

Technological Communication

Big Data Analytics Strategy Framework: A Case of Crowd Management During Hajj

Hanaa Ali Aldahawi

Information Science department, King AbudAlaziz University, Jeddah, Saudi Arabia

ABSTRACT

The objective of the present study was an investigation of applications of big data analytics in Hajj and Umrah for pilgrims, who come to Saudi Arabia every year for tourism and observation of religious rites as per the sacred beliefs of Islam. It has now become a necessity to see more applications of big data analytics in these pilgrimages because of the growing number of people every year. Therefore, crowd control, crowd management and conflict management are essential for reduction of stress, troubles, fatalities, accidents, theft and possible deaths during Hajj and Umrah events. Developing a predictive data analytic model for Hajj and Umrah will improve the efficiency, gross domestic product (GDP), surveillance, revenue generation, opportunities and satisfaction for the pilgrimages. In this paper, review of big data tools was presented along with their use in the decision support system and how it can be used for surveillance and crowd management. A robust big data framework applicable for Hajj and Umrah events was also presented in this paper. This was meant to aid seamless adoption and implementation of big data applications across sectors and government parastatals involved in Hajj and Umrah. The presented framework was also included all the relevant use cases related to these pilgrimages.

KEY WORDS: BIG DATA ANALYTICS, HAJJ, UMRAH, MINISTRY OF HAJJ AND UMRAH, CROWD MANAGEMENT, SAUDI ARABIA.

INTRODUCTION

Hajj and Umrah are both important events for Islamic believers. Hajj is a pilgrimage journey to Mecca and is part of the five pillars of Islam; it is required for every Muslim who is financially and physically capable to make this pilgrimage at least once in their lifetime. Performing Hajj successfully, which is one performed without evil commitment, results in erasing sins. Umrah is also linked to erasing of sins as well as a rewarding spiritual experience. With these reasons as the foundation, millions of Muslims from different parts of the world travel to Saudi Arabia for Hajj and Umrah (Ledhem et al., 2020; Luz, 2020).

These annual events attract so many people that it's now considered the world's largest human gathering (Central Department of Statistics and Information, 2014). With the largest gathering, there are aspects that come into play: crowd management and big data frameworks and analysis. Hajj and Umrah are posing big issues due to the fact that the

number is increasing every year. It is estimated that by 2030, there will be more than 5.4 million people in Hajj and around 30 million in Umrah. These huge numbers result in issues related to crowd management and resource control. Big data analytics is a technique that can be used to manage, control, predict and facilitate smooth success of the Hajj and Umrah. Big data refers to a dataset with a large size and complexity. Data mining is an important concept in the big data era. It is the collection, extraction, analysis and predictive use of data which allows the useful and meaningful applications of data. In this paper, we discuss Hajj and Umrah, crowd control and management, big data tools and framework, and design of big data framework application for Hajj and Umrah, to forecast and foster more successful Hajj and Umrah pilgrimages (Wu et al., 2014; Luz, 2020).

We looked at the literature of all the above-mentioned topics and analysed their contribution to the significance of this study. Crowd control and conflict management are a key challenge during Hajj and Umrah. Crowd control is an essential factor to consider in any mass gathering, and Umrah and Hajj are not exceptions. Thus, whatever the reason for mass gathering, there must be provision for eventualities. Hence, security, peaceful processes and

Article Information:*Corresponding Author: haldahawi@kau.edu.sa

Received 21/10/2021 Accepted after revision 25/12/2021

Published: 31st December 2021 Pp- 1975-1984

This is an open access article under Creative Commons License,

Published by Society for Science & Nature, Bhopal India.

Available at: <https://bbrc.in/> DOI: <http://dx.doi.org/10.21786/bbrc/14.4.88>

reduction in casualties, fatalities and deaths must be the objectives for the authorities managing Hajj and Umrah. The use of big data analytics to predict the expected behaviours of the crowd helps in pre-incident planning to enable responsibilities, training, organisation, operating procedures and the rules of engagement in case of any eventuality (Felemban et al., 2020).

