**ISTANBUL TECHNICAL UNIVERSITY ★ INFORMATICS INSTITUTE**

# A NOVEL FRAMEWORK FOR DISASTER RESILIENT SMART CITIES: USING BIG DATA ANALYTICS

**Ph.D. THESIS**

**Syed Attique SHAH**

**Department of Applied Informatics**

**Geographical Information Technologies Programme**

**Thesis Advisor: Prof. Dr. Dursun Zafer ŞEKER**

**SEPTEMBER 2019**

ISTANBUL TECHNICAL UNIVERSITY ★ INFORMATICS INSTITUTE

# A NOVEL FRAMEWORK FOR DISASTER RESILIENT SMART CITIES: USING BIG DATA ANALYTICS

Ph.D. THESIS

Syed Attique SHAH
(706142008)

Department of Applied Informatics

Geographical Information Technologies Programme

Thesis Advisor: Prof. Dr. Dursun Zafer ŞEKER

SEPTEMBER 2019

**İSTANBUL TEKNİK ÜNİVERSİTESİ ★ BİLİŞİM ENSTİTÜSÜ**

# AFETE DAYANIKLI AKILLI ŞEHIRLER İÇİN ÖZGUN BİR ÇERCEVE: BÜYÜK VERİ ANALİTĞİ KULLANIMI

**DOKTORA TEZİ**

**Syed Attique SHAH**
**(706142008)**

**Bilişim Uygulamaları Anabilim Dalı**

**Coğrafi Bilgi Teknolojileri Programı**

**Tez Danışmanı: Prof. Dr. Dursun Zafer ŞEKER**

**EYLÜL 2019**

Syed Attique SHAH, a Ph.D. student of ITU Informatics Institute Engineering and Technology 706142008 successfully defended the thesis entitled "A NOVEL FRAMEWORK FOR DISASTER RESILIENT SMART CITIES: USING BIG DATA ANALYTICS", which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

**Thesis Advisor :** **Prof. Dr. Dursun Zafer ŞEKER** ..............................
Istanbul Technical University

**Jury Members :** **Prof. Dr. Selcuk PAKER** ..............................
Istanbul Technical University

**Prof. Dr. Bülent BAYRAM** ..............................
Yıldız Technical University

**Prof. Dr. Hande DEMIREL** ..............................
Istanbul Technical Univeristy

**Assoc. Prof. Dr. Ozan ARSLAN** ..............................
University of Kocaeli

**Date of Submission :** **31 July 2019**
**Date of Defense :** **12 September 2019**

***To The Almighty,***
***To My Family,***
***To The Search For Ultimate Meaning,***

## FOREWORD

The journey for acquiring a Ph.D. degree is inconceivable if you lack the appropriate support and guidance of some exceptional people around you. I appreciate all those people who have been there for me during this roller coaster ride of my Ph.D. life.

My advisor Prof. Dr. Dursun Zafer ŞEKER deserves a tremendous amount of respect and appreciation for his unfaltering support and guidance. His invaluable supervision, encouragement and insightful suggestions and annotations make it possible for me to successfully complete this work. I found him not only an advisor with exceptional energy but also a father-like figure, who was able to encourage me and make me understand the importance of hard work and dedication.

I would like to express my very great appreciation to Prof. Dr. Dirk DRAHEIM who has become one of the most influential people in my life. I owe him a debt of gratitude, as I would not have been able to accomplish my research publications without him. I have been fortunate to work with Information Systems Group, Tallinn University of Technology, Estonia under his kind supervision as a visiting Ph.D. student. His nurturing and observant suggestions are immeasurable and they mean a lot to me.

I would like to thank the Turkish Government for sponsoring my Ph.D. studies. I would also appreciate the support and flexibility provided by my home institute (Balochistan University of Information Technology, Engineering, and Management Sciences (BUITEMS)) during my study leave. A great deal of thanks goes to Dr. Sufian HAMEED, Dr. Mazhar RATHORE and Prof. Dr. Sadok Ben YAHIA for their collaborations, advice, paper reviews, and encouragements. I am extremely thankful to Dr. Ihsan ALI and Dr. Suleman KHAN for their invaluable comments and assistance. I would also like to thank my committee members Prof. Dr. Selcuk PAKER and Prof. Dr. Bülent BAYRAM for their input and support in my report presentations. I would also like to thank Prof. Dr. Wen-Chih PENG who supervised me during my internship at Advanced Database System Laboratory, National Chiao Tung University, Taiwan.

I am grateful to my parents and siblings. I could never ask for a more supportive and caring family than the one I have. I can never thank my beloved parents enough who have been always there to support me emotionally during tough times. Special thanks to my brother Mehr SHAH, for taking care of the family and motivating me. I must apologize for not being able to be with them on so many occasions. Last but not least, a big thanks to all my friends who were consistently helpful whenever I needed them. In particular, thanks to Musa KHAN, who brought inspiration and fun to my life.

September 2019 Syed Attique SHAH

**TABLE OF CONTENTS**

**Page**

## ABBREVIATIONS

| | |
|---|---|
| **API** | : Application Programming Interface |
| **BDA** | : Big Data Analytics |
| **CLOTHO** | : Crowd Lives Oriented Track and Help Optimization System |
| **CNN** | : Convolutional Neural Network |
| **CSO** | : Civil Society Organization |
| **D2D** | : Device-to-Device |
| **DMS** | : Disaster Management System |
| **DRSC** | : Disaster Resilient Smart City |
| **DTSOR** | : Disruption Tolerant Secure Opportunistic Routing |
| **FINDER** | : Finding Isolated Nodes using D2D for Emergency Response |
| **FRTN** | : Flying Real-Time Network |
| **GIS** | : Geographic Information System |
| **GPRS** | : General Packet Radio Service |
| **GPS** | : Global Positioning System |
| **HDFS** | : Hadoop Distributed File System |
| **HPC** | : High-Performance Computing |
| **ICT** | : Information and Communication Technology |
| **IFRC** | : International Federation of Red Cross and Red Crescent |
| **IoT** | : Internet of Things |
| **JSON** | : Java Script Object Notation |
| **LAN** | : Local Area Network |
| **LoRaWan** | : Long Range Wide Area Network |
| **LTE** | : Long Term Evolution |
| **MANET** | : Mobile Ad hoc Network |
| **NGO** | : Non-Governmental Organization |
| **NLP** | : Natural Language Processing |
| **OLAP** | : Online Analytical Processing |
| **PAN** | : Personal Area Network |
| **QoE** | : Quality of Experience |
| **QoS** | : Quality of Service |
| **RDD** | : Resilient Distributed Datasets |
| **SMS** | : Short Message Service |
| **UAV** | : Unmanned Aerial Vehicle |
| **UE** | : User Equipment |
| **VGI** | : Volunteered Geographical Information |
| **WAN** | : Wide Area Network |
| **WiMAX** | : Worldwide Interoperability for Microwave Access |
| **WSN** | : Wireless Sensor Network |

## LIST OF TABLES

**Page**

## LIST OF FIGURES

**Page**

**A NOVEL FRAMEWORK FOR DISASTER RESILIENT SMART CITIES: USING BIG DATA ANALYTICS**

## SUMMARY

Big Data Analytics (BDA) and the Internet of Things (IoT) based disaster management is an under-investigated research area, which includes many interesting opportunities and challenges. With IoT's capability of offering a framework of ubiquitous network with interlinked sensors and smart devices, IoT technology possesses the potential to be incorporated in disaster management and can provide a positive impact on every phase of emergency response. BDA, on the other hand, is known to facilitate the real-time processing of IoT and other related data streams and is capable of providing meaningful results for understanding the situations persisting in the disaster-affected areas, hence based on the analytical results the deployment of resources is optimal and effective. Moreover, big data generated in the IoT environments can be used for performing data analytics, monitoring, forecasts and generating alerts for unusual events. Therefore, this thesis focuses on the joint exploitation of BDA techniques and IoT technologies that can lead to the development of an innovative, effective and highly-needed disaster management environment.

Smart city incentives can play a major role in reducing fatalities by providing information and new insights for resourcefully managing the disaster scenarios. The concept of a smart city is being widely considered as an ideal solution to attain high-quality collaborative multimedia services. Cities are becoming equipped with the latest digital infrastructure of networks, sensors, and smart devices that are generating an enormous amount of data; which can contain rich streams of contextual, spatial and temporal information. With the excessive use of smart-phones and other portable mobile technologies equipped with sensors (i.e., GPS receivers, high-resolution cameras, microphones, accelerometers) and with the emergence of social media, the traditional way of data acquisition and management is being challenged. Big sensed data can provide a number of benefits such as situational awareness enhancement, improved allocation of resources and provision of a better source for informing disaster risk reduction strategies and risk assessments.

In this thesis, a novel conceptual framework is proposed for the fusion of BDA and IoT technologies which promises a new and more effective approach for carrying out the core operations of disaster management processes. The goal is to identify the benefits of BDA- and IoT-based disaster management and investigated the state-of-the-art literature conducted regarding BDA and IoT applications for disaster management. The aim of the objective is to contribute to the knowledge and future research of the design and implementation of BDA- and IoT-based disaster-resilient smart cities. The focus is to find how big data technology, combined with some proposed parameters can effectively be utilized to harvest, integrate, process and analyze datasets to provide updated and useful information for disaster managers.

During disasters, social media and micro-blogging services such as Twitter, Facebook, and Foursquare have become major sources for retrieving real-time information that can be used to trigger the alarm or to plan a rescue operation. Social Media datasets are playing a vital role in providing information that can support decision-making in nearly all domains. This is due to the fact that social media is a quick and economical approach to collecting data. It has already been proved that in case of disaster (natural or man-made) that the information extracted from Social Media sites is very critical to Disaster Management Systems for response and reconstruction.

The quality concerns of unstructured social media datasets are being widely considered as a challenging research opportunity. In this thesis, the quality of social media data is evaluated through a proposed filtration mechanism. Two components of the system are assessed to check the process of filtration. The first proposes a framework that provides updated and filtered real-time input data for the disaster management system through social media, and the second consists of a designed web user API for a structured and defined real-time data input process. The objective of this attempt is to propose a framework that can filter and organize data from the unstructured social media sources through recognized methods and bring this retrieved data to the same level as that acquired through structured and predefined mechanisms, such as a web API. Both components are designed such that they can potentially collaborate and produce updated information for a disaster management system to carry out accurate and effective decision-making.

The integration and processing of big data sources (i.e. IoT-based sensors, social media, crowd-sourced online mapping) can lead to a more effective but also a much challenging environment. This thesis contributes by implementing a novel architecture that can be used for managing and integrating different type's big data in light with available literature around the topic. The thesis concentrates on the implementation model that outlines the details of all the operational steps performed in the deployed system. The research proposes and evaluates a novel architecture to detect disaster events in real-time from a Twitter stream and other IoT-based sensory data to track the evolution of such events over time and location.

The implementation model reviewed in this thesis outlines the details of all the operational steps performed in the deployed system within the scope of DRSC. 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. Initially, the data is aggregated from various recognized sources and then normalized. The Z-score normalization using Mean Absolute Deviation to normalize the aggregated datasets. Then the normalized data set are 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 executable 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 the 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, Apache Spark was deployed 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. The implemented system benefits from 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.

The main layer for data analytics and management contains a set of different tools to aggregate, store, process, query and analyze data. Hadoop Ecosystem and Spark-based analytics are carried out to evaluate real-time and offline analysis for IoT and Twitter datasets. An interoperable and efficient storage mechanism is required for the streaming structured and unstructured data. Hadoop Distributed File System (HDFS) 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. Apache Spark, on the other hand, is an open-source general computation engine for Hadoop, by far can fit the bill for time-critical and massive data sized systems. Spark is ideal for interactive queries and also supports the 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.

The proposed scheme mainly targets processing large datasets that require efficient real-time processing, therefore the implemented system was evaluated with regards to data processing and throughput considering the increasing data size. Data filtering and normalization techniques have sufficiently dragged down the processing time and have increased throughput. The study evaluated various cases of Apache Spark, single and dual node MapReduce Hadoop cluster with generic and filtered datasets to compare the performance of various deployed schemes. The evaluation of the system efficiency is measured in terms of processing time and throughput that demonstrates the performance superiority of the proposed architecture.

## AFETE DAYANIKLI AKILLI ŞEHIRLER İÇİN ÖZGUN BİR ÇERCEVE: BÜYÜK VERİ ANALİTĞİ KULLANIMI


### ÖZET

Büyük Veri Analitiği (BDA) ve Nesnelerin İnterneti (IoT) tabanlı afet yönetimi, fırsat ve zorluklarla birlikte henüz üzerinde çok fazla çalışılmamış bir araştırma alanıdır. IoT'nin birbirine bağlı algılayıcılar ve akıllı cihazlar içeren her yerde bulunan bir ağ çerçevesi sunma yeteneği yardımıyla, IoT teknolojisi, afet yönetimine dahil olma potansiyeline sahiptir ve bu durum acil durum müdahalesinin her aşamasında olumlu bir etki sağlayabilir. Öte yandan, BDA'nın IoT ve diğer ilgili veri akışlarının gerçek zamanlı olarak işlenmesini kolaylaştırması nedeniyle felaketten etkilenen bölgelerdeki durumu analitik sonuçlara dayandırarak ortaya koyduğu için anlamlı değerlendirmeler yapmak mümkün olabilmektedir. Ayrıca, IoT ortamlarında üretilen büyük veriler, veri analitiğini gerçekleştirmek, izlemek, tahmin etmek ve olağandışı olaylar için tahminler ve uyarılar üretmek için kullanılabilir. Bu nedenle, bu tez çalışmasında, yenilikçi, etkili ve ihtiyaç duyulan afet yönetimi ortamının geliştirilmesine katkı vereceğini değerlendirilen BDA ve IoT tekniklerinin ortak kullanımına odaklanılmıştır.

Akıllı şehirlere yönelik çalışmalar, felaket senaryolarını kaynakça yönetmek için bilgi ve yeni bilgiler sunarak ölümleri azaltmada önemli bir rol oynayabilir. Akıllı şehir kavramı, yüksek kaliteli multimedya hizmetlerini elde etmek için ideal bir çözüm olarak kabul edilmektedir. Şehirler, her geçen gün çok büyük miktarda bağlamsal, uzamsal ve zamansal bilgi akışları içeren veri üreten ağların, algılayıcıların ve akıllı cihazların en son dijital altyapısıyla donatılmaktadır. Akıllı telefonların ve diğer taşınabilir mobil teknolojilerin aşırı kullanımıyla donatılmış algılayıcılar (yani GPS alıcıları, yüksek çözünürlüklü kameralar, mikrofonlar, ivmeölçerler) ve sosyal medyanın ortaya çıkmasıyla, geleneksel veri toplama ve yönetme yöntemleri yetersiz kalmaktadır. Algılanan büyük veriler, durumsal farkındalığın arttırılması, kaynakların daha iyi tahsis edilmesi ve afet riskini azaltma stratejileri ve risk değerlendirmeleri hakkında bilgi vermek için daha iyi bir kaynağın sağlanması gibi birçok fayda sağlayabilir.

Bu tez çalışmasında, afet yönetimi süreçlerinin temel işlemlerini gerçekleştirmek için yeni ve daha etkili bir yaklaşım sunan BDA ve IoT teknolojilerinin birleşmesi için yeni bir kavramsal bir çerçeve önerilmiştir. Bu amaç doğrultusunda, BDA ve IoT tabanlı afet yönetiminin faydalarını tespit etmek ve afet yönetimi için BDA ve IoT uygulamaları ile ilgili son teknolojiye yönelik çok geniş kapsamlı bir literatür çalışmasını gerçekleştirilmiştir. Çalışmada, bazı önerilen parametrelerle birleştirilen büyük veri teknolojisinin, afet yöneticilerine güncel ve yararlı bilgiler sağlamak için kriz verilerini üretmek, entegre etmek, işlemek ve analiz etmek için etkili bir şekilde kullanılabileceği üzerinde durulmuştur.

Afetler sırasında, Twitter, Facebook ve Foursquare gibi sosyal medya ve mikro blog hizmetleri, uyarı alarmı verebilmek veya bir kurtarma operasyonu planlamak için

kullanılabilecek gerçek zamanlı bilgileri almak için temel veri kaynakları haline gelmiştir. Sosyal Medya veri setleri, hemen hemen tüm alanlarda karar vermeyi destekleyebilecek bilgiler sağlamada hayati bir rol oynamaktadır. Bu, sosyal medyanın veri toplamada hızlı ve ekonomik bir yaklaşım olduğu gerçeğinden kaynaklanmaktadır. Afet durumunda (doğal veya insan yapımı), Sosyal Medya sitelerinden çıkarılan bilgilerin, müdahale ve yeniden yapılanma için Afet Yönetim Sistemlerinde çok kritik veriler olduğu zaten kanıtlanmıştır.

Yapılandırılmamış sosyal medya veri setlerinin kalitesine yönelik kaygılar, zorlu bir araştırma fırsatı olarak kabul edilmektedir. Bu tez çalışmasında, sosyal medya verilerinin kalitesi önerilen bir filtreleme mekanizması ile değerlendirilmiştir. Bu süreci kontrol etmek için sistemin iki bileşeni değerlendirilir. Birincisi, sosyal medya aracılığıyla afet yönetim sistemi için güncellenmiş ve filtrelenmiş gerçek zamanlı giriş verilerini sağlayan ve ikincisi yapılandırılmış ve tanımlanmış bir gerçek zamanlı veri giriş süreci için tasarlanmış bir web kullanıcısı API'sinden oluşan bir çerçeve önerilmektedir. Bu girişimin amacı, yapılandırılmamış sosyal medya kaynaklarından gelen verileri tanınmış yöntemlerle filtreleyebilecek ve düzenleyebilecek ve bu alınan verileri web API gibi yapılandırılmış ve önceden tanımlanmış mekanizmalar yoluyla elde edilenle aynı düzeye getirebilecek bir çerçeve önermektir. Her iki bileşen de, doğru ve etkili bir karar alma süreci afet yönetimine yönelik sistem için güncellenmiş bilgiler oluşturabilecek ve işbirliği yapabilecek şekilde tasarlanmıştır.

Büyük veri kaynaklarının (yani, IoT tabanlı sensörler, sosyal medya, kalabalık kaynaklı çevrimiçi haritalama) entegrasyonu ve işlenmesi daha etkili ama aynı zamanda çok zorlu bir ortama yol açabilir. Bu tez, farklı türdeki büyük verilerin yönetime alınması ve konuyla ilgili literatür ışığında birleştirilmesi için kullanılabilecek yeni bir mimarinin hayata geçirilmesine katkıda bulunmaktadır. Tez, konuşlandırılmış sistemde gerçekleştirilen tüm operasyonel adımların detaylarını gösteren uygulama modeline odaklanmaktadır. Araştırma, felaket olaylarını gerçek zamanlı olarak bir Twitter akışından ve diğer IoT tabanlı duyusal verilerden tespit etmek için bu tür olayların zaman ve konumdaki ilerleyişini izlemek için yeni bir mimari önermekte ve değerlendirmektedir.

Bu tez çalışmasında incelenen uygulama modeli, DRSC kapsamında konuşlandırılmış sistemde gerçekleştirilen tüm operasyonel adımların detaylarını ortaya koymaktadır. Önerilen uygulama modeli dört katmana ayrılmaktadır. Bunlar; 1) Veri Toplama; 2) Veri Yığma; 3) Veri Ön İşleme; 4) Büyük Veri Analitiği ve Servis Platformu. İlk olarak, veriler tanınmış çeşitli kaynaklardan toplanır ve ardından normalleştirilir. Toplanan veri kümeleri için Ortalama Mutlak Standart Sapma kullanılarak Z-skoru normalizasyonu gerçekleştirilir. Ardından normalleştirilmiş veri seti, tanımlayıcı ve mesaj tipi yardımıyla sınıflandırılır. Sınıflandırma aşaması, içeriği veri durumlarına ve etkili işleme formatlarına göre dağıtır. Sınıflandırılmış veriler daha sonra Hadoop ve Spark için anlaşılabilir formata, yani sıra dosyalarına dönüştürülür. Spark Engine ve Hadoop Ecosystem ile donatılmış sistem platformu, verileri öngörülen algoritmalara göre işler. Uygulama, MapReduce mekanizmasıyla Hadoop ekosistemi kullanılarak gerçekleştirilir. MapReduce'un paralel oluşumu HDFS ile konuşlandırılmıştır. HDFS, verileri veri düğümleri arasında eşit bloklar halinde dağıtır. Her bir blok, Harita fonksiyonunu kullanarak her bir düğümün tahsis edilen bloğunda işlem yapmasına izin veren birden fazla veri düğümüne kopyalanır. Veri bloklarını diğer düğümlere dağıtma yetkisine sahip bir ana düğüm daha sonra Reduce işlevini kullanarak tüm düğümlerin sonuçlarını birleştirir. Bağımsız bir Hadoop tabanlı sistem yalnızca çevrimdışı toplu işleme için uygundur. Bu nedenle, Apache Spark gerçek zamanlı veri işleme için dağıtılmıştır. Apache Spark, gerçek zamanlı veri akışlarında daha güçlü işlemler için Hadoop ile birlikte kullanılmıştır. Hem çevrimiçi hem de çevrimdışı veri akışlarını destekleyen Spark Streaming, sistemde veri toplaması için dağıtılır. Uygulanan sistem, Hadoop üzerinden paralel veri işlemeden ve Apache Spark kullanarak gerçek zamanlı veri işlemeden faydalanmaktadır. Bu kombinasyon esnek ve etkili depolama, doğru parametre hesaplama ve hızlı sonuç üretmeyi sağlamıştır.

Veri analizi ve yönetimi için ana katman, verileri toplamak, depolamak, işlemek, sorgulamak ve analiz etmek için bir dizi farklı araç içerir. Hadoop Ekosistemi ve Spark tabanlı analitik, IoT ve Twitter veri setleri için gerçek zamanlı ve çevrimdışı analizleri değerlendirmek üzere gerçekleştirilir. Akış yapılandırılmış ve yapılandırılmamış veriler için birlikte çalışabilir ve verimli bir depolama mekanizması gerekir. Hadoop Dağıtılmış Dosya Sistemi (HDFS), büyük hacimli verileri işlemek için yüksek verimlilikleki donanımında çalışmak üzere tasarlanmış dağıtılmış bir depolama dosya sistemidir. HDFS, herhangi bir Hadoop tabanlı sistem için temel depolama görevi görür. Diğer taraftan Apache Spark, Hadoop için açık kaynaklı bir genel hesaplama motorudur ve zamanla kritik ve büyük veri büyüklüğündeki sistemler için çok uygundur. Spark etkileşimli sorgular için idealdir ve aynı zamanda gerçek zamanlı veri akışlarının işlenmesini de destekler. Çeşitli bilgisayar dillerini (yani Python, Scala, Java) destekleyen ve akışlı veri kümeleriyle makine öğrenmesi veya SQL atamalarını yürütmek için hızlı, esnek ve kullanımı kolay bir bilgi işlem sağlayan zarif API'lere sahip iyi bilinen bir işlem çerçevesidir. Ayrıca, ihtiyaca göre ayarlama imkanı olan farklı işlevler için çok sayıda kütüphaneye (yani, MLlib, GraphX, Spark Streaming, Spark SQL) sahiptir.

Önerilen çerçeve, gerçek zamanlı işlem gerektiren büyük veri kümelerinin işlenmesine odaklanmıştır. Bu nedenle uygulanan sistem artan veri büyüklüğü dikkate alınarak veri işleme ve verim açısından değerlendirilmiştir. Veri filtreleme ve normalleştirme teknikleri, işlem süresini yeterince düşürmüş ve verimi artırmıştır. Çalışma, çeşitli şemaların performansını karşılaştırmak için Apache Spark'ın farklı durumlarıyla birlikte tek ve çift düğümlü MapReduce Hadoop küme örneklerini genel ve filtrelenmiş veri kümeleriyle değerlendirilmiştir. Sistem verimliliğinin değerlendirilmesi, önerilen mimarinin performans üstünlüğünü gösteren işlem süresi ve verim açısından ölçülmüştür.

# 1. INTRODUCTION

Disasters (natural or man-made) can cause great damage to human life, infrastructure, and environment; anywhere at any time. In the last 10 years, a total number of 3,751 natural disasters such as flood, earthquake, landslide, tsunami, etc. are identified by IFRC, world disaster report 2018 [1]. The financial loss associated with these disasters estimates about 1,658 billion USD, and with human casualties' rising around 2 billion people. Moreover, a total of 118 man-made disasters such as nuclear meltdowns, structure failures, transportation accidents, terrorist acts, etc., were reported in 2017 only, resulting in more than 3000 deaths [2].

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].

With the emergence of latest data analytics, service and communication technologies such as BDA, IoT, cloud computing, fog computing etc., disaster management systems are on the way to get equipped with multiple new supportive data sources as well as fast and cost-efficient data processing tools that can potentially be utilized to assist decision-making in all four phases of a disaster (i.e., rescue, response, mitigation and preparedness). During the course of any disaster, appropriate and timely decision-making based on accurate and up-to-date information determines the effectiveness of a disaster management system [3]. Applications demanding real-time operations on their high-speed data streams require fast and large-scale streaming data analytics to achieve desired results [4]. Through indulging diverse data sources such as physical sensing devices and crowd-sourced information, a larger environment can be provided for disaster management systems to make heterogeneous data sources generate multi-dimensional data useful for performing effective analytics hence generating better results and new insights. The growth of communication through Web 2.0; the possible integration of potential heterogeneous data sources (social media, IoT enabled sensors, satellites, smart-phones, authoritative/public data repositories, etc.); and the emergence of the powerful big data analytics tools (Hadoop, Spark, Kafka etc.) with interactive visualization applications (Kibana, Tableau, Plotly etc.) can lead to a paradigm shift in disaster management systems.

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. 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. [5]. Hence, a smart city equipped with the capability of generating early warnings, monitoring, and predicting the disaster 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 [6]. 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 [7].

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 [8]. 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) [9] are becoming the major basis of data for disaster management, as citizens are actively contributing in disaster response with their increasing 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 [10]. A large amount of literature exists that is emphasizing on questions ranging from the overall framework of disaster social media design [11], to models that help emergency responders understand how crisis information is produced and shared by the general public through social media [12], to architectures for data quality assessment and filtration of user-generated content accessed from social media for disaster management.

“Big data” is normally described as the “next big thing in innovation” and truly so, as big data have a revolutionary approach regarding data management. 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) [13]. Big Data Analytics (BDA) examines large datasets from multiple sources for extracting valuable information and insights that can help organizations make informed decisions.

Big data is associated with the ‘5Vs’ characteristics namely; Volume, Variety, Velocity, Veracity, and Value. Volume refers to the huge collection of data that needs to be stored and processed. Variety refers to the heterogeneous nature of the data having different formats such as text, images, video, and geo-data. Velocity refers to the rate at which the data is generated. Veracity comes in to play due to data incompleteness and data inconsistency, so it refers to the trustworthiness of the data. Big data is normally labeled with lack of veracity in data and due to this concern, big data usage is questioned for a critical domain like disaster management, where data needs to be accurate and reliable. The fifth characteristic is Value, which refers to the outcome or the valuable information conveyed by big data through some analytics for achieving a specific business goal. Figure 1.1 shows the big data characteristics along with with their details and the phase at which they could be considered in proposed scheme for this study.


**Figure 1.1** : Big data characteristics in accordance with the proposed scheme.

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 [14]. 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 [15]. 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 applicable systems to efficiently manage and analyze big data for extracting new insights for decision making [16]. Excitingly, data streams from the IoT will test the traditional approaches for data management and will eventually endorse the concept of big data [17]. 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 [18].

Through the effective collaboration of Internet of Things technologies and state-of-the-art big data analytical tools, large volumes of valuable data can be aggregated from multiple data sources and analyzed to generate required results in real time for effective decision-making in many applications. One such mission-critical application is disaster management that demands time-sensitive and high-performance characteristics in order to minimize the possibilities of casualties and infrastructural destruction. The rising challenge to productively aggregate and analyze big data generated from various sources, keeping in view the time constraint and an accuracy restraint of disaster management processes offers an open research opportunity.

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 [19]. 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 [20]. 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 [21]. 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) [22].

## 1.1 Purpose of Thesis

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 to the best of our knowledge 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 research attempts to present an architectural solution that is deployed 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.

## 1.2 Motivations

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 [23]. 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 [24].

There is an increasing and compelling demand from the disaster management community and concerned authorities to be provided with updated 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 [25]. Zheng L, et al [26] 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 [27]. 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 [28] [29] [30]. 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.

Established early warning systems such as IMIS (the early warning system for radioactivity in the environment by the German federal government) [31] are often multi-source systems, but they are neither multi-modal nor do they support the disaster management life-cycle (response, continuity, recovery) [32]. 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.

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 real-time to extract the required information. The DRSC concept necessitates more attention due to its time-sensitivity and high accuracy constraint application owing to the life or death of human lives. Such problem signifies the leading edge of BDA and IoT advancements, which collectively are capable of dealing with the urgency of this problem.

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 [33]. IoT technologies offer the ability of distributed sensing with the potential integration of heterogeneous data, which makes it suitable for disaster management applications [34]. 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.

### 1.3 Scope and Limitations

The thesis tries to fill the research gap that exists in planning and designing BDA applications for a time-sensitive and accuracy-demanding application like disaster management. This thesis can assist researchers and practitioners to understand and implement the concepts of BDA and IoT for preparing, responding and recovering from disasters. Concerned authorities for disaster management such as emergency responders, police, public health, fire department, and NGOs/CSOs can benefit from the proposed state-of-the-art BDA- and IoT-based disaster management architectures to enhance their decision-making for effective rescue and response operations. To the best of our knowledge, the contributions presented in this thesis are novel and the first of it's kind regarding BDA and IoT integration encapsulating any disaster management process.

The primary limitation of this work is the lack of implementation on large-scale infrastructure involving direct data collection from heterogeneous data sources over various data communication mediums. The data collection is an expensive and challenging process due to the involvement of different data sources producing a huge amount of data. At this stage, as it is not feasible for this research work to set up or get direct data access from a smart city incentive, therefore, already available and recognized data sources were used for evaluation of the system. Smart city initiatives have various kind of data sources (i.e., social media, IoT enabled sensors, satellites, smart-phones, authoritative/ public data repositories etc.) but this research only focuses on IoT generated and Twitter datasets. Limitation regarding the twitter datasets analysis is that the findings are based on selected hashtags within certain geographic boundaries. Another limitation is that the implementation of the system only focuses on analysis for early warning alert generation of disasters. However, the reference architecture is considering all the aspects of disaster management i.e., early warning alerts, evacuation planning, monitoring, and prediction.

## 1.4 Thesis Contributions

The major contributions of this thesis can be divided into three main categories as follows:

1. Review on the role of Big Data Analytics in Disaster Management

   As the first contribution, this thesis primarily review the existing BDA related literature within the scope of disaster management to explore the unrecognized opportunities and potential challenges associated for effective, timely and accurate disaster management related decision-making. The aim is to systematically identify future research openings and contribute to the knowledge of design and implementation of BDA-based disaster management environments.

Contribution 1 can be further summarized into the following sub-contributions:

(a) The main benefits and the key requirements of BDA-based disaster management environments are identified and discussed.

(b) Systematic literature review is performed to identify the recent research efforts published with regards to state-of-the-art BDA for disaster management applications.

(c) A thematic taxonomy is devised to categorize the related concepts and essential parameters while promoting an efficient yet feasible solution for BDA-based disaster management.

(d) An innovative and comprehensive conceptual reference model for BDA-based disaster management environments is proposed with the aim to provide a roadmap for future realistic applications.

(e) Few credible use cases considering disaster management operations are presented. The selection of use cases considers the sequence of disaster management operations to present an overall picture of the disaster management environments where IoT and BDA play an important role.

(f) A set of open challenges that need to be explored in the future and addressed for the desired research area are highlighted.

2. Proposing a framework for enhancing real-time big social media data to improve the disaster management process

   The second contribution of this thesis presents a implementation model for the development of an integrated system consisting of social media crowd-sourced component and a designed web API component through which organized and reliable data can be provided for real-time disaster management. This design-science research demonstrates that the concept of social media crowd-sourcing can effectively be used for real-time disaster management and tries to aid the theory of making crowd-sourced data as trustworthy as other data sources. The basic theme of this design is to make the unstructured crowd-sourced data processable so that it can be compared and merged with a structured data sources such as a web API. This contribution comprises of two parts: The first proposes a framework that provides updated and filtered real time input data for the disaster management system through social media, and the second consists of a designed web user API for a structured and defined real time data input process. The aim of this study is to propose a framework that can filter and organize data from the unstructured social media sources through recognized methods and bring this retrieved data to the same level as that acquired through structured and predefined mechanisms, such as a web API. Both components are designed such that they can potentially collaborate and produce updated information for a disaster management system to carry out accurate and effective decision-making.

3. BDA based novel framework for Disaster Resilient Smart Cities

   The final contribution of this proposes and discusses the novel reference architecture and philosophy of a Disaster Resilient Smart City (DRSC). 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). Contribution 3 can be further summarized into the following sub-contributions:

   (a) An innovative and state-of-the-art concept of BDA 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 BDA and IoT, 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.

(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, monitoring, and reporting 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 provided with the performed sequential steps to understand the system efficiently.

(d) Two use cases i.e., 2018 Indonesian earthquake twitter data and twitter data for identifying earthquake and storms in Turkey are taken into consider for the implementation model. The main theme for the use cases is to detect and generate early warning for a disaster and provide useful information for rescue and response.

