graph, in which vertices represent programs and edges represent data channels. Dryad executes operations on the vertices in clusters and transmits data via data channels, including documents, TCP connections, and shared-memory FIFO. All kinds of data are directly transmitted between vertexes [46]. In addition, Dryad allows vertexes to use any amount of input and output data, while MapReduce supports only one input and output set. DryadLINQ [47] is the advanced language of Dryad and is used to integrate the aforementioned SQL-like language execution environment

6. NOSQL databases (non – relational databases)

With the development of the Internet and cloud computing, there need databases to be able to store and process big data effectively, demand for high-performance when reading and writing, so the traditional relational database is facing many new challenges.[48] Various database systems are developed to handle datasets at different scales and support various applications. Traditional relational databases cannot meet the challenges on categories and scales brought about by big data. NoSQL databases (i.e., non traditional relational databases) are becoming more popular for big data storage. [49] Especially in large scale and high-concurrency applications, such as search engines and SNS, using the relational database to store and query dynamic user data has appeared to be inadequate. [50]

D. Data Analysis

The last and most important stage of the big data value chain is data analysis, the goal of which is to extract useful values, suggest conclusions and/or support decision-making. Firstly, the purpose and classification metric of data analytics will be discussed. Subsequently, the application evolution for various data sources and summarize the six most relevant areas will be reviewed. Finally, several common methods that play fundamental roles in data analytics will be intruduced. Data analytics addresses information obtained through observation, measurement, or experiments about a phenomenon of interest. The aim of data analytics is to extract as much information as possible that is pertinent to the subject under consideration. The nature of the subject and the purpose may vary greatly. Some example aims include:

- To extrapolate and interpret the data and determine how to use it,
- To check whether the data are legitimate,
- To give advice and assist decision-making,
- To diagnose and infer reasons for fault, and
- To predict what will occur in the future

In [53] data analytics are classified into three levels according to the depth of analysis: descriptive analytics, predictive analytics, and prescriptive analytics.

Descriptive Analytics: exploits historical data to describe what occurred. For instance, a regression may be used to find simple trends in the datasets, visualization presents data in a meaningful fashion, and data modeling is used to collect, store and cut the data in an efficient way. Descriptive analytics is typically associated with business intelligence or visibility systems.

Predictive Analytics: focuses on predicting future probabilities and trends. For example, predictive modeling uses statistical techniques such as linear and logistic regression to understand trends and predict future outcomes, and data mining extracts patterns to provide insight and forecasts.

Prescriptive Analytics: addresses decision making and efficiency. For example, simulation is used to analyze complex systems to gain insight into system behavior and identify issues and optimization techniques are used to find optimal solutions under given constraints. [54]

III. BIG DATA IN CRISIS PHASES

Crisis

Professor C. Hermann in his article in Administrative Science magazine in June 1963 [55] states that “the crisis is a condition characterized by surprise, a high risk of serious values and short reaction time”. The four phases of crisis are: Prevention, Preparedness, Response and Recovery. These formulate the crisis cycle. There are many interesting approaches about the usage of Big Data in crisis management.

Big Data and Crisis Prevention

Information derived from the analysis of Big Data can help to anticipate crises or at least reduce the risks that would arise from a disaster the major crisis effect. One example is in a big earthquake harm arises in telecommunication networks leading to interruption of communications, also has been observed a large number of blackouts. There exists a need to study this data for optimization of the civil infrastructure to avoid this crisis effects. [56]

Big Data and Crisis Preparedness

Big Data analysis can help significantly to the preparation of crisis management. Through the data analysis can be done recognizing the dangers and to provide a sound strategic approach by the respective managers of the crisis. Big Data analysis can also guide the proactive deployment of resources to fully cope with an impeding type of disaster [57]

Big Data and Crisis Response

Big Data analysis in real time can identify which areas need the most urgent attention from the crisis administrators. With the use of the GIS and GPS systems, Big Data analysis can assist the right guidance to the public to avoid or move away from the hazardous situation. Furthermore analysis from prior crisis could help identify the most effective strategy for responding to future disasters. [58]

