IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 4, 2019, accepted July 6, 2019, date of publication July 11, 2019, date of current version July 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2928233

Towards Disaster Resilient Smart Cities: Can Internet of Things and Big Data Analytics Be the Game Changers?

SYED ATTIQUE SHAH[1,7], DURSUN ZAFER SEKER[2], M. MAZHAR RATHORE[3], SUFIAN HAMEED[4], SADOK BEN YAHIA[5], AND DIRK DRAHEIM[6]

[1]Institute of Informatics, Istanbul Technical University, 34469 Istanbul, Turkey
[2]Civil Engineering Faculty, Department of Geomatics Engineering, Istanbul Technical University, 34469 Istanbul, Turkey
[3]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar
[4]IT Security Labs, National University of Computer and Emerging Sciences (NUCES), Karachi 75160, Pakistan
[5]Department of Software Science, Tallinn University of Technology, 12618 Tallinn, Estonia
[6]Information Systems Group, Tallinn University of Technology, 12618 Tallinn, Estonia
[7]Department of Information Technology, Balochistan University of Information Technology, Engineering and Management Sciences, Quetta 87300, Pakistan

Corresponding author: Syed Attique Shah (shah@itu.edu.tr)

ABSTRACT Disasters (natural or man-made) can be lethal to human life, the environment, and infrastructure. The recent advancements in the Internet of Things (IoT) and the evolution in big data analytics (BDA) technologies have provided an open opportunity to develop highly needed disaster resilient smart city environments. In this paper, we propose and discuss the novel reference architecture and philosophy of a disaster resilient smart city (DRSC) through the integration of the IoT and BDA technologies. The proposed architecture offers a generic solution for disaster management activities in smart city incentives. A combination of the Hadoop Ecosystem and Spark are reviewed to develop an efficient DRSC environment that supports both real-time and offline analysis. The implementation model of the environment consists of data harvesting, data aggregation, data pre-processing, and big data analytics and service platform. A variety of datasets (i.e., smart buildings, city pollution, traffic simulator, and twitter) are utilized for the validation and evaluation of the system to detect and generate alerts for a fire in a building, pollution level in the city, emergency evacuation path, and the collection of information about natural disasters (i.e., earthquakes and tsunamis). The evaluation of the system efficiency is measured in terms of processing time and throughput that demonstrates the performance superiority of the proposed architecture. Moreover, the key challenges faced are identified and briefly discussed.

INDEX TERMS Big data analytics, Internet of Things, smart city, disaster management, Hadoop, spark, smart data analytics, geo-social media analytics, disaster resilient smart city.

I. INTRODUCTION

The intensity of disasters (natural or man-made) has increased in the last few decades. IFRC, world disaster report 2018 [1] identified 3,751 natural disasters such as earthquake, flood, tsunami, etc., that occurred in the last 10 years globally, costing total damage of 1,658 billion USD and affecting over 2 billion people. Moreover, a total of 118 man-made disasters such as nuclear meltdowns, structure failures, transportation accidents, terrorism acts, etc., were reported in 2017 only, resulting in more than 3000 deaths [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Lu Liu.

Disaster Management can be considered as a set of organized processes that incorporates the planning and managing of the activities in any of the disaster phases i.e., mitigation, rescue, response, and recovery. Disaster management activities are carried out through the collaboration of various concerned government and private sector authorities. The main aim of disaster management is the integration of the interrelated processes that can provide efficient means to analyze, monitor and or predict disasters. In order to minimize the possibilities of casualties and environmental destruction, disaster management measures need to be both preventive and reactive. The key functions of disaster management are to trigger early warnings, collect the information in real-time,

accurately estimate the damage, quickly figure out the evacuation routes and effectively manage emergency resource [3].

Traditional disaster management systems are getting outdated as they are becoming inadequate to manage operations with multi-sourced data and to store and analyze huge volumes of disaster data in real-time [4]. With the constraints of accurate and timely decision-making, disaster management and resilience processes require a reliable and effective environment that integrates various state-of-the-art technologies to enhance its performance. Moreover, it is very important to be able to engage any information source critical to the situation in time for emergency responders, especially during the initialization of the crisis response [5].

In this age of technology, the disaster management process can be provided with multiple supportive data sources to acquire information that can be utilized effectively in rescue, response and as well as in the mitigation and preparedness phases of a disaster. Modern disaster management systems need to support various types of data generated from heterogeneous sources, hence requires deploying effective data integration and multi-modal data analysis methods to extract valuable information. Relevant information needs to be collected from various potential data sources (i.e., Sensors, Social Media, Satellites, Smartphones, Authoritative/Historic data repositories, etc.), to increase the scope of situational awareness and acquire new insights for effective decision-making. Fortunately, with the emergence of new technologies such as the Internet of Things (IoT), Big Data Analytics (BDA), Cloud Computing, Fog Computing, etc., the disaster management process automation is getting equipped with more advanced services for timely and accurate operations. The growth of communications through Web 2.0; the latest technological advancements such as social media, smartphones, location-based systems, satellites, in-situ sensors data; and the potential ability to integrate them along with traditional data sources such as authoritative/public data and mass media can lead to new application era for the disaster management systems. The availability and integration of information from heterogeneous data sources and its coordination and understanding with decision makers, emergency responders, governments and also citizens will be the core ideology of this new and highly needed disaster management application model.

The world's population living in urban areas and neighboring localities is projected to rise to around 68% by 2050 [6]. The prompting increase in the population density of urban cities has defied the requirements of better services and suitable infrastructure for its inhabitants. Smart city incentives are considered an ideal solution by experts in both academia and industry to answer the challenges that occur from population growth, environmental pollution, shortage of energy sources, etc. [7]. The concept of Smart City is getting popularity, where various electronic devices and network infrastructure are incorporated together to attain high-quality two-way collaborative multimedia services. Hence, a smart city equipped with the capability of generating early warnings, monitoring, and predicting the disasters can be a game changer in minimizing fatalities by generating the required information and insights for the concerned authorities to intelligently manage the disaster scenarios.

An important component of any smart city is IoT, an infrastructure that allows devices to communicate with each other over the internet. IoT is evolving rapidly and immense value is given to it by various governments, enterprises and academic institutions. In the modern world, the scope and size of IoT are triumphing drastically, endowing new opportunities and also demanding challenges in the world of the internet [8]. Due to the intercommunication among various devices in such systems, a substantial amount of data is generated known as big data. The devices in such systems sense and transfer a large amount of data (Big Data) to the main station after identifying the encompassing activities. Billions of devices in correspondence with a huge population would intercommunicate, leading to the production of overwhelming big data that requires storage and analytics for information acquisition. Moreover, as the interconnected devices in IoT are getting more advanced, a variety of multimedia content (video, audio, still image, etc.) is also becoming available in IoT [9].

Social media platforms are also offering open opportunities for smart city initiatives to extract valuable information for improved decision-making. Users of social media are regarded as ''Human as a sensor'' since they provide real-time information that can offer more insights about a particular incident [10]. Social media enables people to communicate, express views and share contents like text /micro-blog, photos and videos with or without geo-location through an internet-based application. Crowdsourcing and especially volunteered geographical information (VGI) [11] are becoming the major basis of data for disaster management, as citizens are actively contributing in disaster response with easy access to social media and location-enabled reporting tools. Geospatial data, boosted with crowd generated geospatial content in the last few years is more in focus as compared to conventional data sources for disaster/crisis management systems [12]. A large amount of literature exists that is emphasizing on questions ranging from the overall framework of disaster social media design [13], to models that help emergency responders understand how crisis information is produced and shared by the general public through social media [14], to architectures for data quality assessment and filtration of user-generated content accessed from social media for disaster management [15].

''Big data'' is normally described as the ''next big thing in innovation'' and truly so, as big data concept is a revolutionary approach regarding data management and analysis. In literature, the term ''big data'' usually refers to two different concepts, i.e. a) to state the massive size of the data itself, and b) to state the ever-evolving set of techniques and technologies that aid in effective processing and more insightful analysis of large volumes of data. For big data applications, the most important task is to discover hidden values rapidly from datasets having the enormous size that can

possess various types of data (i.e., structured, semi-structured and unstructured) [16]. Big Data Analytics (BDA) examines large datasets from multiple sources for extracting valuable information and insights that can help organizations make informed decisions.

The huge volumes of unstructured data were considered useless a decade ago, but with the advancements of BDA tools; these datasets are being analyzed to acquire valuable information and insights. However, the reliability of captured data, ensuring the privacy of citizens, and lack of understanding and collaboration between volunteer groups and governmental organizations for managing big data are some of the key issues still faced [17]. Traditional data collection methods are very expensive and time-consuming, as it involves tedious field surveys and outdated instruments. Thus, the incorporation of smart technology is needed that can effectively and robustly gather a huge amount of data, perform analytics and predict the future for improved planning and development [18]. With the growing interest of companies, governments, and academia for utilizing the potential benefits of BDA, a great deal of research is going on regarding designing and deployment of systems to efficiently manage and analyze big data for extracting new insights for decision making [19]. Currently, the main sources of big data are the human interactions on the Web 2.0, sensing information on the IoT, operational and transactional data in enterprises and data generated from scientific research, etc. Out of which the big data generated by IoT originate unique characteristics that include heterogeneity between the datasets, a variety of information, unstructured features, noisy data, and high redundancy [20]. Excitingly, data streams from the IoT will test the traditional approaches for data management and will eventually endorse the concept of big data [21].

Developing architectural models that implement the IoT and BDA technologies for disaster management automation and addressing the potential design challenges associated in the same area is an overlooked aspect in current literature. When dealing with a massive amount of distributed data from multiple sources (i.e. social media, sensors, satellites, emergency responders, online news, etc.) the major issues faced are data aggregation, integration, and processing of the multi-source heterogeneous data. For solving data management issues in traditional disaster management systems, there is a need to develop system architectures that support the integration of multi-source data, provide effective communication and fast access, deliver updated and suitable data and assist in the standardization of information [22]. Since traditional methodologies are not suitable to deal with these huge volumes of data from multiple sources, BDA frameworks seem to be the effective solution to extract the required information and new insights from these raw data streams [23]. Big data has the potential for producing a much-advanced version of emergency response, as it has access to critical real-time information that can be helpful for disaster management [24]. Moreover, BDA is capable of processing huge sets of disaster-related data in real-time during any of the four phases of disaster management (i.e., Mitigation, Preparedness, Response, and Recovery) [25].

The major contributions of this paper include:

a) An innovative and state-of-the-art concept of BDA- and IoT-based environment for disaster resiliency in smart city infrastructure is proposed. The proposed concept of Disaster Resilient Smart City (DRSC) urges for the collaboration of IoT and BDA, where IoT has the potential to offer a framework of a ubiquitous network of interlinked sensors and smart devices, and BDA has the potential to facilitate the real-time processing of IoT along with other related data streams to reveal new information, patterns, and insights for effective disaster management. Figure. 1 illustrates the overall scenario of the concept of a BDA- and IoT-based DRSC.

b) A novel reference architecture is presented to demonstrate the general framework for the proposed concept of DRSC, with the aim to provide a roadmap for future expeditions. A complete five-layered architecture is planned for a DRSC environment, which supports large volumes of datasets from multiple data sources for efficient real-time and offline analysis that aids in triggering early warning/alert generation, monitoring, and prediction of disaster situations.

c) A combination of the Hadoop framework and Spark analytical engine is implemented and tested to support real-time and offline processing on various datasets generated from IoT and Twitter. The implementation model of the deployed system is focusing on the alert generation for disasters within the scope of the proposed DRSC framework to understand the system efficiently.

d) The system is evaluated regarding processing time and throughput. The results demonstrate the performance superiority of the system.

e) Finally, the open challenges that can be faced during the deployment of such an environment are identified and discussed briefly.