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

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

## 1.5 Methodology

The methodology to carry out this research includes various steps as depicted in Figure 1.2. The initial step involves identifying and classifying the research problem. Once the research problem was identified a systematic literature review was conducted to extract various related work and existing solutions to understand the problem in depth. Background research was performed on the key topics and their subtopics to gather as many resources as possible to extract the required knowledge using the systematic literature review protocols. Various BDA and IoT state-of-the-art works were identified

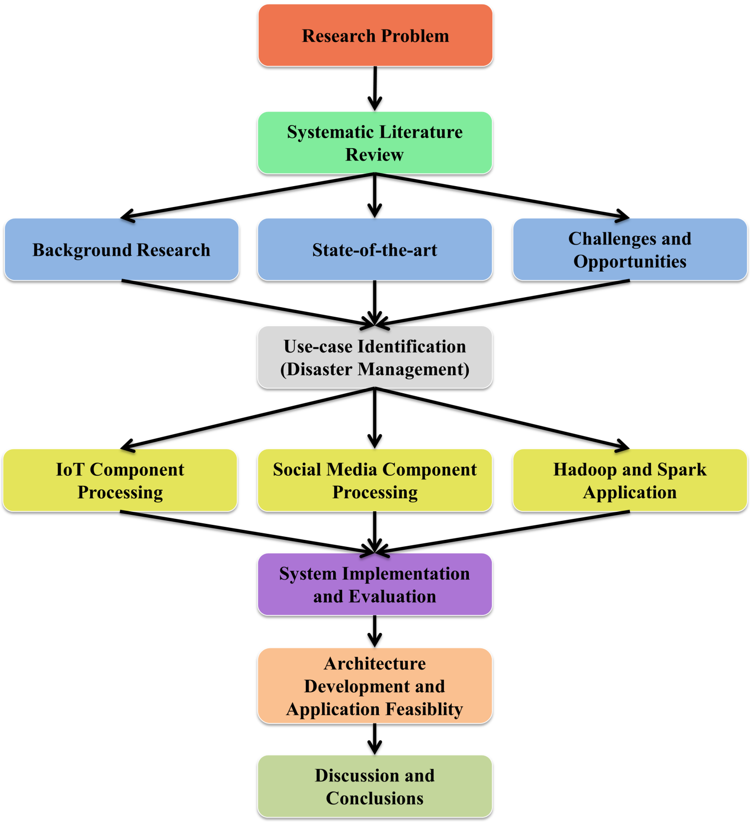


**Figure 1.2** : Research methodology for the thesis.

to check different possibilities for technologies integration and functionalities. The key challenges that can be faced in this specific research were identified and opportunities were discussed for future expeditions. Disaster management was chosen as a use-case to demonstrate the applications of this research work as disaster management processes require a time-sensitive and performance demanding application to save human life. Disaster management also provides various big data challenges in real applications due to the involvement of various data sources, in particular recently social media and IoT sensors usage is in high for various disaster management applications. The data analytics for disaster management requires various data filtration and quality concerns to be solved to generate applicable decisions in a short amount of time. These conditions provide an ideal scenario and a perfect use-case to evaluate the validity of various data quality techniques and big data analytical tools for this research. Disaster management has various applications such as prediction, alert generation and monitoring that can be investigated with the integration of IoT and BDA, but in the context of this research work, only the alert generation for disasters process was considered. The research work was further divided into three different components each directed towards the overall system implementation using different data sources. IoT and social media components were implemented separately using various data sources for alert generation for various kind of disaster i.e., fire, toxic gases, earthquake, and tsunami. Hadoop and Spark frameworks were evaluated with single and dual node machines for various results. After evaluation of the implemented system, an architecture was designed for large-scale application of IoT and BDA for disaster management for smart city initiatives. Lastly, discussion was made on the overall implementation of the architecture identifying its benefits and requirements and potential future directions were highlighted along with the drawn conclusions.

### 1.6 Thesis Organization

This thesis comprises of six chapters in total. Chapter 1 introduces the problem statement along with the purpose and motivations of the research. It describes the scope and key limitations of this work. It also highlights the overall methodology carried out for this research. Chapter 2 discusses the systematic literature review performed to identify the unrecognized opportunities and challenges associated with BDA and IoT for disaster management application. In this chapter, a thematic taxonomy is classified with several related attributes and inspects the prevalent solutions to propose a conceptual reference model for the deployment of BDA- and IoT-based disaster management environments. The aim of this chapter is to systematically identify future research openings and contribute to the knowledge of design and implementation of BDA- and IoT-based disaster management environments. Chapter 3 presents a framework for the filtration and quality insurance of the social media datasets. This chapter discusses about two components related to social media datasets. The first part proposes a framework that provides updated and filtered real-time input data for the disaster management system through social media and the second part consists of a designed web user API for a structured and defined real-time data input process. A novel reference architecture and the philosophy of a Disaster Resilient Smart City (DRSC) through the integration of various data sources including social media (Twitter) and IoT technologies is proposed and discussed in Chapter 4. In the fourth chapter, 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). Chapter 5 presents the proposed scheme implementation and performance evaluation results. Finally, Chapter 6 summarizes the overall research findings and discusses the recommendations for future research.

# 2. THE RISING ROLE OF BIG DATA ANALYTICS AND IOT IN DISASTER MANAGEMENT[1]

## 2.1 Abstract

The recent development of Big Data Analytics (BDA) and the Internet of Things (IoT) technologies create a huge opportunity for both disaster management systems and disaster-related authorities (emergency responders, police, public health, and fire departments) to acquire state-of-the-art assistance and improved insights for accurate and timely decision-making. The motivation behind this research is to pave the way for effective utilization of the available opportunities that the BDA and IoT collaboratively offer to predict, understand and monitor disaster situations. Most of the conventional disaster management systems lack the support for multiple new data sources and real-time big data processing tools that can assist decision makers with quick and accurate results. This chapter highlights the importance of BDA and IoT for disaster management and investigates recent studies directed towards the same. A thematic taxonomy is classified with several related attributes and inspect the prevalent solutions to propose a conceptual reference model for the deployment of BDA- and IoT-based disaster management environments . The reference model with its proposed integrated parameters can provide guidelines to harvest, transmit, manage, and analyze disaster data from various data sources to deliver updated and valuable information for disaster management. Some important use cases from a disaster management perspective is also enumerated. Lastly, the main research challenges that need to be addressed in such an important field of research are highlighted.

In this chapter, the existing BDA and IoT literature is reviewed within the scope of disaster management to explore the unrecognized opportunities and potential

[1]This chapter is based on the paper "Shah, S. A., Seker, D. Z., Hameed, S., Draheim, D. (2019). The Rising Role of Big Data Analytics and IoT in Disaster Management: Recent Advances, Taxonomy and Prospects. IEEE Access, 7, 54595-54614. [Online]. Available: https://ieeexplore.ieee.org/document/8698814."

challenges associated with their collaboration for effective, timely and accurate disaster management related decision-making. The key forte of this chapter is the emphasis that the integration of BDA and IoT technologies can provide promising solutions and new insights for disaster management applications. The aim of this chapter is to systematically identify future research openings and contribute to the knowledge of design and implementation of BDA- and IoT-based disaster management environments. A huge research gap still exists in planning and designing integrated BDA and IoT applications for a time-sensitive and accuracy-demanding application like disaster management. To the best of our knowledge, this paper presents the first survey of its kind regarding BDA and IoT integration encapsulating any disaster management process.

## 2.2 Introduction

The financial loss associated with these disasters estimates about 1,658 billion USD, and with human casualties' rising around 2 billion people [1]. Moreover, disastrous events such as terrorist attacks, oil spills, nuclear meltdowns, transportation accidents, etc., are prominent news channel headlines almost every day. Most of the large metropolitan cities of developing nations with increasing population are highly disaster vulnerable regions of the world. This is because their authorities lack situational information in case of a disaster, as they are largely constrained by shortage of resources [35]. Both natural and man-made disasters require preventive and reactive measures that need to be pre-planned for effective applications to reduce the chances of causalities and environmental/infrastructure damage. Therefore, disaster management systems need to effectively extract affirmative knowledge, monitor and analyze the ground situation, facilitate evacuations and predict the occurrence of disasters. Disaster management related government authorities, researchers and practitioners have been endeavoring to enhance the disaster management processes by considering new ideas from various research gatherings, such as information technology, cartography, health sciences, and environmental sciences. Their ultimate goal is to enhance the data gathering, managing, processing and visualizing phases of disaster management systems for timely and accurate decision-making. This precise and quick decision-making constraints for the disaster management systems require the utilization and integration of several state-of-the-art technologies to support its operations resourcefully.

The concept of smart city is being widely considered as an ideal solution to attain high-quality collaborative multimedia services [36]. Cities are becoming equipped with the latest digital infrastructure of networks, sensors and smart devices that are generating an enormous amount of data; which can contain rich streams of contextual, spatial and temporal information [37]. Smart city incentives can play a major role in reducing fatalities by providing information and new insights for resourcefully managing the disaster scenarios. With the excessive use of smart-phones and other portable mobile technologies equipped with sensors (i.e., GPS receivers, high-resolution cameras, microphones, accelerometers) the traditional way of data acquisition and management is being challenged. Big sensed data can provide a number of benefits such as, situational awareness enhancement, improved allocation of resources and provision of a better source for informing disaster risk reduction strategies and risk assessments [25]. Multiple data sources can generate a large amount of unstructured data to the remote station on request or after identifying the encompassing activities. However, it is quite challenging to process these huge volumes of heterogeneous data in real-time when a disastrous event is triggered [22]. Practices focusing on the discovery, collection, classification, search and distribution of real-time disaster information have the highest priority for an efficient performance in disaster management tasks [26].

Currently, BDA- and IoT-based disaster management is an under-investigated research area, that includes many interesting opportunities and challenges. With IoT's capability of offering a framework of ubiquitous network with interlinked sensors and smart devices [38], IoT technology possess the potential to be incorporated in disaster management and can provide a positive impact on every phase of emergency response [39]. BDA on the other hand, is known to facilitate the real-time processing of IoT and other related data streams [40], and is capable of providing meaningful results for understanding the situations persisting in the disaster-affected areas, hence based on the analytical results the deployment of resources is optimal and effective [41]. Moreover, big data generated in the IoT environments can be used for performing data analytics, monitoring, forecasts and generating alerts for unusual events [42].

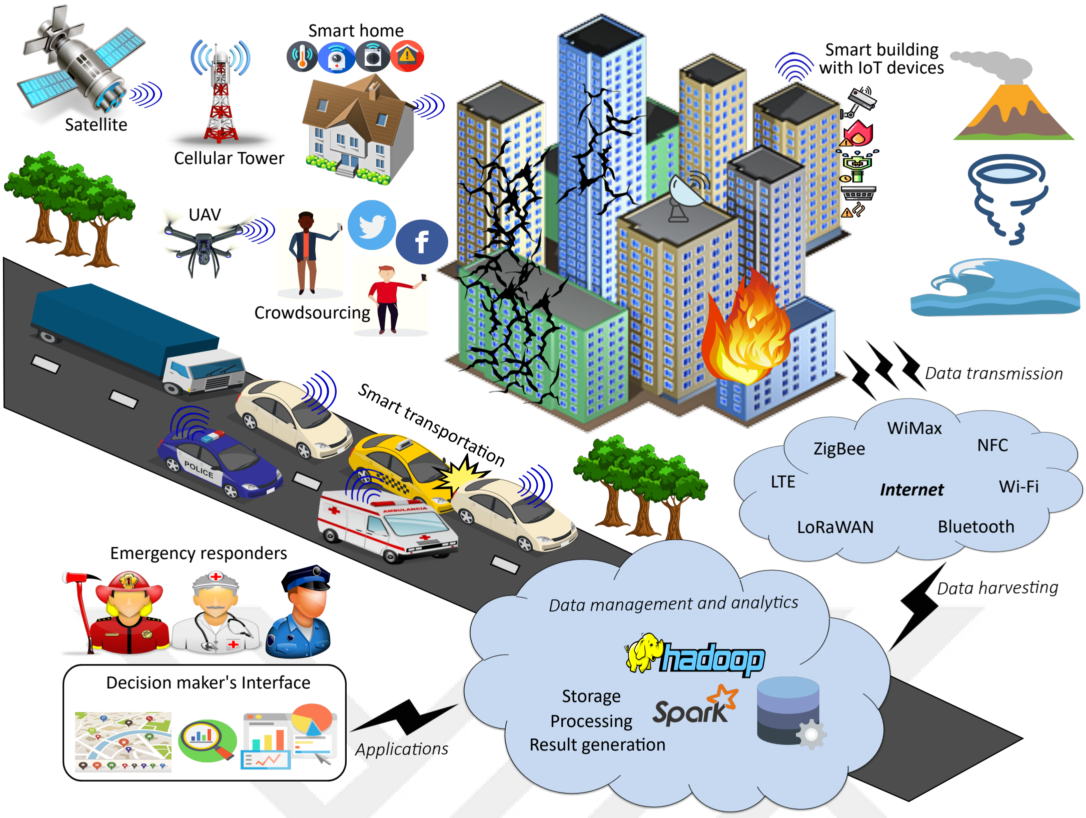


**Figure 2.1** : General illustration of BDA- and IoT-based disaster management environment.

Therefore, we argue that the joint exploitation of BDA techniques and IoT technologies can lead to the development of an innovative, effective and highly-needed disaster management environment. A general illustration of BDA- and IoT-based disaster management environment is demonstrated in Figure 2.1.

## 2.3 Disaster Management and the need for BDA and IoT

In order to understand the uprising role of BDA and IoT in disaster management, it is important to have a clear image of disaster management systems and its operations. In this section, we will first describe the disaster management systems and its applications and requirements. Then we will discuss the benefits that the collaboration of BDA and IoT offers for disaster management and also identify some of its requirements.

### 2.3.1 Disaster Management Systems

Disaster Management can be defined as a systematic approach that involves planning and managing the disaster mitigation, rescue, response and recovery through the collaboration of federal, state, local and private sector entities. The general concept of disaster management can be viewed as a combination of many interrelated processes that aims at providing efficient means to understand, analyze, monitor and predict disaster occurrences. With the rapid advancement in Information and Communication Technology (ICT) from the last two decades, it is now possible to initiate a quick response to any disaster situation in reasonable time and budget.

**Table 2.1** : Main DMS applications and requirements.

| Disaster Status | DMS Applications | DMS Requirements |
|---|---|---|
| Pre-disaster | Disaster Prediction<br>Early Warning<br>Simulation Exercises | Reliability<br>Availability<br>Maintainability<br>Accuracy<br>Usability |
| Post-disaster | Evacuation<br>Rescue Assistance<br>Monitoring / Surveillance<br>Logistics Management | |

Disaster Management System (DMS) is a type of information system that assists the decision makers and responders in acquiring, managing and utilizing the disaster information for timely and effective disaster management. The main components of DMS can be divided into data integration, data mining, and multi-criteria decision-making [43]. DMS can be regarded as highly integrated and complex systems that require application specific design and maintenance. Currently, due to the involvement of various interlinked data nodes and with large scale of data requiring real-time analytics, the designing and implementation of a DMS becomes a multidimensional and complex problem. Disaster management applications can be categorized into pre-disaster and post-disaster phases since they deliver diverse functionalities with different requirements for response time, accuracy and effectiveness. Pre-disaster applications such as disaster prediction, early warning system, and simulation exercises etc., focus on measured and inclusive data analysis. On the other hand, post-disaster applications such as Evacuation, Rescue Operations and Monitoring etc., require spontaneous and accurate results. However, each application of DMSs should support heterogeneous and distributed data sources and allow decision makers to extract useful knowledge in an interactive manner. DMSs must possess the desirable technical factors such as reliability, availability, maintainability, accuracy and usability requirements [44]. As categorized in Table 2.1 each DMS application needs to satisfy the requirements. Through ensuring these requirements the developers can set benchmark quality attributes to verify the performance and measure the effectiveness of the DMS.

### 2.3.2 BDA- and IoT- based disaster management environments

Disaster management systems requires to be shifting to state-of-the-art environments that are supporting multiple data sources and are equipped with latest technologies offering broader range of capabilities for enhanced connectivity, storage, real-time analytics and cost-effective applications. These environments can be successful deployed by indulging BDA and IoT technologies together for disaster related operations. Figure 2.2 presents the benefits that can be achieved through the combination of BDA and IoT for disaster management systems and also identifies the main requirements for deploying a BDA- and IoT-based disaster management environment.

#### 2.3.2.1 Benefits

BDA- and IoT-based disaster management environments can provide a number of benefits within the scope of disaster management. Some of the key benefits are described in the following subsections.

**Connectivity:** Connectivity is required to facilitate the aggregation of huge volumes of data from heterogeneous data sources to high-performance computing infrastructures and further sharing of information with concerned disaster management authorities. Due to the availability of various communication technologies, one of the key benefits of BDA- and IoT-based disaster management environment is to provide reliable connectivity. Connectivity among the interlinked data nodes and DMS acts as the backbone for the insurance of successful operations. As a number of communication technologies are available the overall environment architecture has to be flexible to deal with different communication protocols including local and remote communications [45]. Moreover, with the evolution in post-disaster communication networks, seamless connectivity is provided even with the distraction of other conventional communication networks in post-disaster situations.

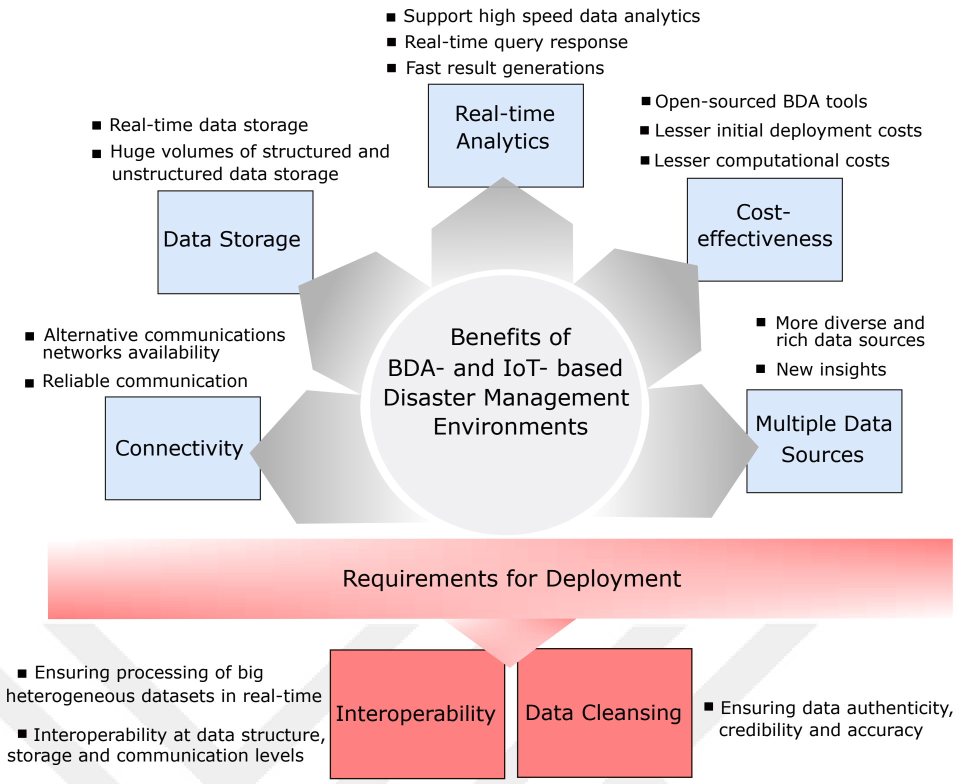


**Figure 2.2** : Benefits and requirements of BDA- and IoT-based disaster management environments.

**Data Storage:** Storage of huge volumes of heterogeneous data in real-time can be challenging in conventional DMSs. With BDA technologies such as Hadoop, large-sized structured or unstructured datasets can effectively be stored on low-cost commodity hardware. Real-time environments having streaming storage capability for IoT devices and other data sources can enhance the entire data processing efficiency and can provide a number of benefits to the designated applications [46]. Moreover, BDA technologies can enable efficient processing with low latency for data analytics while maintaining the storage of massive unstructured datasets.

**Real-time Analytics:** Due to the dynamic and demanding nature of disaster management, real-time analytics is one of the key requirement for current disaster management environments. Connectivity among various data sources results in massive data generation at high speed that can create hurdles in performing real-time analytics. A dedicated technological platform with the software solution capability to perform real-time processing, streaming and in-memory computing is needed to deal with such enormous and high-velocity data [47]. The ability provided by BDA to perform fast analytics with real-time queries is vital to help decision makers obtain required results for an effective emergency response.

**Cost-effectiveness:** BDA tools are mostly open-sourced and it offers a huge cost reduction opportunity as compared to buying proprietary data processing software solutions for disaster management operations. Cost-effectiveness is an important factor for disaster-concerned authorities in developing countries, where disaster management systems are not deployed due to lack of funds. Map-reduce is an ideal solution for cost-effective data storage and useful for decreasing the computational costs of the overall system [48]. Moreover, with the declining costs of hardware and software utilities of IoT deployments, state-of-the-art technologies can be deployed easily with much lesser budget.

**Multiple Data Sources:** In the context of integrating IoT environments equipped with multiple data sources such as cameras, sensors, smartphones etc., with BDA technologies assisting in data processing, a number of data sources can be incorporated to gather new and valuable insights and information. Engaging multiple data sources provide alternative ways to address problems that require multidimensional representations of the data to extract the common patterns for a solution that are inaccessible through a single source of data [49]. With the availability of diverse and rich data sources, BDA- and IoT-based disaster management environments can surpass conventional DMSs data sources.

#### 2.3.2.2 Requirements

The key requirements for deploying BDA- and IoT-based disaster management environments are described in the following subsections.

**Interoperability:** The capability of being able to link, combine and process two or more datasets is known as interoperability. The collected datasets from heterogeneous sources might not align with each other, or it can be difficult to determine the possible relationships among them. During real-time data harvesting and integration, it is important and challenging at the same time to achieve the maximum level of interoperability. Interoperability can be ensured at technical, syntactic, semantic and pragmatic levels [50]. Hence, good practice can be to apply interoperability checks at the data structure, storage, and communication levels through abstraction and virtualization to ensure high reliability.

**Data Cleansing:** Data cleansing is essential for disaster management, as incomplete, error-prone and ambiguous data can lead to more problems and wastage of precious time. Data cleansing parameters determine the accuracy of the analysis carried out on a particular dataset. However, as data cleansing works on a complex relationship model and can require extra computation power and processing time, a balance should be kept between the data cleansing model and the accuracy improvement of the analysis [51]. Moreover, with the growing usage of social media data for disaster management processes, a different kind of unstructured data is emerging that needs to be checked for authenticity, credibility, and accuracy.

## 2.4 Recent Advances

The research on BDA and IoT in the domain of disaster management is still in its infancy. This section reviews the recent research contributions with the aim to identify the key research areas and highlight the latest advancements recognized to enhance the disaster management related processes.

### 2.4.1 BDA for Disaster Management

Big data analytics provides a variety of solutions on huge multi-sourced datasets collected from the disaster area to uncover hidden patterns and understand the situations on the ground so that rescue activities can be carried out effectively and logistics can be managed optimally. One of the main advantages of using BDA is that it enables data scientists to analyze huge volumes of data involving different data sources that may not be collected using traditional tools [52]. BDA depends on various technologies and tools for the execution of huge volumes of structured, semi-structured and unstructured data for analytical processes. Research trends in BDA for disaster management focus on both the content/text and the spatial points of view of the data for analysis and result generations [41].

**Table 2.2** : Comparison between recent BDA-based studies focusing on disaster management.

| [htbp] **Study** | **Year** | **BDA Tools** | **Data Source** | **Text Analytics** | **Spatial Analytics** | **Focus** |
|---|---|---|---|---|---|---|
| [53] | 2018 | Spark | Crowdsourced sensor data | ✓ | ✓ | Improving near-real-time application for flood risk management |
| [54] | 2018 | Kafka, Spark | Twitter | ✓ | ✓ | Providing disaster situational-awareness |
| [55] | 2018 | Spark | Historical database of meteorological data center | ✓ | ✓ | Simulation for Typhoon risk assessment |
| [56] | 2018 | Spark | Historical database of earthquake catalogs | ✓ | ✓ | Earthquake magnitude prediction |
| [57] | 2018 | Kafka, Spark | Mobile communication base station data | - | ✓ | Identifying earthquake emergency through high precision heat map |
| [58] | 2017 | Hadoop, Spark | Satellite Imagery, Sensor data | - | ✓ | Fire response optimization and evacuation planning |
| [59] | 2017 | Hadoop | Social media, Remote sensing, Wikipedia data | ✓ | ✓ | Enhancing disaster coordination and assistance for relief operations |

Despite the limited publications regarding BDA for disaster management, some of the recent research as compared in Table 2.2 shows that a variety of data sources are being utilized with various open-source BDA tools within the scope of disaster management. For instance, for flood risk management an interoperable mechanism was designed by authors in [53] to integrate heterogeneous sensors that enable access and filtering of the data in near-real-time using Spark. The approach used in their study offers a method to enhance near-real-time applications using heterogeneous data streams i.e., crowdsourced and sensor data. In another study [54], a big data crisis mapping system was designed that is able to collect and analyze Twitter data utilizing Kafka and Spark. The system extracts information related to the disaster from the collected geo-tagged tweets by applying classification technique and semantic annotators. This information is then visualized on a web-based dashboard for emergency responders to acquire greater situational awareness in the early stages of the disaster. The authors in [55] specified that Spark-based computation on huge sets of historical data provides better performance for the simulation to identify typhoon risk assessment feasibility. Similarly, in another study [56], the authors used several regression algorithms using Spark to analyze large catalog of earthquake events. They demonstrated very promising results regarding the prediction of earthquake magnitudes in the state of California. In [57], the authors proposed a real-time collection and classification algorithm of mobile phone position data by stream processing environments such as Kafka and Spark to produce a high precision heat map of the population affected by the earthquake. An integrated disaster management system developed through the combination of Hadoop and Spark was presented in [58]. Their proposed system addresses large-scale datasets issues of spatial and temporal perspectives and provides predictive risk analytics for fire response's resource optimization and evacuation planning. A study was conducted [59] to demonstrate a framework that synthesizes multi-sourced data such as social media, remote sensing and Wikipedia to build a flexible solution that provides historical and future disaster analysis involving Hadoop for spatial data mining and text mining.

### 2.4.2 From WSN to IoT for Disaster Management

Wireless Sensor Networks (WSNs) consists of autonomous low-powered sensors nodes that are spread across a specific area and capable of measuring and reporting of environmental conditions (i.e., smoke, temperature, vibration, locations). WSNs have long been used in disaster monitoring related research, such as event monitoring in emergency scenarios [60], natural disaster monitoring [61], and multi-agent system-based disaster management [62]. However, unaided WSNs lack in a multitude of social, technical and economic perspectives for extensive deployment in disaster management [63]. WSN is an integral part of IoT and can benefit from the data management, processing and decision-making characteristics of IoT to provide meaningful interpretations and supporting decisions based on its generated sensed data. From the last few years, research interest in many domains including disaster management is diverted to IoT, as it is predicted that by 2020, IoT will be interconnecting nearly 50 billion new connections [64]. IoT provides a resourceful platform, consisting of various tools and technologies that are supported by communications among various physical and virtual entities to observe, communicate and process data. IoT provides an ideal solution for data gathering in disaster-struck areas, as it offers alternative means of communication carried on low battery-powered and IoT-enabled wireless devices.

Recent research on disaster management is widely considering IoT to provide multi-dimensional and multi-sourced information for timely decision-making. IoT can be effective solution for disaster event detection. IoT offers smart aggregation, integration, and analysis of multi-dimensional and multi-sourced data, which are the main steps for situational awareness for effective decision-making. In a study [65] the authors demonstrated how IoT with semantic web technologies can be successfully deployed for earthquake-related event detection. The proposed system was able to semantically annotate streams that were retrieved from web services gathering IoT-based sensors data for effective earthquake event detection. Another system based on IoT [66] focused on the quick and systematic evacuation of large crowds of people after disasters. Crowd lives oriented track and help optimization system (CLOTHO) aims at reducing the loss of lives by deploying an IoT-based solution that uses a mobile cloud computing platform. The data collection part of the system includes the mobile terminal that is backed by IoT while the storage and data analytics part comprises of a cloud-backed system. Dhafer, et al. [67] proposed an emergency and disaster relief system which is monitored by a cloud-based IoT platform. The system is known as Critical and Rescue Operations using Wearable Wireless sensors networks ($CROW^2$) and integrates heterogeneous wireless devices such as smartphones and sensors with various communication technologies such as WiFi and Bluetooth to support end-to-end network connectivity. This system helps emergency rescuers to be connected with any functioning network or the internet.

### 2.4.3 Post-Disaster Communication Networks

Most of the conventional communication infrastructures get unresponsive in post-disaster scenarios, either due to physical damage or overloaded network congestion. Recent advancements in wireless communication technologies have a lot to offer for post-disaster communications with rapidly deployable, scalable and efficient networks that can ensure the flow of data and provide communication assistance for rescue and response operations.

Device-to-Device (D2D) communication offers an improved Quality of Service (QoS) and high Quality of Experience (QoE) for User Equipments (UEs) to manage radio spectrum and most importantly energy consumption of the devices in the disaster-affected area. A disaster communication architecture based on D2D communication was proposed in [68]. The study has focused on extending the lifetime of energy-constrained networks by employing energy harvesting techniques from radio frequency signals via the base station at the user equipment relay. Similarly, another study [69] focuses on cooperative D2D protocol to ensure smooth connection and expand the average battery life of the devices. The protocol was designed to assist low battery level devices to find neighboring devices having high battery levels so that they can act as relay. This mechanism is aimed at extending the communications for covering the disaster area. In [70], a framework named FINDER (Finding Isolated Nodes using D2D for Emergency Response) was proposed to locate and link the disconnected mobile devices in the disaster area. The study uses a multi-hop D2D communication derived from Ant Colony optimization to improve the message deliver probability and to extend the network lifetime and energy efficiency of the devices.

A MANET can be defined as a temporary distributed network that comprises of a set of mobile nodes with infrastructure less, decentralized and dynamic features. MANETs can provide a practical solution for post-disaster communications. The researchers in [71] reviewed the mobility models, routing algorithms and network simulators for MANETs in disaster scenarios. For post-disaster scenarios, the authors in [72], introduced new schemes for MANETs routing and gateway load balancing. This novel scheme aims at improving communications in affected areas by reducing network congestion. A novel framework named disruption tolerant secure opportunistic routing (DTSOR) was proposed in [73] that ensures smooth and secure communication between high mobility devices during emergency situations. Through performance analysis and simulations, the study claimed that the proposed framework in terms of the packet delivery ratio, network overhead and throughput overtakes many modern data transfer approaches. Another study [74] proposed the concept of hybrid cellular-MANET architecture using available cellular base stations in post-disaster situations. The proposed architecture is responsive to device mobility and possesses the self-organizing feature of MANET.

UAVs provide an open opportunity for quick and easy deployment of cellular base stations as secondary communication infrastructure where required in post-disaster scenarios. For distributing tactical and sensor data over a specific area or connecting on ground devices within range, the data link system of UAVs can be programmed effectively with additional broadcasting jobs. The authors in [75] reviews the latest advancements in UAVs for network-assisted post-disaster management. They identified the key issues and suitable network architectures for UAVs assisted network for disaster management. A study was conducted [76] to investigate the use of UAVs as Aerial Base Stations (ABSs) for disaster communications in a situation where conventional communication infrastructures have totally failed. The study analyzed communication improvements obtained by the ABSs through simulations and found a noticeable increase in effective communication probability when ABSs were deployed in optimal locations. A flying ad-hoc network, named the Flying Real-Time Network (FRTN), was proposed in [77]. The feasibility of this proposed network to provide communication in post-disaster scenarios was presented by illustrating the real-time scheduling of message delivery and simulation-based analysis.

### 2.4.4 Crowdsourcing

Crowdsourcing is boosting the idea of "people as sensors", a concept recently being recognized in the disaster management domain for incorporating new and big datasets that can be processed for retrieving required information with more insights. Crowdsourcing can either be active, where people willingly participate to provide data; or passive, where typically social media platforms are used to collect the data with or without the contributor's knowledge. Active crowdsourcing platforms are deployed by concerned authorities (Government or NGOs/CSOs) to acquire real-time information from disaster-affected people, for improving emergency response and resource allocations. There are a number of web-based and mobile applications for enabling active crowdsourcing in a disaster-affected area. One such platform that facilitates real-time, multimedia supported and collaborative mapping is Ushahidi [78]. Ushahidi platform has been extensively utilized in disasters such as the 2010 Haiti earthquake and the 2011 Japan tsunami. Active crowdsourcing provides more credible data with less noise as compared to data collected from passive crowdsourcing that require different data quality filtrations. However, most of the recent research efforts are focused on passive crowdsourcing for disaster management; as big volumes of data are generated with people tending to report a status/tweet/description, image, video/audio and most importantly precise locations using various social media platforms. This massive data contains critical information (text, image or video), sentiments, personal opinions, and GPS coordinates. Effectively analyzing such data can provide a better situational awareness and enhanced assistance for rescue and response. Moreover, social media offers a suitable solution for establishing communications with affected people, acquiring feedbacks and enhancing empowerment between people and concerned authorities.