In fact, for a huge number of people to be gathered in a given place, crowd control and management must come into play, especially when safety and security measures are considered. We can define crowd “as an assembly of persons large enough to produce a sense of considerable mass, casually gathered together without organised discipline or order”. Hence, to have a disciplined and organised crowd requires crowd control and management. Therefore, crowd management is defined as organised and systematically planning the movement as well as assembly of people in an orderly manner as to achieve the desired objective and purpose of the mass gathering effectively. On the other hand, crowd control refers to providing guidelines for people regarding their attitudes and behaviour in a gathering (Brian and Kingshott, 2014).

Data Analysis: The Hajj activities take place for five days and in specific locations. With millions of people carrying out these activities within a limited period of time and in the same places, crowd management has turned out to be an issue the stakeholders of the event, such as the Ministry

of Hajj and Umrah, have had to battle with over the years. In some cases, there have been occurrences of tragedies. According to Alnuaim and Almasry (2012), the table below is a highlight of disastrous accidents that have occurred during Hajj and Umrah due to poor crowd management.

Even though there have been efforts geared towards improving crowd management over the years, there are still numerous challenges that Saudi Arabia faces. The slow turning of the wheel of progress when it comes to crowd management in Hajj and Umrah is linked to issues such as: Infrastructure problems: the pilgrimage activities are limited to a few places in the holy city of Mecca. The infrastructure such as the road networks connecting these places have some problems. Lack of following the schedule: there are schedules rolled out by the Hajj authorities but some of the pilgrims do not fully follow them, thus creating hindrance in the process. Ever increasing number of pilgrims: each year, there is a drastic increase in the number of pilgrims and the event planners can hardly plan adequately, especially when it comes to numbers. Having more people than planned for leads to issues such as long waits to perform the Hajj and Umrah activities, which, in most cases, end up in disaster.

Pilgrims getting lost: according to Amro and Nijem (2012), in 2011, over 30,000 pilgrims went missing as a result of overcrowding at the Hajj holy locations. Children and foreign pilgrims are mostly affected, and the language barriers sometimes make the problem worse.

Table 1. Data Report of Past Hajj Incidences (Alnuaim & Almasry,2012)

Date	Accidents	Casualties	Place
1975	Fire	Death of 200 pilgrims	Camps for pilgrims near Makkah
1990	Suffocation	Death of 1,426 pilgrims	Inside a pedestrian tunnel
1994	Stampede	Death of 270 pilgrims	Al-Jamarat in Mina
1998		Death of 118 pilgrims	
2001		Death of 35 pilgrims	
2003		Death of 14 pilgrims	
2004		Death of 251 pilgrims	
2006		Death of 346 pilgrims	
2015		Death of 2,411 pilgrims	Mina

Figure 1: Big Data Representation (Adapted from Ezeogu et al., 2019)



Limited guidance at the ritual places: lack of adequate guidance even in direction leads to problems. Research on the areas of crowd management has improved over the years. There are more studies geared towards improving movement and assembly of people. Some of the changes that can be attributed to the continuous growth in research include small cars not being allowed in Mecca to limit congestion. The expansion of the railway network through construction of the monorail Al Mashaer Al Mugaddassah Metro Line in 2010 to link Mina, Arafat and Muzdalifah has resulted in reducing transportation problems for pilgrims (Reffat, 2012; Still et al., 2020).