The remainder of this paper is organized as follows. Section II outlines the motivation for the research in detail. Section III thoroughly describes the proposed architecture and its subsection layers i.e., Data Resource, Data Transmission, Data Aggregation, Data Analytics and Management, and Application and Support Service. The implementation model for the deployed environment and its subsection i.e., Data Harvesting, Data Aggregation, Data Pre-Processing, and Big Data Analytics and Service Platform are also presented in the same section. Section IV presents the data analytics results and discussion along with the applied critical threshold, system implementation and system efficiency evaluation details. Section V highlights the key challenges that need to be addressed for future research undertakings. Finally, Section V concludes the paper.

II. MOTIVATIONS

There is an increasing and compelling demand from the disaster management community and concerned authorities

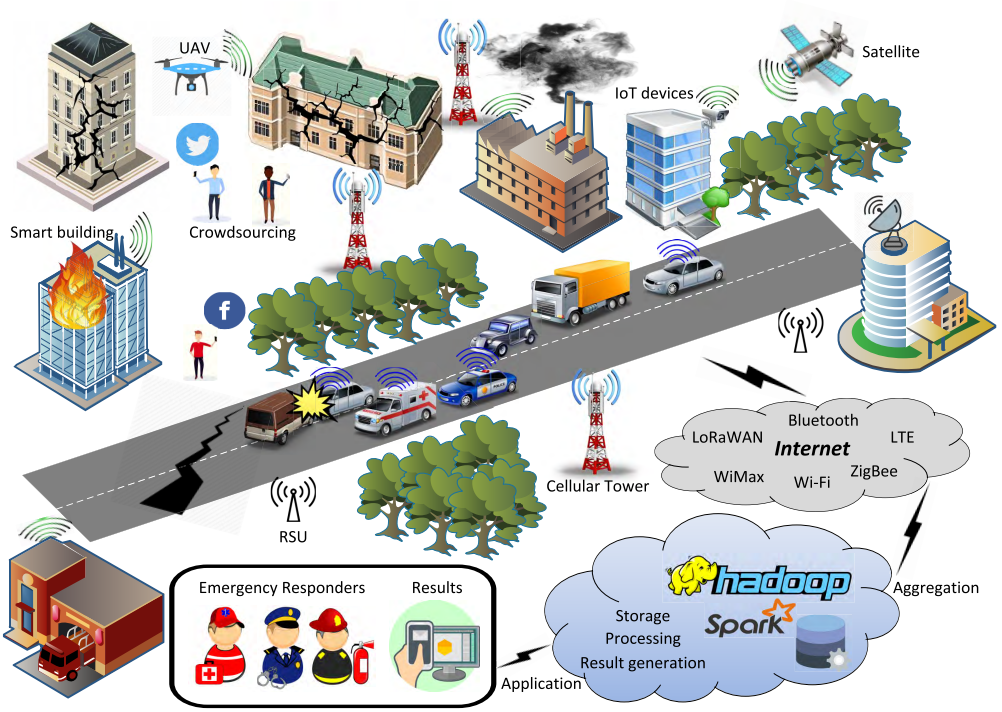

FIGURE 1. Illustrative scenario of a BDA- and IoT-based disaster resilient smart city.

to be provided with latest and accurate information for disaster management processes using any possible data source. Moreover, disaster response needs more improved operations and lack of (big) data availability for supply networks is a major limitation [26]. Zheng L, et al [27] state *"the techniques to efficiently discover, collect, organize, search, and disseminate real-time disaster information have become national priorities for efficient crisis management and disaster recovery tasks"*. It is challenging for the traditional disaster management systems to collect, integrate and process large volumes of data from multiple sources in real-time [28]. Moreover, the constraint of generating results in a small amount of time for emergency rescue and response, growing big data management issues and limited computational power makes the current traditional disaster management inadequate for the efficient and successful application. Previous studies have widely discussed the importance of timely, operational and accurate information for disaster management processes [29]–[31]. During the initial stages of a disaster, the responsible authorities need to make accurate and fast decisions. These decisions can only be successfully implemented if they are provided with quality information from different sources covering multiple dimensions.

Apart from the conventional data sources (i.e., field surveys, satellite imagery, archived databases) for disaster management a number of new potential data sources needs to be evaluated. One of the potential data sources for disaster management includes IoT-based sensors. IoT based sensors provide multi-dimensional data that can help in collecting the required information (readings of temperature, radiation, toxic gases, etc.) in case of any disaster. IoT driven platforms can provide disaster management systems such as early warning system with time critical, scalable and interoperable services [32]. IoT technologies offer the ability of distributed sensing with the potential integration of heterogeneous data, which makes it suitable for disaster management applications [33]. Another emerging and yet underused big data source for disaster management is social media. A smart city needs to consider social media to enhance communications with citizens, acquire feedbacks and encourage empowerment between citizens and authorized organizations. Though dealing with social media data requires an applied research approach, however, the importance of basic research for introducing the latest technology aided platforms and addressing the emerging architectural level issues for fast and effective processing of social media-generated data particularly for disaster applications cannot be neglected.

In one of our previous work [34], we identified the key benefits of BDA- and IoT-based disaster management systems and also investigated the recent literature conducted regarding various applications encompassing BDA and IoT for disaster management. We concluded that there are numerous benefits and many unexplored open opportunities allied with BDA and IoT technologies for the time-sensitive and accuracy-demanding application of disaster management. The growth of big data, the advancement of BDA tools and the expansion of the IoT are boosting the concept of smart cities. Smart cities are getting equipped with multiple data

sources to effectively help the citizens in their daily life activities. To deploy any smart city initiative, advance data sensing capabilities with highly efficient communication network play a major role. However, for a smart city to become a DRSC it needs to execute effective aggregation and storage of huge volumes of data, integrate heterogeneous datasets and perform analytics in both real-time and offline to extract the required information. These challenges signify the leading edge of BDA and IoT advancements, which collectively are capable of dealing with the urgency of this problem.

Disaster management applications necessitate more attention due to their time-sensitivity and high accuracy constraints owing to the life or death of human lives. Disaster management can be divided into various applications i.e., early warning/alert generation, response, evacuation, monitoring, and prediction. This study tries to implement and evaluate the alert generation application for disasters through a proposed architecture. Established early warning systems such as IMIS (the early warning system for radioactivity in the environment by the German federal government) [35] are often multi-source systems, but they are neither multi-modal nor do they support the disaster management life-cycle (response, continuity, recovery) [36]. Furthermore, they do not exploit today's available state-of-the-art technologies (such as Hadoop and Spark) and are, therefore, limited with respect to dealing with existing and emerging big data challenges.

BDA frameworks are used to analyze various applications of the smart city, however the time sensitive and accuracy demanding disaster/crisis/emergency management applications are still to be evaluated. There are very few research resources in the area of the smart city and disaster resilience and BDA- and IoT-based DRSC is rarely been investigated. Moreover, the requirement of an efficient and scalable compact environment for a BDA- and IoT-based DRSC has not been fully met yet. Therefore, this study attempts to present an architectural solution that is designed and evaluated for a DRSC and able to work with different data sources supported by state-of-the-art big data analytical tools. The motivation behind our effort is to provide innovative and effective BDA- and IoT-based DRSC architecture that considers heterogeneous data sources and real-time processing for more instant and insightful results. The aim of this research is to integrate different aspects of BDA and IoT for effective utilization of multi-source big data and to gain from the opportunities they offer for effective disaster management.

III. BDA- AND IOT-BASED DISASTER RESILIENT SMART CITY

In this section, we first propose a novel conceptual reference architecture that aims at providing an effective platform for storage, mining, and processing of various data sources including IoT generated and crowdsourced big data. Then we present the detailed information regarding the implementation model of our deployed system to illustrate its overall operations and functions.

A. PROPOSED REFERENCE ARCHITECTURE OF BDA- AND IOT-BASED DRSC

Several BDA and IoT architectures focusing on various operations and attributes in smart city concepts are found in the literature. For example, real-time data was utilized for BDA in an IoT-based smart city for the smart transportation system in [37]. In [38], a healthcare architecture is proposed that uses BDA on data from dedicated IoT devices. Similarly, in [39] the authors proposed an architecture for smart urban planning based on BDA and utilizing IoT datasets. In another study [40], a framework was proposed for weather data analysis using BDA and IoT to extract meaningful information from huge volumes of unstructured data. In [41], BDA and IoT based classification extension system design were proposed for monitoring water conditions in real-time. However, to the best of our knowledge, no architecture has entirely focused on integrating BDA and IoT for any kind of disaster management or resilience in smart city projects.

There is a great scope to validate and evaluate various BDA and IoT technologies for a mission-critical application such as disaster management. To benefit from the state-of-the-art applications and value-added capabilities presented by BDA and IoT with disaster management in perspective, we propose a novel disaster resilience smart city reference architecture that can be assisted with the advanced capabilities collaboratively offered by BDA and IoT. Based on the abstraction and identification of the various technological domains, the proposed architecture of IoT and BDA for a DRSC in this study can either be considered as i) a reference model for data abstraction that defines relationships among IoT and Big Data entities for DRSC and; ii) a standardized framework for assembling overall DRSC system entities.

The following challenging characteristics are taken into consideration during the design process.

- The architecture should be open to any potential data source that can provide additional insights to the results.
- The architecture needs to ensure the effective transmission of data over the communication networks.
- The architecture needs to guarantee the flawless storage of structured and unstructured data that can be either real-time or historical data.
- The architecture should be scalable to handle different data processing algorithms and analytical packages.
- The architecture should be able to present the processed results to the decision makers in an interactive manner and if necessary share the results with other subsequent applications.

As shown in Figure. 2 the proposed architecture is split into five layers, i.e., 1) Data Resource; 2) Data Transmission; 3) Data Aggregation; 4) Data Analytics and Management, and 5) Application and Support Services layer. The following subsections thoroughly describe each layer of this envisioned architecture in detail.

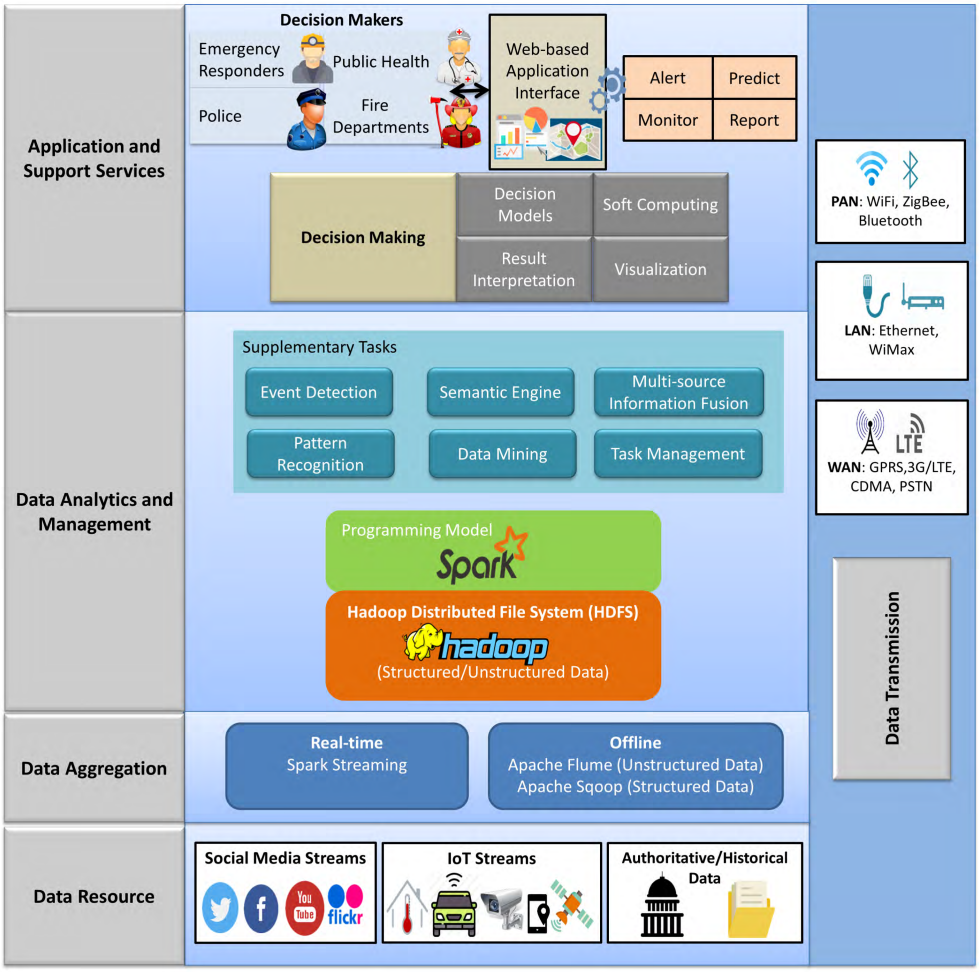

FIGURE 2. Proposed reference architecture for BDA- and IoT-based disaster resilient smart city.