Recently social media is one of the most emerging big data source for disaster management research. Particularly, with the increasing use of smartphones in the last few years, geospatial data generated from social media platforms are more in demand over conventional data sources for disaster management [10]. The concept of Volunteered Geographical Information (VGI) [79] is being widely used for disaster management, as citizen engagement in disaster response is increasing. Kusumo et al. [80] examined the benefits of using VGI for spatially planning the evacuation shelters. They used Jakarta floods as a case study and their analysis showed that 35.6% of the shelter locations desired by residents matched with the locations of the government evacuation shelters. In another study [81], VGI extracted from social media was used for real-time rainfall and flooding events detection through user-generated high-quality eyewitnesses in shape of texts and photos by applying deep learning approaches. Flood events in various cities such as Paris, London and Berlin were targeted as case studies and analysis was performed through spatio-temporal clustering and visualization techniques enabled by a web map application.

As mentioned earlier, the reliability of passive crowdsourced data has been difficult to evaluate. Limited research is available until now in terms of quality assessment methods on the data produced by social media platforms. One such study [82] on the credibility assessment of users reporting about various disasters, compared the user profiles and their geographic references, with the classification of tweets through Naive Bayes models. The datasets of this study were extracted from past earthquake events in Myanmar and Italy. The study found similar geographic granularity and identified 88 to 99% precision of information contained in the collected Tweets. The need to effectively extract meaningful information from huge sets of data generated by social media platforms in a lesser amount of time for effective disaster management is an emerging issue. BDA seems to be the choice in the recent research efforts to deal with such issues. The authors in [83] used Hadoop platform and machine learning techniques to perform sentiment analysis on big social data. Support vector machine algorithm was used for the sentiment classifications and an interactive visualization mechanism was deployed to provide information for prompt decision-making.

## 2.5 Taxonomy of BDA- and IoT-based Disaster Management

In this section, we present the thematic taxonomy of BDA- and IoT-based disaster management. The taxonomy identifies and categorizes key attributes essential for the development of BDA- and IoT-based disaster management environment. For the development of this taxonomy, we followed an iterative approach as suggested by

Nickerson et al. [84]. Due to the involvement of multi-disciplinary topics i.e., Big Data Analytics, Internet of Things and Disaster Management, we had to consider different dimensions and characteristics that are important to identify and valuable for the development of BDA-and IoT-based disaster management environments. The development of this taxonomy was a continuous process, involving refinement at various stages to sufficiently satisfy the qualitative attributes of being, (a) concise (a limited number of dimensions that are important, because extensive classifications are difficult to understand), (b) comprehensive (includes all main dimensions of objects of interest), (c) extendable (open to include new dimensions), (d) explanatory (provides valuable descriptions of the nature of the objects under study) [84]. This thematic taxonomy was classified by conducting an extensive and inclusive review of the related literature, with the aim to unearth the main attributes of BDA- and IoT-based disaster management environments. In this effort to capture the vastness and variety of multi-disciplinary topics involved, we identified some of the key attributes on the bases of their significance, consideration and association with BDA- and IoT-based disaster management environments. This taxonomy can provide guidance to researchers to understand the foundations for the development of such environments and future acquisitions. As illustrated in Figure 2.3, the taxonomy at the top level is categorized into the following seven attributes.

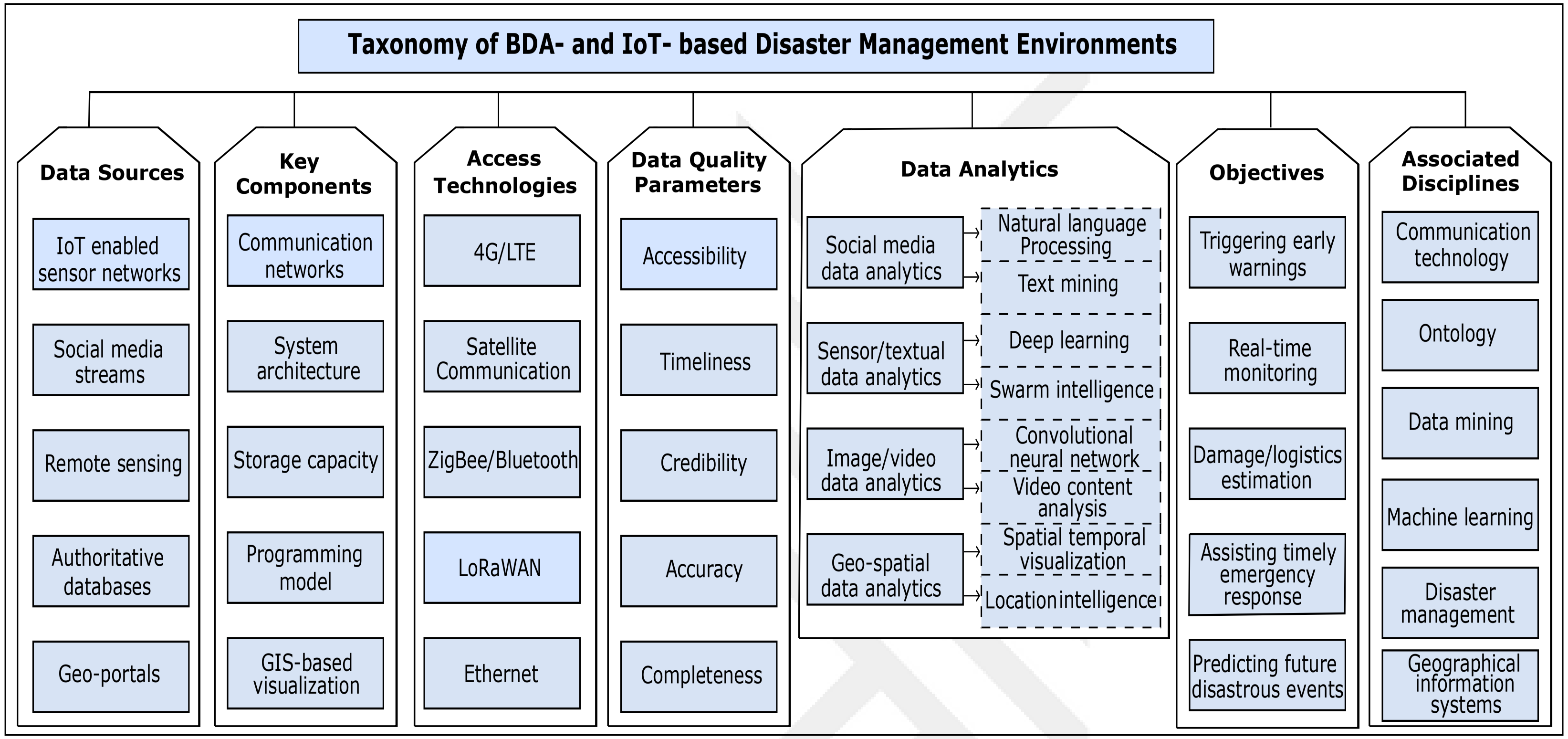


**Figure 2.3** : Taxonomy of BDA- and IoT based disaster management environments.

### 2.5.1 Data Sources

The key characteristic of BDA- and IoT-based disaster management environment is its diverse and rich data sources. Table 2.3 presents the details of potential data sources for disaster management. The main potential data sources include social media streams, the integrated networks of IoT enabled sensors, remote sensing, authoritative or public historical databases and geo-portals. The data generated from these sources is of diverse descriptive nature (i.e., location, temperature, humidity, orientation, event description, image, audio/video etc.) and hence involves different data formats. Moreover, most of the data captured are unstructured and require some pre-processing techniques prior to any kind of analytics. It is very important to understand the significance of disaster-related data, that needs to be accessible, accurate and complete, and also support real-time processing. The main challenges associated with disaster-related data are:

- Identifying and aggregating disaster-related heterogeneous data from IoT/big data infrastructure.
- Extracting useful information from huge volumes of collected heterogeneous and unstructured data, that requires data pre-processing and event detection techniques.
- Interpreting and visualizing data in near real-time.

**Table 2.3** : Summary of data sources.

| Data Source Category | Potential Data Sources | Available Information | Possible Data Formats | Data Structure | Data Processing Type |
|---|---|---|---|---|---|
| Social media | Twitter, Facebook, YouTube | Text, Image, Audio/Video | JSON, CSV, JPEG, MPEG-2 | Unstructured | Streaming |
| IoT-enabled sensors | Surveillance cameras, Seismometer, Thermometer, Hygrometer | Status data, Location-based readings | JSON, GeoJSON, XML, CSV, MPEG-2 | Semi-structured | Real/near real time |
| Remote sensing | Satellite imagery, UAV imagery | Aerial imagery, GPS measurements | IMG, GRID, GeoTIFF | Semi-structured | Batch |
| Historical data warehouses | Government, Humanitarian/ NGO's Databases | Demographic, Health records, Recorded maps, Digital archives, Surveys | CSV, GFF, XML, JSON | Structured | Batch |
| Geo-portals | National Geo-Portals, Open Global Geo-Portals | Spatial Data | KML, SHP, GeoJSON, TIN, OSM | Structured | Batch |

### 2.5.2 Key Components

The realization of a BDA- and IoT-based disaster management environment depends on the availability of some critical components. Communication networks are one of the key components and act as a backbone in developing the environment. A combination of various communication networks and protocols provide the overall network infrastructure for data transmission and facilitate connectivity among numerous data sources. Disaster information networks should be assembled by combining various wired, wireless and satellite network so that a “never-die-network" can be ensured for both normal and disaster occurrence cases [85]. System architecture provides the blueprint that determines the overall structure and behavior of a system. A well-designed system architecture is a cornerstone to tackle the conceptual and practical issues that can be faced with a complex system involving big data and IoT. A pre-planned conceptual model provides well-thought-out solutions for the successful integration of heterogeneous components for an accurate and effective disaster management application. Due to data acquisition from heterogeneous sources at a rapid rate, the need for effective data storage and management of these huge datasets is obligatory, while ensuring availability and reliability at the same time. The main challenge is to differentiate and store large-sized data (i.e., images and videos) accessed in the real-time, from the small-sized data (i.e., log and text files) accessed in batches, acquired from sensors and static databases. Programming model represents the core characteristics of any big data framework and plays an important role in determining the performance of big data processing engines. It is important to select a programming model that functions in real-time with high performance and reliability. There are various programming models currently available, i.e. MapReduce, SQL-based, functional and statistical models having different advantages and applications. In any disastrous situation, emergency responders and decision makers require quick and accurate location-based descriptions, suggestions and predictions easy to understand and interact. GIS-based visualization tools provide a user-friendly and interactive interface for mapping datasets that can demonstrate the overall picture and offer new insights to the decision makers.

### 2.5.3 Access Technologies

A reliable, robust, energy efficient and disaster resilient data transmission network acts as the backbone for any disaster management system. Access technologies from a disaster communication perspective should provide reliable connectivity and optimized services for effective data transmission between data generating devices and back-end servers.It is very important to ensure the flow of data and the safety and connectivity of the network in order to acquire situational awareness in case of a disaster event [86]. A collection of various communication network topologies is required to obtain an autonomous BDA- and IoT-based disaster management environment. Some of the main access technologies that are useful for disaster communication are 4G/LTE, satellite communication, ZigBee, Bluetooth, LoRaWan and Ethernet. LTE (Long Term Evolution, also called 4G) provides communications with wide area mobility, improved interactivity and on the go multimedia services. 4G/LTE technologies are widely used by major telecom operators globally. With its high speed and low latency features users are able to operate applications such as social networks, maps navigation, browsing, etc. in addition to traditional voice calls and SMS services. These cellular mobile communications with its wide access to the people can be utilized for early warnings and disaster alerts. Satellite communications are not vulnerable to damage from disasters, which make them the reliable communication infrastructure in full-fledged disasters. However, the main concerns are the cost of satellite bandwidth, low throughput and large latency. Nevertheless, satellite communications can be a cost-effective solution for sever disaster than establishing a new communications infrastructure in disastrous areas. Short-range wireless technologies such as ZigBee and Bluetooth can be effective in establishing communication networks within a small disaster-affected area. LoRaWAN is emerging as the new communication technology for smart city applications. LoRaWAN ensures interoperability between various operators and offers low-power and low-cost mobile communications that can be beneficial for disaster communications. The importance of wired communication technologies (i.e., Ethernet, PSTN) cannot be neglected in disaster communication networks. High-speed communications can be achieved with dedicated fiber-based connection lines to enable transmission of data within the Local Area Networks (LAN) of various disaster management authorities.

### 2.5.4 Data Quality Parameters

Incomplete, ambiguous, error-prone and noisy data can cause serious issues in data analytics and hence in decision making for disaster response. Data quality parameters determine the accuracy and productivity of the analysis performed on a particular dataset. Data quality dimensions such accessibility, timeliness, credibility, accuracy and completeness are vital for disaster management processes. Accessibility determines the mode in which the data is accessed from the source and whether the data has any legal constraints on usage. Timeliness describes the movement of data, i.e., real-time or static and whether the data needs to be updated. Credibility is to ensure that the data is verified and its source is identified. Accuracy is to check whether the data is free of any redundancy and is explicitly related to the scenario. Completeness determines the clarity and understandability of the data according to the situation.

### 2.5.5 Data Analytics

The state-of-the-art big data analytical tools are one of the key technologies that assist the concept and operation of the BDA- and IoT-based disaster management environments. The heterogeneous data sources, producing huge volumes of multi-dimensional and multi-modal data requires powerful data analytics for productive execution. To develop an efficient and real-time data execution enabled system for disaster management processes various big data analytical tools need to be employed. A combination of advanced big data analytical tools, i.e., Hadoop Ecosystem and Spark can be utilized to analyze huge sets of data accurately and efficiently with suitable algorithms and techniques. Data analytics varies according to the data types captured from the heterogeneous data sources and the desired results. Following are some data analytics types and prescribed methods that can provide new insights and quick results for effective rescue and response based decision-making.

#### 2.5.5.1 Social media data analytics

With the extensive use of social media applications, users in real time generate huge amounts of unstructured but potentially useful datasets. These datasets need to be checked for reliability, credibility, and authenticity prior to any kind of analytics aimed at extracting actionable information. Sequenced information processing operations i.e., filtering, categorizing, extracting and summarizing can be the best approach to deal with these issues [87]. Natural Language Processing (NLP) techniques can be used to search, classify, and compile textual descriptions acquired from social media user in a disaster response scenario [88]. Text mining is another useful analytic technique to extract valuable structured data from huge volumes of unstructured text. Social media datasets can be evaluated with a number of standard text mining techniques to collect the required information about a specific disaster [89]. Text mining basically regulates semantics, keywords, labels, tags, and themes in the shape of separate files and formats for extracting key pieces of information.

#### 2.5.5.2 Sensor/textual data analytics

Another data source generating huge volumes of data is IoT-based sensors. These big sensed datasets play a vital role for making spontaneous and effective decisions for disaster rescue and response. The function specific and geographically distributed sensors can provide valuable information and insights through powerful analytics. Deep learning algorithms operate on hierarchical learning process to extract high-level and complex abstractions as data representations. Deep learning is an important big data analytics tool as it effectively analyses huge amounts of unsupervised data even being unlabeled [90]. IoT-based sensors are complex to manage and aggregating data is hard usually due to the lack of decentralized control. Swarm intelligence can be useful to resolve complex issues with IoT-based sensor systems having dynamic properties and limited computation power [91].

#### 2.5.5.3 Image/video data analytics

The real-time streams of high-quality images and video content, from surveillance cameras, UAVs and citizens with mobile devices are providing decision-relevant situational information on causalities and damaged buildings, roads, bridges, etc. With the advances in machine learning and vision techniques for analyzing image/video datasets, rescue operations, planning evacuation routes, damaged infrastructure surveys, and other disaster management activities can be greatly assisted. Convolutional neural network (CNN) is a class of deep neural networks, commonly used to extract topological properties from visual imagery. It is simple and robust to operate, as it automatically learns visual feature sets from the training data. A study [92] on investigating the potentials of CNN for aerial imagery demonstrated that CNN are useful for object detection and correctly locates the areas that match to categories in which the CNN was trained for. Moreover, [93] used CNN on video analysis for fire detection and concluded that CNN achieves better classification performance than some of other conventional methods for fire detection. Video content analysis (VCA) enables automatic video analyzing to search, identify, classify, and determine temporal and spatial events. Using video content analysis, [94] proposed a warning system for flood event detection on feeds from surveillance cameras.

#### 2.5.5.4 Geo-spatial data analytics

Geo-spatial data or data with location component is considered as the most essential input element in latest technologies. The geo-spatial data-sets needs to be analyzed to gain information about disaster locations as it occurs, identify the area and people that require urgent assistance and locate appropriate areas for shelters to name the least. With the advent of satellite remote sensing, location-based sensors and smartphones equipped with GPS, a huge volume of geo-spatial data is generated. Spatial temporal data visualization comprises of powerful tools that supports analysis of geo-spatial data over time through interactive visualization. Spatial-temporal data visualization greatly assists decision-making in all the phases of disaster management [58]. Location intelligence offers unique insights, reveal hidden patterns and information based on geo-spatial data for better decision-making. Location intelligence is effectively used to detect the spatial and temporal distribution of flood risks [95] and for waste collection solution to improve cities management systems [96].

### 2.5.6 Objectives

The convergence of BDA and IoT technologies can set a new meaning to the overall objectives of disaster management. One of the main objective of this system is early warning generation, that can save lives and reduce infrastructure damage. Real-time disaster monitoring involves the extraction of information from the system to make informed and timely decisions. It is important for the system to accurately estimate the damages caused and logistics required. Moreover, it should figure out the evacuation routes quickly in emergency response. Effective and timely decision-making needs a reliable, fast processing and data resourceful system that integrates different state-of-the-art technologies to improve its operations. Predicting future disastrous events is becoming a reality with the evolution in the latest technologies, such as low-powered sensor networks, reliable wireless technology, sophisticated algorithms and advanced data analytics.

### 2.5.7 Associated Disciplines

It is important to identify and merge the concepts of all the related disciplines that are used to design, develop and manage BDA- and IoT-based disaster management environment. The general perception that BDA and IoT-based environments only requires technical skill is wrong, as interdisciplinary approaches are required in their domain-specific applications. Professionals having a specific set of skills and experience of communication technology, data mining, machine learning, ontology, disaster management and geographical information systems need to collaborate in designing an operational architecture that can fulfill the objectives of BDA- and IoT-based disaster environments.

## 2.6 Reference Model

As discussed, the integration of BDA and IoT technologies can provide a resourceful platform for acquiring, storing, processing big disaster-related data and generating the required results for timely and accurate decision-making. To effectively utilize the value-added capabilities and opportunities offered by BDA and IoT within the scope of disaster management, we introduce a novel reference model derived from the classified taxonomy and related literature. Based on the identification and abstraction of correlated technical and theoretical knowledge, this novel reference model presents the overall functionality and configuration for disaster management environments. The main theme of this reference model is to provide guidelines for developers to ensure effective decision-making through such disaster management environments. Multiple IoT based BDA architectures focusing on general applications are found in the literature [97] [98] [99] [100]. Most of these architectures are focusing on overall operations in a smart city concept and there is a lack of disaster management specific

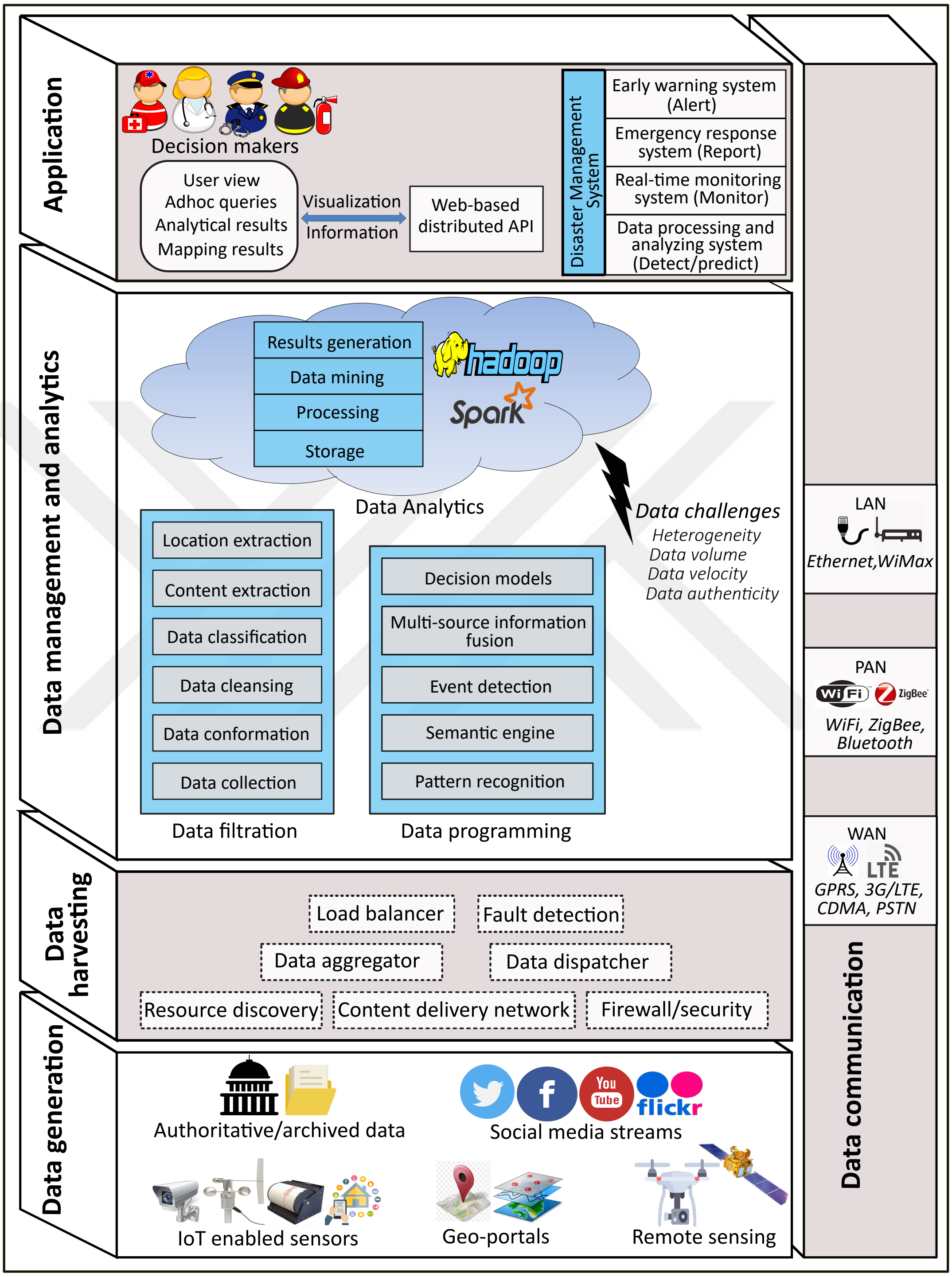


**Figure 2.4** : The proposed reference model for BDA- and IoT-based disaster management environments.

architectural models in the existing literature. Feasibility of defining a standardized framework that deploys IoT and BDA for time critical and performance demanding application like disaster management is far from reality. However, it is theoretically feasible to direct the designing process in this new and dynamic environment towards the deployment of a realistic architecture.

For designing the proposed reference model, this study has adopted design science research method [101] to present the high-level model for an environment where the integration of potential big data sources and operations of various tools and techniques can ensure effective disaster management. This study followed the seven design science research guidelines i.e. "*Design as an Artifact, Problem Relevance, Design Evaluation, Research Contribution, Research Rigor, Design as a Search Process, and Communication of Research*" specified by Hevner AR, et al [102]. The proposed reference model is an artifact that utilizes big data analytics and IoT for effective disaster management (*Design as an Artifact*). The model supports processing huge sets of heterogenous data (structured/unstructured) in real time that is highly demanded by current disaster management systems (*Problem Relevance*). The implementation of different components in the proposed model can be justified by numerous performance measures published during actual deployments (*Design Evaluation*). The innovative design that assembles various data sources and the integration of state-of-the-art artifacts for effectively utilizing the benefits that can be gained through BDA in real-time for disaster management is the key contribution of this study. Moreover, this research highlights new challenges and parameters when deploying IoT and BDA in disaster management (*Research Contribution*). This study relies on a systematic literature review on the advanced topics of IoT and BDA for disaster management. The formation and assessment of the artifacts are established from the recognized knowledge base from multiple academic fields (*Research Rigor*). Critical feedback and continuous literature study were carried out throughout the design of the framework, which led to many iterations and modifications (*Design as a Search Process*). Involving both the linked academic community and related field professionals to highlight any defects in the final design resulted in more improvements (*Communication of Research*).

Moreover, after thorough analysis of several related architectures, we found out that the following key points need to be considered during the design process of any disaster management environment that is involving BDA and IoT technologies.

- The architecture needs to be scalable to indulge new data sources that can provide valuable information and insights.
- Flawless communication over the network or alternative networks in case of any transmission failure or destruction.
- Effective storage of structured and unstructured data that are either collected through real-time streams or historical data batches.
- Flexible to accommodate various computation intelligence techniques, algorithms and analytical packages.
- Able to share the results to other systems or applications and present the information in an interactive manner to the decision-makers.

The design of some disaster management environments may vary depending on its application scope and size (i.e., industrial/building disaster management vs urban disaster management) and the nature of required results based on urgency, performance, compatibility and scalability. However, this reference model can provide a standardized framework for considering and assembling the overall disaster management system entities involving many BDA and IoT entities. As shown in Figure 2.4 the model supports multi-sourced data that is enabled by the cutting-edge BDA and IoT technologies and techniques. The model consists of five layers, which are briefly discussed in the subsequent sections.

### 2.6.1 Data Generation

The data generation layer consists of all the potential data sources that are useful for developing situational awareness and providing new insights for the incident. Apart from traditional disaster management data sources (i.e., field survey, GIS-based data) a massive amount of valuable data is generated by human and physical sensing resources that can be utilized for disaster management processes to enhance operations and gain new insights. IoT-based devices provide factual data while crowd-sourced data from social media streams provides real-time information but in an unstructured format. Remote sensing data are essential in disaster management, particularly in response and monitoring phases as they provide a large area of coverage and location observation. Geo-portals contain the open-source spatial data regarding the incident area, which is useful for mapping and visualization. An important data source is the authoritative/archived data owned by the government and NGOs, which contains historical and survey data reports that intend to be embedded for effective analysis.

### 2.6.2 Data Harvesting

Data harvesting is triggered by the disaster event to engage all the dedicated and available data sources. It is important to tackle big data close to the source, especially in emergency response systems, with the intent to decrease irrelevant content, subsequently assisting real-time processing and improving access time for information [103]. The resource discovery component identifies the availability and accessibility of diverse and distributed data sources that are relevant to disaster management. Content delivery networks allocate specific tasks to a distributed system and improve response time. Load balancer is responsible to ensure maximized throughput, increased capacity and reliability of applications. Fault detection mechanism identifies hardware or software failures and saves time in troubleshooting. Data aggregators need to be utilized fittingly, as data is collected from some sources (e.g., authoritative/archived data, geo-portals) in large repositories in the form of batches, while at a rapid rate and in real time from other sources (e.g., social media streams, IoT-enabled sensors). Data dispatcher is responsible for dispatching processed information or queries from the system back to IoT devices or users on social media, for demanding more information and sending alerts or safety precautions. Furthermore, it is important to impose a security mechanism at the data harvesting layer before transferring the data by installing firewalls on the channel.

### 2.6.3 Data Communication

Data communication is the core layer and is responsible for transmission of data in all the proposed layers, using available communication technologies. Depending on the compatibility with the data source, various communication technologies with allocated gateways, categorized in different network type (i.e., LAN, WAN, PAN) can be integrated to enable smooth transmission in an efficient and secure manner. Wired local area communication networks (i.e., Ethernet), along with wireless local area networks (i.e., WiMAX (IEEE 802.16)) are used to provide short distances connections (e.g., office building, airport, and hospital). Wide area networks, such as general packet radio service (GPRS), Code-division multiple access (CDMA), long term evolution (LTE) and even public switched telephone network (PSTN) are suitable in the transmission of data over large areas. Zigbee (IEEE 802.15.4), Bluetooth (IEEE 802.15.1) and Wi-Fi (IEEE 802.11p) are effective short-range communication technologies that are compatible with high-level communication protocols. Notably, in emergencies, wireless communication especially satellite communication, Wi-Fi and WiMAX, has been the most effective means of communication [104].

### 2.6.4 Data Management and Analytics

Data management and analytics is the core layer responsible for performing data filtration, programming and analytics operations. Initially, the filtration process starts with the collection of datasets from heterogeneous sources. Data conformation categorizes only potentially relevant datasets required for the incident to save processing time. Data quality parameters such as accuracy, consistency and reliability are checked in data cleansing. Classification of information retrieved from data (i.e., current status, casualties/injured reported, impact area maps, images or videos reported, instructions suggested) is performed. About a specific incident the required content is extracted from classification. The location co-ordinates attached to the data readings, maps and geo-tagged social media posts are extracted for mapping. This method of filtering and categorizing data will help in managing data for analysis and reduce storage space, hence decrease the computational overhead for data analytics.

A set of data programming tasks are proposed to ensure effective analysis according to the results required for disaster management. The programming tasks are based on the decision model, which is the template that defines how the essential goals are perceived, organized and processed to reach a specific decision. Technologies for multi-source information fusion combine essential information from massive heterogeneous multi-source data. Event detection in real time that is backed by social media and multiple sensor data is critical for disaster management. Semantic engine is used for effective knowledge management by searching, extracting and categorizing unstructured information. Pattern recognition provides machine learning ability to detect the configuration of features and identify the required information from textual or image/spatial datasets.

Analyses related to disasters are time-critical in nature and with huge volumes of streaming data at the back-end, demand significant computational power for accurate and high-speed processing. These constraints demand for processing the resource data through a combination of cutting-edge powerful big data analytics tools. A state-of-the-art solution for this environment would be a combination of the Hadoop Ecosystem and the Spark analytics engine. Hadoop is considered the backbone of any big data architecture. It is an open-source software platform that supports enormous data storage and processing. It is a much cheaper and effective solution than running a dedicated data center. While, Spark, an open-source in-memory data processing framework is suitable for interactive data queries and enables processing of real-time data streams with the combination of its application-specific libraries. Spark, can be used with Hadoop data source as a programming model for processing. Moreover, a combination of different machine learning algorithms, natural language processing and data mining techniques can be used for further analysis. The obvious aim of deploying state-of-the-art data analytic tools is to facilitate the decision-making process with a continuous flow of reliable and updated information extracted from multiple resources.

### 2.6.5 Applications

The huge sets of valuable data resources backed by powerful data analytics, enables the application layer to implement an interface that allows interactive reporting and visualization of information to non-technical decision makers (i.e., emergency responders) in real-time. BDA application services can integrate with different disaster management expert systems designed to alert, report, monitor and detect/predict disaster situations (i.e., early warning systems and emergency response systems). The application layer should operate on a web-based access control API to prevent unauthorized access. The application interface needs to support different visual tools for generating reports in an interactive manner.

### 2.7 Use Cases

This section presents some of the important use cases for IoT and BDA enabled disaster management with the aim to highlight the capability and importance of the said technologies. The selection of use cases considers the sequence of disaster management operations to present an overall picture of the disaster management environments where IoT and BDA play an important role. As presented in Table 5, most of the use cases are focusing on the collaborative deployment of multiple sensors (i.e., weather station sensors, cameras, GPS, wearable sensors, smartphones, etc). These different types of sensors provide huge volumes of heterogeneous data (i.e., textual, image/video and spatial) through IoT. However, with supportive BDA applications, it is possible to process the collected datasets that enables much richer and effective systems. Moreover, the mode of processing is more towards real-time applications, which makes sense due to the involvement of BDA and IoT based applications.

**Table 2.4** : Comparison of IoT and BDA assisted disaster management use cases.

| Use Cases | Benefits | Study | IoT Devices | BDA Applications | Data Type |
|---|---|---|---|---|---|
| Early Warning | Timely and exact cautionary alert acknowledgement of a disastrous event | [33] | Multiple sensors | Text mining, Semantic analytics | Textual |
| | | [105] | Remote sensing, Observatory sensors | Data mining, Spatial data analytics | Textual, Image/video |
| | | [106] | Smartphones | Spatial data analytics | Textual, Spatial |
| Evacuation | Quick and accurate identification of evacuation routes for occupants following a disaster | [107] | Vehicular sensors | Streaming analytics | Textual |
| | | [108] | Smartphones | Location Intelligence | Spatial |
| | | [109] | Mobile-nodes | Map-matching visual analytics | Spatial |
| Monitoring | Better situational awareness enabling faster and effective response/recovery operations at a lower cost | [110] | Cameras | Image-processing, Visual analytics | Image/video |
| | | [111] | Multiple sensors | Convolutional neural network (CNN) | Textual, Image/video |
| | | [112] | Multiple sensors | Data mining | Textual |
| Prediction | Forecasting potential disaster risks in advance | [113] | Multiple sensors | Machine learning | Textual |
| | | [114] | Multiple sensors | Deep learning | Textual |
| | | [115] | Multiple sensors | K-mean clustering | Textual |

### 2.7.1 Early Warning

Early warning for disasters (i.e., floods, landslides, tsunami, forest fires, storms etc.) can prevent loss of lives and minimize the disaster's impact costs. A warning notified with sufficient time before the disaster will allow people to evacuate the area and help the emergency responders to organize and take the necessary precautionary actions. Structure of the early warning system is determined by the goals it desires to achieve, considering timely and accurate information processing regarding an upcoming event. Early warning systems for environmental disaster management are mostly involving IoT [33]. Early warning systems receive the data from real-time sensors, process the information and provide an interactive warning service for more information. However, big data challenges need to be solved during the development and application of such IoT-based information systems [105]. Moreover, with the evolution and widespread use of different of IoT devices such as Smartphones, it is possible to deploy their embedded sensors (GPS, accelerometers, gyroscopes, etc.) to monitor and provide valuable data for early warning systems [106].