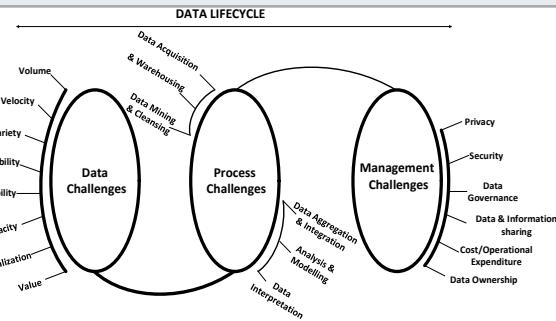
Mahmassani and Sheffi (1981) came up with a one-of-a-kind proposal on the study of the behaviour of people in certain situations. As a result of this proposal, Dr

Felemban's centre is working on the development of a bracelet that would help in tracking pilgrims as well as be a source of guidance for them throughout the various Hajj and Umrah activities. Installation of over 800 surveillance cameras is also a result of research in crowd management. This is slowly helping the stakeholders to improve the safety measures even as the number of pilgrims increases each year. Research on the structural issues of the sacred sites in relation to crowd management problems has also been vital, among other scientists, proposed that the geometry of the sacred sites structures should be refined. This proposal is being implemented, and one of the examples is the reconstruction of the Jamaraat Bridge to help accommodate a larger number of pilgrims performing the Hajj ritual that involves the stoning of the devil, which is carried out on the bridge (Algadhi et al. 2002; Pin et al., 2011).

Furthermore, in 2012, author carried out a study on crowd management of pilgrims with the use of thermography. Through this, they were able to provide a realistic demonstration of behaviour of pilgrims during three Hajj activities. In addition, a study by Maciej et al. (2011) took another approach, by introducing a crowd management system that was based on optical data flow. In this approach, the system was to consider two lines of threads: one being the analysis of behaviour for detection of situations that could be considered dangerous in a crowd, and the second one being the detection of any delays or hold ups in a given area (Khozium et al., 2012; Still et al., 2020).

Additionally, other studies on crowd management in Hajj and Umrah have been geared towards intelligence-based detection and management of congestion, development of an intelligent computational real-time virtual environment model, a multi-agent approach to the issue of crowd simulation and modelling and use of a wireless sensor network deployment model in monitoring pilgrims in cases such as evacuation (Stefania et al., 2007; Reffat, 2012; Felemban et al., 2020).

Figure 2: Conceptual Classification of Big Data Life Challenges. (Adapted from Sivarajah, 2016)



Big Data Concept: Big data analytics was first developed by (Chen et al., 2012) when they pointed out the relationship between business intelligence and analytics, and the connection they have with mining of data and analysis of statistics. The large number of pilgrims in Hajj and Umrah events bring with it a lot of data, which, if effectively and efficiently captured, stored, managed and analysed, can help the stakeholders involved in planning the Hajj to make

the event safer, more secure and successful. Thus, data, be it audio, video or textual from millions of people, is a lot, especially in recent times of high-volume data consumption; it is considered big data. Therefore, big data is simply defined as a dataset with a size that is more than what a typical software tool can take in terms of capturing, storing, managing and analysing. Big data describes innovative techniques and technologies to capture, store, distribute, manage and analyse larger sized datasets with diverse structures. According to Halevi (2012) and Apollos et al. (2019), in describing big data concept, the five Vs ideology is applied (Manyika et al., 2011; Apollos et al., 2019).

Studies on big data have focused on looking at challenges in the understanding of the whole big data concept. This includes decisions on what data is generated as well as collected, privacy issues and consideration of ethics while mining the data. Tole (2013) indicated that one of the challenges that businesses and other sectors are facing is building a viable solution for big data; therefore, it continues to be a learning process with creation, innovation and implementation of new approaches every now and then. Computing architectures are proposed to replace the von-Neumann traditional computing architecture. The idea behind the new architecture, computation-in-memory using new nanomaterial such as memristor, is to improve storage capacity and speed of computation in the era of big data (Lazer et al., 2009; Boyd & Crawford, 2012; Crawford, 2013; Hargittai, 2015 Ezeogu et al., 2019; Felemban et al., 2020).