1) DATA RESOURCE LAYER

This layer deals with the recognition all the potential data sources and collection of data generated by them. It contains all the potential IoT based data sources for DRSC such as in-situ sensors, RFID tags sensing, GPS, surveillance cameras, smartphones, satellite remote sensing etc. Moreover, DRSC can benefit to a large scale from a number of data resources, that can be taken aboard, such as social media streams and authoritative/historical data held by government or other disaster management organizations. Depending on the type of the source, the data can be about location, orientation, temperature, humidity, situation description, image or audio/video etc. Moreover, the collected data can be both structured and unstructured as illustrated in Table 1. These data sources generate different data types and formats. Hence integrating them for processing is a challenging task. The main data formats that can efficiently be processed in this proposed framework are (XML, CSV, JSON, ARFF, JPEG, and MPEG-2). Moreover, different data converters can be used to integrate various types of data prior to the data processing phase. The data sources are connected to a local data access middle layer or a remote station where the generated data are collected and integrated to be communicated via the data transmission layer.

2) DATA TRANSMISSION LAYER

Data transmission layer acts as the core component in any smart city architecture as it is providing the communication channels throughout the environment. The transmission layer is responsible to connect the data sources to the data aggregation layer and provide communication channels through out the environment in a secure and efficient manner. It is recommended to establish the disaster information networks by considering all the available options in the form of wired, wireless, or satellite networks to ensure a ''never-die-network environment'' [42]. The transmission can be on wired or wireless medium categorized by Local Area Network (LAN), Wide Area Network (WAN) and Personal Area Network (PAN). The transmission layer is supported by a combination of access transmission communication technologies such as ZigBee, Bluetooth, Wi-Fi, Ethernet, WiMax, NFC and RFID; and network transmission communication technologies such as CDMA, GPRS, 3G/LTE, and 5G.

TABLE 1. Structured and unstructured data in the context of disaster management.

Structured Data Examples	•Digitally archived incident related data (Reports on damages, causalities, injuries, etc.) •Data resources approved by government authorities (Demographic, Health records, etc.) •Location-based GPS third-party verified spatial data •Sensory data with meta-data (temperature, humidity, wind speed, precipitation etc.)
Unstructured\Semi-structured Data Examples	•Crowd-sourced data, including micro-blogs/text descriptions about the incident •Multi-media data (Pictures and videos) shared on social media related to the disaster •Public surveillance and private CCTV video recordings •Satellite imagery data of the affected area •Electronic/Online news data from different channels and web-sources

3) DATA AGGREGATION LAYER

With the possible inclusion of many heterogeneous data sources (i.e. IoT sensors, social media streams, satellites, electronic media, geo-portals, authoritative data), the system's reliability and value for effective decision-making increase undoubtedly, but on the other hand, it can also increase system vulnerability and complexity. The Data Aggregators are responsible to collect all the data under one multi-source database through different transmission mediums. Data can be gathered in the form of structured and unstructured data separately, using Apache Flume and Apache Sqoop respectively. Moreover, Spark Streaming can be utilized for real-time data collection. Apache Flume [43] is an open-source tool which provides a distributed and reliable service for collecting, aggregating and transferring huge volumes of unstructured data. It can aggregate and channelize unstructured data from various sources to HDFS directly. It is fault tolerant, robust and simple with many recovery mechanisms that use extensible data model for online analytic applications. Apache Sqoop [44] on the other hand is also an open-source tool but designed for extracting bulk data from structured databases (i.e. Relational database, NoSQL database, Data warehouses) to HDFS. Spark Streaming is ideal for real-time data aggregation from sources like Twitter and IoT based data streams. A combination of these tools, through a data pipeline can be utilized to collect the desired data. In this phase, the essential Meta data information such as data source, content, time stamps, location, etc. can also be identified.

4) DATA ANALYTICS AND MANAGEMENT LAYER

The main layer for data analytics and management contains a set of different tools to aggregate, store, process, query and analyze data. A combination of different BDA frameworks (i.e., Hadoop Ecosystem and Spark) can be reviewed to develop a real-time and efficient system for disaster management processes. An interoperable and efficient storage mechanism is required for the streaming structured and unstructured data. Hadoop Distributed File System (HDFS) [45] is a distributed storage file system designed to operate on commodity hardware with higher efficiency to handle large volumes of data. HDFS acts as the underlying storage for any Hadoop based system. Its main advantage is scalability, from a single server to thousands of machines and each capable of using local storage and computation. It consists of two types of nodes, i.e. NameNode denoted as "Master" and numerous DataNodes denoted as "Slaves". Namenode is responsible for managing the hierarchy of life system and director namespace that acts as metadata while DataNodes stores the actual data content. The data content is split into blocks and each block is replicated across different DataNodes for redundancy. Reports of all the existing blocks are sent to the NameNode periodically for monitoring and record. Along with HDFS based storage, a variety of programming models can be used for processing and analyzing big data, depending on the final results and business requirements. In this big data environment, it is very important to execute queries rapidly and retrieve results in the shortest time possible. Apache Spark [46] an open-source general computation engine for Hadoop, by far can fit the bill for a time critical and massive data sized systems. Spark is ideal for interactive queries and also supports processing of real-time data streams. It is a well-recognized processing framework with elegant APIs that supports various computer languages (i.e. Python, Scala, Java) and ensures fast, flexible and easy-to-use computing to execute machine learning or SQL assignments with streaming datasets. Moreover, it has a vast set of libraries (i.e. MLlib, GraphX, Spark Streaming, Spark SQL) for different functions with the possibility of adjusting and tuning according to the requirement.

A set of various supplementary tasks can be performed to accomplish the required analysis and to provide accurate and timely results to the decision-makers. Event detection is very critical in disaster management and needs to be operational to identify any disastrous event that occurs. Event detection backed by IoT sensor data and social media streams can detect any incident within the first few seconds of its occurrence [47]. Pattern recognition mechanism offers the machine learning ability to detect the useful patterns of information from textual or spatial data sets crucial for disaster management [48]. Semantic engine can be utilized for effective information management, i.e., categorizing, searching and extracting of unstructured information. A number of data mining techniques can be utilized to discover new, effective and otherwise hidden patterns of insights from the available information. Multi-source information fusion technologies help to integrate the required data from heterogeneous data

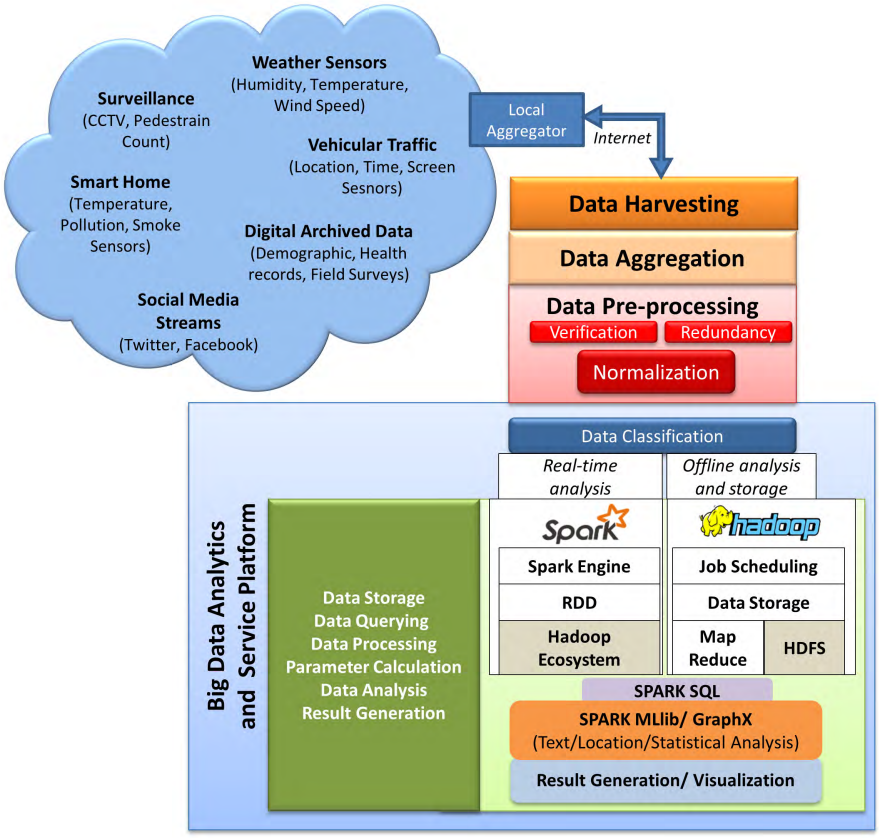

FIGURE 3. Implementation model of the deployed system.

sources. Task management helps to identify workloads on different entities in the system and effectively managing system's operations.

5) APPLICATION AND SUPPORT SERVICES

An interactive web-based application interface can provide decision makers (Emergency responders, Public health, Police, Fire Department) with the required results. The results can be queried and displayed with different visualization tools accessed through web-based APIs. Software solutions that provide a web-based user interface and does not require manually scripted queries can be utilized for result generation and visualization for decision-makers. Furthermore, the decision making process can be integrated with various services such as decision models, soft computing, result interpretation and visualization technologies depending on the requirements for a specific application. The obvious aim of the big data analytics platform is to boost the decision-making process with a steady flow of the required information and new patterns for more insights. The decision-making process can be supported by defining decision models that contain the steps of how the required goals are distinguished, structured and processed to carry out a particular decision. The decision-making process can then be provided with the generated results by using defined decision models, result interpretation tools and soft computing methods. Various visualization tools such as Kibana, Tableau, Plotly, etc., can be used to provide an interactive and user-friendly interface for decision-makers. Moreover, the proposed big data analytical services environment should be able to integrate with the traditional disaster management systems to provide results according to their configurations and requirements.

B. IMPLEMENTATION MODEL

The implementation model that outlines the details of all the operational steps performed in our deployed system focusing on alert generation within the scope of DRSC is presented in Figure. 3. The proposed implementation model is divided into four layers, i.e., 1) Data Harvesting; 2) Data Aggregation; 3) Data Pre-Processing; 4) Big Data Analytics and Service Platform. The following subsections explain each layer of the implementation model in detail.

1) DATA HARVESTING

Initially, a number of potential data sources (i.e., weather sensors, smart home-generated data, vehicular traffic, social media streams) that provide valuable information in the context of disaster management are identified. Data is primarily collected through local data aggregators of each respective data source. Local data aggregators convert the analog data

TABLE 2. Dataset details.

S#	Datasets	Description	Size	No. of Parameters	Targeted Application	Source
1	Fire	Fire Dynamic Simulator (FDS) developed by NIST, USA contains temperature data of a building	500 MB	06	Fire detection alert	[49]
2	Pollution	499 gas sensors placed within Aarhus city, Denmark to measure gases including (CO), (SO2), (NO2), (O3), and particulate matter	Raw data: 32GB Structural data: 570MB	07	Pollution level monitoring for toxic gases alert	[50]
3	Traffic	Road traffic simulator and signals optimizer	400 MB	06	Emergency evacuation route blockage alert	[51]
4	Twitter	Twitter data for the month of September, 2018	41 GB	04	Generating alert and collecting crowd-sourced information about disasters	[52]

into machine-readable digital form. Data harvesting process transfers the data from local aggregators that are collecting data from sensors environments measuring the real-world situations. The data harvesting is a challenging process due to the involvement of heterogeneous data sources producing huge amount of data. Therefore, we assume that potential sensors already deployed by various centers for different applications provide the data for our system. These data resource centers collect real-time data from heterogeneous sensors already deployed in smart cities. Hence, we are skipping the data harvesting mechanism in our proposed model and considering the recognized data sources as mentioned in Table 2, which consists of the information about all the utilized datasets, including dataset description, size, number of parameters, application and the reference of the data sources.