Big Data and Crisis Recovery

When the recovery activation will gradually start, the infrastructure would provide a big data source. The Big Data analysis sharing useful information for recovery procedures about volunteer coordination and logistics during the crisis. [59]

IV. CONCLUSION

In this paper, the usefulness of the analysis of Big Data management in crises and disasters was presented. A brief analysis of the collection data sources during the crisis, the technological means and the tool storage and processing of Big Data. The challenges arising from the review concerns the important research fields of the Social media data usage in crisis management. In this context, a system-engineering approach of a big data management system into four phases, data generation, data acquisition, data storage, and data analytics was also outlined. The era of big data is upon us, bringing with it an urgent need for advanced data acquisition, management, and analysis mechanisms. In the big data acquisition phase, typical data collection technologies were investigated during each stage of the data

ISBN: 978-1-61804-320-7

life cycle the management of big data is the most demanding issue. Many challenges in the big data system need further research attention. Big data research remains in its embryonic period. Research on typical big data applications, is required that can improve the efficiency of government sectors, and promote the development of human science and technology, while it is also required to accelerate big data progress. Furthermore there are interesting challenges in data mining in crisis and disasters management. Algorithms need to be developed for completing tasks such as pattern mining for discovering interesting associations and correlations, clustering and trend analysis, to understand the nontrivial changes and trends, and classification to prevent future reoccurrences of undesirable phenomena. Finally several security challenges in storage and transmission of data need to be under constant investigation, in order to address newly emerging threats.

REFERENCES

[1] S. Kailser, F. Armour, J. A. Espinosa and W. Money, "Big Data: Issues and Challenges Moving Forward," 46th Int. Conf. System Sciences, pp. 995,

[2] S. Kailser, F. Armour, J. A. Espinosa and W. Money, "Big Data: Issues and Challenges Moving Forward," 46th Int. Conf. System Sciences, pp. 996-997,

[3] S. Kailser, F. Armour, J. A. Espinosa and W. Money, "Big Data: Issues and Challenges Moving Forward," 46th Int. Conf. System Sciences, pp. 996-997,

[4] S. Mehrotra, X. Qiu, Z. Cao, and A. Tate, "Technological Challenges in Emergency Response","(Periodical style)," IEEE, pp. 6 July/August 2013 https://www.computer.org/intelligent.

[5] S. Mehrotra, X. Qiu, Z. Cao, and A. Tate, "Technological Challenges in Emergency Response","(Periodical style)," IEEE, pp. 6 July/August 2013 https://www.computer.org/intelligent

[6] F. Gallagher. (2013). The Big Data Value Chain [Online]. Available: http://fraysen.blogspot.sg/2012/06/big-data-value-chain.html

[7] E. B. S. D. D. Agrawal et al., "Challenges and opportunities with big Data" A community white paper developed by leading researchers across the united states," The Computing Research Association, CRA White Paper, Feb. 2012.

[8] D. Fisher, R. DeLine, M. Czerwinski, and S. Drucker, "Interactions with big data analytics," Interactions, vol. 19, no. 3, pp. 50 59, May 2012.

[9] What is Big Data, IBM, New York, NY, USA [Online]. Available: http://www-01.ibm.com/software/data/bigdata/ 2013

[10] J. Manyika et al., Big data: The Next Frontier for Innovation, Competition, and Productivity. San Francisco, CA, USA: McKinsey Global Institute, 2011, pp. 1_137.

[11] H. Hu, Y. Wen, T. Chua and X. Li, "Toward Scalable System for Big Data Analytics: a Technology Tutorial," (Periodical style)," IEEE pp 8 July 2014 657-659.

[12] M. Chen, S. Mao, and Y. Liu , " Big data : a survey", Mobile Networks and Applications, vol.19, no.2, pp.181-183, 2014

[13] H. Hu, Y. Wen, T. Chua and X. Li, "Toward Scalable System for Big Data Analytics: a Technology Tutorial," (Periodical style)," IEEE pp 8 July 2014 659-663.