Big Data Life Cycle: Today, big data analysis is done using intelligence plans. According to Manyika et al. (2011), big data can contribute positively towards the transparency, performance, creation and innovative decision-making wherever and whenever it is used effectively. According to the author, as opposed to the past when capturing, storing, managing and analysing of big data was difficult, today it is easier thanks to the development of big data systems. With these systems, visualising, detection of trends and development of algorithms to predict occurrence of issues is possible and easier (Mawed et al., 2017; Khalid & Zebaree, 2021).

Computing technologies are used when it comes to capturing, storing managing and analysing big data. However, it is important to note that a lot of challenges come with the big data: data challenges, process challenge and management challenges. These are the stages in a big data life cycle. Akerkar (2014) and Zicari (2014) categorise the challenges of big data into data, process and management challenges, depending on the stage of the data life cycle a set of data is at. The data challenges are the ones that have to do with the data's characteristics including its volume, velocity, variety, veracity, quality, volatility, dogmatism and discovery. Process challenges, on the other hand, are those related to techniques on how the data are captured, integrated and transformed as well as the procedure of selecting the correct analysis model and how to generate results. Lastly, management challenges look at security, privacy, ethical and governance issues of big data. To ensure that decision made by people utilising big data

are evidence-based, choosing efficient data processing or analysing methods is necessary (Gandomi and Haider, 2015; Felemban et al., 2020).

The benefits and potential that comes with using big data is limitless, but the technologies, skills and tools that are available to help with big data analysis can prove to be restrictive. Labrinidis and Jagadish (2012) defined big data analysis as the methods or processes used in the examination and attainment of intellect or viable information from large datasets. Big data analysis, commonly referred to as BDA, is more than just tracing or capturing, categorising, understanding and quoting data (Davenport & Dyché, 2013). They go further to indicate that choosing the right analytical method is crucial, as it also influences the information one can extract from the data. The types of analytics include:

Descriptive Analytics: scrutinises data and information to define the current state of on a situation or organisation, patterns and exceptions, producing ad hoc reports, standard reports and alerts (Joseph & Johnson, 2013).
Inquisitive Analytics: as the name suggests, this type of analytics focuses on using data to support or reject propositions (Bihani & Patil, 2014).
Predictive Analytics: here, data is used to forecast and statistically model future possibilities (Waller & Fawcett, 2013).
Prescriptive Analytics: this is more about optimising and randomly testing data to evaluate how a business or situation is performing (Joseph & Johnson, 2013).
Pre-Emptive Analytics: this involves using the data to develop the capacity to take precautions that may lead to undesirable occurrences. An example is analysing data to use it to identify possible risks and develop recommendation to avoid or deal with the identified risks (Szongott et al., 2012).

To deal with big data analysis, fields of machine learning (ML) and deep learning (DL) framework were developed and expanded to deal with BDA. ML explores predictive features or patterns in big data, thus allowing prediction of what could happen in future. On the other hand, DL plays a major role in the extraction of data that are useful and meaningful. DL, which was inspired by the human brain, was introduced in the 1940s but had not been determined until 2006 when Hinton and Salakhutdinov (2006) introduced the layer-wise-greedy-learning method as a way of overcoming the deficiency of neural network (NN) method. The NN method worked optimally through trapping into the optima local point that is exacerbated with lack of adequate training data in terms of size.

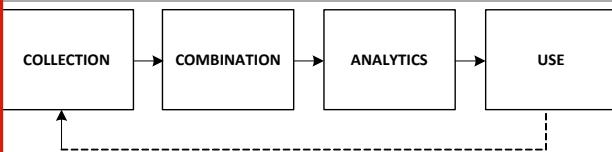
To deal with this challenge, Hinton proposed that unsupervised learning should be used prior to layer-by-layer training. There are four categories of DL algorithms: convolutional neural networks (CNN), auto-encoder, sparse coding and restricted Boltzmann machines. In the case of Hajj and Umrah, there is limited literature on big data and big data analytics even with the large datasets these events come with. There are several technologies being developed that are meant to help the stakeholders to take advantage of this data and use it for better, safer and more organised Hajj and Umrah events (Vanani and Majidian, 2019).