### 2.7.2 Evacuation

The instantaneous and accurate identification of evacuation paths following a natural disaster is critical to saving the lives of the occupants. Quickly understanding the damage situations through appropriate data and processing techniques can lead to effective evacuation. In the event of a large-scale disaster, ensuring minimum road congestions are important for evacuation plans. The evacuees need to be guided towards safe and least congested routes to decrease the evacuation time. For the transportation network, real-time road situations need to be considered to compute and identify the maximum flow capacity of the roads [107]. Evacuees can also contribute for identifying blocked and congested roads using their smartphones and share the information with each other through short-range wireless communications. This approach can not only navigate the evacuees to safe places but also help in aggregating disaster-related information [108] [109].

### 2.7.3 Monitoring

Disaster monitoring service aims at providing effective response and recovery operations in both pre- and post-disaster situations. Due to the increasing usage of state-of-the-art technologies such as IoT and BDA, disaster monitoring service is getting faster, reliable, effective and more situational aware. Conventional disaster monitoring systems are often costly and time-consuming as they appoint only gauge sensors that could only measure one-dimensional physical parameters. However, advanced disaster monitoring systems are involving multiple data sources to geographically detect and visually monitor disaster events at a lower cost. For instance, through image-based automated monitoring, surveillance cameras are transformed into visual sensors [110]. This approach of visual sensing provides spatiotemporal information that can be utilized for a reliable automated remote monitoring of floods. With the aid of deep learning methods, multiple data sources such as map-based web services, sensors, and video cameras can be incorporated to perform real-time monitoring [111]. Moreover, disaster monitoring can benefit from the convergence of different technologies to analyze huge sets IoT extracted data with data mining techniques and identify emerging risks and changes in weather for potential disasters [112].

### 2.7.4 Prediction

Disaster occurrence is out of human control; however, through the deployment of various state-of-the-art smart technologies, we can predict, mitigate and even prevent the loss of human lives and infrastructure. Research in the field of disaster prediction has shifted from statistical and theoretical submissions to successful real-world applications. With the advancement in IoT and BDA technologies, disaster prediction systems are getting great success to minimize the adversity caused by disaster such as floods, wildfires, hurricanes, tsunamis etc. Promising innovations in technology such as IoT based IP-based sensor networks and evolving techniques of machine learning are being deployed for disaster predictions [113]. Within the scope of smart cities, disaster predictions are becoming a reality through deep learning techniques backed by

IoT big data [114]. Moreover, the convergence of IoT big data and high-performance computing (HPC) can provide the capability of real-time disaster predictions [115].

### 2.8 Open Research Challenges

This section highlights the main open research challenges that need to be explored in the future to have better understanding and development related knowledge of the desire research area.

**Disaster Data Quality:** The quality of the collected data is very critical for disaster management, as noisy, incomplete and error-prone data can lead to serious problems and wastage of precious time in a disaster scenario. This factor is an additional overhead for BDA- and IoT-based disaster management environments, which requires to be solved prior to any kind of analysis. Data quality parameter plays a major role in determining the accuracy of the analysis carried out on a particular dataset. Table 2.5 describes five proposed parameters of data quality that are commonly recognized and are suitable to formulate the disaster data in the filtration process. Each dataset needs to satisfy the conditions describe against the specified parameter in order to be eligible for further processing. Many filtration algorithms and data format converters are being proposed on a regular basis; however, it remains an open research challenge for disaster management where data quality should have the highest priority.

**Where is Disaster Dataset's Metadata?** Metadata extraction from multiple heterogeneous data sources for a time-sensitive and data quality critical application like disaster management is an important challenge. The essential metadata information about the datasets, i.e., data source, content, time stamps, spatial reference are very important to be identified, in the context of this environment. With effective extraction of metadata, a lot of data quality concerns and integration issues can also be solved at the grassroots level, and reliable datasets can be provided for the disaster management operations.

**Multi-sourced Disaster Data Aggregation:** Collecting disaster-related data from heterogeneous sources and integrating that voluminous data in real time is

**Table 2.5** : Disaster data quality parameters.

| **Parameter** | **Conditions** |
|---|---|
| Accessibility | Is the data access public or proprietary?<br>Whether the data has any rights or legal concentrates on usage?<br>Whether the data needs a special aggregator to collect? |
| Timeliness | Whether the data is gathered on run-time or through the historical database<br>Does the data require to be updated in intervals? |
| Credibility | Is the data source identified?<br>Can any organization or system verify the source? |
| Accuracy | Is the data related to the incident/crisis/disaster in any sense?<br>Is the data free from data redundancy? |
| Completeness | Is the data clear and understandable?<br>Can the data be classified to gain the desired results? |

a challenging activity. Moreover, data needs to be collected from multiple geographically distrusted servers which in return make the aggregation process more difficult. Data aggregator normally handles the collection and integration of similar data to tackle the data redundancy problem and minimize resource consumption. However, the data aggregation problems raise with the increasing number of data sources, demanding more storage and computation power.

**But Which Data Analytics Application?** Selecting the type of analysis to be performed on the newly acquired big datasets within the scope of disaster response or management can be a challenging task. The choice of a particular analysis method will determine the effectiveness and performance of the overall environment and hence will eventually affect decision making. Moreover, the desirable analysis and results may demand a combination of different analytical methods that can increase system workload and affect performance. Another challenge is to identify and analyze what data sets can support smooth and effective processing in real-time and hence provide accurate results.

**Time Constraint for Quick Response:** Due to huge data volumes, it is quite difficult to extract quality information in a limited time for effective decision-making to emergency responses. The data processing is time-consuming, as it involves multi-sourced data harvesting, filtering, and categorizing; that can take a lot of time even with advanced big data analytical tools. It is an important challenge for the existing techniques and tools to preprocess data and generate the required results in a specified amount of time to provide quick emergency response and save lives.

**Architectural Challenges:** Due to the lack of a defined model for BDA- and IoT-based disaster management environment in the existing literature, detailed observations of different related reference models is required. The architecture for such an environment needs to be flexible to accommodate all the data sources, consistent to configure different network topologies for data communication and supportive to fetch the required results for effective decision-making. Moreover, the architecture should be designed to keep the environment resilient so it can handle any type of disruption caused by disasters. With multi-sourced data, it is challenging to design a generic data model that integrates heterogeneous data while being flexible, effective and secure.

**Fault Tolerance during Disasters:** Fault tolerance is the ability of the system to work effectively even in the case of a hardware or software failure. With heterogeneous and distributed data source environments there is always a chance for some hardware devices and sensors to fail because of physical damage or disruption in communication channels, particularly in a disastrous situation. The BDA- and IoT-based environment having distributed components, predominantly its data sources can be affected by the disaster as well. Hence, disaster resilient system architecture needs to be planned, so that the data can be effectively channelized and processed even in the course of any destruction. Data sources are critical in the successful deployment of the environment and should be able to generate data with infrastructure impairment and power blackouts. Hence backup power consumption mechanism and data management capabilities; such as a redundant backup system or cloud-based distributed storage system with distributed computing facilities needs to be established.

**Privacy and Security:** Privacy concerns have been a serious issue in both big data and IoT domains, as open personal information is widely utilized which, if misused can lead to threats such as profiling, tracking, theft, and discrimination [116]. Big data usually contains some sort of confidential information related to people or government and hence high-level security is required as the data moves over different types of networks. Social media data sources can increase privacy concerns as its data sets contain personal details and location of the users. These data sets can be very sensitive in crisis like civil wars and resistance movements. Additionally, open source big data analytics tools and most of the technologies in the Hadoop Ecosystem lack sufficient security mechanism [117]. Managing the access control of the big disaster datasets is vital to safeguard against any malicious use of data, hence proper security mechanism is required to ensure data protection.

**Standardization Challenge:** Standardization of IoT in general and big data, in particular, is still in its infancy. Standards can promote system efficiency, foster technological changes and provide recognized guidelines for policy, governance and future research. As disaster management requires various systematic solutions, it can be difficult to develop standards initially. However, standards such as communication protocols, security protocols, meta-data and data aggregation standards are the core activities that need to be formalized to increase the value of disaster management environments and services.

## 2.9 Conclusion

This chapter, identified the benefits of BDA- and IoT-based disaster management and investigated the state-of-the-art literature conducted regarding BDA and IoT applications for disaster management. The study classified the related literature by presenting a thematic taxonomy that unearths the main attributes of BDA- and IoT-based disaster management environments. We also presented a thorough overview of the overall architectural deployment of BDA- and IoT-based disaster management environments through a reference model having dedicated layers, such as data generation, harvesting, communication, management and analytics, and applications. This chapter discussed and compared some indispensable use cases to show the role of BDA and IoT in different disaster management phases. Moreover, the key requirements for the successful deployment of the environment and the challenges that need to be resolved are sketched out. It is conclude that this survey can be used as a guideline to understand the overall functionalities for productive utilization of the opportunities associated with BDA and IoT towards the construction of an effective disaster management environment.

# 3. A FRAMEWORK FOR SOCIAL MEDIA DATA ANALYTICS FOR DISASTER MANAGEMENT

## 3.1 Abstract

Social Media datasets are playing a vital role providing information that can support decision-making in nearly all domains. This is due to the fact that social media is a quick and economical approach for collecting data. It has already been proved that in case of disaster (natural or man-made) the information extracted from Social Media sites is very critical to Disaster Management Systems for response and reconstruction.

This chapter comprises of two components, the first part proposes a framework that provides updated and filtered real time input data for the disaster management system through social media and the second part consists of a designed web user API for a structured and defined real time data input process. This research contributes to the discipline of design science for the information systems domain. The aim of this chapter is to propose a framework that can filter and organize data from the unstructured social media sources through recognized methods and bring this retrieved data to the same level as that acquired through structured and predefined mechanisms, such as a web API. Both components are designed such that they can potentially collaborate and produce updated information for a disaster management system to carry out accurate and effective decision-making.

## 3.2 Introduction

During disasters (e.g., floods, earthquake, storms, large fire, etc.), people tend to report and share their observations, findings and suggestions on various social media sites. However, it is still a challenge for researchers as how to automatically filter out useful information and make that information search-able and accessible for emergency services.

Social Media applications are considered very useful to collect information in case of any disaster because it is the fastest and the cheapest source to provide effective, updated and relevant information for decision making. The practical use of such applications is making a vast number of academic studies to research on many aspects of social media in Disaster Management [118] Social Media provides its user the opportunity to contribute and disseminate valuable information, be it in the shape of text, pictures, audio and video; that is necessary for disaster management processes and communications [119].

Current research states that the communication services such as Short Message Services (SMS) or social media (Facebook, Twitter) have the ability to improve the regular and updated transmission of valuable information and provide the effective resources of information in all of the disaster management life cycle phases and aid in developing a disaster resilient community [120] [121]. In case of any disaster the emergency service authorities should be able to access the social media networks and blogs to identify the source and scale of the disaster and develop the recovery plans according to the affected communities' requirements. In addition, authorities should be able to observe online communities to detect mounting trends and possible hot-spots that can substantiate as indicators for disaster [120]. When it comes to managing disasters efficiently, the main thing for government and emergency agencies is to be provided with accurate, updated and complete information; otherwise it can have serious consequences if the information is provided incorrect and late [122] [123].

Real-time Geospatial Information Systems use social media as crowd sourcing virtual network to map social feeds using geotag metadata with longitude latitude coordinates. GIS systems with all its hardware based sensors can be combined with Social media as the basic theme is to create a compressive source of information to understand the disaster situation accurately so that the emergency responders and the general public can be added to improve overall awareness [124] [125]. However, there are some challenges also associated with social media data collection methods; such as variable quality of the data, intelligently managing the big volume of social media feeds, requirement of manual checking and verification of the data, accurately Geo-parsing map information and the need to find right balance between time wasting false positives and responsive alters.

Some related work already done shows the practicality of this research are the recognition of frameworks such as Twitcident [126] and SensePalce2 [127]. A very good example of web API for collecting citizen reactions in case of an earthquake and then model earthquake activity according to that accessed information is a website named as “Did You Feel It?” by U.S Geological Survey (USGS) (http://earthquake.usgs.gov/earthquakes/dyfi/). The concept of crowd souring is a corresponding research area that is aimed to work as a virtual sensor to collect data from every potential source possible and to support activities from basic mapping (e.g., Open Street Map) to be a source for providing data to the disaster manage-ment systems (e.g., CrisisMappers.org, Ushahidi).

### 3.3 Social Media in relation with Disaster Management

Social Media are applications that are totally depended on the user generated content or applications in which the user generated contents and activities play an important role in increasing the overall value of that application or service [128]. On the other hand, a Disaster is a sudden event that seriously affects the normal routine conditions of a community or society. It has not only an economic and environmental impact, but also an important humanitarian component. Disasters could be natural calamities such as earthquakes, tornadoes or hurricanes, but also man-made destructive activities such as terrorist attacks or industrial accidents. Disaster management can be modeled into four phases, namely mitigation, pre-paredness, response, and reconstruction. Having a good strategy for each of the phases is essential for an efficient disaster management. In order to accomplish this, managers need proper information about the different activities within each of the four disaster management phases [129]. With social media, information is now accessible in real-time, so those activities can be planned more accurately. With disaster management models and its phases a lot of research work needs to be done to map where and how social media information can be used to improve the decision making. Conducted research has identified that the use of social media is more increased and even surpassed the use of other conventional communication methods such as fixed phones after a disastrous incident. Social Media sites like Facebook, Twitter and YouTube can be very handy when tsunamis, earthquakes, floods and other natural disasters strike to collect the real time data.

According to Crystal Washington [130] "Social media is the application that -

- Provides valuable information to those in a disaster area pre and post disaster (via the Internet, if available, or SMS updates).
- Drives awareness to those outside the affected areas, generating volunteers and/or donors.
- Connects displaced family and friends.
- Provides information about unclaimed property, and in worst case scenarios, bodies.
- Offers information about aid, centers and other resources available to those affected."

Five discrete uses for social media in disaster management are identified by [127] as,

a) "to disseminate information to the public (e.g., for alerts)".

b) "to gather information from the public (e.g., crowdsourcing)".

c) "to coordinate with crisis management professionals".

d) "to monitor activities of crisis management professionals".

e) "as input to situational assessment for crisis management".

Instead of categorizing the existing research into disaster management phases directly, we added the social media application because they represent a more fine-grained perspective of social media within a disaster. Moreover, in applying the social media applications to traditional disaster management phases allows us to integrate literature about potential social media activities into the disaster management model. This will act as a theoretical lens to classify the existing research into disaster management phases.

## 3.4 Proposed Research Framework

This research work contributes to the discipline of design science for the information systems domain. Based on reviewed literature and detailed perception of related technologies, we propose a model that can provide updated and essential input data for the disaster management system through two different potential crowd-sourced data platforms; the social media component and a designed web user API component.

The proposed model is divided in two sections considering the method involved for data input. As seen in Figure 3.1, the first section is the web user API, named as "Disaster Analytic API" which is a designed web template to take the data according to some structured parameters already defined and to somewhat en-force the users to follow the designed format for feeding data. This second section is the "social media system component" which is taking in the unstructured social media data and applying selective recognized methods to filter the data and present it for further process to the disaster management system component. This model emphasis on the quality of real time data gained through crowd sourcing, which is normally considered noisy and unfit to use for accurate decision making processes. This model can be a good example to compare and measure data quality gained through a designed API for structured data and real time data gained through social media. There is no such comparison currently found in the literature and this area needs to be focused to make the coward sourcing data gathering mechanism more effective and open to both structured (web data entry API) and unstructured (social media) for new actionable insights.

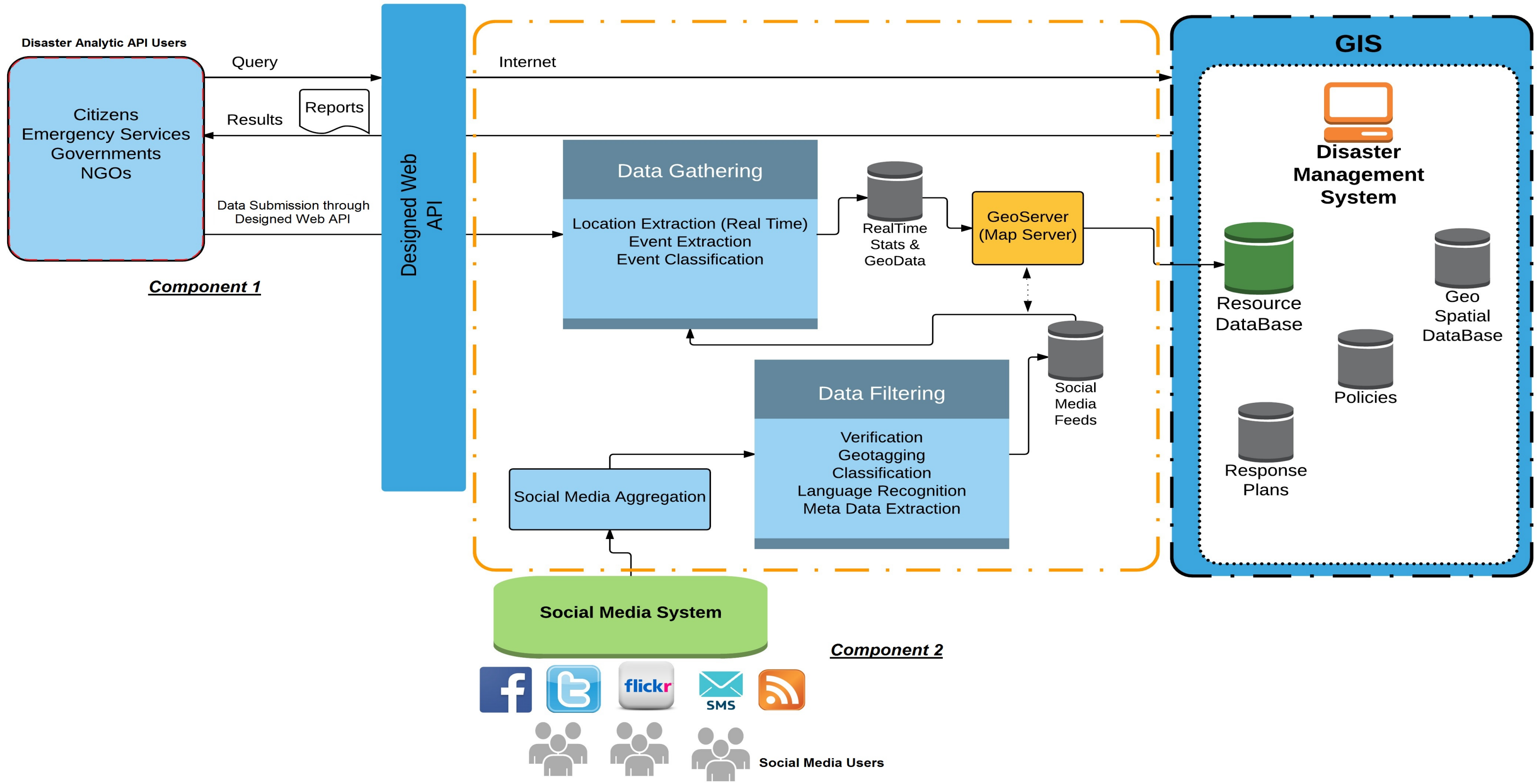


**Figure 3.1** : Proposed research model for structured and unstructured datasets.

### 3.4.1 Designed Web User API Component

This component named as “Disaster Analytic” is a proposed designed web API for receiving information in case of any disaster in a structured data entry template, which covers all the sections such as location, time, scale, description, attachments (photo or video) etc. for data to be actionable. Additionally, Disaster Analytic can also manage queries from the user requesting reports regarding the disaster situations. A rough outline for the interface is proposed in Figure 3.2. The potential users of this web API can be citizens or volunteers, governments’ officers, emergency services and NGOs who can provide data under defined format and query, search or request for a report through the disaster management system.

The main aim for the design and implementation of this web user API is making sure that accurate and machine readable information is received. It also should be user friendly, and can be operated without much setup or training involved. Every user need to be identified and authorized login should be created. There are many different types of format constraints to be followed in order to get the structured and machine readable information. The notable constraints can be User’s Log in ID (who), Type of Incident (what), Photo of the identified incident (what), Scale (what), Date and Time (when), location of the incident (where), Classification and description (how). To make more user friendly, constraints such as date and time need to be provided automatically so that user should be able to select the options rather than make manual entry in the form. The user can request the forms directly from the disaster management system whenever they want and are able to see the recent verified social media feedbacks and trends being shared for regarding the disaster.

### 3.4.2 Social Media System Component

We are living in an information age where people tend to report, discuss and share the ground facts, observations, and their experiences on different Social Media forums even if it’s regarding a specific disaster. The aim of this Social Media component is to filter and extract the required data gained through crowd sourcing, and make it beneficial for disaster management.

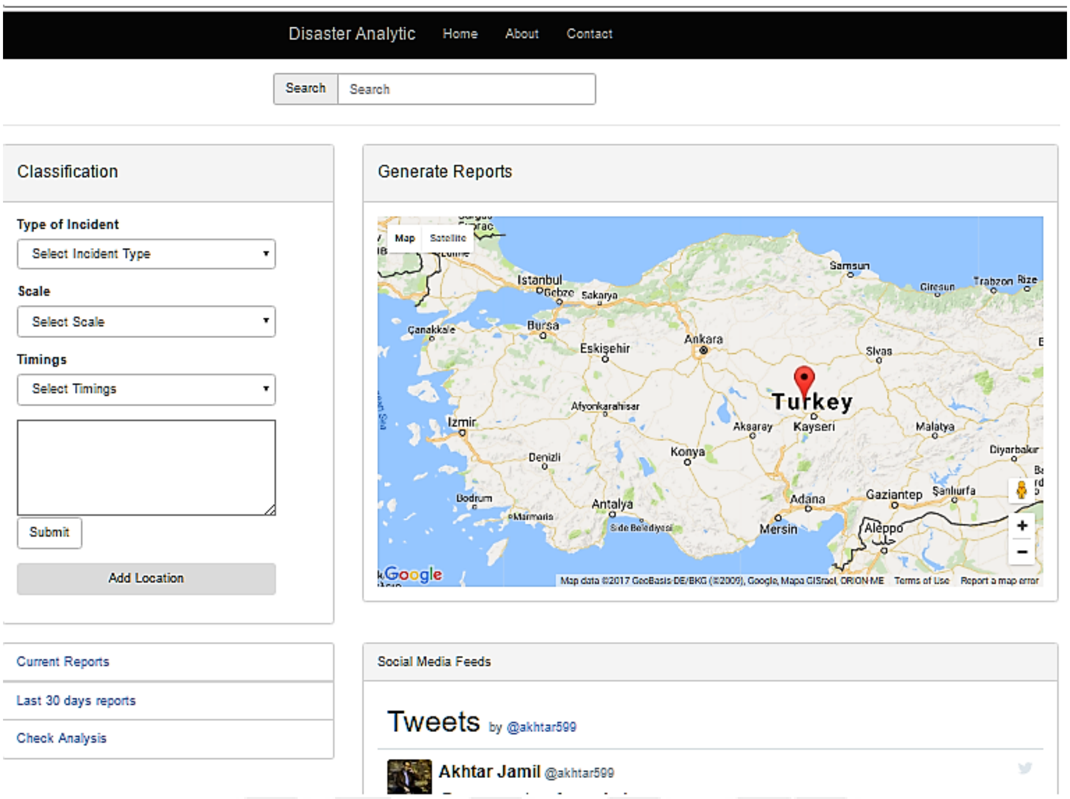


**Figure 3.2** : Designed web API (Disaster Analytic).

The data filtering process as shown in figure 3.3 is following the semantic methods as it supports filtering and extraction of social media feeds to recognize only the suitable feeds relevant to the incident and provides means to organize information about the incidents for performing real time analysis. After intensive literature review some notable filtering methods are selected and their details are presented in sequence in the following sections.

#### 3.4.2.1 Data capturing

Whenever an incident is reported the system is triggered to get the real time data through Social Media. Through this framework the data is captured from different Social Media sources and translated into an incident profile. The main aim of this incident profiling is to produce a profile that can provide the valid raw data for data filtering and extraction. “An incident profile is a set of weighted attribute-value pairs that describe the characteristics of the incident” [126]. These attributes contain data regarding the incidents and might have specific weights to highlight the importance of each attribute according to the type of incident occurring. Incident profiles are kept open to change as it needs to be updated according to the modifications that

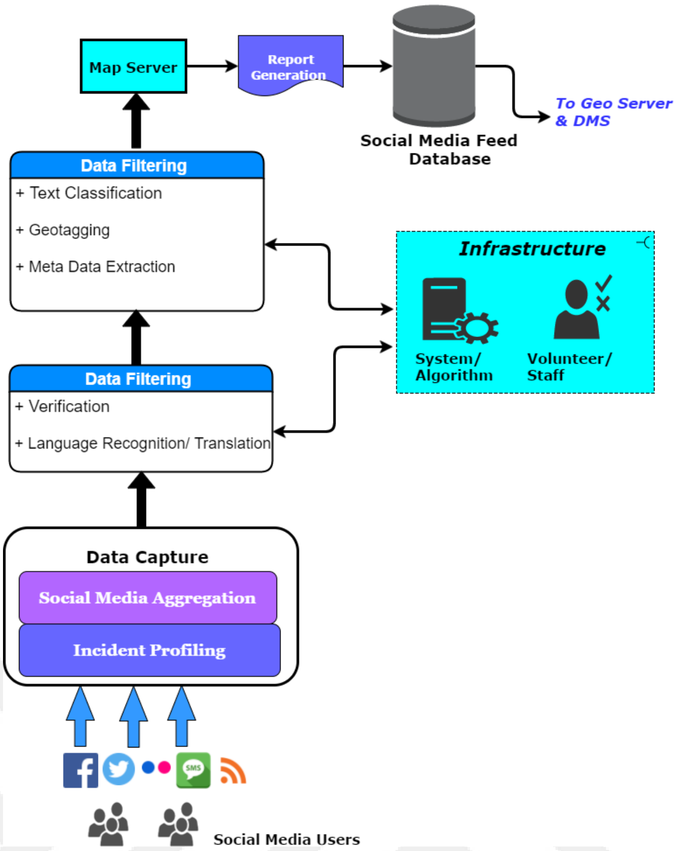


**Figure 3.3** : Proposed social media data filtration component.

may occur during the incident. Keeping in view the incident profile designed for the disaster, social media aggregation is per-formed to capture any type of feed that fulfills the criteria. These feeds are then reported and processed for extracting the useful information for the system. Some common attributes for an incident profiling are mentioned in Table 3.1.

**Table 3.1** : Incident profiling attributes.

| **Incident Profiling Attributes** | **Description** |
|---|---|
| Classification | Earthquake, Fire, Flood, Power Failure, Accident |
| Address | State, City, Zip Code, Coordinates (if extractable) |
| Descriptive location | Any location name used |
| Time of occurrence | Time Stamp |
| Scaling | Size or value reported |
| Image reporting | Any Image or Video reported |
| Observations | Comments regarding the incident |

#### 3.4.2.2 Verification

In this phase of data filtration the authenticity and reliability of the crowd-sourced data are checked because of the anonymous source through which it is collected.

This remains an issue with social media data and a lot of algorithms are designed to cope with this concern. Privacy concern is also an issue and need to be addressed with proper procedures. A lot of social media sites have already taken their users in satisfaction regarding location and feeds in their user agreements. Verification can also be performed manually by volunteers or stuff, but can be time consuming and laborious.

#### 3.4.2.3 Language recognition

Regions having multiple languages can get social media feeds in different languages and hence affect the source data by adding noise. This issue highlights the need for improved filtering techniques to translate feeds in a common language (i.e., English). Neuro-Linguistic Programming (NLP) techniques can be used to manage multilingual situations.

#### 3.4.2.4 Metadata extraction

Metadata Extraction is a vital part of these social feeds as it provides additional information on each feed. The metadata table can contain information about the originator of the feed, and furthermore to verify the source it may contain the pro-file picture, number of followers and feeds, location and time-stamp when sharing the feed. Such type of Metadata can strengthen the reliability and accuracy of the data provided for processing and decision making.

#### 3.4.2.5 Geotagging

Through a designed API we can get map source data that provide street and building level locations using coordinates, but on the other hand social media feeds generally describe locations on regional levels if satellite navigation system is not activated. As we know that the precise location of the incident plays an important role in disaster management therefore to enable spatial exploration of social media feeds a geotagging model should be used to display the narrative of the feeds at its exact location on the map. The latitude/longitude coordinates of the us-er profile can be used to get the desired location.

#### 3.4.2.6 Text classification

This phase is the core part of feature extraction as it indicates the extracted con-tents and provides the required information for reporting or to be used in a disaster management system. It may contain reports regarding casualties, possible threats and damages about the incident. The classification can further be categorized ac-cording to the feeds if the publisher was witness, hearing the news, observing or smelling something. Handcrafted rules are used for the classification that can work in both the attribute-value pairs and the plain words stated in the social media feeds [126]

Some common attributes, keeping in mind the filtered social media feeds for classification, are described in Table 3.2.

**Table 3.2** : Text classification attributes

| Classification Type | Description |
|---|---|
| Impact area | Map highlighting the potentially affected areas |
| Status | Current status of the incident |
| Threats | Possible future threats |
| Related news | Ongoing news about the situation |
| Casualties and Injured | Number of casualties and injuries reported |
| Image or video reporting | Any Images of the incident received |
| Respond time | Possible time to respond to rescue |
| Instructions | Any precautions needed |

## 3.5 Structured and Unstructured Data Transformation

It is important to identify the structured and unstructured data types so that data pre-processing techniques can be developed accordingly. Unstructured data is any information that is not organized in a specific pattern or does not have any planned data model while structured data is organized information that can be easily stored and mapped into specific fields. The main drawback is that around 80 percent of all potentially useful big data is in unstructured format and needs some filtration mechanism to extract the required information. Unstructured data has always been ignored and considered as “dead data”; however with the evolution in big data technologies, valuable information and useful insights can be generated by processing unstructured data.

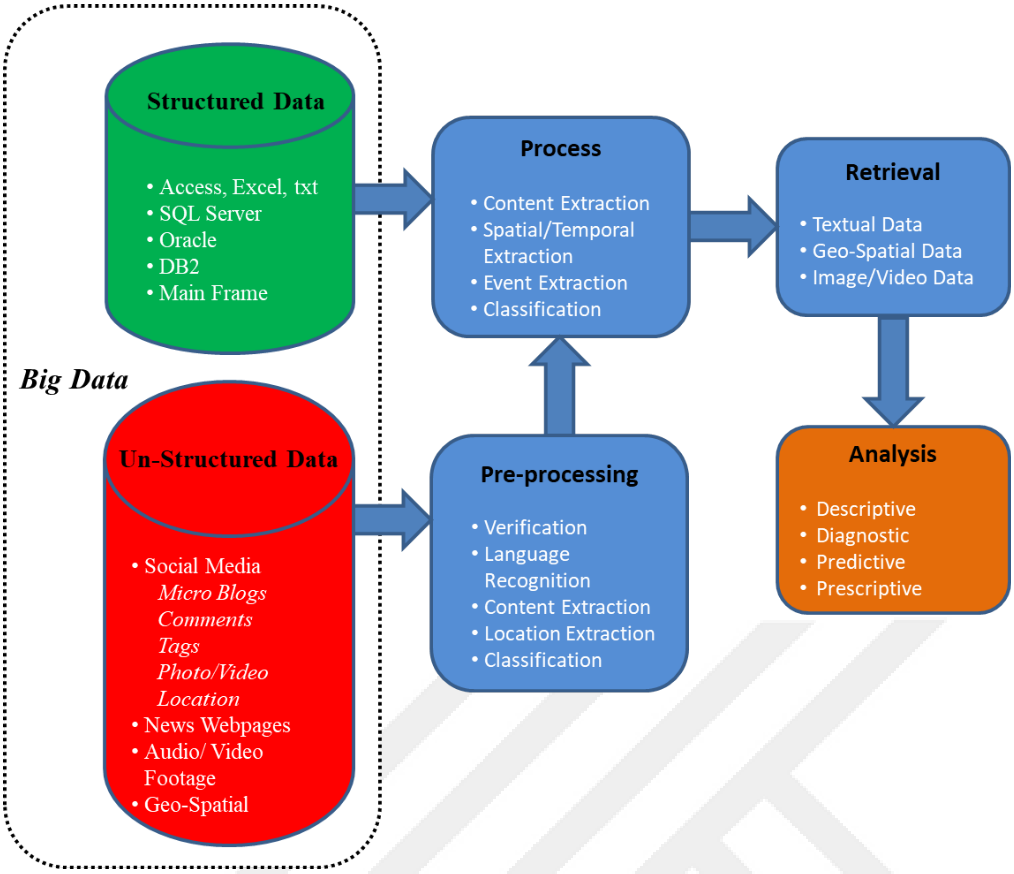


**Figure 3.4** : Big disaster data transforming levels.

Figure 3.4 shows the transformation of unstructured data after filtering through the proposed pre-processing steps to be used along with structured data. The identified datasets available after executing categorization and modeling are ready to be retrieved in three different data formats i.e. textual, Geo-spatial and image/video data. Here data modeling means abstraction of application logic through a programming model that enables data analysis applications. Data integration in this phase is vital as it deals with heterogeneity found in the data structures and semantics for forming a uniformly interpretable data structure for effective large-scale analysis. A set of parameters are proposed that needs to be followed during the acquisition and filtration process. Social media data being unstructured needs to be evaluated through a filtration process so that it can be executed with structured datasets.