[14] M. Chen, S. Mao, and Y. Liu , " Big data : a survey", Mobile Networks and Applications, vol.19, no.2, pp.181-183, 2014

[15] Wahab MHA, Mohd MNH, Hanafi HF, Mohsin MFM (2008) Data pre-processing on web server logs for generalized association rules mining algorithm. World Acad Sci Eng Technol 48:2008

[16] A. Nanopoulos , Y. Manolopoulos , M. Zakrzewicz and T. Morzy , (2002) Indexing web access-logs for pattern queries. In: Proceedings of the 4th international workshop on web information and data management. ACM, pp 63–68

[17] K. Joshi , A. Joshi ,Y. Yesha, (2003) On using a warehouse to analyze web logs. Distrib Parallel Databases 13(2):161–180

[18] J. Cho and H. Garcia-Molina, ``Parallel crawlers,'' in Proc. 11th Int. Conf. World Wide Web, 2002, pp. 124_135

[19] C. Castillo, "Effective web crawling," ACM SIGIR Forum, vol. 39, no.1, pp. 55_56, 2005.

[20] S. Choudhary et al., "Crawling rich internet applications: The state of the art.'' in Proc. Conf. Center Adv. Studies Collaborative Res.(CASCON), 2012, pp. 146-160.

[21] (2013, Oct. 31). Robots [Online]. Available: http://user-agentstring. Info/list-of-ua/bots

[22] A. K. Jain, et al., Biometrics: Personal Identification in Networked Society. Norwell, MA, USA: Kluwer, 1999.

[23] Introduction to Disaster Management, Virtual University for Small States of the Commonwealth (VUSSC) pp 97-129, July 2011

[24] Introduction to Disaster Management, Virtual University for Small States of the Commonwealth (VUSSC) pp 97-129 July 2011

[25] Introduction to Disaster Management, Virtual University for Small States of the Commonwealth (VUSSC) pp 97-129 July 2011

[26] V. Hristidis, S. Chen, T. Li, S. Luis and Y. Deng,"Survey of Data Management and Analysis in Disaster Situations. (Periodical style)," ELSEVIER June 2010 https://www.elsevier.com/locate/jss

[27] M. Chen, S. Mao, and Y. Liu , " Big data : a survey", Mobile Networks and Applications, vol.19, no.2, pp.171-209, 2014.

[28] S.Chaudhuri, ''What Next? A Half-Dozen Data Management Research Goals for Big Data and the Cloud''(Periodical style)," Proceedings of the 1st symposium on Principles of Database Systems, ACM, 2012.

[29] M. Chen, S. Mao, and Y. Liu , " Big data : a survey", Mobile Networks and Applications, vol.19, no.2, pp.184-185, 2014.

[30] U. Troppens, R. Erkens,W. Mueller-Friedt, R.Wolafka, and N. Haustein, Storage Networks Explained: Basics and Application of Fibre Channel
SAN, NAS, ISCSI, FCoE. NewYork,NY, USA:Wiley, 2011.

[31] H. Hu, Y. Wen, T. Chua and X. Li, "Toward Scalable System for Big Data Analytics: a Technology Tutorial," (Periodical style)," IEEE pp 8 July 2014 665-666.

[32] R.Cattell Scalable sql and nosql data stores. ACM SIGMOD Record 39(4):12–27, 2011

[33] McKusick MK, Quinlan S. "Gfs: eqvolution on fastforward". ACM Queue 7(7):10, 2009

[34] R. Chaiken, B. Jenkins, Larson, P-A°. Ramsey B, D. Shakib, S. Weaver, J. Zhou Scope: "easy and efficient parallel processing of massive data sets. Proc VLDB Endowment "1(2):1265–1276, 2008

[35] D. Beaver, S. Kumar, Li HC, J .Sobel, P. Vajgel et al (2010) Finding a needle in haystack: facebook's photo storage. In OSDI, vol 10. pp 1–8