2) DATA AGGREGATION

Data aggregation process is performed to categorize the collected data for the effective extraction of the required values. Data aggregation process ensures the accessibility of the required data values from the available data sets and assembles it for further analysis. Our proposed model is open to various data sources (i.e., weather sensors, smart home-generated data, vehicular traffic, social media streams, digitally archived data). The collection of different data sources is considered as a Data Resource (DR) that provides the required data to the system. The DR contains Datasets sets (DS) (i.e., temperature, smoke, gas, etc.) comprising of Values (V) with their respective recorded Time (t). Table 3 shows the categorized illustration of the datasets that can be mathematically presented as in Equation 1. This categorization of DR helps in evaluating the required DS with respect to specific timings for a given scenario.

$$DS_m = \sum_{t=1}^{n} V_{m,t}$$

$$DR = \sum_{i=1}^{m} DS_i$$

Hence,

$$DR = \sum_{i=1}^{m} \sum_{t=1}^{n} V_{i,t} \tag{1}$$

TABLE 3. Data resource categorized illustration.

DR =		t_1	t_2	t_3	...	t_n
	DS_1=	$V_{1,1}$	$V_{1,2}$	$V_{1,3}$	...	$V_{1,n}$
	DS_2=	$V_{2,1}$	$V_{2,2}$	$V_{2,3}$	...	$V_{2,n}$
	DS_3=	$V_{3,1}$	$V_{3,2}$	$V_{3,3}$	...	$V_{3,n}$
	.	.	.	.	.	
	.	.	.	.	.	
	DS_m=	$V_{m,1}$	$V_{m,2}$	$V_{m,3}$		$V_{m,n}$

3) DATA PRE-PROCESSING

The datasets are initially pre-processed to remove incomplete, ambiguous, and redundant data. The raw datasets usually contain outlying, unfeasible or missing values that can lead to ambiguous results. Hence initially, the datasets need to be inspected for such issues to ensure that the atomicity of the data is retained. This layer cleanses the data by dealing with incomplete and noisy data. Data filtration steps define the data quality parameters for huge volumes of unstructured and structured data. This layer ensures the verification and credibility of the data source through its meta-data. The collected data contains a significant amount of redundant data; therefore, redundancy checks, that could be either syntactic and or semantic, remove unnecessary data to minimize the storage and processing load. Data pre-processing techniques need to be applied prior to any kind of data analytics. Data pre-processing also referred to as normalization, applies various data transformation techniques to compile the data values so that they fall within a prescribed range i.e. $[0 \sim 1]$. When integrating different data sources, normalization plays a key role in scaling the wide and short-ranging values to a common range for better data analysis. In our proposed algorithms, we required a common threshold value for some diverse datasets. Therefore, a normalization technique that can preserve the significance of each value including outliers was required.

We used the Z-score normalization using Mean Absolute Deviation to normalize the aggregated datasets. Z-score normalization [53] also referred to as zero-mean normalization technique is widely used to normalizes the dataset input values using Mean and either Standard Deviation (σ) or Mean Absolute Deviation (MAD). We opted for Z-score normalization with Mean Absolute Deviation (MAD) instead of Standard Deviation (σ) as it has been shown to be more

robust to outlier values and hence reduce outliers effect on the results. Mathematically it can be shown as,

$$\begin{aligned} S_A &= \frac{1}{n}(|V_1 - \bar{A}| + |V_2 - \bar{A}| + |V_3 - \bar{A}| + \ldots + |V_n - \bar{A}|) \\ V_i' &= \frac{V_i - \bar{A}}{S_A} \\ N_i &= \frac{1}{5}(V_i') \\ NDS &= N_i + 0.5 \end{aligned} \tag{2}$$

where $(\bar{A})$ is the mean of the attribute dataset and V_n represents the values in the dataset. S_A shows the final MAD value of that particular attribute data set. The normalization of values through Z-score normalization using MAD can be derived mathematically as shown in Equation 2. Where V_i represents the old values and V_i' is the new normalized value of an attribute dataset. The values after the z-score normalization lies between $[-2 \sim 2]$. To convert the values to an interval scale of $[0 \sim 1]$, we first divided all values by 5 to get the 1-point range. As the mean is still 0 at this stage therefore we added 0.5 to all values producing the final normalized values ranging from $[0 \sim 1]$. Normalized Data set (NDS) against each respective DS is then considered for further analysis. The pseudocode for the normalization process is proposed in Algorithm 1.

Algorithm 1 Data Normalization

BEGIN

Input: Datasets of each data values *(DS)*

Output: Dataset of normalized values *(NDS)*

Steps:

1: **FOR EACH** $i = 1$ to n **LOOP** ▷ i is ID of dataset
2: Calculate the Mean $(\bar{A})$ for each dataset
3: Calculate the Mean Absolute Deviation (S_A) for each dataset
4: Find the Z-score normalization (V_i') for each dataset value ▷ $V_i' = \frac{V_i - \bar{A}}{S_A}$
5: Divide all values by 5 ▷ to get 1-point range
6: Add 0.5 to all values ▷ to get values at scale of $[0 \sim 1]$
7: Return the normalized values in new datasets *(NDS)*
8: **CONTINUE**$(n+1)$;
9: **END LOOP**

END

We also focused on normalizing the Twitter dataset (TDS) considering alert generation process that can be achieved with the number of tweets in a specific location about a specific disaster event. Based on number of the geo-located tweets and textual content analysis, an alert generation process can be initiated. Moreover, with the twitter dataset input also compressed to $[0 \sim 1]$ scale regarding the number of location tweets and hashtag tweets, a wider set of possible solutions can be achieved with the integration of the threshold settings for various other applications. We retrieved Tweets from a specific disaster-affected location containing useful hashtags that are referring to the respective disaster and then sort the tweets according to their time-stamps. The algorithm that generates alerts is based on the number of disaster-related hashtags within the number of geo-located tweets gathered from the targeted location in a specified amount of time. Initially, the total numbers of tweets gathered from the target location are identified denoted as (T_L). Then, the total number of tweets with the related hashtags within (T_L) are filtered and denoted as (T_H). For example, for an earthquake scenario in Istanbul, Turkey, (T_L) will be the total number of tweets collected within the geo-coordinates of Istanbul. Then the number of earthquake-related hashtags or keywords (i.e., Earthquake, Deprem (Turkish for Earthquake)) are filtered out as (T_H) from (T_L) in fixed time intervals (t) (i.e., 5 mins). To normalize the Twitter dataset (TDS) in hand to a scale of $[0 \sim 1]$, we used the Equation 3.

$$TDS_t = \frac{T_H}{T_L} \tag{3}$$

4) BIG DATA ANALYTICS AND SERVICE PLATFORM

Large volumes of data require combination of state-of-the-art big data analytical tools that can efficiently process the datasets for both real-time and offline analysis. As shown in the proposed architecture, a combination of the Hadoop ecosystem and Spark engine is utilized to meet these requirements. Initially, the data is classified with the help of the identifier and the message type. The classification phase distributes the contents according to their data status and formats for effective processing. The classified data is then converted to Hadoop and Spark understandable format i.e., sequence files. The system platform equipped with the Spark Engine and Hadoop Ecosystem process the data according to the prescribed algorithms. The implementation is attained by using the Hadoop ecosystem with MapReduce mechanism. Parallel formation of MapReduce is deployed with HDFS. HDFS distributes the data in equal blocks among the data nodes. Each block is copied on more than one data node allowing each node to perform processing on its allocated block by using Map function. A master node with the authority of distributing data blocks to other nodes then concatenates the results from all the nodes by using Reduce function. A standalone Hadoop based system is only suitable for offline batch processing. Therefore, we deployed Apache Spark for real-time data processing. Apache Spark is used along with Hadoop for more powerful operations on real-time streams of data. Spark Streaming that supports both online and offline data streams is deployed for data aggregation in the system. Spark Engine works with Resilient Distributed Datasets (RDDs) which is an efficient in-memory (RAM) cluster computing abstraction. Spark provides fast, flexible, fault tolerant and advanced data analytics operations. By default Hadoop implementation in programmed in Java, so we used Java language for programming and also opted for the use of Java version of Spark. In our system, we are benefiting from the parallel data processing through Hadoop and real-time data processing by using Apache Spark. This combination provides flexible

and effective storage, accurate parameter calculation and fast result generation. SparkSQL [54] is a SQL based declarative languages that perform big data analysis tasks. It is Spark's module to query data inside Spark core programs. For data query we used SparkSQL as it gives fast response to queries even if scaling to thousands of nodes with spark engine. It enables extension with advanced analytics algorithms such as machine learning and graph processing. One of the key advantages of using Spark is the advance libraries it offers for analytics. Spark MLlib [55] is a machine learning framework that works with the Spark core engine. It is quite famous with data scientist due to its simplicity, language compatibility, scalability, Spark based speed performance and easy integration with other tools. It allows data scientist to forget about the infrastructure and configuration complexities and to only focus on their data related issues and models. Spark MLlib is a general-purpose library, which offers several optimized machine learning algorithms (e.g., classification, clustering, filtering, collaborative) and provides the flexibility to amend and extend the algorithms for specialized use cases. Spark GraphX [56] constitutes an interactive graph computation engine that manipulates graphs and executes graph and data parallel systems. It provides a library of graph-based algorithms (i.e. triangle counting, counted components, PageRank) for different graphs manipulation operations. Once we get the results from the big data analytics and service platform, the generated results are then visualized for better understanding.

IV. DATA ANALYTICS: RESULTS AND DISCUSSION

This section presents the defined critical threshold, analysis results, system implementation and efficiency evaluation details to perform and understand the feasibility of the study. The system developed with a combination of the Hadoop ecosystem and Spark engine is considered as the main station. The link is established from smart systems and twitter streams to the main station for aggregation of the real-time and offline data. As discussed before, due to the limited data access, at this level it is not possible to directly aggregate data from various potential data sources, therefore, existing smart systems' and twitter datasets are utilized for analysis. The aim of the analysis is to demonstrate how multiple heterogeneous data sources can be used in a DRSC concept to achieve the desired results. In the remainder of this section, we first explain the critical threshold used for the various applications. Then the analysis results and discussion against each IoT generated datasets and geocoded Twitter datasets are presented. Lastly, the system implementation and efficiency evaluation details are presented that illustrates the proposed system is efficient and scalable for applications.

A. DEFINING THE CRITICAL THRESHOLD

The critical threshold (CT) can be defined as a particular value or boundary limit which if exceeded alters the results or generate an alert. Various CTs are set for different datasets according to the application requirements in this study. CTs are defined manually for each dataset accordingly, such as temperature CT for fire detection and alert generation, toxic gases level CT for pollution monitoring, etc. CT values can be defined in binary, float or percentage format, such as 55 degree Celsius for fire detection and 200 $gram/meter^3$ gas level for toxic gases alert generation. The CT values are set based on the atmospheric conditions of the application environment. CTs needs to be carefully defined as the effectiveness of results depends on it. Table 4 contains the CT values established for different applications used in this study.

TABLE 4. Defined critical threshold for different IoT applications.

Application	Critical Threshold (CT)
Temperature	55°C
Smoke	200 g/m^3
Gases	200 g/m^3
Traffic	125 Vehicles

TABLE 5. Defined twitter critical threshold for various alert message level.