## 3.6 Conclusion

This chapter presents a design model for the development of an integrated system consisting of social media crowd-sourced component and a designed web API component through which organized and reliable data can be provided for real-time disaster management. This design-science research demonstrates that the concept of social media crowd-sourcing can effectively be used for real-time disaster management and tries to aid the theory of making crowd-sourced data as trustworthy as other data sources. The basic theme of this design is to make the unstructured crowd-sourced data process-able so that it can be compared and merged with a structured data sources such as a web API. The effectiveness of real-time crowd-sourced disaster management systems has been proven but there are many gaps and challenges in this research domain. The design science to model integrating frameworks plays a key role for providing the basis for interdisciplinary research to be carried out.

# 4. TOWARDS DISASTER RESILIENT SMART CITIES: CAN INTERNET OF THINGS AND BIG DATA ANALYTICS BE THE GAME CHANGERS?[1]

## 4.1 Abstract

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 chapter, a novel reference architecture and philosophy of a Disaster Resilient Smart City (DRSC) is proposed and discussed through the integration of 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.

## 4.2 Introduction

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

[1]This chapter is based on the paper "Shah, S. A., Seker, D. Z., Rathore, M. M., Hameed, S., Yahia, S. B., Draheim, D. (2019). Towards Disaster Resilient Smart Cities: Can Internet of Things and Big Data Analytics Be the Game Changers?. IEEE Access, 7, 91885-91903. [Online]. Available: https://ieeexplore.ieee.org/document/8759905."

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 [131]. The prompting increase in the population density in urban cities has given rise to the requirement of proper services and a better infrastructure for its inhabitants. 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. 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. [5] [132]. Hence, a smart city equipped with the capability of generating early warnings, monitoring, and predicting the disaster 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.

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 real-time to extract the required information. The DRSC concept necessitates more attention due to its time-sensitivity and high accuracy constraint application owing to the life or death of human lives. Such problem signifies the leading edge of BDA and IoT advancements, which collectively are capable of dealing with the urgency of this problem.

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.

## 4.3 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.

### 4.3.1 Proposed Reference Architecture of BDA- and IoT-based DRSC

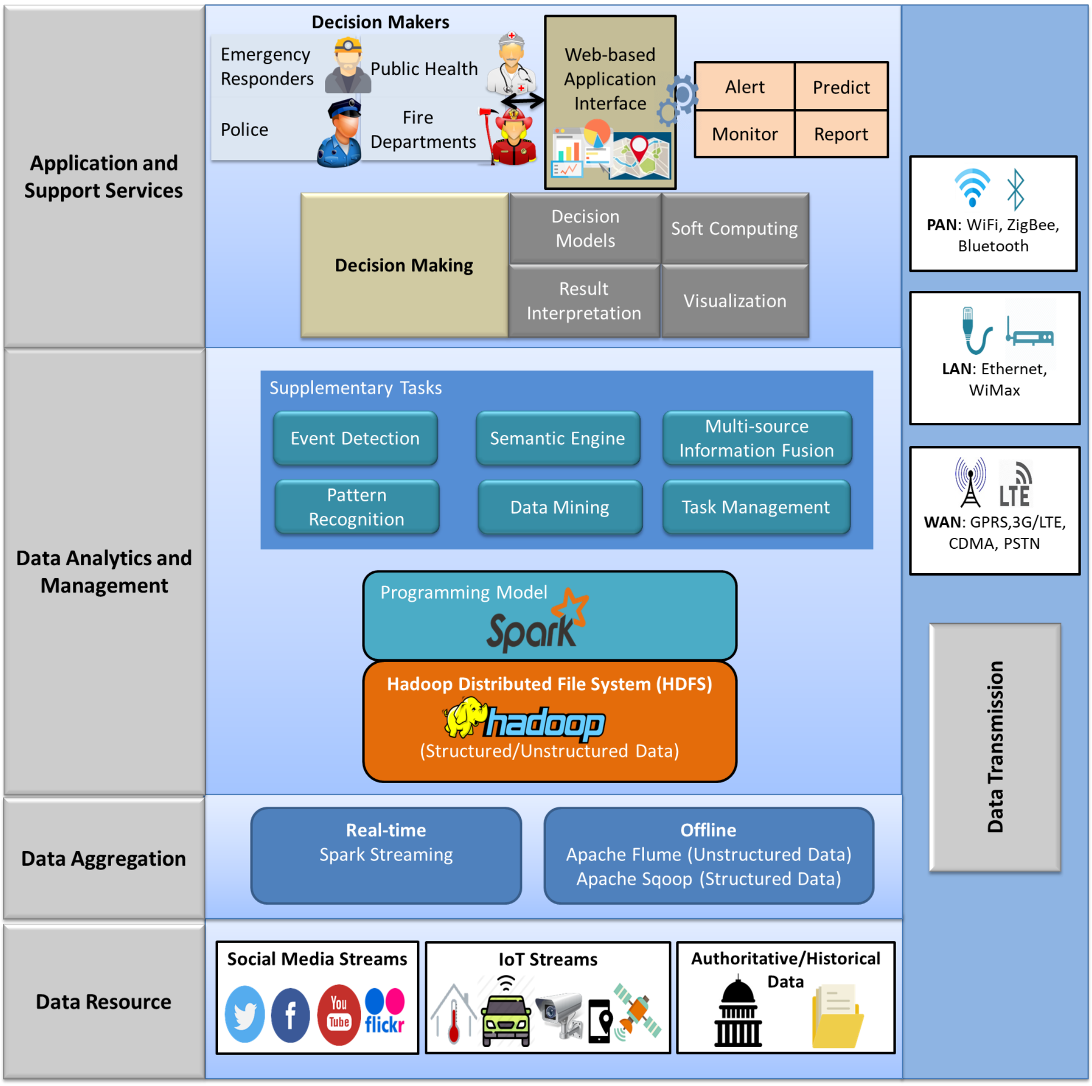


**Figure 4.1** : Proposed reference architecture for BDA- and IoT-based Disaster Resilient Smart City(DRSC).

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 [133]. In [134], a healthcare architecture is proposed that uses BDA on data from dedicated IoT devices. Similarly, in [135] the authors proposed architecture for smart urban planning based on BDA and utilizing IoT datasets. In another study [136] a framework was proposed for weather data analysis using BDA and IoT to extract meaningful information from huge volumes of unstructured data. In [137], 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 time critical and performance demanding application like 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 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 4.1 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.

#### 4.3.1.1 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 4.1. 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.

#### 4.3.1.2 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" [85]. 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

**Table 4.1** : Structured and unstructured data in the context of disaster management.

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

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.

#### 4.3.1.3 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 [138] 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 [139] 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.

#### 4.3.1.4 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) [140] 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 [141] 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 [142]. 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 [65]. 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 sources. Task management helps to identify workloads on different entities in the system and effectively managing system's operations.

#### 4.3.1.5 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.

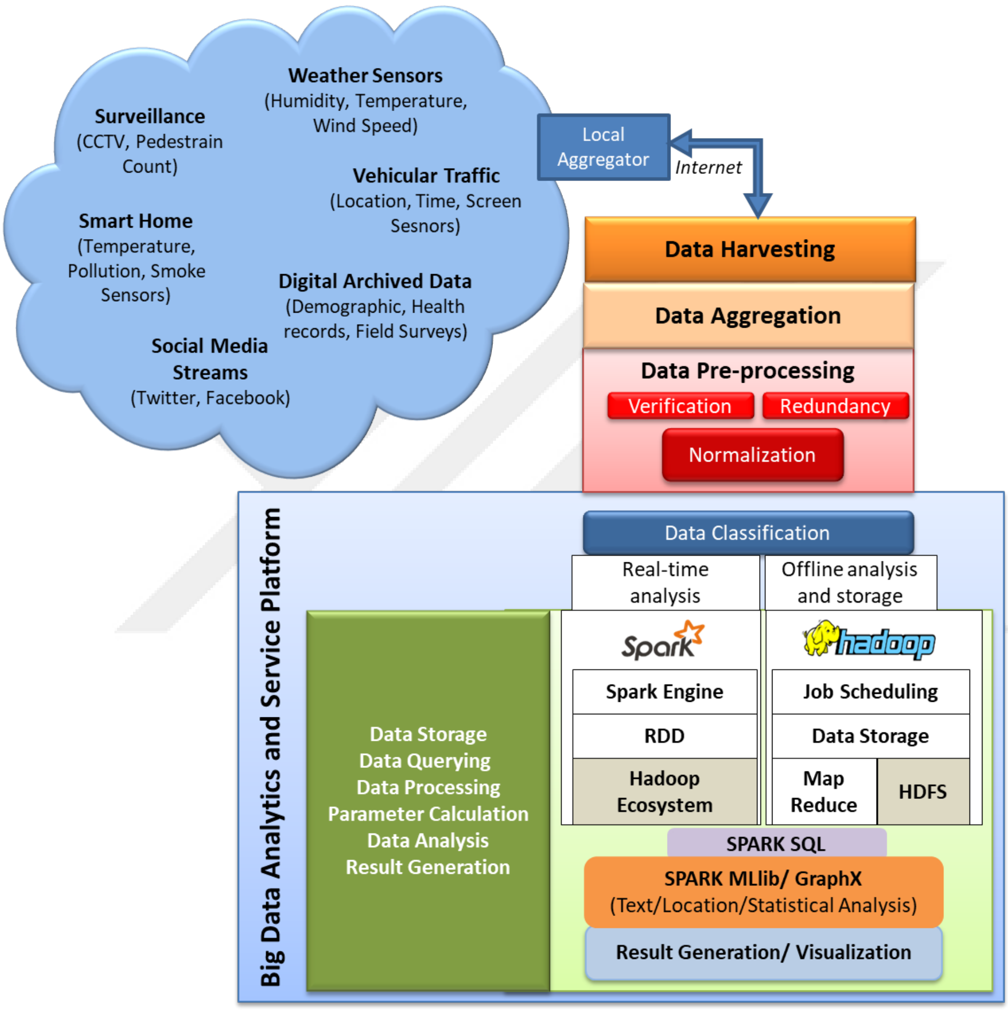


**Figure 4.2** : Implementation model of the deployed system.

### 4.3.2 Implementation Model

The implementation model that outlines the details of all the operational steps performed in our deployed system within the scope of DRSC is presented in Figure 4.2. 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.

#### 4.3.2.1 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 within the scope 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 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 4.2, 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.

**Table 4.2** : Dataset details for the data used in the proposed model.

| 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 | [143] |
| 2 | Pollution | 499 gas sensors placed within Arhus city, Denmark to measure gases including (CO), (SO2), (NO2), (O3), and particulate matter. | Raw data: 32GB<br>Structural data: 570MB | 07 | Pollution level monitoring | [144] |
| 3 | Traffic | Road traffic simulator and signals optimizer | 400 MB | 06 | Emergency evacuation path planning | [145] |
| 4 | Twitter | Twitter data for the month of September, 2018 | 41 GB | 04 | Collecting crowd-sourced information about disasters | [146] |

#### 4.3.2.2 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 4.1. 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,

$$DR = \sum_{i=1}^{m} \sum_{t=1}^{n} V_{i,t} \tag{4.1}$$

**Table 4.3** : Data Resource categorized illustration.

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

#### 4.3.2.3 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 [147] also referred to as zero-mean normalization technique is widely used to normalizes the dataset input values using Mean and either Standard Deviation ($\sigma$) or Mean Absolute Deviation (MAD). We opted for Z-score normalization with Mean Absolute Deviation (MAD) instead of Standard Deviation ($\sigma$) 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,

$$\begin{aligned} S_A &= \frac{1}{n}(|V_1 - \bar{A}| + |V_2 - \bar{A}| + |V_3 - \bar{A}| + ... + |V_n - \bar{A}|) \\ V_i^{'} &= \frac{V_i - \bar{A}}{S_A} \\ N_i &= \frac{1}{5}(V_i^{'}) \\ NDS &= N_i + 0.5 \end{aligned} \tag{4.2}$$

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 4.2. 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:

**FOR EACH** *(i)* = 1 to (n) **LOOP** //*(i)* is ID of dataset
Calculate the Mean $(\bar{A})$ for each dataset
Calculate the Mean Absolute Deviation $(S_A)$ for each dataset
Find the Z-score normalization $(V_i^{'})$ for each dataset value // $V_i^{'} = \frac{V_i - \bar{A}}{S_A}$
Divide all values by 5 //to get 1-point range
Add 0.5 to all values //to get values at scale of [0 $\sim$ 1]
Return the normalized values in new datasets *(NDS)*
**CONTINUE** ((n)+1);
**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 4.3.

$$TDS_t = \frac{T_H}{T_L} \tag{4.3}$$

#### 4.3.2.4 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 [148] 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 [149] 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 [150] 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.

### 4.4 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.

### 4.4.1 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 4.4 contains the CT values established for different applications used in this study.

**Table 4.4** : Defined Critical Threshold (CT) 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 4.5** : 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, we established the Twitter Critical Threshold ($T_{CT}$) as shown in Table 4.5. The alert message status depends on the TDS value derived from the Equation 4.3 according to a respective scenario in a given time-frame.

### 4.4.2 Analysis Results for IoT datasets

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 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 how the system detects these events and generate alerts.

---

**Algorithm 2** Fire Alert

---

**BEGIN**<br>
Input: Temperature *(T)* and Smoke *(S)* Readings<br>
Output: Fire Alert/No-Fire<br>
Steps:<br>
**FOREACH** (n) Reading of Temperature *(T)* and Smoke *(S)* **LOOP**<br>
*(T_Avg)*= $\frac{\sum_{t=n-3}^{n} T_t}{3}$<br>
*(S_Avg)*= $\frac{\sum_{t=n-3}^{n} S_t}{3}$<br>
**IF** (*(T_Avg)*>CT)<br>
*(T_Report:) = TRUE*<br>
**ELSE**<br>
*GoTo*($T_{n+1}$)<br>
**ENDIF**<br>
**IF** ((S_Avg)>CT)<br>
(S_Report:) = *TRUE*<br>
**ELSE**<br>
*GoTo*($S_{n+1}$)<br>
**ENDIF**<br>
**END LOOP**<br>
**IF** ((T_Report) && (S_Report) = *TRUE*)<br>
*GENERATE* (Fire_Alert);<br>
**ELSE**<br>
**CONTINUE**((n)+1);<br>
**ENDIF**<br>
**END**

---

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) [143], is used to generate various

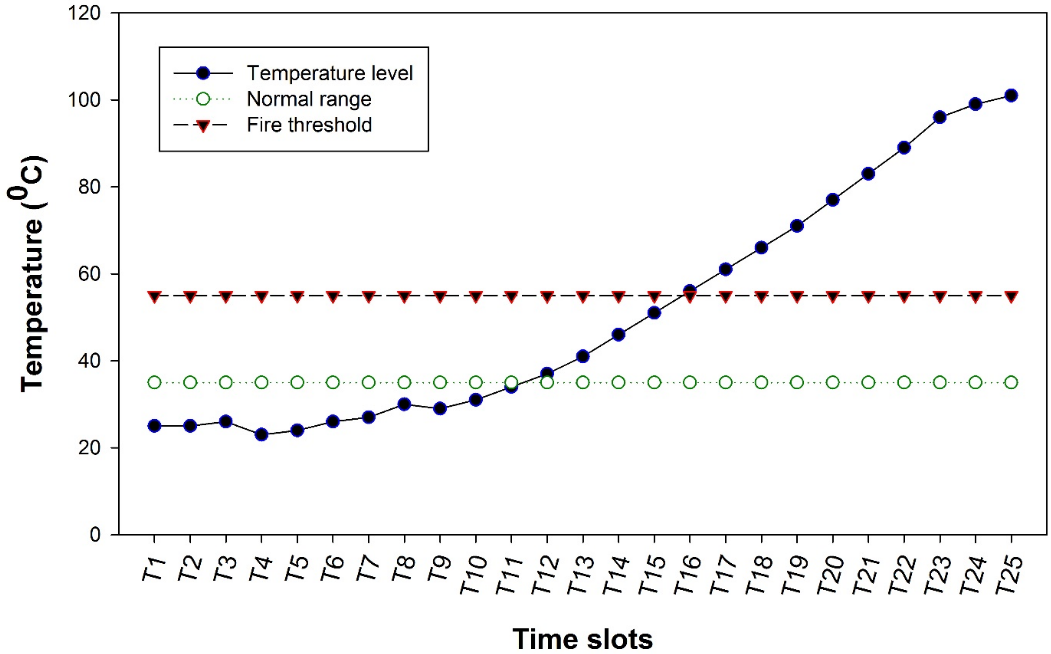


**Figure 4.3** : Fire monitoring through temperature analysis in a building.

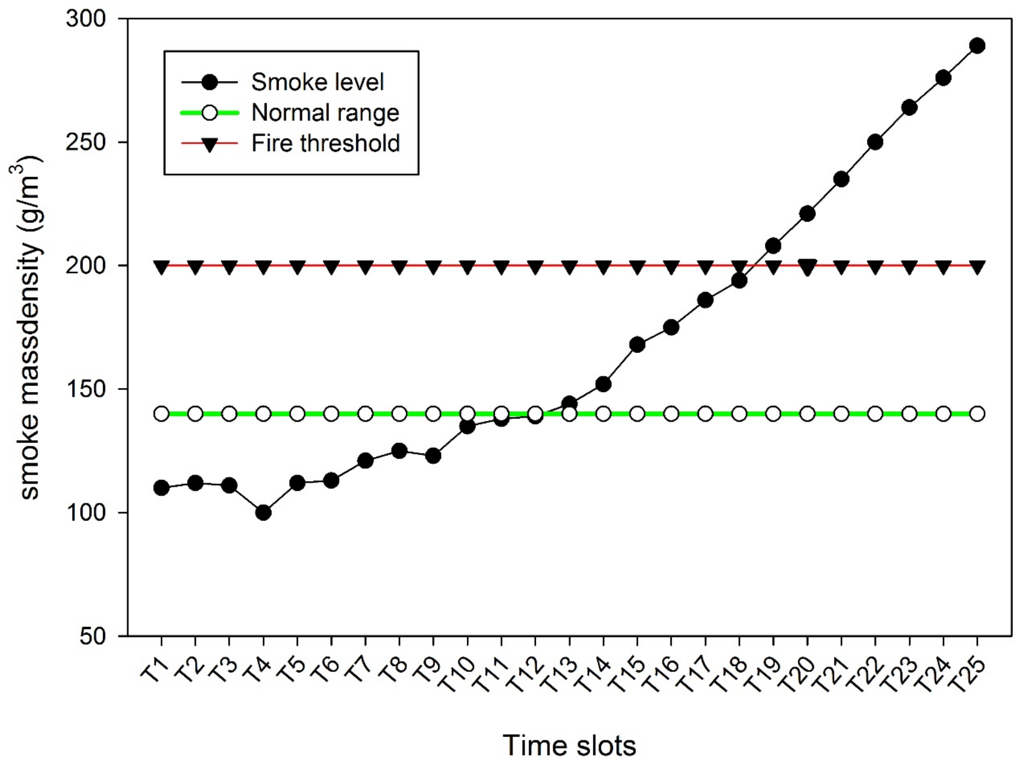


**Figure 4.4** : Smoke monitoring through smoke density in a building.

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 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 4.3 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 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 4.4, 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 [144] of Arhus city of 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 4.5 shows 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 4.6

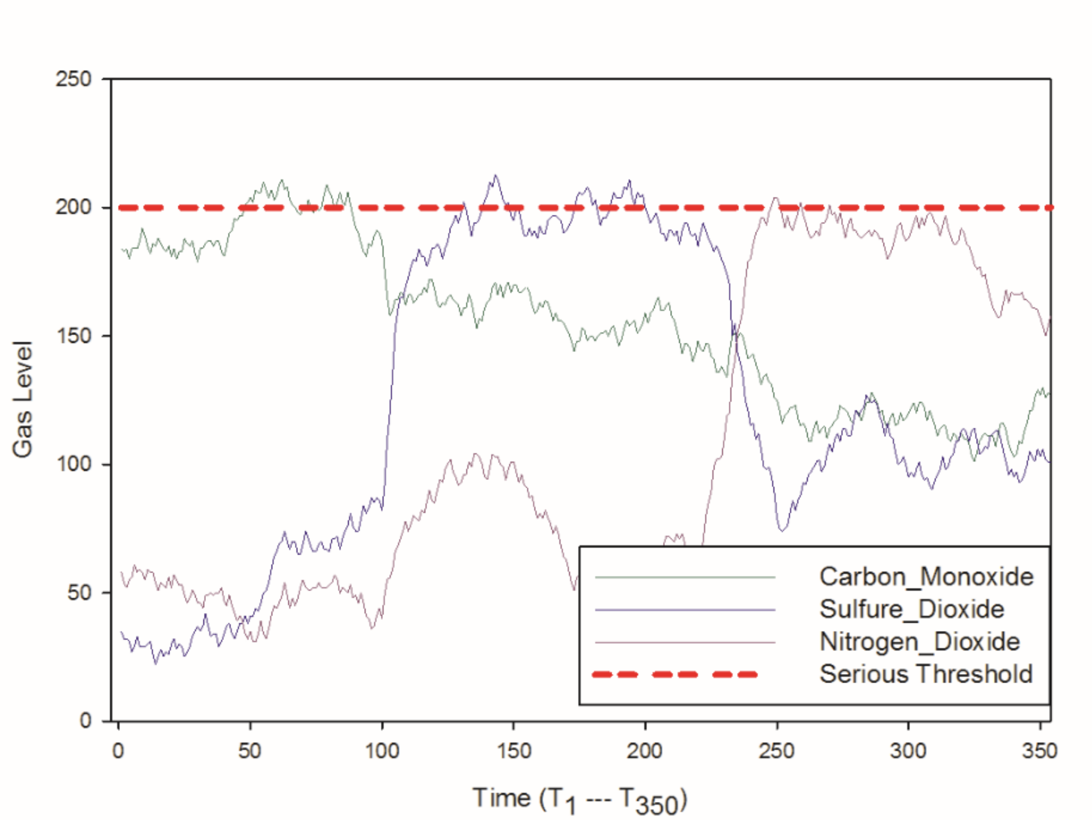


**Figure 4.5** : Pollution monitoring in a city through various gases.

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.

---

**Algorithm 3** Pollution Level Alert

---

**BEGIN**

Input: Air Quality Metrics *(M)*

Output: Pollution Alert/not-polluted

Steps:

**FOREACH** Gas_Readings (R) *(R_CO, R_SO2, R_NO2, R_O3)* in *(M)* **LOOP**
**IF** *(R_CO, R_SO2, R_NO2, R_O3)* >CT)
*Rep* =1
*Pollution_Alert();*
**ELSE**
*Rep* =0
*GoTo (Next_R)*
**ENDIF**
**END LOOP**

**END**

---

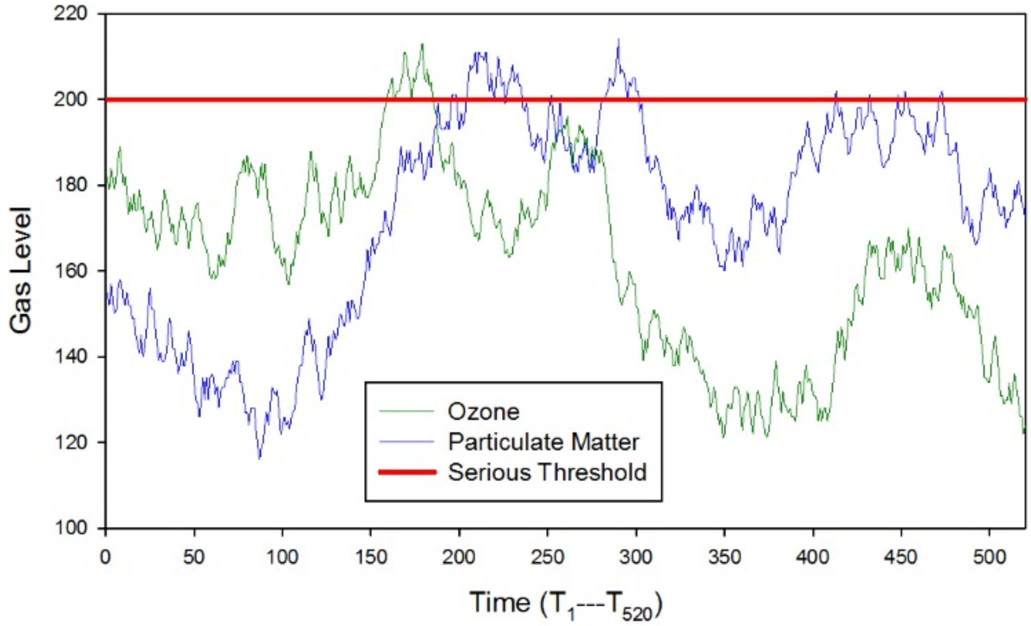

**Figure 4.6** : Pollution monitoring in a city through Ozone and particulate matter's level.

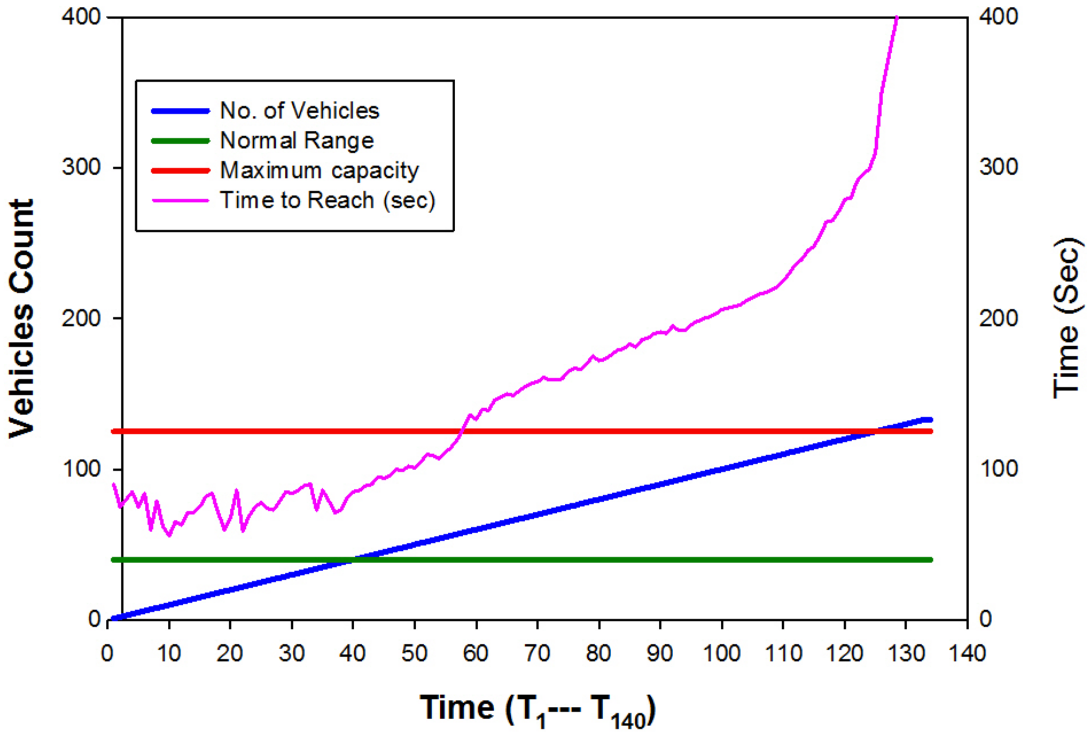

**Figure 4.7** : Traffic blockage analysis on a road.

**Algorithm 4** Route Blockage Alert

**BEGIN**

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

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

Steps:

Identify Time interval *(T) // (T)* is time to reach

Identify *(R) // (R)* is Route towards destination

**FOREACH** Reading *(Num_vehicles)* at *(T)* on *(R)* **LOOP**

**IF** *(Num_vehicles)* >CT)

*GoTo(Next_Reading)*

**ELSE**

*Blockage_Alert();*

*Alternative_Route (Assign New_Route (R))*;

**ENDIF**

**END LOOP**

**END**

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 [145]. 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 4.7. 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 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.

### 4.4.3 Analysis Results for Twitter datasets

#### 4.4.3.1 Case study 1: Indonesia

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 magnitude 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 [151].

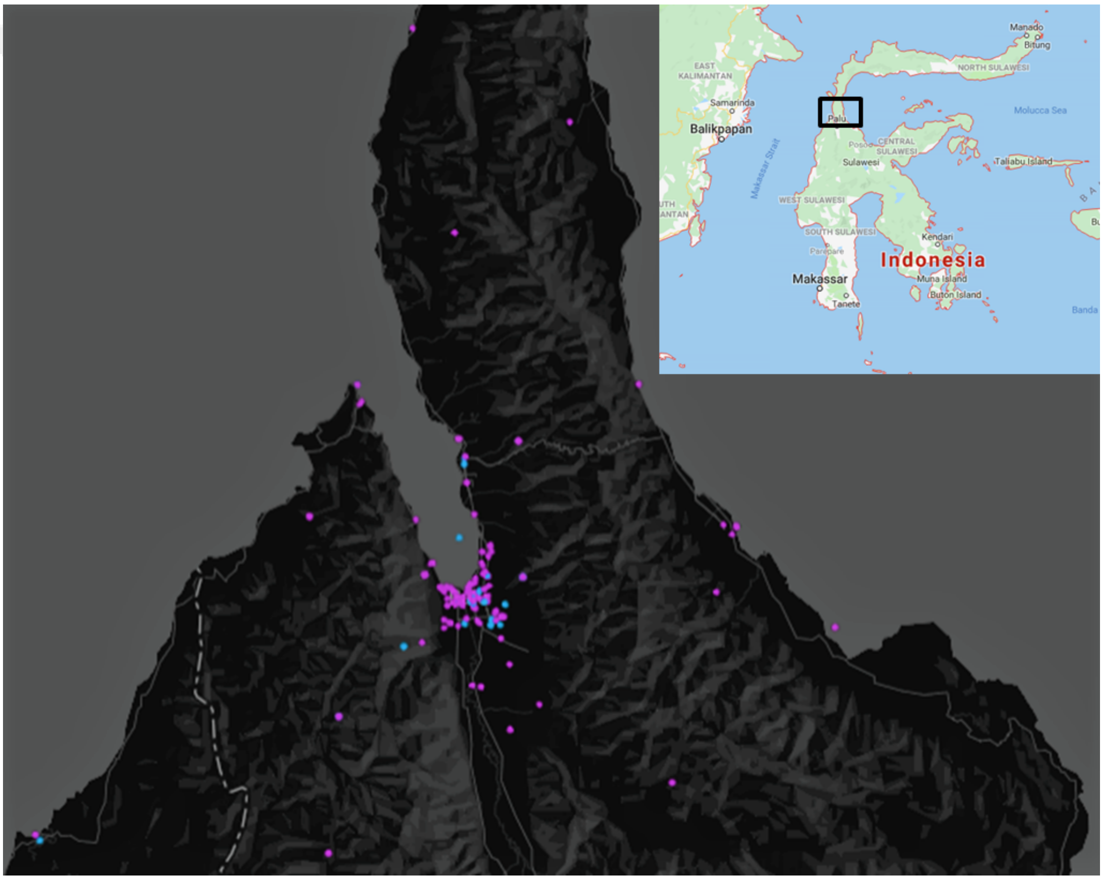


**Figure 4.8** : Overall Geocoded Tweet map of Palu, Indonesia from 28th to 29th September 2018.

For the Twitter-based analysis, we acquired data from the Twitter stream grab [146], 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 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.