Big Data State, Approaches And Activities: In the literature review, we expounded on the concept of big data. However, in this section, the focus is on big data current state, approaches and activities including possible advantages of using big data in Hajj and Umrah events. Big data poses privacy and reliability requirement as essential in the data management. The task to develop a system that can be scaled to handle data influx and provide sufficient capacity for analysis is a key challenge in big data. This particular challenge includes both hardware resources requirement and architecture, and also processing requirements. Hence, in this section, the current state of big data is briefly discussed along with the approach and activities.

Figure 3: Big Data Approach (Adapted from IBM Student Guide, 2018)



Figure 4: Big Data Activities



Current State of Big Data: Big data has many applications. The intelligent use of big data within the health sector can save over \$300 billion, as seen in precision medicine diagnostics. In addition, as surveyed by Archenaa and Anita (2015), big data analytics is applied in health care and government for patient centric services, detecting spreading diseases earlier, improving treatment method, reducing unemployment rate, providing quality education and addressing needs urgently (McPadden et al., 2018; Keshavarz et al., 2021).

Furthermore, the potentials of big data allow integrating geographical information system and business intelligence, hence building an effective predictive analytic system which can be of very good use and advantage for crowd control

and management. A technology-driven process allows firms to analyse, report and make critical informed decisions. Big data and business analytics market was predicted to increase up to \$203 billion by 2020. It has become the trending practice for organisations to construct and extract valuable information. This is achieved by using open-source applications such as Apache Hadoop, Spark, NoSQL and many others. Hadoop dominates about 60% of the market space (Keshavarz et al., 2021).

With use of big data in business intelligence with ML and AI technologies, we can track of employees, assets, equipment and collect and interpret data to make informed decision, thus resulting in personalisation and better user experience. As reported in McPadden et al. (2018), integrating data-lake and analytic platform used to provide real-time access to health care allows for continuous monitoring of patient in real-time analytics. Big data-lake makes parallelised computation more accessible. Hadoop allows big data storage and batch processing analysis, hence enabling distributed data storage and scalable processing capacity. However, Hadoop has with fewer tools available for streaming data and real-time analysis.

Predictive analytics, intelligent security, internet of things (IoT), edge computing and self-services, in-memory computing and business intelligent application, decision support system and geospatial information system are all different areas and trends on which big data has had a great impact. The current state and trends of big data exploration have various advantages and applications in the industries, government, medical, engineering and social events. Big data is now driving innovations by studying interdependencies between humans, events, institution, processes and then using the insights gained to improve decisions (Keshavarz et al., 2021).

This insight is known as big data visualisation. Visualisation of data is a key purpose of gathering data, which provides insight to the large pool of data. The data visualisation pipeline stages consist of simulating, preparing, mapping, rendering and interpreting. The data flowing from the simulating stage go through preparing, mapping, rendering and interpreting stages, and then return to the simulating stage. The mapping is very essential in the visualisation pipeline in big data analytics. Thus, data visualisation helps to identify inherent patterns and infer correlation and causal relationships (Xuefeng et al., 2020).

The purpose of visualisation is to provide suitable methods and instruments to explore data and information. It enables the extraction of useful information from complex or/and voluminous datasets through the use of interactive graphics and imaging, thus steering the dataset and seeing what may have appeared invisible. Parallel implementation for big data visualisation is used to address scalability issues using many core GPUs, distributed clusters or hybrid architectures (Xuefeng et al., 2020).