	Alert Message Status	Range
T_{CT}	Negative	[0 – 0]
	Informational	[0 – 0.09]
	Warning	[0.1 – 0.39]
	Critical/Emergency	[0.4 – 1]

Since we also have normalized Twitter dataset (TDS) considering tweet counts and their time-stamps, we established the Twitter Critical Threshold (T_{CT}) as shown in Table 5. The alert message status depends on the TDS value derived from the Equation 3 according to a respective scenario in a given time-frame.

B. ANALYSIS RESULTS

In order to exploit the proposed architecture for IoT generated datasets, we considered three main incidents that normally happen in our daily life and are suitable within the context of alert generation for disaster management. The application of these incidents are 1) Detecting fire in a building; 2) Monitoring overwhelming nature of pollution in the city; 3) Identifying road blockage due to any natural disaster or accident for assisting emergency evacuation. We elaborated in detail how the system detects these events and generate alerts.

The building (factory, office, house, etc.) temperature data is monitored for every room in order to identify the fire accident in the building. The fire simulator developed by NIST, called Fire Dynamic Simulator (FDS) [49], is used to generate various fire events in the building. We analyzed the temperature and smoke readings with their rising rates to identify a fire event or no event. Then, we set critical threshold for temperature and smoke readings for the fire event as proposed in Algorithm 2. The rising rate is calculated as the rising temperature and smoke values per minute. The algorithm calculates the average of the last 3 values with each new temperature and smoke value respectively. If the

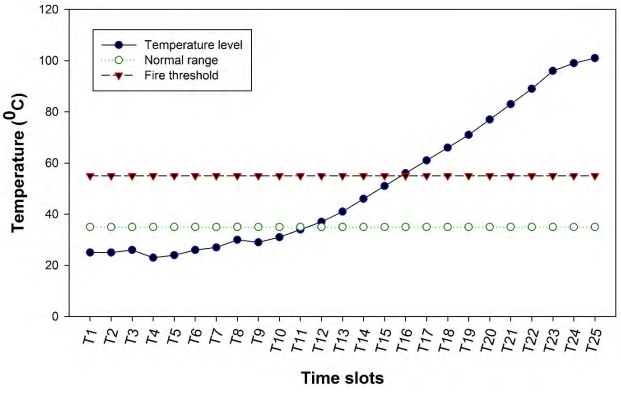

FIGURE 4. Fire monitoring through temperature analysis in a building.

Algorithm 2 Fire Alert

BEGIN

Input: Temperature *(T)* and Smoke *(S)* Readings
Output: Fire Alert/No-Fire
Steps:

1: **FOREACH** n Reading of Temperature *(T)* and Smoke *(S)* **LOOP**
2: $T\_Avg = \frac{\sum_{t=n-3}^{n} T_t}{3}$
3: $S\_Avg = \frac{\sum_{t=n-3}^{n} S_t}{3}$
4: **IF** ($T\_Avg >$ CT)
5: $T\_Report := TRUE$
6: **ELSE**
7: $GoTo(T_{n+1})$
8: **ENDIF**
9: **IF** ($S\_Avg >$ CT)
10: $S\_Report := TRUE$
11: **ELSE**
12: $GoTo(S_{n+1})$
13: **ENDIF**
14: **END LOOP**
15: **IF** (T_Report && $S\_Report = TRUE$)
16: *GENERATE* (Fire_Alert);
17: **ELSE**
18: **CONTINUE**(n+1);
19: **ENDIF**

END

average of the temperature and smoke readings exceeds their allocated CT values respectively, then the event is reported positive. If both temperature and smoke values result in positive reports, then the algorithm generates a fire alert. This method is proposed to confirm the occurrence of the fire event with different sensors data and to reduce the chances of false alarm in case of malfunction of one sensor.

Figure. 4 shows the temperature scenario (in degrees Celsius) while considering no-fire event and then abruptly the fire occurs. Till time T10, there is no event, thus, the temperature is lower and its changing behavior is quite predictable, which is also lower. Afterward, the temperature level

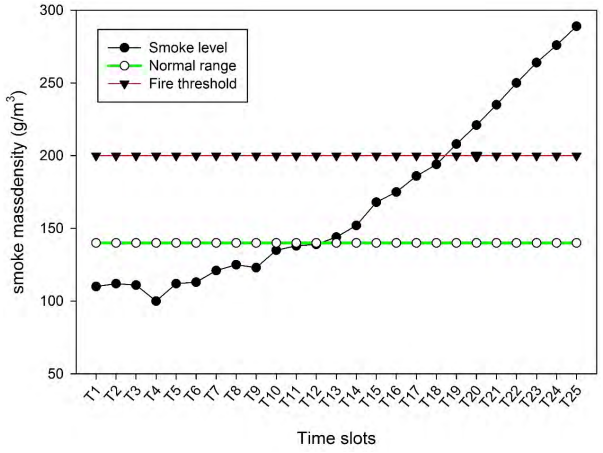

FIGURE 5. Smoke monitoring through smoke density in a building.

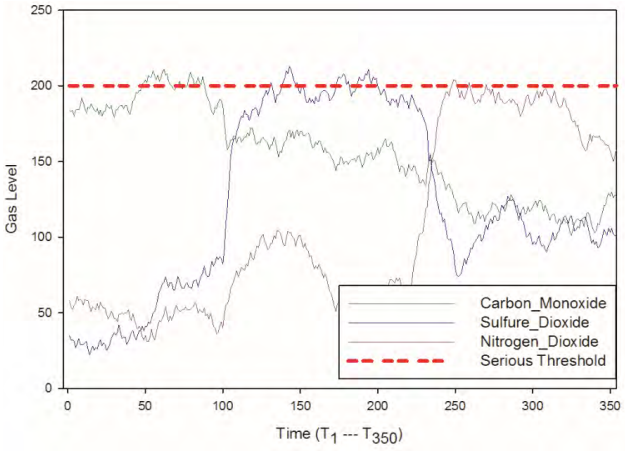

FIGURE 6. Pollution monitoring in a city through various gases.

upsurges gradually from the normal range. Hence, the system started analysis using temperature rising rate and noticed that the rising level is quite higher than before. So, the system presumed that it is fire. However, when its level increased from the critical threshold for temperature, the fire event is confirmed to report a true status. Similarly, Figure. 5, shows the smoke scenario measured in $gram/meter^3$. When both scenario returns true status the fire alert is generated and notified to take further necessary actions.

Correspondingly, we have also taken the pollution data [50] of Aarhus city in Denmark to generate alert for the invincible nature of pollution and toxic gasses in the city. The data is collected through 499 gas sensors placed within the city to measure toxic gases including carbon monoxide (CO), sulfur dioxide (SO2), nitrogen dioxide (NO2), ozone (O3), and other Particulate matter. Once any of these gases level exceeds from the normal range, it can be dangerous for citizens, especially children, elderly people, and allergy or asthma patients. Thus, the system generates alerts to the citizens if it exceeds the established CT indicating higher toxic gases level in the air. Algorithm 3 shows the pseudocode for the pollution level alert generation process. Figure. 6 shows

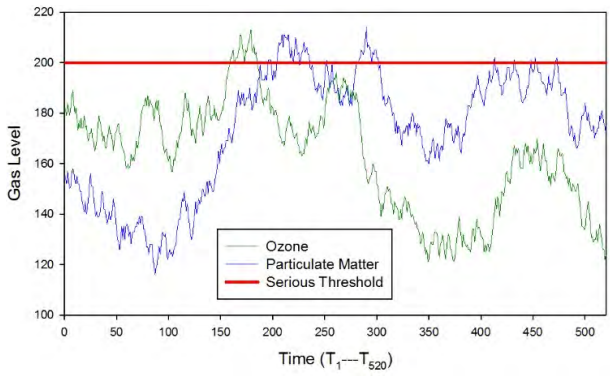

FIGURE 7. Pollution monitoring in a city through ozone and particulate matter's level.

Algorithm 3 Pollution Level Alert

BEGIN

Input: Air Quality Metrics *(M)*

Output: Pollution Alert/not-polluted

Steps:

1: **FOREACH** Gas_Readings *R (R_CO, R_SO2, R_NO2, R_O3)* in *(M)* **LOOP**
2: **IF** *(R_CO, R_SO2, R_NO2, R_O3)* > CT)
3: *Rep* =1
4: *Pollution_Alert();*
5: **ELSE**
6: *Rep* =0
7: *GoTo (Next_R)*
8: **ENDIF**
9: **END LOOP**

END

various time slots when the toxic gases i.e., carbon monoxide (CO), sulfur dioxide (SO2), and nitrogen dioxide (NO2) exceed from the serious threshold. Whereas, Figure.7 elaborates the changing behavior of ozone and particulate matters. At time T1 to T50, most of the time the ozone level is more than 200 in the air, which is dangerous for citizens. Accordingly, the system generates alerts to the people to take precautionary measures or avoid going outside.

Furthermore, for emergency evacuation path planning and real-time traffic analysis, to identify road blockage and accidents, we used the manually modified version of Volkhin road traffic simulator [51]. We took pairs of locations and the traffic data among them, including a number of vehicles moving in between each of the pairs and their speed. Road blockage is identified when the number of vehicles exceeds from the threshold and the 'time to reach' is exponentially increased. Algorithm 4 depicts the pseudocode for the route blockage alert process. The analysis result of the road blockage is depicted in Figure. 8. Till time T40, the number of vehicles between two the specified points is minimum. Consequently, the 'time to reach' is least and fluctuates based on the average speed of vehicles. However, whenever the vehicle

Algorithm 4 Route Blockage Alert

BEGIN

Input: Traffic Data with *(Num_vehicles)* and Time interval *(T)*

Output: Route Status *(Blocked and New Route)*

Steps:

1: Identify Time interval *(T)* ▷ *(T)* is time to reach
2: Identify *(R)* ▷ *(R)* is Route towards destination
3: **FOREACH** Reading *(Num_vehicles)* at *(T)* on *(R)* **LOOP**
4: **IF** *(Num_vehicles)* > CT)
5: *GoTo(Next_Reading)*
6: **ELSE**
7: *Blockage_Alert();*
8: *Alternative_Route (Assign New_Route (R))*;
9: **ENDIF**
10: **END LOOP**

END

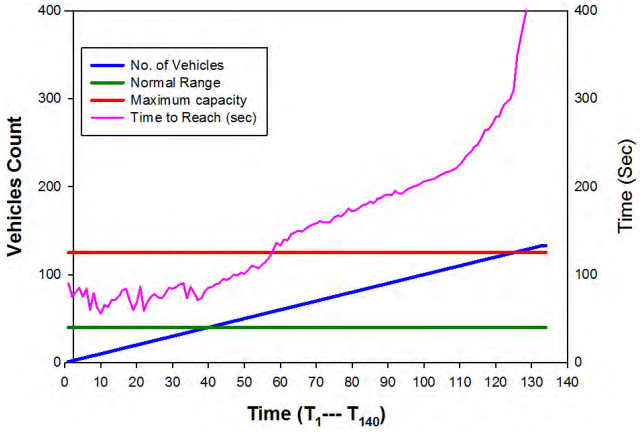

FIGURE 8. Traffic blockage analysis on a road.

count rises from the normal range, the 'time to reach' starts increasing accordingly as both are proportional to each other. Once the number of vehicles crosses the serious threshold limit (i.e., the maximum capacity of vehicles on the road), the value 'time to reach' parameter boosted exponentially. This boosting time value and the number of vehicles are two indicators of road blockage to assist emergency evacuation path planning.

In order to analyze the proposed architecture for Twitter datasets we focused on 2018 earthquake followed by tsunami occurred at Palu, Sulawesi, Indonesia. On 28 September 2018 at 18:02:44 local time, a large earthquake of 7.5 magnitudes struck the island of Sulawesi, Indonesia. Following the earthquake, a tsunami struck Palu city, sweeping houses, and buildings on its way. The death toll is estimated to be more than 3,000 people [57].