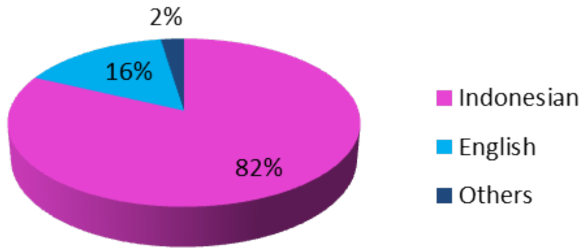


**Figure 4.9** : Major languages used for all geocoded tweets within Palu, Indonesia.

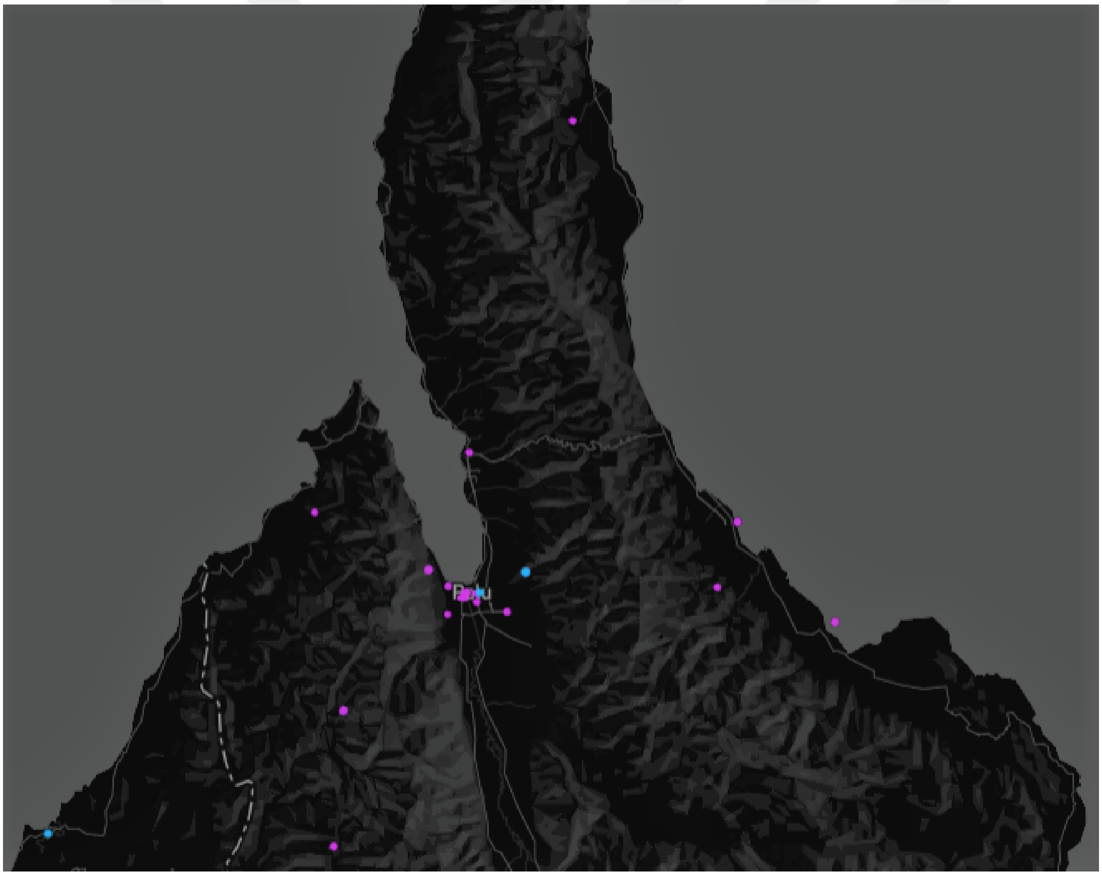

**Figure 4.10** : Geocoded Tweets found with #Earthquake and #Gempabumi in Palu, Indonesia.

A total of 981 geocoded tweets were collected within the specified range from 28th to 29th September 2018 as shown in Figure 4.8. Most of the twitter users do not enable the geo-location option while tweeting due to privacy concerns [152] and less than 5 percent of tweets have geo-coordinates attached with them [153]. Hence, the lesser number of tweets can be justified. The tweets were mostly tweeted in the Indonesian language (about 82%) as shown in Figure 4.9. The final results were mapped using MAPD [154] 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 desired results in a compact manner. We analyze all the geocoded tweets through hashtags, considering the main natural disasters i.e. (Earthquake and Tsunami).

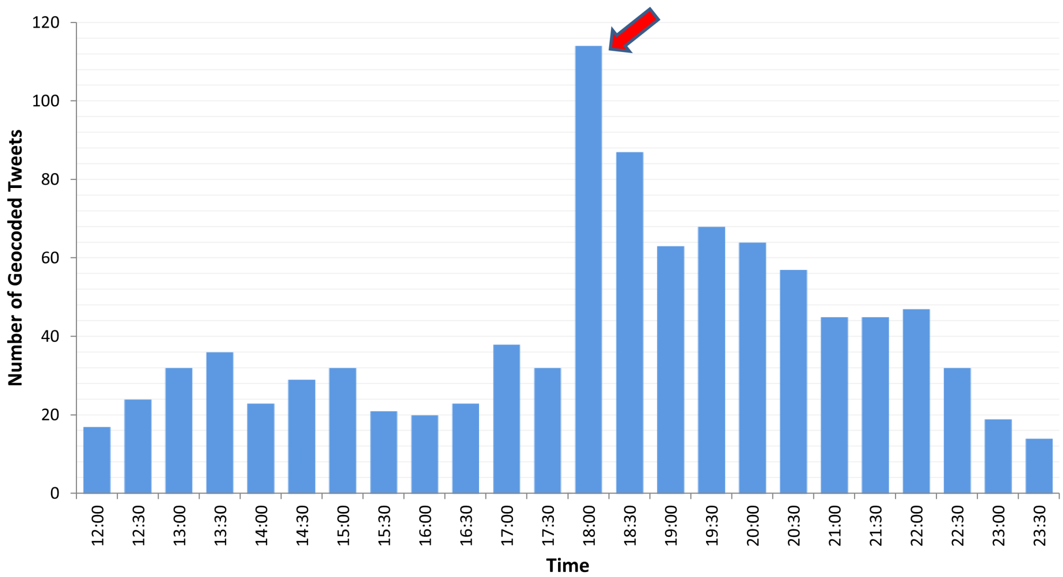


**Figure 4.11** : Total number of Tweets on 28th September with 30 minutes intervals.

A total number of 104 tweets were identified with hashtags of Earthquake and Gempabumi (Indonesian for earthquake). Figure 4.10 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. As can be seen in Figure 4.11 the most number of tweets for the day were reported between 18:00 to 18:30 and surely the earthquake struck on 18:02:44 has triggered this increase. Moreover, we tried to filter tweets with hashtags #Earthquake and #Gempabumi reported from time 17:50 to 18:30 and found out that within the first 10 minutes of the earthquake around 30 percent of the tweets contained the hashtags #Earthquake and #Gempabumi. Figure 4.12 compares the tweets with the defined hashtags and tweets without the defined hashtags around the earthquake timings. We also tried to form a word cluster of the words from the English language tweets being reported during the first 6 hours of the earthquake. As can be seen in Figure 4.13 the word magnitude is the most used word in the reported tweets to mention the magnitude of the earthquake i.e., 7.5 magnitude.

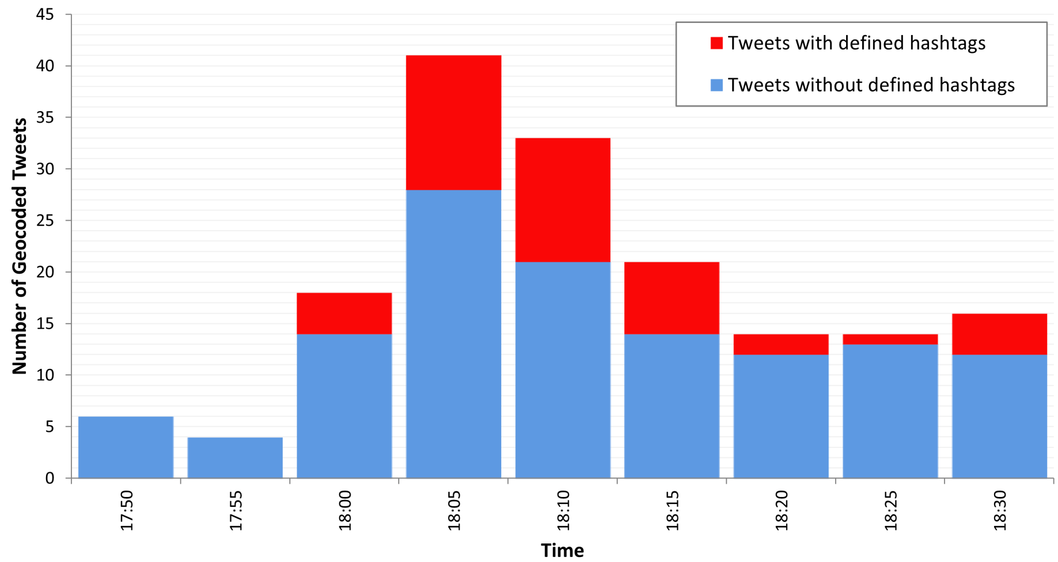


**Figure 4.12 :** Number of Tweets with and without defined (#Earthquake and #Gempabumi) hashtags.


**Figure 4.13 :** Word cluster with the main words in the filtered tweets.

After the earthquake, in approximately 30 minutes a six metre-high tsunami arrived to Palu city causing damage that was more devastating [155]. This scenario presents a very good case study to identify the role of twitter using alert generation thresholds and collecting information. The overall work-flow of Twitter data analysis is illustrated in

Figure. 4.14. As it is mentioned a range of 30 km was considered surrounding the center of Palu city to consider geocoded tweets within the specified coordinates. The tweets were the filtered with respect to the defined hashtags and the time-stamps they were reported on. With each 5 minutes span the number of tweets within and without hashtags were calculated using the proposed Equation 4.3 and the Twitter-based critical threshold range as shown in Table 4.5, so that crowdsourced early warning alerts are generated for such a situation. Moreover, once the alert is triggered we can also extract useful textual information and multimedia content i.e., images and videos from the filtered datasets in real-time.

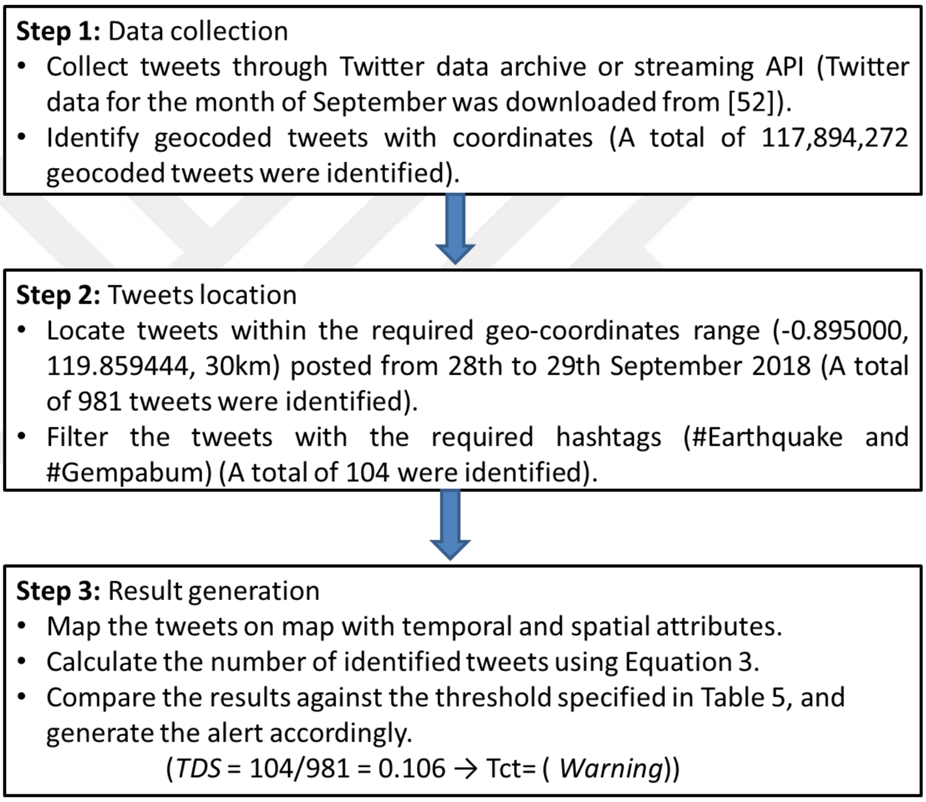


**Figure 4.14** : The workflow of Twitter data analysis.

#### 4.4.3.2 Case study 2: Turkey

In this section, we provide the application of geo-social data analysis component under the proposed architecture, aimed at detecting various disaster events in Turkey. To analysis the Geo-social media data component on the proposed architecture, we acquired data from the Twitter stream grab [146], 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. Initially, we collected 128

GB twitter data over 3 months (February, March, and April 2018) out of which we found more than 330 million geocoded tweets overall. We only focused on geocoded tweets from Turkey and hence filtered out tweets by country. A total of 17,531,215 geocoded tweets were discovered within Turkey from 1st February to 30th April 2018. The tweets were mostly tweeted in the Turkish language (about 81%) as shown in Figure. 4.15. The final results were mapped using MAPD [154] for temporally visualizing data.

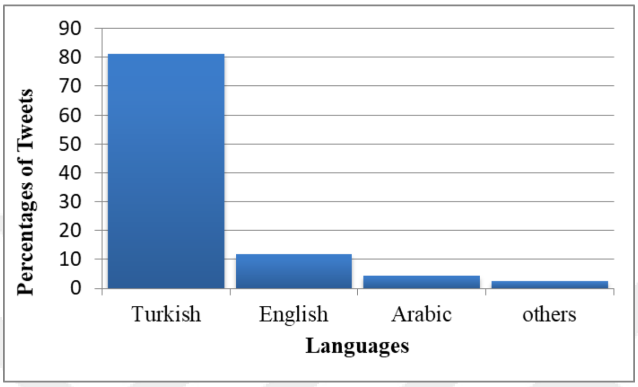


**Figure 4.15** : Major languages used for all geocoded tweets within Turkey

Figure. 4.16 shows the overall geocoded tweet map obtained after mapping the datasets for Turkey. Analyses are conducted on geocoded twitter data obtained from 1st February to 30th April 2018. As nearly 81% of tweets are in the Turkish language, we, therefore, analyzed only Turkish keywords instead of other languages. Moreover, it can be acknowledged through this map that most of the tweets are shared from urban areas and populous cities like Istanbul, Izmir, and Ankara.

The cross filtering capability of Twitter to analyze any activity with a given hashtag provides a great opportunity to acquire the desired results in a compact manner. We analyze all the geocoded tweets through hashtags, considering the main natural disasters that frequently occur in Turkey, such as earthquakes (Deprem) and storms (Firtina).

Figure. 4.17 shows the geocoded tweet map filtered with #Deprem (earthquake) in Turkey. A total of 1,663 geocoded tweets are extracted with #Deprem (earthquake). We found the most the tweets are tweeted from western coastal areas near the Aegean Sea and eastern Anatolian Faults of Turkey. These regions are parallel to tectonic active areas and all major earthquakes from the last 112 years in Turkey tend to occur in this region [156] [157]. Moreover, we found that majority of earthquakes during the

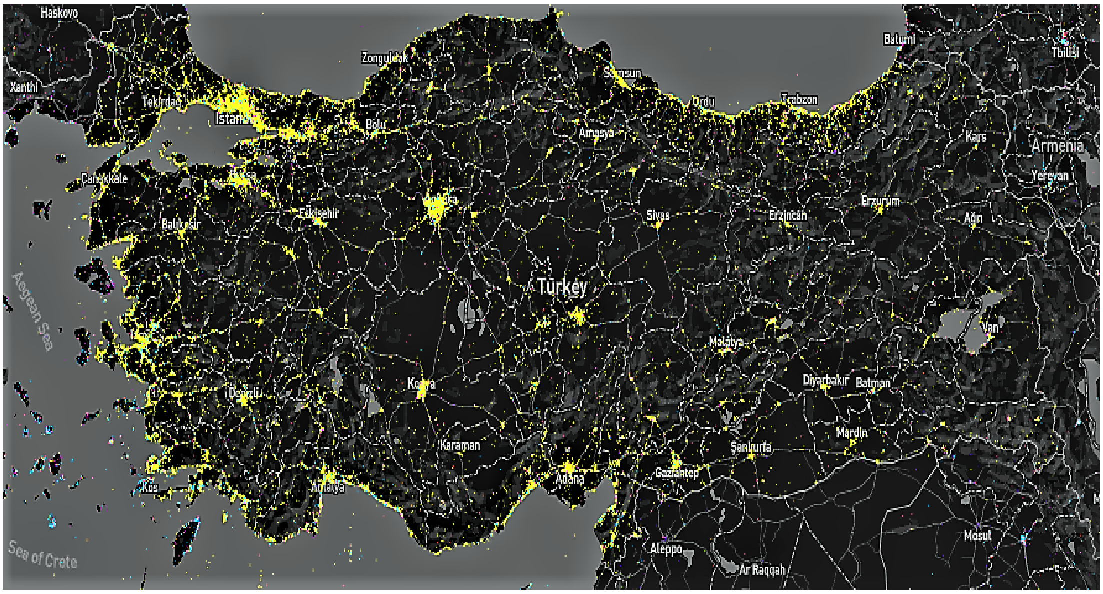


**Figure 4.16** : Overall geocoded Tweet map of Turkey from 1st February to 30th April 2018.

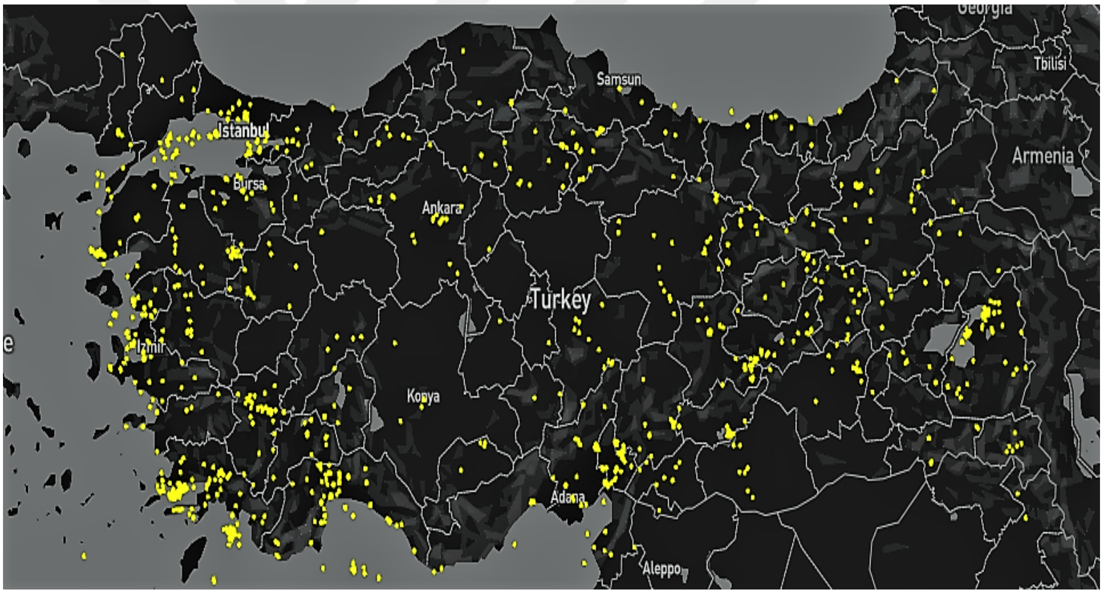


**Figure 4.17** : Geocoded Tweets found with #Deprem (earthquake) in Turkey.

studied 3 months that could be felt (>3 magnitudes), were reported in tweets within few minutes of its occurrence. Similarly, Figure. 4.18 shows the geocoded tweet map filtered with #Firtina which translates to both Storm and Hurricane in Turkish. A total of 68 geocoded tweets are extracted with #Firtina. Understandably, most of the tweets are tweeted from the western coastal area near the Aegean Sea while some are shared from the northeastern coastal area near the Black Sea.

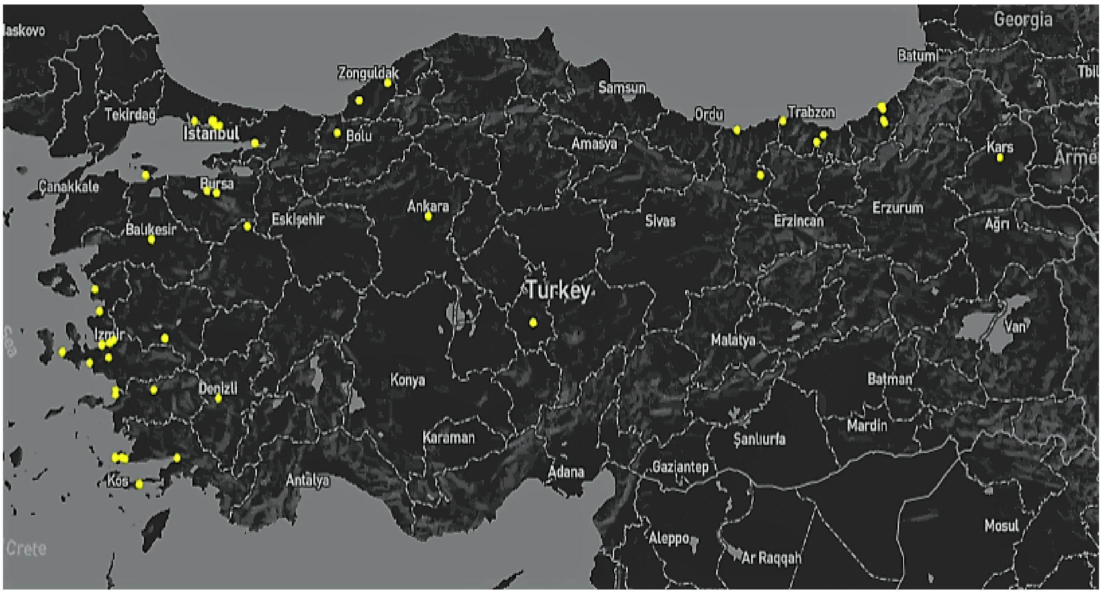


**Figure 4.18** : Geocoded Tweets found with #Firtina (Storm) in Turkey.

## 4.5 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.

### 4.5.1 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.

### 4.5.2 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.

### 4.5.3 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.

### 4.5.4 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 [116]. 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 [158]. Additionally, there is lack of adequate security tools for a number of technologies in the Hadoop Ecosystem [117]. 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 [159]. 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.

### 4.5.5 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.

### 4.5.6 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.

### 4.5.7 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.

## 4.6 Conclusion

The aim of this chapter is to contribute to the knowledge and guide the 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. Variety of datasets including IoT-based smarty city and twitter datasets are analyzed for showing the validity and evaluation of the proposed DRSC concept.

# 5. PROPOSED SCHEME IMPLEMENTATION AND PERFORMANCE EVALUATION

## 5.1 Abstract

A Disaster Resilient Smart City (DRSC) environment is a complete architecture and can be implemented according to the needs and system capability for any smart city initiative. For research purpose, this study tries to evaluate the performance of a small-scale implementation with the already mentioned data filtration and normalization techniques. The chapter displays the overall processing and throughput results with different data sizes. Processing time and throughput of the proposed scheme are compared with different perspectives to evaluate the system efficiency and performance. Single node MapReduce with filtered datasets is compared with dual node MapReduce Hadoop cluster to verify the implemented techniques. Moreover Apache Spark implementation with Hadoop is also tested. At last, a comparison is made with other published works to demonstrate the contribution of the proposed schemes.

## 5.2 System Application

Hadoop ecosystem and Spark engines can be deployed on any commodity hardware. The main 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 5.1.

## 5.3 System Evaluation

Since, this study focused on processing large datasets that requires efficient real-time processing, therefore the system was evaluated with regards to data processing and

**Table 5.1** : The hard 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 |

throughput considering the increasing data size. Figure 5.1 shows the processing time efficiency result corresponding to the increasing dataset size on various data integration points. It is expectable that with the increasing data, the processing time also rises. However, with the proposed scheme, the rise in the processing time is quite lower corresponding to the huge rise in the data size. Figure 5.2 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 the study in hand 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 due to parallel processing implementation using MapReduce programing paradigm and number of simultaneous nodes of Hadoop. As the study also processed a huge set of tweets for alert generation process, therefore processing time for Twitter dataset is shown in Figure 5.3. 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.

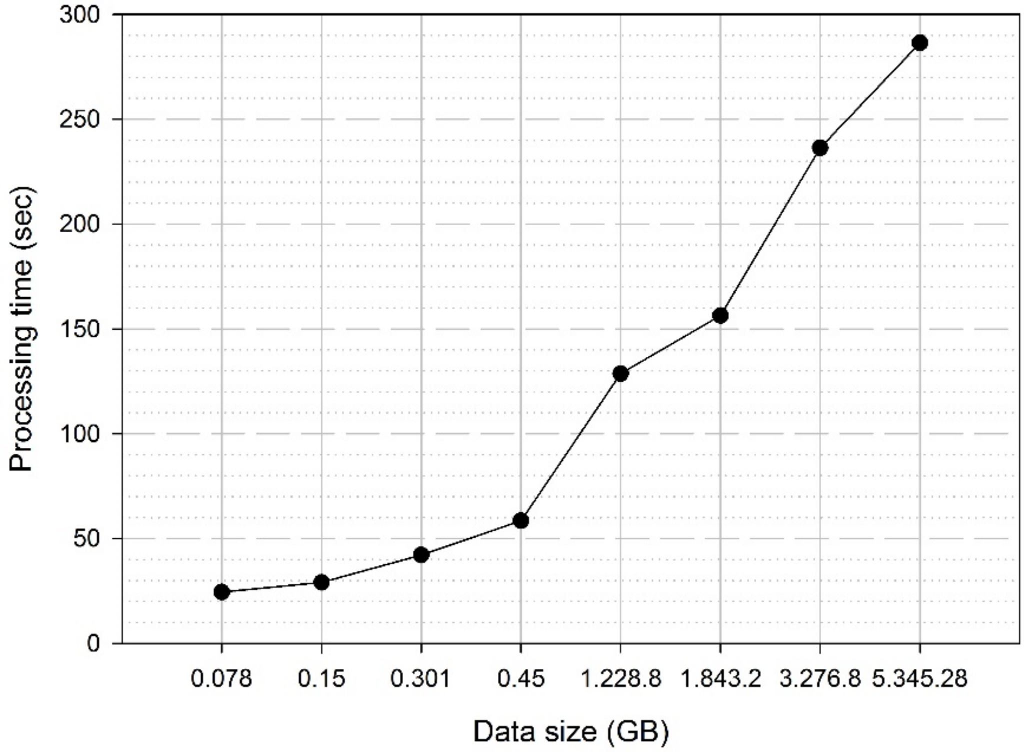


**Figure 5.1** : System's processing time efficiency with increasing dataset size.

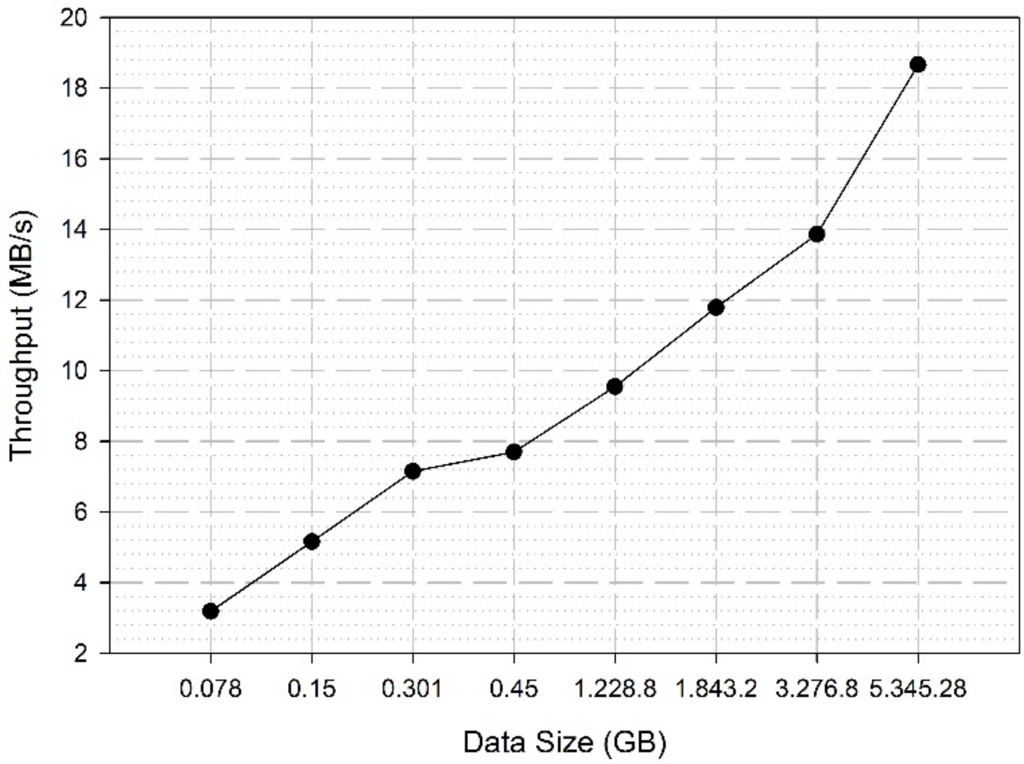


**Figure 5.2** : System's throughput efficiency with increasing dataset size.

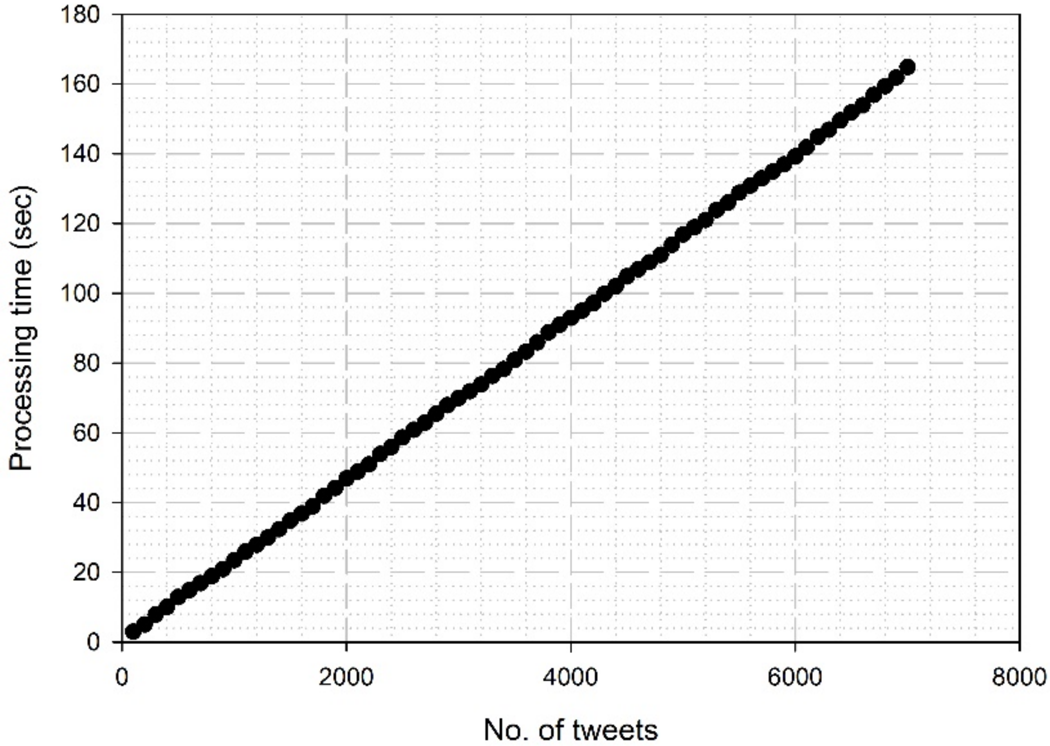


**Figure 5.3** : Efficiency of the system with respect to processing time for increasing number of tweets.

Data filtering and normalization techniques have sufficiently dragged down the processing time and have increased throughput. This is due to the removal of noisy data according to the defined scheme of the filtration and normalization processes. With the established Apache Spark and dual node Hadoop cluster, the study opted to evaluate the processing and throughput of filtered and generic (non-filtered) datasets. Figure 5.4 shows the comparison between Apache Spark and dual node Hadoop cluster with filtered and generic datasets respectively. As depicted with the filtered datasets the processing time for both Apache Spark and dual node Hadoop have significantly reduced. On the other hand, Figure 5.5 reveal that the throughput has increased for both Apache Spark and dual node Hadoop cluster with the filtered datasets. As a single processor follows a constant throughput as it cannot divide its tasks for parallel data processing. However, Hadoop adheres a multicore application and applies the parallel execution of tasks to ensure maximum benefit from the available cores. Due to this ability, the throughput is directly proportional to the data size and increases with it.

So, with smaller datasets, the throughput is also lesser that makes Hadoop not feasible for smaller datasets requiring higher throughput.

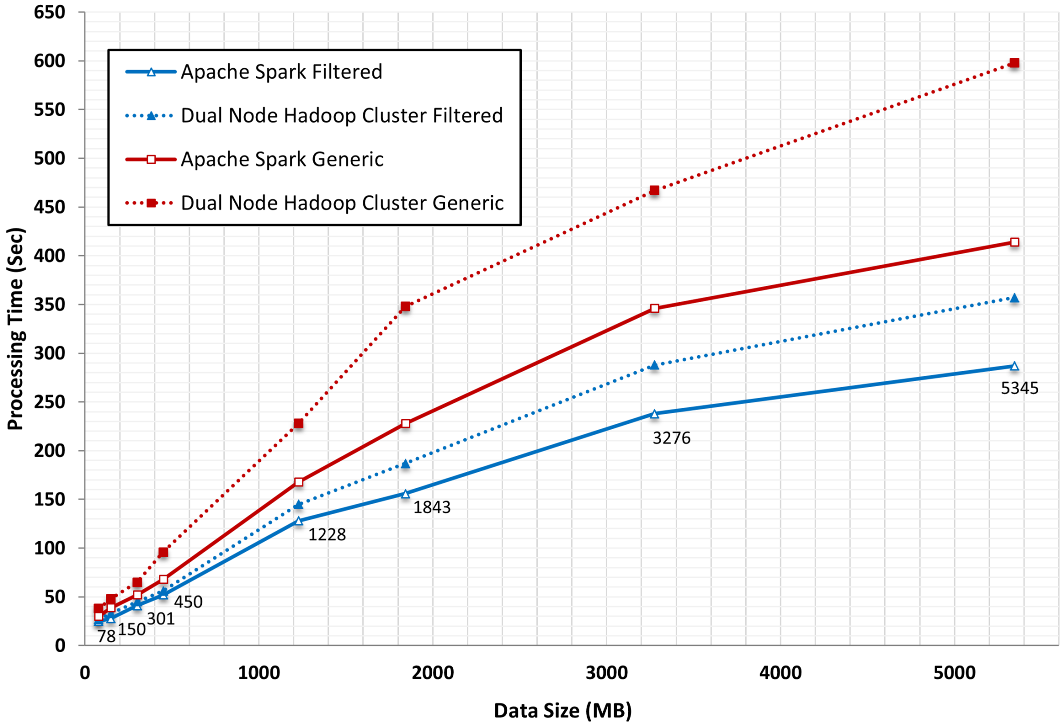


**Figure 5.4** : Processing time comparison between Apache Spark and dual node Hadoop cluster with filtered and generic datasets.