Hence, data preparation and rendering during visualisation are achieved using modern GPUs and distributed visualisation frameworks with Hadoop/Spark that

leverages GPUs to compute data aggregates with kernel density estimation. Another example of data visualisation framework is the open-source MapD, which uses GPUs to accelerate the processing of large and complex datasets that can process billions of rows of data in milliseconds. In addition, MAP model and MAP-Vis framework as proposed in Xuefeng et al. (2020) realises millisecond-level multidimensional data querying with good interactive visualisation, and both the MAP model and MAP-Vis framework provide high scalability for processing and online visualisation. They use SPARK as a pre-processing tool and HBase as a distributed storage platform (Perrot et al., 2015; Root & Mostak, 2016; Felemban et al, 2020; Khalid & Zebaree, 2021).

Big Data Approaches and Activities: Big data approaches entail the predictive model design and data mining tools used for big data deployment. Big data processing can be performed by batch processing and stream processing. Batch processing collects and stores data in batches, in order to analyse and generate results, while stream processing is suitable for real-time feedback and requires response time constraints. Stream processing reduces computational time. This big data approach is essential in health care analytics for real-time extraction of needed information from a large amount of patient data, for alerting emergencies and complications; thus, it has been helpful for real-time diagnostics, decision and intervention for urgent care. Figure 3 shows an illustration of big data driving business strategies.

Meanwhile, big data also entails different activities, which includes data collection, combination, analytics and use, as shown in Figure 4 (Klievink et al., 2017). Although, big data is not a technology itself, it is a collection of large datasets that cannot be easily handled by conventional data processing technology due to its great variety, velocity and large volume (Kankanhalli et al., 2016; Klievink et al., 2017). It involves visualisation of data using a more specialised computational tools, storage and analytic technologies.

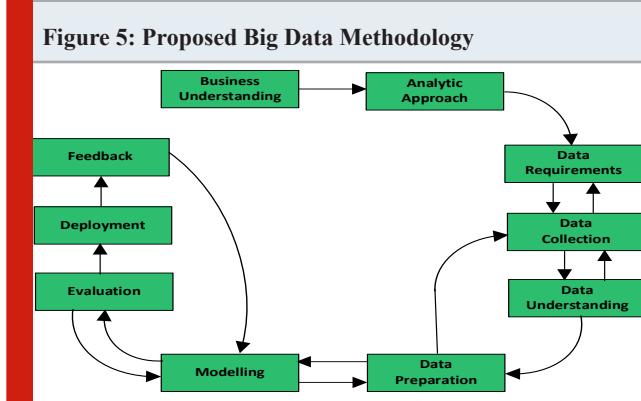
Big Data Advantages in Hajj: There are many advantages of big data in Hajj like: increase perception of risk and more effort in the hospitality management of the visitors, reduce provocation during activities, enhance visitors experience, safety and security, preparation and planning, management and organisation principle and crowd management and response including spontaneous disturbance.

Big Data Methodology: Big data methodology provides methods and analysis of the rules or procedures of inquiry for constructing a model to predict, discover underlying patterns and gain insight in data.

The development methodologies in big data process following a software development life cycle are as follows: feasibility and business understanding; systems analytic approach; data requirement, collection and understanding; data preparation; modelling; evaluation; deployment and feedback. The objectives of a big data project depend on the purpose and must strategically be motivated with strong

executive support, and must meet a business need. Large-scope big data projects may benefit from having managers and a co-project manager, to develop and be executed by project team. If the project will be outsourced, then a process needs to be developed for creating a request for proposals and then evaluating the proposals submitted.

Figure 5: Proposed Big Data Methodology



If the development will occur in-house, development tools and technical issues need to be resolved. The feasibility analysis should have determined if the project could be completed in-house. The project manager should identify tasks that must be completed, resources that are needed and project deliverables. Deliverables are especially important for monitoring the progress of the project. Furthermore, milestones are identified to help non-technical managers monitor a big data project, while the chief information officer (CIO) of the organisation and one or more business managers usually monitor the progress of a large-scope or high visibility big data project. Meanwhile, big data deployment in Hajj objectives are as follows:

- 1. Visitors (Pilgrimage) relationship management:** shopping frequency, locational pricing adjustment, identity recognition, performance targeted advertisement campaigns, visited locations, disability care management and many more.
- 2. Improved data accuracy and management:** this includes the operational maintenance, support, data analysis and tracking and monitoring of data.