For the Twitter-based analysis, we acquired data from the Twitter stream grab [52], a Twitter data archive containing data from 2012 to 2018. The data sets are collected on a monthly basis, each having size of more than 40 GB and are provided in JSON format. Originally, we collected 41 GB

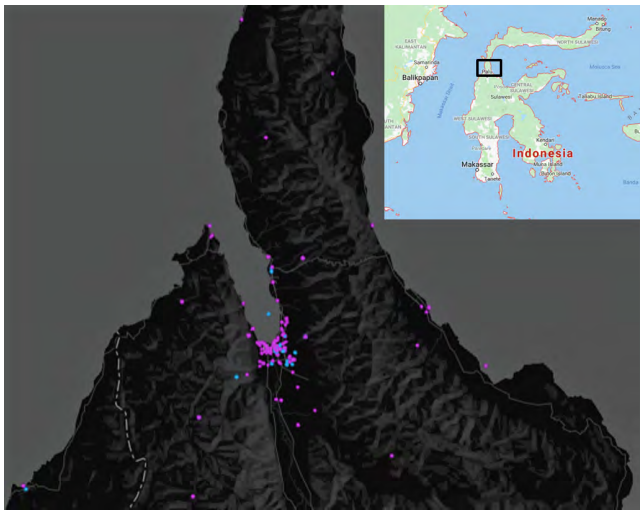

FIGURE 9. Overall geocoded tweet map of palu, indonesia from 28th to 29th september 2018.

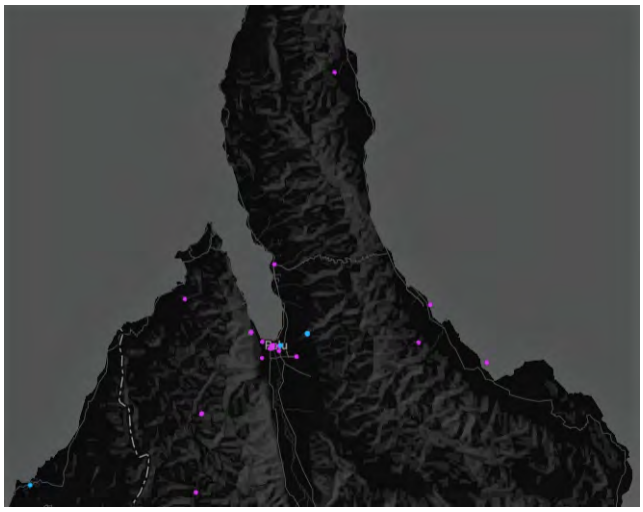

FIGURE 10. Geocoded tweets found with #earthquake and #gempabumi in palu, indonesia.

of Twitter data for the month of September 2018. Initially, we found a total number of 117,894,272 geocoded tweets without any geo-coordinate filtration. Since we only wanted to focus on the Palu city; therefore, we filtered out the tweets within the geo-coordinates of Palu city.

A total of 981 geocoded tweets were collected within the specified range from 28th to 29th September 2018 as shown in Figure. 9. Most of the twitter users do not enable the geo-location option while tweeting due to privacy concerns [58] and less than 5 percent of tweets have geo-coordinates attached with them [59]. Hence, the lesser number of tweets can be justified. The tweets were mostly tweeted in the Indonesian language (about 82%) as shown in Figure. 11. The final results were mapped using MAPD [60] for temporally visualizing data. The cross filtering capability of Twitter to analyze any activity with a given hashtag provides a great opportunity to acquire the

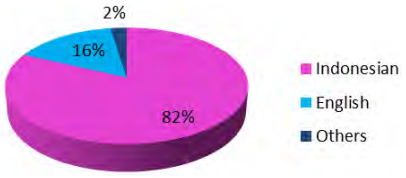

FIGURE 11. Major languages used for all geocoded tweets within Palu, Indonesia.

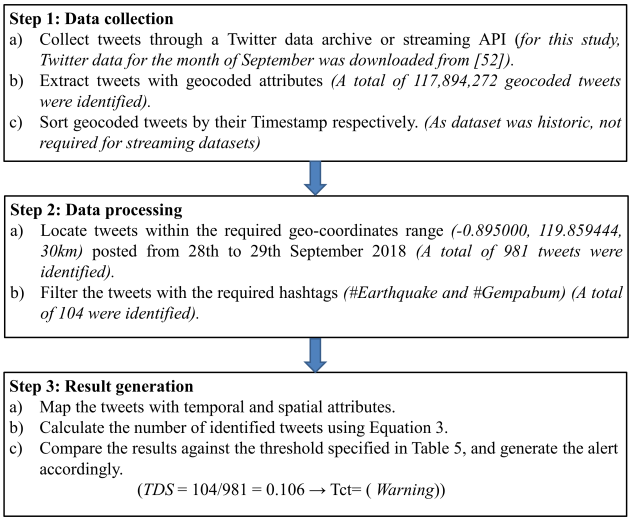

FIGURE 12. The workflow of Twitter data analysis.

desired results in a compact manner. We analyze all the geocoded tweets through hashtags, considering the main natural disasters i.e. (Earthquake and Tsunami). A total number of 104 tweets were identified with hashtags of Earthquake and Gempabumi (Indonesian for earthquake). Figure. 10 shows the geocoded tweet map filtered with #Earthquake and #Gempabumi within Palu city. Interestingly, these tweets were reported within a few minutes of the earthquake occurrence. After the earthquake, in approximately 30 minutes a six metre-high tsunami arrived to Palu city causing damage that was more devastating [61]. This scenario presents a very good case study to identify the role of twitter using alert generation thresholds. With the proposed Equation 3 and the Twitter-based critical threshold range as shown in Table 5, we can generate warning alerts for such a situation. The overall workflow of Twitter data analysis is illustrated in Figure. 12.

C. SYSTEM IMPLEMENTATION AND EVALUATION

The system is implemented on Hadoop single node environment assisted by different Spark libraries operated on Ubuntu 14.04 LTS with machine specifications as coreTM i5 supported by 3.2 GHz x 4 processors and 8 GB of RAM. The main hardware and software configurations used to implement the proposed system are shown in Table 6.

Since, we focused on processing large datasets that requires efficient real-time processing, therefore we evaluated our system with regards to data processing and

TABLE 6. The hardware and software configurations of the system.

Item	Version
Processor	Core(TM) i5-3470 3.2GHz
Hard Disk	SATA 7200RPM HDD
RAM	8GB
OS	Ubuntu 14.04 LTS
Apache Spark	Spark 2.3.1
Apache Hadoop	Hadoop 2.6.5

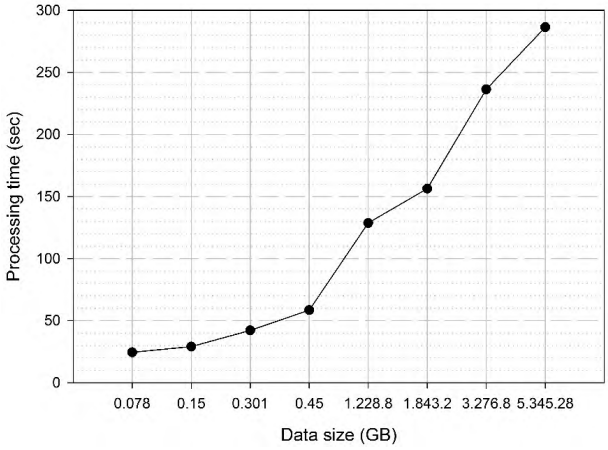

FIGURE 13. Efficiency of the system with respect to processing time with increasing dataset size.

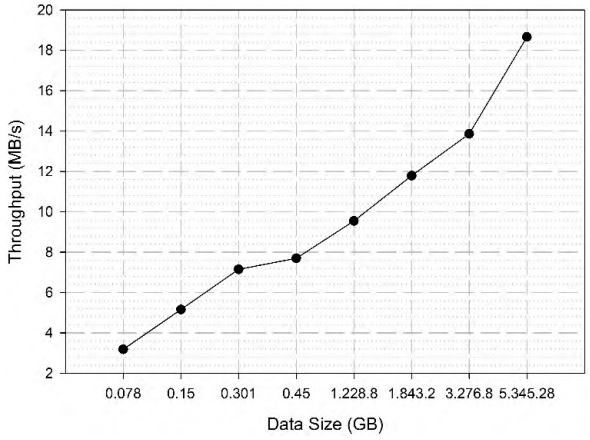

FIGURE 14. Efficiency of the system with respect to throughput with increasing dataset size.

throughput considering the increasing data size. Figure. 13 shows the processing time efficiency result corresponding to the increasing dataset size. It is expectable that with the increasing data, the processing time also rises. However, with our system, the rise in the processing time is quite lower corresponding to the huge rise in the data size. Figure. 14 shows the throughput analysis result of the system. The throughput result shows the number of MBs processed by the system in a given timeframe. The system shows promising throughput tendency with increasing data size. In addition, with our Hadoop implementation, the throughput of the system is increasing the function of data size. This increasing throughput with the data size is the major achievement

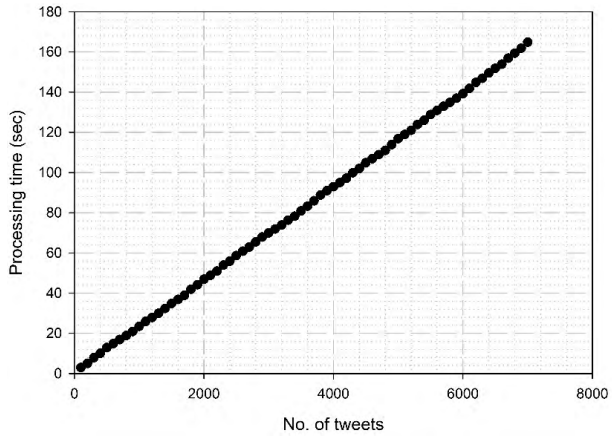

FIGURE 15. Efficiency of the system with respect to processing time for increasing number of tweets.

due to parallel processing implementation using MapReduce programing paradigm and number of simultaneous nodes of Hadoop. As we also processed a huge set of tweets for alert generation process, therefore processing time for Twitter dataset is shown in Figure. 15. Here, the system processed the tweets in accordance to the time sequence they were reported (i.e., milliseconds). The number of tweets are the 117,894,272 geocoded tweets that were identified without any geo-coordinate filtration.

V. RESEARCH CHALLENGES

In this section, the main research challenges that can be associated with the DRSC environment are discussed. The study has highlighted a variety of challenges that may be encountered during the designing and implementation phases and can reduce the efficacy of the environment. These research challenges can also identify some promising future research directions for further exploration and development of the DRSC environments.

A. FAULT TOLERANCE

In a disastrous situation, with multiple data sources, the probability for various hardware components to fail is high due to physical damage, exhausted batteries or failure of communication channels. In a DRSC environment data sources should be able to provide data even with blackouts and infrastructure impairment to maintain system availability. Backup power consumption mechanism and alternative communication channel establishment need to be guaranteed. Moreover, the environment needs to be equipped with capabilities such as regular backups and cloud-based storage mechanism with distributed computing support that can be used in case the primary system goes down.

B. INTEROPERABILITY

Data is acquired from various real-time and static data sources having different data formats. It is challenging to integrate large volumes of heterogeneous data that possibly can be of low quality due to high data redundancy. The required

information is hard to filter from this massive quantity of noise and ambiguous data as a whole. It is more challenging to integrate these heterogeneous datasets according to the system's requirements. To deal with data heterogeneity issues, sampling and filtering techniques need to be trained to acquire the highest level of semantic interoperability and data quality. Due to the diversity of data sources, interoperability issues is an open challenge that can be tackled if interoperability is assured on the data generation, structure, storage, coding, and software/hardware operations level.

C. META DATA

For a time-sensitive and data quality critical application like disaster management, metadata plays a vital role in identifying and managing the data sets. The collection and management of metadata for heterogeneous big data sources especially in disaster situations is an important challenge. Generating and maintaining metadata in big data paradigm is very difficult due to multiple data sources and data formats. While some of the data sets already possess some kind of metadata attached to them, most lack it. Additionally, it becomes more complex as many data sources i.e. numerous in-situ sensors are operated for different purposes by the government and private organizations. The key metadata features that need to be identified for the disaster-related data sets in the context of DRSC environment are data source, content, time stamps, spatial reference, data identification numbers. Through metadata, a number of data quality concerns and integration related issues can be removed and authentic datasets can be presented for analysis.