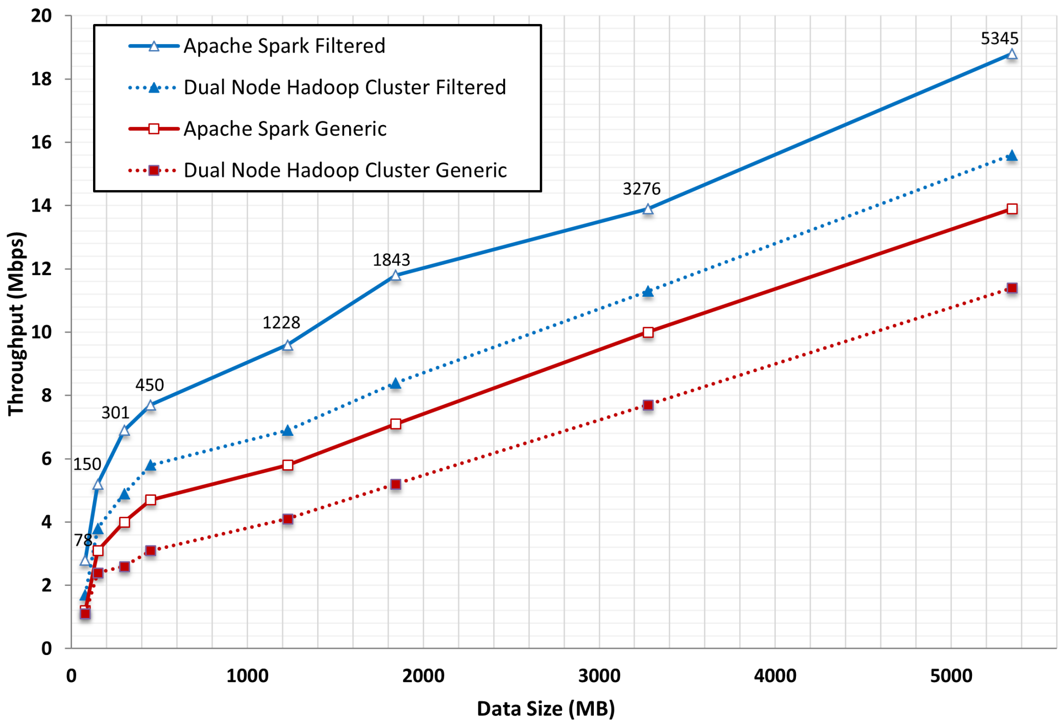


**Figure 5.5** : Throughput comparison between Apache Spark and dual node Hadoop cluster with filtered and generic datasets.

In this research, MapReduce algorithm is also utilized for testing purpose to process the datasets. As MapReduce operates by dividing the processing into two stages i.e., Map and Reduce. Each stage contains Key and Value as their input and output respectively. The users depending on the type of data can assign keys and values accordingly. These Key and Value pairs are actually the record entity that MapReduce job obtains for processing. Figure 5.6 [160] shows the general funcationality of a MapReduce program. By default, Map input gets line offset as the key and the content of the line is considered the value with TextInputFormat. As an example of a MapReduce operation for temperature dataset Algorithm 1 shows the Mapper function while Algorithm 2 depicts the Reducer function for temperature dataset.

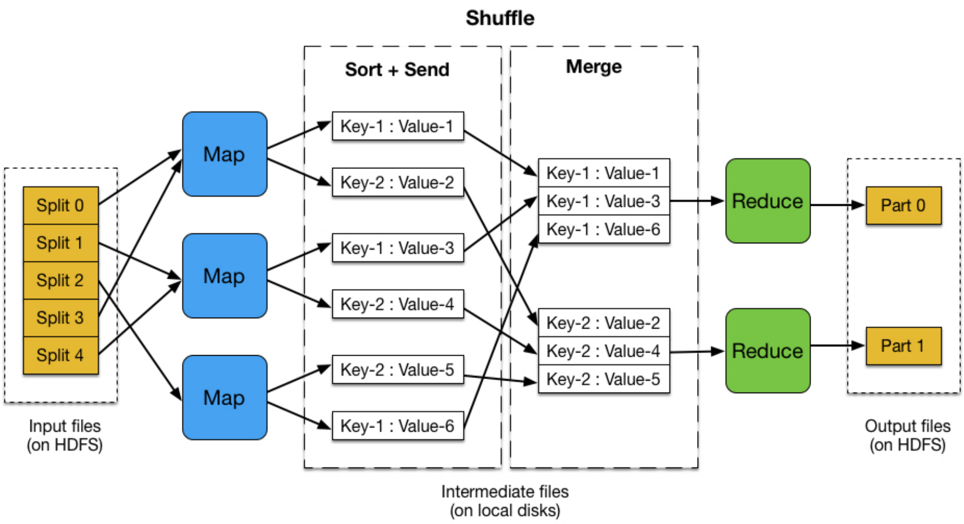


**Figure 5.6** : MapReduce function illustration.

---

**Algorithm 5** Mapper for Temperature Dataset

---

**BEGIN**
**Input**
Key: *Line-offset*
Value: *Row*
**Output**
Key:= *TimeStamp*
Value:= *TempReading*
Steps:
1: TimeStamp, TempReading:= line.split('\t')
2: Key:=TimeStamp
3: Value:=TempReading

Yield (Key, Value)
**END**

---

**Algorithm 6** Reducer for Temperature Dataset

**BEGIN**
**Input**
Key: *TimeStamp*
Value: *TempReading*
**Output**
Key:= *NotifiedTimeStamp*
Value:= *NotifiedTempReading*
Steps:
Initialize CT (Critical Threshold)
**FOREACH** *(TempReading )* at (TimeStamp) **LOOP**
**IF** *(TempReading )* >CT)
List.append (TempReading)
Key:= TimeStamp
Value:= ListReading
Yield (Key, Value)
**ELSE**
*Next_Reading();*
**ENDLOOP**
**END**

The study also tested the filtered datasets with implemented MapReduce on a single and dual node Hadoop cluster. Figure 5.7 shows the processing time for single node MapReduce Hadoop and dual node MapReduce Hadoop for filtered and generic datasets. Similarly, Figure 5.8 shows the throughput for the same scenario. The throughput gradually increases with the data size. As single core executions tend to have an unchanged throughput level as it does not support parallel computation and since cannot distribute its tasks. On the other hand, multicore executions such as Hadoop uses available multiple cores at the same time and since attain the maximum usage of the available processing power. Due to this phenomenon, the throughput increases with the dataset along with the usage of computing core. However, throughput tends to be less with smaller datasets as the cores are not fully utilized. Similarly if the computing capacity has reached its maximum tenancy the throughput level will be on a constant level as well.

As shown the results reveal that for both Spark and MapReduce the throughput has increased with the increase in data size. It can be noted that with the filtered datasets the processing time and throughput of a single node are much better than the same generic datasets on dual node MapReduce Hadoop cluster. The proposed scheme reduced the processing time by 19.8 percent and increased the throughput by 14.9 percent of filtered single node MapReduce Hadoop comparative to generic dual node MapReduce Hadoop cluster. Thus, implementing the proposed scheme can save time and expenditure for multiple nodes.

In short, with proper filtration and normalization techniques, larger datasets can have enhanced processing time and throughput for analysis. We evaluated the processing time and throughput with a large dataset (5345 MB) and examined quite good enhancements in the proposed system.

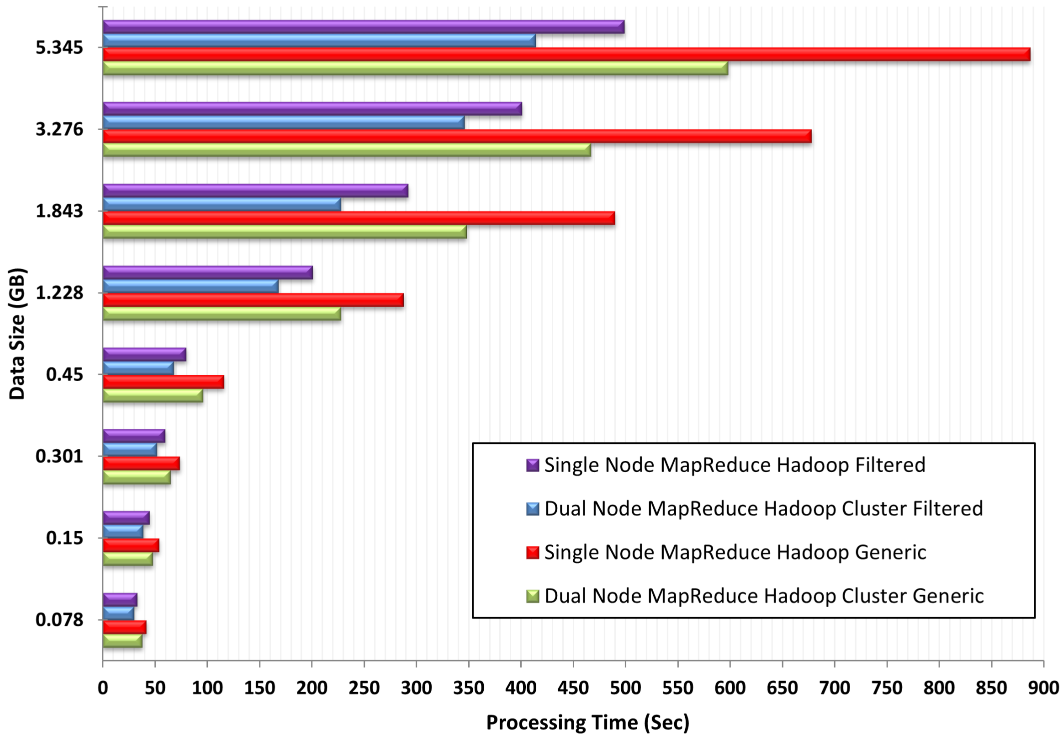


**Figure 5.7** : Processing time for single node MapReduce Hadoop and dual node MapReduce Hadoop for filtered and generic datasets.

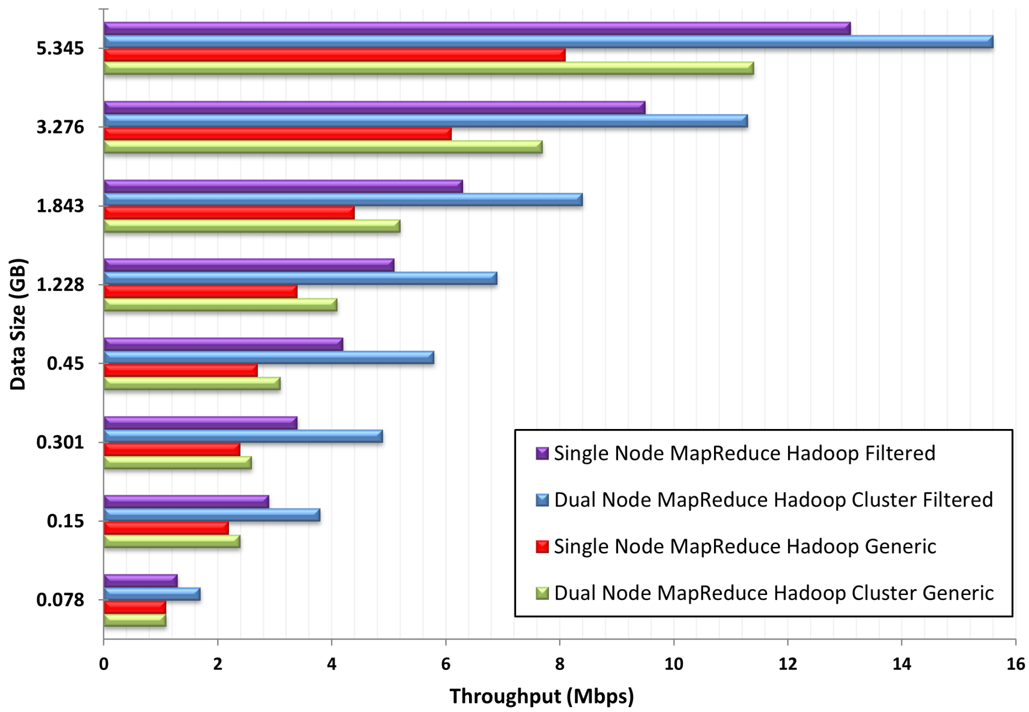


**Figure 5.8** : Throughput for single node MapReduce Hadoop and dual node MapReduce Hadoop for filtered and generic datasets.

These results show a clear advantage even if the proposed scheme is compared with other published related work, both in terms of processing time and throughput. For example, Silva et.al [132] proposed a big data analytics embedded smart city architecture to analysis major types of datasets on Hadoop ecosystem. They have used Kalman filter to achieve data filtration in their proposed framework. Kalman filter works as an optimal estimator that is useful in remving noise from the datasets. Similarly, in [40] a smart city architecture was implemented using Hadoop with Spark, Storm and voltDB for real-time processing of IoT datasets to produce analytical results. They performed the data filtration through data classification techniques. The proposed scheme in this study is compared with these two studies as their implementation model is based on single and dual node Hadoop with Spark application and the datasets are similar, although the data filtration techniques varies.

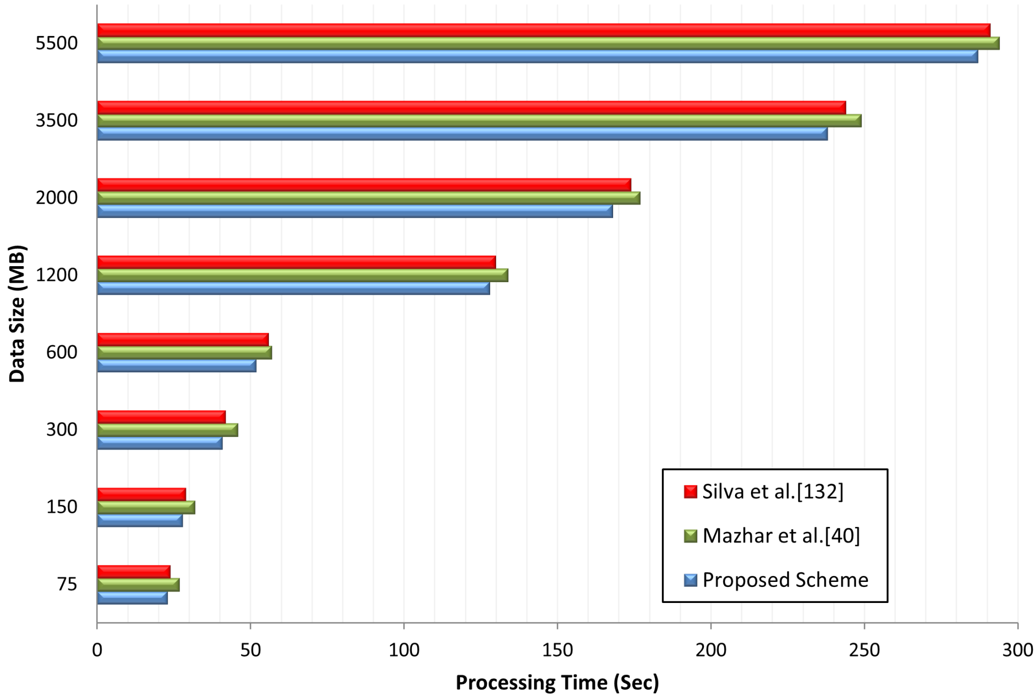


**Figure 5.9** : Processing time comparison with other studies.

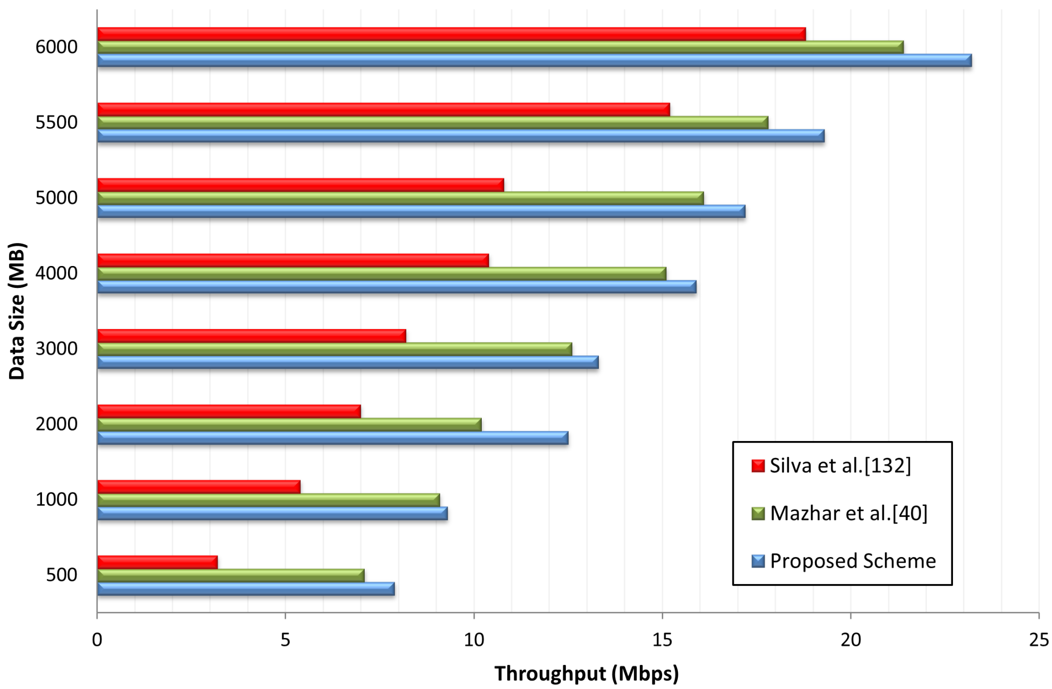


**Figure 5.10** : Throughput comparison with other studies.

The proposed scheme has attained the optimal processing rate with respective to the datasize as compared to the published works [132] [40]. As can be seen in Figure 5.9 the processing time of the processing is much lesser with both smaller and large datasets. Similarly, as depicted in Figure 5.10, the proposed scheme is leading from the other published schemes all the way through but the lead has increased with 4000 MB datasize in terms of data throughput.

## 5.4 Summary

The performance evaluation of the proposed scheme reveals some interesting results. Apache Spark and dual node Hadoop cluster clearly deserves to be implemented even on small scale environments for better results. The data filtration and normalization techniques needs to be carefully selected as it is noted that with the filtered datasets the processing time and throughput of a single node are much better than with the same generic datasets on dual node MapReduce Hadoop cluster. With the removal of noisy and ambiguous data from the analysis dataset through proper filtration and normalization phases, the overall performance of the proposed architecture has improved. The comparison with other existing schemes shows performance superiority, which is the major achievement of the proposed system.

## 6. CONCLUSIONS AND RECOMMENDATIONS

The fusion of BDA technologies and IoT promises a new and more effective approach for carrying out the core operations of disaster management processes. With state-of-the-art big data analytical tools and well-managed IoT, we can not only harvest large volumes of valuable data from multiple data sources but can also generate required results in real-time for effective decision-making.

The evolution in Internet of Things (IoT) technologies, offering integration of heterogeneous sensors and smart devices using advanced network services; and with the expansion of social big data, offering the potential to extract useful insights from a huge set of unstructured data, a disaster-resilient smart city framework can be developed to predict, extract affirmative knowledge, monitor and analysis the disasters occurrence. 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.

In the context of integrating IoT environments equipped with multiple data sources such as cameras, sensors, smartphones, etc., and BDA technologies assisting in data processing, a number of data sources can be incorporated to gather new and valuable insights and information. Engaging multiple data sources provide alternative ways to address problems that require multidimensional representations of the data to extract the common patterns for a solution that is inaccessible through a single source of data. With the availability of diverse and rich data sources, BDA- and IoT-based disaster management environments can surpass conventional DMSs data sources.

One such data source is social media platforms, as they are considered very useful to collect information in case of any disaster because it is the fastest and the cheapest source to provide effective, updated and relevant information for decision making. The practical use of such applications is making a vast number of academic studies to research on many aspects of social media in disaster management. Social media provides its user the opportunity to contribute and disseminate valuable information, be it in the shape of text, pictures, audio, and video; that is necessary for disaster management processes and communications.

In section 3.4 of this thesis, a design model is presented for the development of an integrated system consisting of social media crowd-sourced component and a designed web API component through which organized and reliable data can be provided for real-time disaster management. This design-science research demonstrates that the concept of social media crowd-sourcing can effectively be used for real-time disaster management and tries to aid the theory of making crowd-sourced data as trustworthy as other data sources. The basic theme of this design is to make the unstructured crowd-sourced data process-able so that it can be compared and merged with structured data sources such as a web API. The effectiveness of real-time crowd-sourced disaster management systems has been proven but there are many gaps and challenges in this research domain. The design science to model integrating frameworks plays a key role in providing the basis for interdisciplinary research to be carried out.

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. Moreover, a lot of research is still required to productively model and implement BDA and IoT paradigms, keeping in view the time constraint and accuracy demands of disaster management processes. The aim of this thesis is to contribute to the knowledge and future research of the design and implementation of BDA- and IoT-based disaster-resilient smart cities. This research can provide references for other researchers and industries for future acquisitions in the domain of smart cities and disaster management.

This thesis provides a thorough outline of how BDA 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. In section 4.3.2, the Hadoop ecosystem with Spark is utilized to implement the complete system for a variety of datasets including IoT-based smarty city and twitter datasets. These separate 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. The proposed scheme mainly targets processing large datasets that require efficient real-time processing, therefore the implemented system was evaluated with regards to data processing and throughput considering the increasing data size. Data filtering and normalization techniques have sufficiently dragged down the processing time and have increased throughput. The study evaluated various cases of Apache Spark, single and dual node MapReduce Hadoop cluster with generic and filtered datasets to compare the performance of various deployed schemes. The evaluation of the system efficiency is measured in terms of processing time and throughput and is compared with other studies that demonstrates the performance superiority of the proposed architecture.

For future work, the study anticipates the addition of various other 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, **Technical Report**, International Federation of Red Cross, `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, **Technical Report 1**, Swiss Re Institute, `http://media.swissre.com/documents/sigma1{\_}2018{\_}en.pdf`.

[3] **Hristidis, V.**, **Chen, S.C.**, **Li, T.**, **Luis, S. and Deng, Y.** (2010). Survey of data management and analysis in disaster situations, *Journal of Systems and Software*, *83*(10), 1701–1714, `http://linkinghub.elsevier.com/retrieve/pii/S0164121210001329`.

[4] **Mohammadi, M.** *et al.* (2018). Deep Learning for IoT Big Data and Streaming Analytics: A Survey, *IEEE Communications Surveys and Tutorials*, `1712.04301`.

[5] **Silva, B.N.**, **Khan, M. and Han, K.** (2018). Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities, *Sustainable Cities and Society*, *38*(August 2017), 697–713, `https://doi.org/10.1016/j.scs.2018.01.053`.

[6] **Zeng, D.**, **Guo, S. and Cheng, Z.** (2011). The web of things: A survey, *Journal of Communications*, *6*(6), 424–438.

[7] **Huang, X.**, **Xie, K.**, **Leng, S.**, **Yuan, T. and Ma, M.** (2018). Improving Quality of Experience in multimedia Internet of Things leveraging machine learning on big data, *Future Generation Computer Systems*, *86*, 1413–1423, `https://doi.org/10.1016/j.future.2018.02.046`.

[8] **Birregah, B.**, **Top, T.**, **Perez, C.**, **Châtelet, E.**, **Matta, N.**, **Lemercier, M. and Snoussi, H.** (2012). Multi-layer crisis mapping: A social media-based approach, *Proceedings of the Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, WETICE*, 379–384.

[9] **Goodchild, M.F.** (2007). Citizens As Sensors: the World of Volunteered Geography, *GeoJournal*, *69*(4), 211–221.

[10] **Haworth, B. and Bruce, E.** (2015). A Review of Volunteered Geographic Information for Disaster Management, *Geography Compass*, *9*(5), 237–250.

[11] **Houston, J.B.**, **Hawthorne, J.**, **Perreault, M.F.**, **Park, E.H.**, **Goldstein Hode, M.**, **Halliwell, M.R. and ... Griffith, S.A.** (2015). Social media and disasters: A functional framework for social media use in disaster planning, response, and research, *Disasters*, *39*(1), 1–22.

[12] **Simon, T.**, **Goldberg, A. and Adini, B.** (2015). Socializing in emergencies - A review of the use of social media in emergency situations, *International Journal of Information Management*, *35*(5), 609–619.

[13] **Wang, Z.**, **Mao, S.**, **Yang, L. and Tang, P.** (2018). A survey of multimedia big data, *China Communications*, *15*(1), 155–176.

[14] **Crowley, J.** (2013). Connecting grassroots and government for disaster response, **Technical Report**, Commons Lab, Wilson Center.

[15] **Rathore, M.M.**, **Paul, A.**, **Ahmad, A. and Rho, S.** (2015). Urban Planning and Building Smart Cities based on the Internet of Things using Big Data Analytics, *Computer Networks*, *101*(2016), 63–80, `http://dx.doi.org/10.1016/j.comnet.2015.12.023`.

[16] **Rodríguez-Mazahua, L.**, **Rodríguez-Enríquez, C.A.**, **Sánchez-Cervantes, J.L.**, **Cervantes, J. and ... Alor-Hernández, G.** (2016). A general perspective of Big Data: applications, tools, challenges and trends, *Journal of Supercomputing*, *72*(8), 3073–3113.

[17] **Zaslavsky, A.**, **Perera, C. and Georgakopoulos, D.** (2013). Sensing as a Service and Big Data, *arXiv preprint arXiv:1301.0159*, `http://arxiv.org/abs/1301.0159`, `1301.0159`.

[18] **Chen, M.**, **Mao, S. and Liu, Y.** (2014). Big data: A survey, *Mobile Networks and Applications*, *19*(2), 171–209.

[19] **Meissner, A.**, **Luckenbach, T.**, **Risse, T.**, **Kirste, T. and Holger Kirchner** (2002). Design Challenges for an Integrated Disaster Management Communication and Information System, *The First IEEE Workshop on Disaster Recovery Networks (DIREN 2002)*, *24*, `http://www.l3s.de/{~}risse/pub/P2002-01.pdf`.

[20] **Davenport, T.**, **Barth, P. and Bean, R.** (2012). How big data is different, *MIT Sloan Management Review*, *54*(1), 43.

[21] **Mehrotra, S.**, **Qiu, X.**, **Cao, Z. and Tate, A.** (2013). Technological challenges in emergency response, *IEEE Intelligent Systems*, *28*(4), 5–8.

[22] **Akter, S. and Wamba, S.F.** (2017). Big data and disaster management: a systematic review and agenda for future research, *Annals of Operations Research*, 1–21.

[23] **Baham, C.**, **Hirschheim, R.**, **Calderon, A.A. and Kisekka, V.** (2017). An Agile Methodology for the Disaster Recovery of Information Systems Under Catastrophic Scenarios, *Journal of Management Information Systems*, *34*(3), 633–663, `https://doi.org/10.1080/07421222.2017.1372996`.

[24] **Kapucu, N.** (2006). Interagency Communication Networks During Emergencies, *The American Review of Public Administration*, *36*(2), 207–225, `http://journals.sagepub.com/doi/10.1177/0275074005280605`.

[25] **Watson, H.**, **Finn, R.L. and Wadhwa, K.** (2017). Organizational and Societal Impacts of Big Data in Crisis Management, *Journal of Contingencies and Crisis Management*, *25*(1), 15–22.

[26] **Zheng, L.** *et al.* (2013). Data mining meets the needs of disaster information management, *IEEE Transactions on Human-Machine Systems*, *43*(5), 451–464.

[27] **Neville, K.**, **O'Riordan, S.**, **Pope, A.**, **Rauner, M.**, **Maria Rochford**, **Madden, M.**, **Sweeney, J. and ... O'Brien, C.** (2016). Towards the development of a decision support system for multi-agency decision-making during cross-border emergencies, *Journal of Decision Systems*, *25*, 381–396, `http://dx.doi.org/10.1080/12460125.2016.1187393`.

[28] **Ortmann, J.**, **Limbu, M.**, **Wang, D. and Kauppinen, T.** (2011). Crowdsourcing Linked Open Data for Disaster Management, *Proceedings of Terra Cognita, the 10th International Semantic Web Conference*, January, pp.11–22.

[29] **Alamdar, F.**, **Kalantari, M. and Rajabifard, A.** (2016). Towards multi-agency sensor information integration for disaster management, *Computers, Environment and Urban Systems*, *56*, 68–85, `http://dx.doi.org/10.1016/j.compenvurbsys.2015.11.005`.

[30] **de Albuquerque, J.P.**, **Herfort, B.**, **Brenning, A. and Zipf, A.** (2015). A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management, *International Journal of Geographical Information Science*, *29*(4), 667–689.

[31] **Draheim, D.**, **Horn, M. and Schulz, I.** (2004). The Schema Evolution and Data Migration Framework of the Environmental Mass Database IMIS, *Proceedings of SSDBM 2004 – The 16th International Conference on Scientific and Statistical Database Management*, IEEE Press, pp.341–344.

[32] **Security, I.T.C.I.. and resilience** (2012). *ISO 22313:2012 – Societal security – Business continuity management systems – Guidance*, International Organization for Standardization.

[33] **Poslad, S.**, **Middleton, S.E.**, **Chaves, F.**, **Tao, R.**, **Necmioglu, O. and Bugel, U.** (2015). A Semantic IoT Early Warning System for Natural Environment Crisis Management, *IEEE Transactions on Emerging Topics in Computing*, *3*(2), 246–257.

[34] **Fang, S.**, **Xu, L.D.**, **Member, S.**, **Zhu, Y.**, **Ahati, J.**, **Pei, H.**, **Yan, J. and Liu, Z.** (2014). An Integrated System for Regional Environmental Monitoring and Management Based on Internet of Things, *IEEE Transactions on Industrial Informatics*, *10*(2), 1596–1605.

[35] **Wei, Y.**, **Huang, C.**, **Li, J. and Xie, L.** (2016). An evaluation model for urban carrying capacity, *Habitat International*, *53*, 87–96, `http://dx.doi.org/10.1016/j.habitatint.2015.10.025`.

[36] **Albino, V.**, **Berardi, U. and Dangelico, R.M.** (2015). Smart cities: Definitions, dimensions, performance, and initiatives, *Journal of Urban Technology*, *22*(1), 1–19, `http://dx.doi.org/10.1080/10630732.2014.942092`.

[37] **Kitchin, R.** (2014). The real-time city? Big data and smart urbanism, *GeoJournal*, *79*(1), 1–14.

[38] **Gubbi, J.** *et al.* (2013). Internet of Things (IoT): A vision, architectural elements, and future directions, *Future Generation Computer Systems*, *29*(7), 1645–1660, `1207.0203`.

[39] **Yang, L.**, **Yang, S.H. and Plotnick, L.** (2013). How the internet of things technology enhances emergency response operations, *Technological Forecasting and Social Change*, *80*(9), 1854–1867, `/dx.doi.org/10.1108/BIJ-10-2012-0068`.

[40] **Rathore, M.M.** *et al.* (2016). Urban planning and building smart cities based on the Internet of Things using Big Data analytics, *Computer Networks*, *101*, 63–80.

[41] **Wang, J.** *et al.* (2016). Big Data Analytics for Emergency Communication Networks: A Survey, *IEEE Communications Surveys Tutorials*, *PP*(99), 1, `COMST.2016.2540004`.

[42] **Ahmed, E.**, **Yaqoob, I.**, **Hashem, I.A.T.**, **Khan, I.**, **Ahmed, A.I.A.**, **Imran, M. and Vasilakos, A.V.** (2017). The role of big data analytics in Internet of Things, *Computer Networks*, *129*, 459–471.

[43] **Peng, Y.**, **Zhang, Y.**, **Tang, Y. and Li, S.** (2011). An incident information management framework based on data integration, data mining, and multi-criteria decision making, *Decision Support Systems*, *51*(2), 316–327.

[44] **Tuncay, B.** (2009). Identifying requirements for a disaster-monitoring system, *Disaster Prevention and Management: An International Journal*, *18*(2), 86–99.

[45] **Lloret, J.**, **Tomas, J.**, **Canovas, A. and Parra, L.** (2016). An Integrated IoT Architecture for Smart Metering, *IEEE Communications Magazine*, *54*(12), 50–57.

[46] **Cai, H.**, **Xu, B.**, **Jiang, L. and Vasilakos, A.V.** (2017). IoT-Based Big Data Storage Systems in Cloud Computing: Perspectives and Challenges, *IEEE Internet of Things Journal*, *4*(1), 75–87.