Big data project is an expensive task. It is developed to assist people and the organisation in making effective decisions, predictive analysis, customisation, effective data management and improving the lifestyles and utility of the people. However, the development process and approach may encounter failure. An inadequate development process is one without critical and analytical details of the requirements and feasibility of the big data project. Many factors can contribute to this failure which include the following: Poor and inadequate feasibility and preparation, poor and inadequate documentation and tracking, bad leadership and inexperience project manager, incompetent or lack of technical skills for execution of tasks and projects, inaccurate cost analysis and estimation, ineffective communication at every level of management, choice of big data framework and tools used, organisation culture and Competing priorities (Qi et al, 2020).

Big Data Framework Application In Hajj And Umrah: The framework for big data application in Hajj and Umrah events is presented in Figure 6. In this section, we describe the various components within this framework.

Big Data Tools: Big data tools are used for designing and implementing big data projects. The characteristics of these tools are as follows: applications are written in high-level language code, work is performed in a cluster of commodity machines, data is distributed in advance that brings computation to the data, data is replicated for increased availability and reliability and scalable and fault tolerant.

Figure 6: Big Data Framework for Hajj

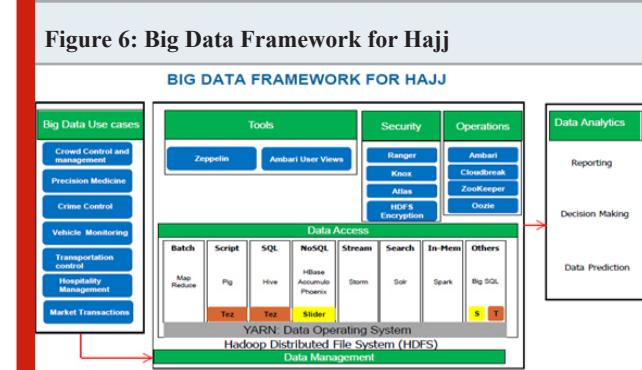


Figure 7: Big Data Tools (Adapted from IBM Student Guide, 2018)



Big data space is a large growing ecosystem. Many tools exist for big data deployment. These tools can be seen in Figure 7. Hadoop is most widely used tool, and it is also an Apache open-source software framework, which is very reliable, scalable and includes distributed computing of massive data. It is developed in Java and consists of three sub projects: MapReduce, Hadoop Distributed File System (HDFS) and Hadoop common. Hadoop is developed in Java and hides the underlying system details and complexity from the user. It is meant for heterogeneous commodity hardware. Its related projects include Hbase, Zookeeper and Avro. Thus, data mining involves using data-intensive and computing-bound algorithms with high processing units to extract information at the required time.

The common use cases for data analytics are extract/transform/load, text mining, index building, pattern recognition, predictive model, risk assessment and collaborative filtering. In the Hajj and Umrah big data

application, it collaborates variety of data and networks to monitor the cases using sensor networks. Thus, sensors are used to monitor people's movements and services, purchase transactions, inventory management, asset tracking, vehicle monitoring and transportation organisation. This design application will require efficient collection, storage and analysis of data. The design framework is an event-driven data processing design that will trigger a certain action when a condition is met. In addition, during the convergence prayer, it will be necessary to carry out stream data processing to monitor events and ensure effective crowd management.

The data access tools for big data design include MapReduce, Pig, Hive, Accumulo, Hbase, Storm, Solr, Spark, Big Sql, Tez and Slider. These are the components that provide data access capabilities (IBM Student Guide, 2018).