D. PRIVACY AND SECURITY

Privacy concern has been a serious issue in big data analytics, as it mostly utilizes personal information (i.e. financial, health records, location) to produce the required results. Personal information is exposed to scrutiny, which is increasing concerns about profiling, segregation, theft, and tracking [62]. For example, social media datasets contain personal information and location of the users, which can be used by malicious agents for harmful purposes, especially in a crisis like civil wars. The end users of IoT are faced with various security and privacy issues that limit IoT's usage and productivity [63]. Additionally, there is lack of adequate security tools for a number of technologies in the Hadoop Ecosystem [64]. Even with the availability of huge and richly detailed data, the threat of security either perceived or imminent can cause serious damage to the trust on data aggregation and sharing [65]. Applying suitable security mechanisms and access control checks on disaster-related data is important to ensure protection against malicious use and sustain data integrity, availability, and confidentiality.

E. TIME CONSTRAINT

Time is critical in disaster management as a quick response can save lives. Engaging huge volumes of heterogeneous data to extract desirable results in a limited time for emergency response is quite difficult. The data quality process itself involves complex processes like data aggregation, filtration, and normalization that can take plenty of time even with advanced tools. Moreover, unstructured data can add to the problem, demanding different filtration methods depending on the particular format. It is a big challenge for the existing techniques and tools to generate quality data from huge volumes of heterogeneous data according to the decision maker's requirement in a specified amount of time.

F. STANDARDIZATION

Standards are useful to endorse system efficiency, adopt technological and administrative changes, and provide legitimate guidelines for usage, policy, and future research. With the growing usage of BDA and IoT technologies, there is a big need and scope for communication standards, data integration standards and security standards to be re-examined. It is very challenging to define and follow standards for different evolving technologies keeping in mind the prerequisite of disaster management to be provided with accurate solutions in near real-time.

G. GIS-BASED VISUALIZATION

Mapping and visualization is the most important part of the DRSC environment, as decision-makers and emergency responders need quick and accurate predictions, insights and ground facts that are easy to interact with and understand. Big data analytics and visualization tools should work flawlessly to acquire effective results in real-time. Generally, the big data analytics interface is designed for technical users, so an additional tool is used for a user-friendly look and visualization. A Geographical information system (GIS) provides an interactive interface for mapping and analyzing spatial data. With the emergence of 3D and touch screen interactive technologies, visualization increases the processing time and hence demands additional system resources. Designing GIS-based visualization supported by big data analytics is an interesting research area which needs to be further investigated for user-friendliness and performance.

VI. CONCLUSION

The collaboration of the latest BDA and IoT technologies provides a more proficient environment for heterogeneous data sources to generate multi-dimensional data that is useful to perform effective analytics for extracting the required information used in disaster management applications. This approach can result in quick and effective situational awareness and hence help in reducing the impact of the disaster. A huge research gap still exists in BDA and IoT system planning and designing for a time-sensitive and performance demanding application like disaster management. The aim of this study is to contribute to the knowledge and guide future research regarding the design and implementation of BDA- and IoT-based disaster resilient smart cities. This study proposed a conceptual architecture for a novel Disaster Resilient Smart City concept by integrating BDA and

IoT. It provides a thorough outline of how BDA and IoT combined with some proposed parameters can effectively be implemented to aggregate, pre-process, and analyze data to provide updated and useful information for disaster managers. Hadoop ecosystem with Spark is utilized to implement the complete system for alert generation of disasters. Variety of datasets including IoT-based smarty city and twitter datasets are analyzed for showing the validity and evaluation of the proposed DRSC concept. The goal is to acquire full benefits that BDA and IoT collaboratively offer so that an improved disaster-resilient smart city concept equipped with the strengths of both the technologies can be designed and implemented. For future work, we anticipate the addition of other applications such as evacuation, monitoring, and prediction of disaster incorporating different data sources such as remote sensing, UAV imaginary, online news media and surveillance cameras for more in-depth analysis and better situational awareness. A Disaster Resilient Smart City (DRSC) environment would allow rapid and effective analysis backed with multi-sourced data for generating an early warning to citizens and assisting in the prevention, monitoring, and recovery from catastrophic situations. This study can provide references for researchers and industries for future acquisitions in the domain of smart cities and disaster management.

REFERENCES

[1] IFRC. (2018). *Executive Summary World Disasters Report: Leaving No One Behind*. [Online]. Available: https://media.ifrc.org/ifrc/wp-content/uploads/sites/5/2018/10/B-WDR-2018-EXECSUM-EN.pdf

[2] SwissRe. (2018). *Natural Catastrophes and Man-Made Disasters in 2017: A Year of Record-Breaking Losses*. [Online]. Available: http://media.swissre.com/documents/sigma1_2018_en.pdf

[3] V. Hristidis, S.-C. Chen, T. Li, S. Luis, and Y. Deng, "Survey of data management and analysis in disaster situations," *J. Syst. Softw.*, vol. 83, no. 10, pp. 1701–1714, 2010. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0164121210001329

[4] C. Baham, R. Hirschheim, A. A. Calderon, and V. Kisekka, "An agile methodology for the disaster recovery of information systems under catastrophic scenarios," *J. Manage. Inf. Syst.*, vol. 34, no. 3, pp. 633–663, 2017. doi: 10.1080/07421222.2017.1372996.

[5] N. Kapucu, "Interagency communication networks during emergencies: Boundary spanners in multiagency coordination," *Amer. Rev. Public Admin.*, vol. 36, no. 2, pp. 207–225, 2006. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0275074005280605

[6] *World Urbanization Prospects: The 2018 Revision*, Dept. Econ. Social Affairs Population Division, New York, NY, USA, 2019.

[7] B. N. Silva, M. Khan, and K. Han, "Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities," *Sustain. Soc.*, vol. 38, pp. 697–713, Apr. 2018. doi: 10.1016/j.scs.2018.01.053.

[8] D. Zeng, S. Guo, and Z. Cheng, "The Web of things: A survey," *J. Commun.*, vol. 6, no. 6, pp. 424–438, 2011.

[9] X. Huang, K. Xie, S. Leng, T. Yuan, and M. Ma, "Improving quality of experience in multimedia Internet of Things leveraging machine learning on big data," *Future Gener. Comput. Syst.*, vol. 86, pp. 1413–1423, Sep. 2018. doi: 10.1016/j.future.2018.02.046.

[10] B. Birregah, T. Top, C. Perez, E. Châtelet, N. Matta, M. Lemercier, and H. Snoussi, "Multi-layer crisis mapping: A social media-based approach," in *Proc. Workshop Enabling Technol., Infrastruct. Collaborative Enterprises (WETICE)*, Jun. 2012, pp. 379–384.

[11] M. Goodchild, "Citizens as sensors: The world of volunteered geography," *GeoJournal*, vol. 69, no. 4, pp. 211–221, 2007.

[12] B. Haworth and E. Bruce, "A review of volunteered geographic information for disaster management," *Geogr. Compass*, vol. 9, no. 5, pp. 237–250, 2015.

[13] J. B. Houston, J. Hawthorne, M. F. Perreault, E. H. Park, M. G. Hode, M. R. Halliwell, S. E. T. Mcgowen, R. Davis, S. Vaid, J. A. Mcelderry, and S. A. Griffith, "Social media and disasters: A functional framework for social media use in disaster planning, response, and research," *Disasters*, vol. 39, no. 1, pp. 1–22, 2015.

[14] T. Simon, A. Goldberg, and B. Adini, "Socializing in emergencies—A review of the use of social media in emergency situations," *Int. J. Inf. Manage.*, vol. 35, no. 5, pp. 609–619, 2015.

[15] S. A. Shah, D. Z. Şeker, and H. Demirel, "A framework for enhancing real-time social media data to improve the disaster management process," in *Advances in Cartography and GIScience* (Lecture Notes in Geoinformation and Cartography), M. Peterson, Ed. Cham, Switzerland: Springer, Jul. 2017, pp. 75–84. [Online]. Available: http://link.springer.com/10.1007/978-3-319-57336-6_6

[16] Z. Wang, S. Mao, L. Yang, and P. Tang, "A survey of multimedia big data," *China Commun.*, vol. 15, no. 1, pp. 155–176, Jan. 2018.

[17] J. Crowley, "Connecting grassroots and government for disaster response," Commons Lab, Wilson Center., Tech. Rep. SSRN 2478832, 2013.

[18] M. M. Rathore, A. Ahmad, A. Paul, and S. Rho, "Urban planning and building smart cities based on the Internet of Things using big data analytics," *Comput. Netw.*, vol. 101, no. 4, pp. 63–80, Jun. 2016. doi: 10.1016/j.comnet.2015.12.023.

[19] L. Rodríguez-Mazahua, C. A. Rodrîguez-Enríquez, J. L. Sánchez-Cervantes, J. Cervantes, J. L. García-Alcaraz, and G. Alor-Hernández, "A general perspective of big data: Applications, tools, challenges and trends," *J. Supercomput.*, vol. 72, no. 8, pp. 3073–3113, 2016.

[20] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Netw. Appl.*, vol. 19, no. 2, pp. 171–209, Apr. 2014.

[21] A. Zaslavsky, C. Perera, and D. Georgakopoulos, "Sensing as a service and big data," 2013, *arXiv:1301.0159*. [Online]. Available: https://arxiv.org/abs/1301.0159

[22] A. Meissner, T. Luckenbach, T. Risse, T. Kirste, and H. Kirchner, "Design challenges for an integrated disaster management communication and information system," in *Proc. 1st IEEE Workshop Disaster Recovery Netw. (DIREN)*, vol. 24, 2002, pp. 1–7. [Online]. Available: http://www.l3s.de/ risse/pub/P2002-01.pdf

[23] T. H. Davenport, P. Barth, and R. Bean, "How big data is different," *MIT Sloan Manage. Rev.*, vol. 54, no. 1, p. 43, 2012.

[24] S. Mehrotra, X. Qiu, Z. Cao, and A. Tate, "Technological challenges in emergency response," *IEEE Intell. Syst.*, vol. 28, no. 4, pp. 5–8, Jul./Aug. 2013.

[25] S. Akter and S. F. Wamba, "Big data and disaster management: A systematic review and agenda for future research," *Ann. Oper. Res.*, pp. 1–21, Aug. 2017. [Online]. Available: http://link.springer.com/10.1007/s10479-017-2584-2

[26] H. Watson, R. L. Finn, and K. Wadhwa, "Organizational and societal impacts of big data in crisis management," *J. Contingencies Crisis Manage.*, vol. 25, no. 1, pp. 15–22, 2017.

[27] L. Zheng, C. Shen, L. Tang, C. Zeng, T. Li, S. Luis, and S.-C. Chen, "Data mining meets the needs of disaster information management," *IEEE Trans. Human-Mach. Syst.*, vol. 43, no. 5, pp. 451–464, Sep. 2013.

[28] K. Neville, S. O'Riordan, A. Pope, M. Rauner, M. Rochford, M. Madden, J. Sweeney, A. Nussbaumer, N. McCarthy, and C. O'Brien, "Towards the development of a decision support system for multi-agency decision-making during cross-border emergencies," *J. Decis. Syst.*, vol. 25, pp. 381–396, Jun. 2016. doi: 10.1080/12460125.2016.1187393.

[29] J. Ortmann, M. Limbu, D. Wang, and T. Kauppinen, "Crowdsourcing linked open data for disaster management," in *Proc. 10th Int. Semantic Web Conf. Terra Cognita*, Jan. 2011, pp. 11–22.

[30] F. Alamdar, M. Kalantari, and A. Rajabifard, "Towards multi-agency sensor information integration for disaster management," *Comput., Environ. Urban Syst.*, vol. 56, pp. 68–85, Mar. 2016. doi: 10.1016/j.compenvurbsys.2015.11.005.