[36] Chang F, Dean J, Ghemawat S, HsiehWC,Wallach DA, Burrows M, Chandra T, Fikes A, Gruber RE (2008) Bigtable: a distributed storage system for structured data. ACM Trans Comput Syst (TOCS) 26(2):4

[37] R. Cattell (2011) Scalable sql and nosql data stores. ACM SIGMOD Record 39(4):12–27

[38] M .Burrows (2006) The chubby lock service for loosely-coupled distributed systems. In: Proceedings of the 7th symposium on Operating systems design and implementation. USENIX Association, pp 335–350

[39] Lakshman A, Malik P (2009) Cassandra: structured storage system on a p2p network. In: Proceedings of the 28th ACM symposium on principles of distributed computing. ACM, pp 5–5

[40] M. Chen, S. Mao, and Y. Liu , " Big data : a survey", Mobile Networks and Applications, vol.19, no.2, pp.187-190, 2014.

[41] S. Sagiroglou and D.Sinanc, "Big data : a review," in Proceedings of the International Conference on Collaboration Technologies and Systems (CTS'13), pp 42-47, IEEE, San Diego, Calif, USA, May 2013.

[42] S. Sagiroglou and D.Sinanc, "Big data : a review," in Proceedings of the International Conference on Collaboration Technologies and Systems (CTS'13), pp 42-47, IEEE, San Diego, Calif, USA, May 2013.

[43] J. H. Howard et al., ``Scale and performance in a distributed le system,''ACM Trans. Comput. Syst., vol. 6, no. 1, pp. 51-81, 1988.

[44] D .Laney (2001) 3-d data management: controlling data volume, velocity and variety. META Group Research Note, 6 February

[45] P. Zikopoulos, C. Eaton., (2011) Understanding big data: analytics for enterprise class hadoop and streaming data. McGraw-Hill

[46] IsardM, BudiuM, Yu Y, Birrell A, Fetterly D. Dryad: distributed data-parallel programs from sequential building blocks.ACM SIGOPS Oper Syst Rev 41(3):59–72. 2007

[47] Yu Y, Isard M, Fetterly D, Budiu M, Erlingsson U´ , Gunda PK, Currey Dryadlinq: a system for general-purpose distributed data-parallel computing using a high-level language. In: OSDI, vol 8. pp 1–14, 2008.

[48] J.Han, E. Haihong, G.lee, J.Du. "Survey on NoSQL database" Pervasive Computing and Applications (ICPCA), 2011 6th International Conference. Pp 363-366, IEEE, 26-28 Oct. 2011

[49] M. Chen, S. Mao, and Y. Liu , " Big data : a survey", Mobile Networks and Applications, vol.19, no.2, pp.186, 2014

[50] J.Han, E. Haihong, G.lee, J.Du. “Survey on NoSQL database” Pervasive Computing and Applications (ICPCA), 2011 6th International Conference. Pp 363-366, IEEE, 26-28 Oct. 2011

[51] Karger D, Lehman E, Leighton T, Panigrahy R, Levine M, Lewin D (1997) Consistent hashing and random trees: distributed caching protocols for relieving hot spots on the world wide web. In: Proceedings of the twenty-ninth annual ACM symposium ontheory of computing. ACM, pp 654–663.

[52] M. Chen, S. Mao, and Y. Liu , ” Big data : a survey”, Mobile Networks and Applications, vol.19, no.2, pp.186, 2014

[53] G. Blackett. Analytics Network-O.R. Analytics [Online]. Available:http://www.theorsociety.com/Pages/SpecialInterest/AnalyticsNetwork_analytics.aspx, 2013.

[54] H. Hu, et al., “Toward Scalable System for Big Data Analytics: a Technology Tutorial,”,” IEEE pp 671-672. July 2014

[55] C. Hermann,’’ Some Consequences of Crisis which Limit the Viability of Organizations’’ Administrative Science Quarterly (pp 61-82), 1963.

Big Data and Disaster Management, JST/NSF joint workshop, pp, 7, 8, 20 July 2011