[47] **Storey, V.C. and Song, I.Y.** (2017). Big data technologies and Management: What conceptual modeling can do, *Data and Knowledge Engineering*, *108*(February), 50–67.

[48] **Basanta-Val, P., Fernandez-Garcia, N., Audsley, N.C., Gray, I. and Wellings, A.J.** (2016). Architecting Time-Critical Big-Data Systems, *IEEE Transactions on Big Data*, *2*(4), 310–324.

[49] **Li, J. and Liu, H.** (2017). Challenges of Feature Selection for Big Data Analytics, *IEEE Intelligent Systems*, *32*(2), 9–15, `1611.01875`.

[50] **Janssen, M., Estevez, E. and Janowski, T.** (2014). Interoperability in big, open, and linked data-organizational maturity, capabilities, and data portfolios, *Computer*, *47*(10), 44–49.

[51] **Hu, H.** *et al.* (2014). Toward scalable systems for big data analytics: A technology tutorial, *IEEE Access*, *2*, 652–687.

[52] **Mikalef, P., Pappas, I.O., Krogstie, J. and Giannakos, M.** (2018). Big data analytics capabilities: a systematic literature review and research agenda, *Information Systems and e-Business Management*, *16*(3), 547–578.

[53] **de Assis, L.F.F., Horita, F.E., de Freitas, E.P., Ueyama, J. and de Albuquerque, J.P.** (2018). A service-oriented middleware for integrated management of crowdsourced and sensor data streams in disaster management, *Sensors (Switzerland)*, *18*(6).

[54] **Avvenuti, M., Cresci, S., Del Vigna, F., Fagni, T. and Tesconi, M.** (2018). CrisMap: a Big Data Crisis Mapping System Based on Damage Detection and Geoparsing, *Information Systems Frontiers*, *20*(5), 993–1011.

[55] **Lin, S., Fang, W., Wu, X., Chen, Y. and Huang, Z.** (2018). A Spark-Based High Performance Computational Approach for Simulating Typhoon Wind Fields, *IEEE Access*, *6*, 39072–39085.

[56] **Cort, G.A., Esteban, A.M., Shang, X. and Alvarez, F.M.** (2018). Computers and Geosciences Earthquake prediction in California using regression algorithms and cloud-based big data infrastructure, *Computers & Geosciences*, *115*(May 2017), 198–210.

[57] **Liu, X., Li, X., Chen, X., Liu, Z. and Li, S.** (2018). Location correction technique based on mobile communication base station for earthquake population heat map, *Geodesy and Geodynamics*, *9*(5), 388–397.

[58] **Wang, D., Guo, D. and Zhang, H.** (2017). Spatial Temporal Data Visualization In Emergency Management, *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Emergency Management using - EM-GIS'17*, 1–7, `http://dl.acm.org/citation.cfm?doid=3152465.3152473`.

[59] **Huang, Q., Cervone, G. and Zhang, G.** (2017). A cloud-enabled automatic disaster analysis system of multi-sourced data streams: An example synthesizing social media, remote sensing and Wikipedia data, *Computers, Environment and Urban Systems*, *66*, 23–37.

[60] **Erd, M., Schaeffer, F., Kostic, M. and Reindl, L.M.** (2016). Event monitoring in emergency scenarios using energy efficient wireless sensor nodes for the disaster information management, *International Journal of Disaster Risk Reduction*, *16*, 33–42.

[61] **Chen, D.**, **Liu, Z.**, **Wang, L.**, **Dou, M.**, **Chen, J. and Li, H.** (2013). Natural disaster monitoring with wireless sensor networks: A case study of data-intensive applications upon low-cost scalable systems, *Mobile Networks and Applications*, *18*(5), 651–663.

[62] **Sardouk, A.**, **Mansouri, M.**, **Merghem-Boulahia, L.**, **Gaiti, D. and Rahim-Amoud, R.** (2013). Crisis management using MAS-based wireless sensor networks, *Computer Networks*, *57*(1), 29–45.

[63] **Ray, P.P.**, **Mukherjee, M. and Shu, L.** (2017). Internet of Things for Disaster Management: State-of-the-Art and Prospects, *IEEE Access*, *5*(i), 18818–18835, `arXiv:1011.1669v3`.

[64] **Collier, S.E.** (2017). The Emerging Enernet: Convergence of the Smart Grid with the Internet of Things, *IEEE Industry Applications Magazine*, *23*(2), 12–18.

[65] **Greco, L.**, **Ritrovato, P.**, **Tiropanis, T. and Xhafa, F.** (2018). IoT and semantic web technologies for event detection in natural disasters, *Concurrency Computation*, *30*(21), 1–9.

[66] **Xu, X.**, **Zhang, L.**, **Sotiriadis, S.**, **Asimakopoulou, E.**, **Li, M. and Bessis, N.** (2018). CLOTHO: A Large-Scale Internet of Things-Based Crowd Evacuation Planning System for Disaster Management, *IEEE Internet of Things Journal*, *5*(5), 3559–3568.

[67] **Ben Arbia, D.**, **Alam, M.**, **Kadri, A.**, **Ben Hamida, E. and Attia, R.** (2017). Enhanced IoT-Based End-To-End Emergency and Disaster Relief System, *Journal of Sensor and Actuator Networks*, *6*(3), 19, `http://www.mdpi.com/2224-2708/6/3/19`.

[68] **Ali, K.**, **Nguyen, H.X.**, **Vien, Q.T.**, **Shah, P. and Chu, Z.** (2018). Disaster Management Using D2D Communication with Power Transfer and Clustering Techniques, *IEEE Access*, *6*, 14643–14654.

[69] **Saxena, N.**, **Agiwal, M.**, **Ahmad, H. and Roy, A.** (2018). D2D-Based Survival on Sharing: For Enhanced Disaster Time Connectivity, *IEEE Technology and Society Magazine*, *37*(3), 64–73, `https://ieeexplore.ieee.org/document/8456879/`.

[70] **Thomas, A. and Raja, G.** (2018). FINDER: A D2D based critical communications framework for disaster management in 5G, *Peer-to-Peer Networking and Applications*.

[71] **Rosas, E.**, **Hidalgo, N.**, **Gil-Costa, V.**, **Bonacic, C.**, **Marin, M.**, **Senger, H. and ... Marin, O.** (2016). Survey on Simulation for Mobile Ad-Hoc Communication for Disaster Scenarios, *Journal of Computer Science and Technology*, *31*(2), 326–349.

[72] **Mahiddin, N.A.**, **Sarkar, N.I. and Cusack, B.** (2017). An Internet Access Solution: MANET Routing and a Gateway Selection Approach for Disaster Scenarios, *The Review of Socionetwork Strategies*, *11*(1), 47–64, `http://link.springer.com/10.1007/s12626-017-0004-3`.

[73] **Pushpalatha, K. and Karthikeyan, M.** (2018). A generalized framework for disruption tolerant secure opportunistic routing during emergency situations using MANETs, *Cluster Computing*, http://link.springer.com/10.1007/s10586-018-1849-0.

[74] **Verma, H. and Chauhan, N.** (2015). MANET based emergency communication system for natural disasters, *International Conference on Computing, Communication and Automation, ICCCA 2015*, 480–485.

[75] **Erdelj, M.**, **Natalizio, E.**, **Chowdhury, K.R. and Akyildiz, I.F.** (2017). Help from the Sky: Leveraging UAVs for Disaster Management, *IEEE Pervasive Computing*, *16*(1), 24–32.

[76] **Peng, G.**, **Xia, Y.**, **Zhang, X. and Bai, L.** (2018). UAV-Aided Networks for Emergency Communications in Areas with Unevenly Distributed Users, *Journal of Communications and Information Networks*, *3*(4), 23–32.

[77] **Micheletto, M.**, **Petrucci, V.**, **Santos, R.**, **Orozco, J.**, **Mosse, D.**, **Ochoa, S.F. and Meseguer, R.** (2018). Flying real-time network to coordinate disaster relief activities in urban areas, *Sensors (Switzerland)*, *18*(5), 1–20.

[78] **Meier, P.** (2012). Crisis mapping in action: How open source software and global volunteer networks are changing the world, one map at a time, *Journal of Map and Geography Libraries*, *8*(2), 89–100, 9809069v1.

[79] **Goodchild, M.F.** (2007). Citizens as sensors: The world of volunteered geography, *GeoJournal*, *69*(4), 211–221, arXiv:1404.3341.

[80] **Kusumo, A.N.**, **Reckien, D. and Verplanke, J.** (2017). Utilising volunteered geographic information to assess resident's flood evacuation shelters. Case study: Jakarta, *Applied Geography*, *88*, 174–185, https://doi.org/10.1016/j.apgeog.2017.07.002.

[81] **Feng, Y. and Sester, M.** (2018). Extraction of Pluvial Flood Relevant Volunteered Geographic Information (VGI) by Deep Learning from User Generated Texts and Photos, *ISPRS International Journal of Geo-Information*, *7*(2), 39, http://www.mdpi.com/2220-9964/7/2/39.

[82] **Zahra, K.**, **Ostermann, F.O. and Purves, R.S.** (2017). Geographic variability of Twitter usage characteristics during disaster events, *Geo-Spatial Information Science*, *20*(3), 231–240, http://doi.org/10.1080/10095020.2017.1371903.

[83] **Toujani, R.**, **Chaabani, Y.**, **Dhouioui, Z. and Bouali, H.** (2018). *The Next Generation of Disaster Management and Relief Planning: Immersive Analytics Based Approach*, volumeiLRN CCIS, Springer International Publishing, http://link.springer.com/10.1007/978-3-319-93596-6.

[84] **Nickerson, R.C.**, **Varshney, U. and Muntermann, J.** (2013). A method for taxonomy development and its application in information systems, *European Journal of Information Systems*, *22*(3), 336–359, 1305.0982.

[85] **Shibata, Y.**, **Uchida, N. and Shiratori, N.** (2014). Analysis of and proposal for a disaster information network from experience of the Great East Japan Earthquake, *IEEE Communications Magazine*, *52*(3), 44–50.

[86] **Chung, K. and Park, R.C.** (2016). P2P cloud network services for IoT based disaster situations information, *Peer-to-Peer Networking and Applications*, *9*(3), 566–577.

[87] **Imran, M.**, **Castillo, C.**, **Diaz, F. and Vieweg, S.** (2015). Processing Social Media Messages in Mass Emergency, *ACM Computing Surveys*, *47*(4), 1–38, `http://arxiv.org/abs/1407.7071http://dl.acm.org/citation.cfm?doid=2775083.2771588`, `1407.7071`.

[88] **Ragini, J.R.**, **Anand, P.M. and Bhaskar, V.** (2018). Big data analytics for disaster response and recovery through sentiment analysis, *International Journal of Information Management*, *42*(September 2017), 13–24.

[89] **Huang, Q. and Xiao, Y.** (2015). Geographic Situational Awareness: Mining Tweets for Disaster Preparedness, Emergency Response, Impact, and Recovery, *ISPRS International Journal of Geo-Information*, *4*(3), 1549–1568, `http://www.mdpi.com/2220-9964/4/3/1549/`.

[90] **Najafabadi, M.M.**, **Villanustre, F.**, **Khoshgoftaar, T.M.**, **Seliya, N.**, **Wald, R. and Muharemagic, E.** (2015). Deep learning applications and challenges in big data analytics, *Journal of Big Data*, *2*(1), 1–21.

[91] **Zedadra, O.**, **Guerrieri, A.**, **Jouandeau, N.**, **Spezzano, G.**, **Seridi, H. and Fortino, G.** (2018). Swarm intelligence-based algorithms within IoT-based systems: A review, *Journal of Parallel and Distributed Computing*, *122*, 173–187.

[92] **Sevo, I. and Avramovic, A.** (2016). Convolutional neural network based automatic object detection on aerial images, *IEEE Geoscience and Remote Sensing Letters*, *13*(5), 740–744, `1512.05227`.

[93] **Frizzi, S.**, **Kaabi, R.**, **Bouchouicha, M.**, **Ginoux, J.M.**, **Moreau, E. and Fnaiech, F.** (2016). Convolutional neural network for video fire and smoke detection, *IECON Proceedings (Industrial Electronics Conference)*, 877–882.

[94] **San Miguel**, **P., M.J. and Ruiz, C.R.**, (2016). A Flood Detection and Warning System Based on Video Content Analysis, International Symposium on Visual Computing. Springer, Cham, 2016., volume10073 LNCS, pp.65–74.

[95] **Szewrański, S.**, **Chruściński, J.**, **van Hoof, J.**, **Kazak, J.K.**, **Świader, M.**, **Tokarczyk-Dorociak, K. and Zmuda, R.** (2018). A location intelligence system for the assessment of pluvial flooding risk and the identification of stormwater pollutant sources from roads in suburbanised areas, *Water (Switzerland)*, *10*(6), 1–16.

[96] **Gutierrez, J.M., Jensen, M., Henius, M. and Riaz, T.** (2015). Smart Waste Collection System Based on Location Intelligence, *Procedia Computer Science*, *61*, 120–127.

[97] **Sun, Y., Song, H., Jara, A.J. and Bie, R.** (2016). Internet of Things and Big Data Analytics for Smart and Connected Communities, *IEEE Access*, *4*(4), 766–773, `http://ieeexplore.ieee.org/document/7406686/`.

[98] **Marjani, M., Nasaruddin, F., Gani, A., Karim, A., Hashem, I.A.T., Siddiqa, A. and Yaqoob, I.** (2017). Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges, *IEEE Access*, *5*, 5247–5261, `2017`.

[99] **Rehman, M.H.U., Ahmed, E., Yaqoob, I., Hashem, I.A.T., Imran, M. and Ahmad, S.** (2018). Big Data Analytics in Industrial IoT Using a Concentric Computing Model, *IEEE Communications Magazine*, *56*(2), 37–43.

[100] **Lim, C., Kim, K.j. and Maglio, P.P.** (2018). Smart cities with big data : Reference models , challenges , and considerations, *Cities*, *82*(May), 86–99, `https://doi.org/10.1016/j.cities.2018.04.011`.

[101] **Peffers, K., Tuunanen, T., Rothenberger, M.A. and Chatterjee, S.** (2007). A Design Science Research Methodology for Information Systems Research, *Journal of Management Information Systems*, *24*(3), 45–77, `http://www.tandfonline.com/doi/full/10.2753/MIS0742-1222240302,z0022`.

[102] **Hevner, A.R., March, S.T., Park, J., Ram, S. and Ram, S.** (2004). Design Science in Information Systems Research, *MIS Quarterly*, *28*(1), 75–105.

[103] **Oteafy, S.M. and Hassanein, H.S.** (2018). Big Sensed Data: Evolution, Challenges, and a Progressive Framework, *IEEE Communications Magazine*, *56*(7), 108–114.

[104] **Ran, Y.** (2011). Considerations and suggestions on improvement of communication network disaster countermeasures after the wenchuan earthquake, *IEEE Communications Magazine*, *49*(1), 44–47.

[105] **Fang, S.** *et al.* (2015). An integrated information system for snowmelt flood early-warning based on internet of things, *Information Systems Frontiers*, *17*(2), 321–335.

[106] **Zambrano, A.M., Perez, I. and Esteve, M.** (2019). Technologies of Internet of Things applied to an Earthquake Early Warning System, *Future Generation Computer Systems*, *75*(2017), 206–215, `http://dx.doi.org/10.1016/j.future.2016.10.009`.

[107] **Khan, M.U.S., Khalid, O., Huang, Y., Ranjan, R., Zhang, F., Cao, J., Veeravalli, B. and ... Zomaya, A.Y.** (2017). MacroServ: A Route Recommendation Service for Large-Scale Evacuations, *IEEE Transactions on Services Computing*, *10*(4), 589–602.

[108] **Fujihara, A. and Miwa, H.**, (2014). Disaster evacuation guidance using opportunistic communication: The potential for opportunity-based service, Big Data and Internet of Things: A Roadmap for Smart Environments, Springer, Cham, pp.425–446, `http://link.springer.com/10.1007/978-3-319-05029-4`, `arXiv:1502.01815v3`.

[109] **Komatsu, N.**, **Sasabe, M.**, **Kawahara, J. and Kasahara, S.** (2018). Automatic evacuation guiding scheme based on implicit interactions between evacuees and their mobile nodes, *GeoInformatica*, *22*(1), 127–141.

[110] **Lo, S.W.**, **Wu, J.H.**, **Lin, F.P. and Hsu, C.H.** (2015). Visual sensing for urban flood monitoring, *Sensors (Switzerland)*, *15*(8), 20006–20029.

[111] **Pan, J.**, **Yin, Y.**, **Xiong, J.**, **Luo, W.**, **Gui, G. and Sari, H.** (2018). Deep Learning-based Unmanned Surveillance Systems for Observing Water Levels, *IEEE Access*, *6*, 1–1.

[112] **Kim, J.C.**, **Jung, H. and Chung, K.** (2016). Mining Based Urban Climate Disaster Index Service According to Potential Risk, *Wireless Personal Communications*, *89*(3), 1009–1025.

[113] **Furquim, G.**, **Filho, G.P.**, **Jalali, R.**, **Pessin, G.**, **Pazzi, R.W. and Ueyama, J.** (2018). How to improve fault tolerance in disaster predictions: A case study about flash floods using IoT, ML and real data, *Sensors (Switzerland)*, *18*(3), 1–20.

[114] **Kök, I.**, **Şimşek, M.U. and Özdemir, S.** (2018). A deep learning model for air quality prediction in smart cities, *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017*, *2018-January*, 1983–1990.

[115] **Sood, S.K.**, **Sandhu, R.**, **Singla, K. and Chang, V.** (2018). IoT, big data and HPC based smart flood management framework, *Sustainable Computing: Informatics and Systems*, *20*, 102–117, `https://doi.org/10.1016/j.suscom.2017.12.001`.

[116] **Tene, O. and Polonetsky, J.** (2012). Privacy in the Age of Big Data: a time for big decisions, *Stanford Law Review Online*, *64*, 63–69, `http://www.stanfordlawreview.org/sites/default/files/online/topics/64-SLRO-63{\_}1.pdf{\%}5Cnpapers3://publication/uuid/F1C87BD7-F850-4414-B368-BC6C2EB96091`.

[117] **Kim, G.H.**, **Trimi, S. and Chung, J.H.** (2014). Big-data applications in the government sector, *Communications of the ACM*, *57*(3), 78–85, `http://dl.acm.org/citation.cfm?doid=2566590.2500873`, `1204.6078`.

[118] **Veil, S.R.**, **Buehner, T. and Palenchar, M.J.** (2011). A Work-In-Process Literature Review: Incorporating Social Media in Risk and Crisis Communication, *Journal of Contingencies and Crisis Management*, *19*(2), 110–122.

[119] **Turel, O. and Serenko, A.** (2012). The benefits and dangers of enjoyment with social networking websites, *European Journal of Information Systems*, *21*(5), 512–528.

[120] **Chan, J.C.** (2012). The Role of Social Media in Crisis Preparedness, Response and Recovery, *Vanguard - Singapure*, 26, `http://www.oecd.org/governance/risk/TheroleofSocialmediaincrisispreparedness,responseandrecovery.pdf`.

[121] **Ahmed, A.** (2011). U s m d m, *Thirty Second International Conference on Information Systems, Shanghai 2011*, 1–11.

[122] **Erskine, M.A. and Gregg, D.G.** (2012). Utilizing Volunteered Geographic Information to Develop a Real-Time Disaster Mapping Tool: A Prototype and Research Framework, *CONF-IRM 2012 Proceedings*, 5–1.

[123] **Ostermann, F.O. and Spinsanti, L.** (2011). A Conceptual Workflow For Automatically Assessing The Quality Of Volunteered Geographic Information For Crisis Management, *14th AGILE International Conference on Geographic Information Science*, 1–6, `http://publications.jrc.ec.europa.eu/repository/handle/111111111/22883`.

[124] **Adam, N.R.**, **Shafiq, B. and Staffin, R.** (2012). Spatial computing and social media in the context of disaster management, *IEEE Intelligent Systems*, *27*(6), 90–97.

[125] **Yin, J.**, **Karimi, S.**, **Lampert, A.**, **Cameron, M.**, **Robinson, B. and Power, R.** (2015). Using Social Media to Enhance Emergent Situation Awareness: Extended Abstract, *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, (Ijcai), 4234–4238, `https://delvalle.bphc.org/pluginfile.php/13922/mod{\_}resource/content/2/YIN-IS2012.pdf`.

[126] **Abel, F.**, **Hauff, C.**, **Houben, G.J.**, **Stronkman, R. and Tao, K.** (2012). Twitcident: Fighting fire with information from Social Web streams, *Proceedings of the 21st international conference companion on World Wide Web - WWW '12 Companion*, 305–308, `http://www.scopus.com/inward/record.url?eid=2-s2.0-84861035766{\&}partnerID=tZOtx3y1`.

[127] **MacEachren, A.M.**, **Jaiswal, A.**, **Robinson, A.C.**, **Pezanowski, S.**, **Savelyev, A.**, **Mitra, P.**, **Zhang, X. and Blanford, J.** (2011). SensePlace2: GeoTwitter analytics support for situational awareness, *VAST 2011 - IEEE Conference on Visual Analytics Science and Technology 2011, Proceedings*, 181–190.

[128] **Kaplan, A.M. and Haenlein, M.** (2010). Users of the world, unite! The challenges and opportunities of Social Media, *Business Horizons*, *53*(1), 59–68.

[129] **Jayaraman, V.**, **Chandrasekhar, M. and Rao, U.** (1997). Managing the natural disasters from space technology, *Acta Astronauti*, *40*(2), 291–325.

[130] **Crystal Washington**, The Role of Social Media During Natural DisastersCrystal Washington, https://crystalwashington.com/the-role-of-social-media-during-natural-disasters/.

[131] **United Nations** (2018). World Urbanization Prospects : The 2018 Revision, **Technical Report**, United Nations.

[132] **Nathali Silva, Bhagya, M.K. and Han, K.** (2017). Big data analytics embedded smart city architecture for performance enhancement through real-time data processing and decision-making, *Wireless Communications and Mobile Computing*, *2017*.

[133] **Rathore, M.M.**, **Paul, A.**, **Hong, W.H.**, **Seo, H.C.**, **Awan, I. and Saeed, S.** (2018). Exploiting IoT and big data analytics: Defining Smart Digital City using real-time urban data, *Sustainable Cities and Society*, *40*(August 2017), 600–610.

[134] **Babar, M.**, **Rahman, A.**, **Arif, F. and Jeon, G.** (2018). Energy-harvesting based on internet of things and big data analytics for smart health monitoring, *Sustainable Computing: Informatics and Systems*, *20*, 155–164, https://doi.org/10.1016/j.suscom.2017.10.009https://linkinghub.elsevier.com/retrieve/pii/S2210537917302238.

[135] **Babar, M. and Arif, F.** (2017). Smart urban planning using Big Data analytics to contend with the interoperability in Internet of Things, *Future Generation Computer Systems*, *77*, 65–76, http://dx.doi.org/10.1016/j.future.2017.07.029.

[136] **Onal, A.C.**, **Berat Sezer, O.**, **Ozbayoglu, M. and Dogdu, E.** (2018). Weather data analysis and sensor fault detection using an extended IoT framework with semantics, big data, and machine learning, *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017*, *2018-January*, 2037–2046.

[137] **Arridha, R.**, **Sukaridhoto, S.**, **Pramadihanto, D. and Funabiki, N.** (2017). Classification extension based on IoT-big data analytic for smart environment monitoring and analytic in real-time system, *International Journal of Space-Based and Situated Computing*, *7*(2), 82–93.

[138] **Apache Software Foundation**, Apache Flume, https://flume.apache.org/.

[139] **The Apache Software Foundation**, Apache Sqoop, http://sqoop.apache.org/.

[140] **Shvachko, K.**, **Kuang, H.**, **Radia, S.**, **Presented, R.C. and Pokluda, A.** (2010). The Hadoop Distributed File System, *Mass storage systems and technologies (MSST), 2010 IEEE 26th symposium on*, 1–10, http://www.alexanderpokluda.ca/coursework/cs848/CS848PaperPresentation-AlexanderPokluda.pdf.

[141] **Zaharia, M.**, **Chowdhury, M.**, **Franklin, M.J.**, **Shenker, S. and Stoica, I.** (2010). Spark : Cluster Computing with Working Sets, *HotCloud'10 Proceedings of the 2nd USENIX conference on Hot topics in cloud computing*, 10, arXiv:1011.1669v3.

[142] **Zhang, P.**, **Deng, Q.**, **Liu, X.**, **Yang, R. and Zhang, H.** (2017). Emergency-Oriented Spatiotemporal Trajectory Pattern Recognition by Intelligent Sensor Devices, *IEEE Access*, *5*, 3687–3697.

[143] **McGrattan, K.**, **Hostikka, S.**, **Floyd, J.**, **Baum, H.**, **Rehm, R.**, **Mell, W. and McDermott, R.** (2004). Fire dynamics simulator (version 5), technical reference guide. Volume I: Mathematical Model. NIST Special Publication 1018-5. National Institute of Standards and Technology. Gaithersburg, USA., **Technical ReportVersion 5**, National Institute of Standards and Technology, http://pyrosim.pl/files/Za??cznik2-walidacja.pdf.

[144] **Bischof, S.**, **Karapantelakis, A.**, **Sheth, A.**, **Mileo, A. and Barnaghi, P.** (2014). Semantic Modelling of Smart City Data, *W3C Workshop on the Web of Things Enablers and services for an open Web of Devices*, 1–5, http://www.w3.org/2014/02/wot/papers/karapantelakis.pdf.

[145] **Volkhin**, GitHub open source code RoadTrafficSimulator_ Road traffic simulator and signals optimizer in CoffeeScript & HTML5, https://github.com/volkhin/RoadTrafficSimulator.

[146] **Team, A.**, The Twitter Stream Grab : Internet Archive, https://archive.org/details/twitterstream.

[147] **Kotsiantis, S.B.**, **Kanellopoulos, D. and Pintelas, P.E.** (2006). Data preprocessing for Supervised Leaning, *International Journal of Computer Science*, *1*(2), 1–7.

[148] **Armbrust, M.**, **Ghodsi, A.**, **Zaharia, M.**, **Xin, R.S.**, **Lian, C.**, **Huai, Y. and ... Franklin, M.J.** (2015). Spark SQL: Relational Data Processing in Spark, *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data - SIGMOD '15*, 1383–1394, http://dl.acm.org/citation.cfm?id=2723372.2742797.

[149] **Meng, X.**, **Bradley, J.**, **Yavuz, B.**, **Sparks, E.**, **Venkataraman, S.**, **Liu, D. and ... Talwalkar, A.** (2016). MLlib: Machine Learning in Apache Spark, *The Journal of Machine Learning Research*, *17*(1), 1235–1241, http://arxiv.org/abs/1505.06807.

[150] **Xin, R.S.**, **Gonzalez, J.E.**, **Franklin, M.J.**, **Stoica, I. and AMPLab, E.** (2013). GraphX: A Resilient Distributed Graph System on Spark, *First International Workshop on Graph Data Management Experiences and Systems*, 2, 1402.2394.

[151] **AHA Center** (2018). SITUATION UPDATE NO. 15-FINAL M 7.4 EARTHQUAKE & TSUNAMI Sulawesi, Indonesia, **Technical Report 5**, https://ahacentre.org/situation-update/

situation-update-no-15-sulawesi-earthquake-\
\26-october-2018/.

[152] **Wang, S.**, **Sinnott, R. and Nepal, S.** (2018). P-GENT: Privacy-Preserving Geocoding of Non-Geotagged Tweets, *Proceedings - 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications and 12th IEEE International Conference on Big Data Science and Engineering, Trustcom/BigDataSE 2018*, 972–983.

[153] **Wang, Z.**, **Ye, X. and Tsou, M.H.** (2016). Spatial, temporal, and content analysis of Twitter for wildfire hazards, *Natural Hazards*, *83*(1), 523–540.

[154] **MAPD**, GitHub - omnisci_mapd-core_ The MapD Core database, https://github.com/omnisci/mapd-core.

[155] **Kate Ravilious**, Terrawatch: why did the quake in Sulawesi cause a tsunami? | World news | The Guardian, https://www.theguardian.com/world/2018/oct/02/terrawatch-why-did-the-quake-in-sulawesi-cause\
\-palu-tsunami.

[156] **Gökkaya, K.** (2016). Geographic analysis of earthquake damage in Turkey between 1900 and 2012, *Geomatics, Natural Hazards and Risk*, *7*(6), 1948–1961, https://doi.org/10.1080/19475705.2016.1171259.

[157] **Cambaz, M.D. and Mutlu, A.K.** (2016). Regional Moment Tensor Inversion for Earthquakes in Turkey and Its Surroundings: 2008–2015, *Seismological Research Letters*, *87*(5), 1082–1090, http://srl.geoscienceworld.org/lookup/doi/10.1785/0220150276.

[158] **Hameed, S.**, **Khan, F.I. and Hameed, B.** (2019). Understanding Security Requirements and Challenges in Internet of Things (IoT): A Review, *Journal of Computer Networks and Communications*, *2019*, 1–14, http://arxiv.org/abs/1808.10529https://www.hindawi.com/journals/jcnc/2019/9629381/, doi: 10.1155/2019/9629381, 1808.10529.

[159] **Sicari, S.**, **Rizzardi, A.**, **Grieco, L.A. and Coen-Porisini, A.** (2015). Security, privacy and trust in Internet of things: The road ahead, *Computer Networks*, *76*, 146–164, http://dx.doi.org/10.1016/j.comnet.2014.11.008.

[160] **Prof. Jimeng Sun**, CSE 6250 Big Data for Healthcare | MapReduce Basics, http://sunlab.org/teaching/cse6250/fall2018/hadoop/mapreduce-basic.html{\#}mapreduce.

**CURRICULUM VITAE**

**Name Surname:** Syed Attique Shah

**Place and Date of Birth:** Quetta, Pakistan - 10/12/1988

**E-Mail:** shah@itu.edu.tr, syed.attique@gmail.com

**EDUCATION:**

- **B.Sc.:** 2006 – 2010, BS (Information Technology), Balochistan University of Information Technology, Engineering & Management Sciences (BUITEMS), Pakistan
- **M.Sc.:** 2010 – 2013, MS (Information Technology), Balochistan University of Information Technology, Engineering & Management Sciences (BUITEMS), Pakistan

**PROFESSIONAL EXPERIENCE AND REWARDS:**

- 2014 – Till date Assistant Professor (Study Leave for PhD), B.U.I.T.E.M.S, Quetta, Pakistan.
- 2018 Granted fully funded scholarship by Dora Plus study mobility program for doctoral students at Tallinn University of Technology, Estonia.
- 2018 Granted National Chiao Tung University Taiwan Elite Internship Program.
- 2017 Secured the Erasmus+ Programme Scholarship to continue doctoral studies for Fall 2017 at Tallinn University of Technology, Estonia.
- 2017 Selected for International Cartographic Association (ICA) Scholarship of 1000 Euro grant for presenting research paper as young scientist of cartography and GIScience in ICC2017 Washington, DC USA.
- 2016 Granted fully funded Student Exchange JTRE Program Scholarship by the University of Tokyo, Japan.
- 2014 Granted fully funded Türkiye Bursları Scholarship for Ph.D. from 2014-2019 at Istanbul Technical University, Turkey.
- 2013 Designated with Excellence Award for the year 2013 best teacher in Faculty of ICT, BUITEMS, Pakistan.
- 2013 Designated with Gold Medal by the Chancellor BUITEMS for MS (Information Technology).

**PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:**

- **Shah, S. A.**, Seker, D. Z., Rathore, M. M., Hameed, S., Yahia, S. B., Draheim, D. (2019). Towards Disaster Resilient Smart Cities: Can Internet of Things and Big Data Analytics Be the Game Changers?. IEEE Access *(Impact factor: 4.098)*, 7, 91885-91903. [Online]. Available: https://ieeexplore.ieee.org/document/8759905. DOI: 10.1109/ACCESS.2019.2928233.

- **Shah, S. A.**, Seker, D. Z., Hameed, S., Draheim, D. (2019). The Rising Role of Big Data Analytics and IoT in Disaster Management: Recent Advances, Taxonomy and Prospects. IEEE Access *(Impact factor: 4.098)*, 7, 54595-54614. [Online]. Available: https://ieeexplore.ieee.org/document/8698814. DOI: 10.1109/ACCESS.2019.2913340.

- **Shah, S. A.**, Şeker, D. Z., Demirel, H. (2017). A Framework for Enhancing Real-Time Social Media Data to Improve the Disaster Management Process. 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

**OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:**

- **Syed Attique Shah**, Dursun Zafer Şeker, Caner Güney “Integrating Data Source Possibilities for Geo-Business Intelligence”. Poster accepted for GEOSPATIAL WORLD FORUM, World Trade Center, Rotterdam, Netherlands, May 23-26, 2016. http://geospatialworldforum.org/2016/poster-presenters.html

- Rehman, Abdul, Asad Ali, Riaz ul Amin, and **Attique Shah**. "VANET thread based message trust model." In Eighth International Conference on Digital Information Management (ICDIM 2013), pp. 58-60. IEEE, 2013.

- Ahmed, A., **Shah, A.**, Sani, K. A., Bukhari, A. H. S. (2012). Particle swarm optimization for N-queens problem. Journal of Advanced Computer Science and Technology, 1(2), 57-63.