[31] J. P. de Albuquerque, B. Herfort, A. Brenning, and A. Zipf, "A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management," *Int. J. Geograph. Inf. Sci.*, vol. 29, no. 4, pp. 667–689, 2015.

[32] S. Poslad, S. E. Middleton, F. Chaves, R. Tao, O. Necmioglu, and U. Bugel, "A semantic IoT early warning system for natural environment crisis management," *IEEE Trans. Emerg. Topics Comput.*, vol. 3, no. 2, pp. 246–257, Jun. 2015.

[33] S. Fang, L. Da Xu, Y. Zhu, J. Ahati, H. Pei, J. Yan, and Z. Liu, "An integrated system for regional environmental monitoring and management based on Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 1596–1605, May 2014.

[34] S. A. Shah, D. Z. Seker, S. Hameed, and D. Draheim, "The rising role of big data analytics and IoT in disaster management: Recent advances, taxonomy and prospects," *IEEE Access*, vol. 7, pp. 54595–54614, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8698814/. doi: 10.1109/ACCESS.2019.2913340.

[35] D. Draheim, M. Horn, and I. Schulz, "The schema evolution and data migration framework of the environmental mass database IMIS," in *Proc. 16th Int. Conf. Sci. Stat. Database Manage. (SSDBM)*, Jun. 2004, pp. 341–344.

[36] *Societal Security–Business Continuity Management Systems–Guidance*, Standard ISO 22313:2012, 2012.

[37] M. M. Rathore, A. Paul, W.-H. Hong, H. C. Seo, I. Awan, and S. Saeed, "Exploiting IoT and big data analytics: Defining smart digital city using real-time urban data," *Sustainable Cities Soc.*, vol. 40, pp. 600–610, Jul. 2018.

[38] M. Babar, A. Rahman, F. Arif, and G. Jeon, "Energy-harvesting based on Internet of Things and big data analytics for smart health monitoring," *Sustain. Comput., Informat. Syst.*, vol. 20, pp. 155–164, Dec. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2210537917302238. doi: 10.1016/j.suscom.2017.10.009.

[39] M. Babar and F. Arif, "Smart urban planning using Big Data analytics to contend with the interoperability in Internet of Things," *Future Gener. Comput. Syst.*, vol. 77, pp. 65–76, Dec. 2017. doi: 10.1016/j.future.2017.07.029.

[40] A. C. Onal, O. B. Sezer, M. Ozbayoglu, and E. Dogdu, "Weather data analysis and sensor fault detection using an extended IoT framework with semantics, big data, and machine learning," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2017, pp. 2037–2046.

[41] R. Arridha, S. Sukaridhoto, D. Pramadihanto, and N. Funabiki, "Classification extension based on IoT-big data analytic for smart environment monitoring and analytic in real-time system," *Int. J. Space-Based Situated Comput.*, vol. 7, no. 2, pp. 82–93, 2017.

[42] Y. Shibata, N. Uchida, and N. Shiratori, "Analysis of and proposal for a disaster information network from experience of the Great East Japan earthquake," *IEEE Commun. Mag.*, vol. 52, no. 3, pp. 44–50, Mar. 2014.

[43] Apache Software Foundation. *Apache Flume*. Accessed: Jan. 18, 2019. [Online]. Available: https://flume.apache.org/

[44] The Apache Software Foundation. *Apache Sqoop*. Accessed: Jan. 18, 2019. [Online]. Available: http://sqoop.apache.org/

[45] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The Hadoop distributed file system," in *Proc. IEEE 26th Symp. Mass Storage Syst. Technol. (MSST)*, May 2010, pp. 1–10. [Online]. Available: http://www.alexanderpokluda.ca/coursework/cs848/CS848 Paper Presentation - Alexander Pokluda.pdf

[46] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark: Cluster computing with working sets," in *Proc. 2nd USENIX Conf. Hot Topics Cloud Comput. (HotCloud)*, 2010, p. 95.

[47] P. Zhang, Q. Deng, X. Liu, R. Yang, and H. Zhang, "Emergency-oriented spatiotemporal trajectory pattern recognition by intelligent sensor devices," *IEEE Access*, vol. 5, pp. 3687–3697, 2017.

[48] L. Greco, P. Ritrovato, T. Tiropanis, and F. Xhafa, "IoT and semantic Web technologies for event detection in natural disasters," *Concurrency Comput.*, vol. 30, no. 21, p. e4789, 2018.

[49] R. McDermott, K. McGrattan, and S. Hostikka, "Fire dynamics simulator (version 5) technical reference guide," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep. NIST Special Publication, 1018-5, 2008.

[50] S. Bischof, A. Karapantelakis, C.-S. Nechifor, A. Sheth, A. Mileo, and P. Barnaghi, "Semantic modelling of smart city data," in *Proc. W3C Workshop Web Things Enablers Services Open Web Devices*, 2014, pp. 1–5. [Online]. Available: http://www.w3.org/2014/02/wot/papers/karapantelakis.pdf

[51] Volkhin. *Road Traffic Simulator and Signals Optimizer in CoffeeScript & HTML5*. Accessed: Mar. 14, 2019. [Online]. Available: https://github.com/volkhin/RoadTrafficSimulator

[52] Archive Team. *The Twitter Stream Grab: Internet Archive*. Accessed: Mar. 14, 2019. [Online]. Available: https://archive.org/details/twitterstream

[53] S. B. Kotsiantis, D. Kanellopoulos, and P. E. Pintelas, "Data preprocessing for supervised leaning," *Int. J. Comput. Sci.*, vol. 1, no. 2, pp. 111–117, 2006.

[54] M. Armbrust, R. S. Xin, C. Lian, Y. Huai, D. Liu, J. K. Bradley, X. Meng, T. Kaftan, M. J. Franklin, A. Ghodsi, and M. Zaharia, "Spark SQL: Relational data processing in spark," in *Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD)*, 2015, pp. 1383–1394. [Online]. Available: http://dl.acm.org/citation.cfm?id=2723372.2742797

[55] X. Meng, J. Bradley, B. Yavuz, E. Sparks, S. Venkataraman, D. Liu, J. Freeman, D. Tsai, M. Amde, S. Owen, D. Xin, R. Xin, M. J. Franklin, R. Zadeh, M. Zaharia, and A. Talwalkar, "MLlib: Machine learning in apache spark," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 1235–1241, 2016. [Online]. Available: http://arxiv.org/abs/1505.06807

[56] R. S. Xin, J. E. Gonzalez, M. J. Franklin, and I. Stoica, "GraphX: A resilient distributed graph system on spark," in *Proc. 1st Int. Workshop Graph Data Manage. Experiences Syst.*, 2013, Art. no. 2.

[57] AHA Center, "SITUATION UPDATE No. 15—Sulawesi Earthquake," Sulawesi, Indonesia, Tech. Rep. 5, 2018. [Online]. Available: https://ahacentre.org/situation-update/situation-update-no-15-sulawesi-earthquake-26-october-2018/

[58] S. Wang, R. Sinnott, and S. Nepal, "P-GENT: Privacy-preserving geocoding of non-geotagged tweets," in *Proc. 17th IEEE Int. Conf. Trust, Secur. Privacy Comput. Commun., 12th IEEE Int. Conf. Big Data Sci. Eng. Trustcom/BigDataSE*, Aug. 2018, pp. 972–983.

[59] Z. Wang, X. Ye, and M.-H. Tsou, "Spatial, temporal, and content analysis of Twitter for wildfire hazards," *Natural Hazards*, vol. 83, no. 1, pp. 523–540, 2016.

[60] MAPD. *GitHub Omnisci Mapd-Core. The MapD Core Database*. Accessed: Apr. 12, 2019. [Online]. Available: https://github.com/omnisci/mapd-core

[61] K. Ravilious. *Terrawatch: Why Did the Quake in Sulawesi Cause a Tsunami? | World News | The Guardian*. Accessed: Apr. 8, 2019. [Online]. Available: https://www.theguardian.com/world/2018/oct/02/terrawatch-why-did-the-quake-in-sulawesi-cause-palu-tsunami

[62] O. Tene and J. Polonetsky, "Privacy in the age of big data: A time for big decisions," *Stanford Law Rev. Online*, vol. 64, pp. 63–69, Feb. 2012. [Online]. Available: http://www.stanfordlawreview.org/sites/default/files/online/topics/64-SLRO-63_1.pdf%5Cnpapers3://publication/uuid/F1C87BD7-F850-4414-B368-BC6C2EB96091

[63] S. Hameed, F. I. Khan, and B. Hameed, "Understanding security requirements and challenges in Internet of Things (IoT): A review," *J. Comput. Netw. Commun.*, vol. 2019, Jan. 2019, Art. no. 9629381. [Online]. Available: https://www.hindawi.com/journals/jcnc/2019/9629381/. doi: 10.1155/2019/9629381.

[64] G.-H. Kim, S. Trimi, and J.-H. Chung, "Big-data applications in the government sector," *Commun. ACM*, vol. 57, no. 3, pp. 78–85, 2014. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2566590.2500873

[65] S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini, "Security, privacy and trust in Internet of Things: The road ahead," *Comput. Netw.*, vol. 76, pp. 146–164, Jan. 2015. doi: 10.1016/j.comnet.2014.11.008.

SYED ATTIQUE SHAH received the M.S. degree in IT from the Balochistan University of Information Technology, Engineering and Management Sciences, Quetta, Pakistan. He is currently pursuing the Ph.D. degree from the Institute of Informatics, Istanbul Technical University, Istanbul Turkey. He was an Assistant Professor with the Department of Information Technology, BUITEMS, Quetta Pakistan. His research interests include big data analytics, cloud computing, information management and the Internet of Things.

DURSUN ZAFER SEKER received the Ph.D. degree in geomatics from Istanbul Technical University, Istanbul, Turkey, in 1993. Since 2004, he has been a Full Professor with the Department of Geomatics Engineering, Istanbul Technical University. His expertise is on photogrammetry, remote sensing, coastal zone management, watershed management and spatial data modelling and analysis from both the theoretical and empirical viewpoint. In these fields, he has been involved with several research projects both national and international, where these projects were interdisciplinary. He has authored more than 80 SCI international papers and more than 250 conference proceedings.

M. MAZHAR RATHORE received the master's degree in computer and communication security from the National University of Sciences and Technology, Pakistan, in 2012. He received the Ph.D. degree in computer science and engineering with Kyungpook National University, South Korea in 2018. He is currently working as a Postdoctoral Researcher with the College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar. His research interests include big data analytics, the Internet of Things, smart systems, network traffic analysis and monitoring, remote sensing, smart city, urban planning, intrusion detection, and computer and network security. He is ACM professional member. He is serving as a Reviewer for various IEEE, ACM, Springer, and Elsevier journals.

SUFIAN HAMEED received the Ph.D. degree in networks and information security from University of Göttingen, Germany. He was an Assistant Professor with the Department of Computer Science, National University of Computer and Emerging Sciences, Pakistan. He also leads the IT Security Labs at NUCES. The research lab studies and teaches security problems and solutions for different types of information and communication paradigms. His research interests include network security, web security, mobile security and secure architectures, and protocols for cloud and the IoTs.

SADOK BEN YAHIA received the habilitation degree to lead researches in computer sciences from the University of Montpellier, in 2009. He has been a Professor with the Technology University of Tallinn (TalTech), since 2019. His research interests include on combinatorial aspects in big data and their applications to different fields, e.g., data mining, combinatorial analytics (e.g., maximum clique problem, minimal transversals), smart cities (e.g., information aggregation & dissemination, traffic prediction). He is currently a member of the steering committee of the International Conference on Concept Lattices and their Applications (CLA) as well as the International French Spoken Conference on Knowledge Extractions and Management.

DIRK DRAHEIM received the Ph.D. from Freie Universität Berlin and a habilitation from Universität Mannheim, Germany. He is a Full Professor in information system and the Head of the Information Systems Group with Tallinn University of Technology, Estonia. Under his supervision the Information Systems Group conducts research in large- and ultra-large-scale IT systems. He is an initiator and a leader of numerous digital transformation initiatives.