MapReduce: It provides frameworks that manage the complexity of parallelisation. Input is split into pieces as HDFS blocks or splits, then the worker nodes process the individual pieces in parallel and store the results in its local file system, where a reducer accesses it. **Spark:** It is a large-scale data processing framework that can be used to write several applications in several languages like Java, Scala, Python and R. The advantage of Spark is that it also allows data to be operated in memory, is easy to use and runs programs faster than MapReduce in memory computation. It can combine SQL, streaming and complex analytics. Furthermore, it runs in variety of environment and also with diverse data access sources such as Hadoop, Mesos, Cloud, Hbase, S3 and Cassandra. **Storm:** Storm is Apache Hadoop stream-processing framework that tracks and determines successful completion of tree of tuples triggered by every spout tuple.

Solr: Apache Solr is a fast open-source enterprise search platform built on Apache Lucene Java search library; it allows full-text indexing search, providing highly reliable, scalable, fault tolerant, centralised configuration and automated failover and recovery. **Hive:** Hive was developed originally by Facebook; it allows abstraction of data on top of non-relational semi-structured data. It provides an SQL-like interface for the user. **Hive** uses a serialisation/deserialisation interface to read data from table and then write back in any custom format. **Big SQL:** Big SQL builds on Apache Hive foundation, and it uses the native C/C++ MPP engine. It is an SQL-processing engine for Hadoop cluster that provides SQL on Hadoop interface. It requires no new SQL syntax nor proprietary storage format. **Pig:** It is a procedural/dataflow open-source programming language originally developed by Yahoo. It is a high-level programming language for data transformation and is better than Hive for unstructured data. Its native language is Pig Latin. The compiler translates it into sequence of MapReduce programs. **HBase:** It is a NoSQL data store and is a distributed and scalable big data store used when random and real time read/write access is needed in big data. Thus, it enables handling of large tables of data running on clusters of commodity hardware. It is modelled after Google's Big Table and provides Big Table-like capabilities on top of Hadoop and HDFS.

Falcon: It is for managing data life cycle in Hadoop clusters for backups, archival of data, feed retention and much more. It is a data governance engine that defines schedules and monitors data management policies. It allows Hadoop admins to centrally define data pipeline, which is used to auto-generate workflows in Oozie. **Ranger:** It is used for data security control over the Hadoop platform. It manages policies for access to files, folders, table, databases and columns. These policies can be set for individual users and groups.

Zookeeper: It provides centralised service for maintaining configuration information. It is fast, reliable and ordered. Distributed applications can use it to store and mediate updates to important configuration information.

Design of Big Data Analytics in the Hajj: The big data analytics in the Hajj can be designed for different reasons such as: fraud detection in credit card transactions, credit issuance, increase speed of detection, crowd control and risk management, 360° view of the pilgrimage, planning and monitoring to reduce congestion, real time analysis to weather, identify traffic patterns to reduce transportation cost and save time of arrival and exit, predict trends and prepare for demand for the Hajj attendees and optimise pricing and promotions. Use cases in the Hajj to be tackled: Precision medicine, transportation/congestion control, Hospitality management, market transactions, crowd control and conflict management and crime detection and prevention.

Big Data Use Cases: The use cases are the data sources generated by sensors, smart devices, social networks, survey, regulatory agencies and IoT equipment. These use cases are included in the data production phase identified for data management and analysis. In this Hajj framework, the use cases are every identifiable factor that will be involved in the pilgrimage; in addition, people will fall sick, people with disabilities will attend and health care services and monitoring of patients are essential in critical cases. Hence, precision medicine which involves real time monitoring will be used.

Transportation and congestion control to identify best route and traffic congestion control while attending and leaving the prayer centre is very essential. This is best managed and identified using big data. Hospitality management ensures effective accommodation, care and services provided to the visitors during Hajj. Therefore, crowd control and management, conflict resolution and market transactions are all events that must be identified, including credit card frauds during business transactions. An effective crime detection and control is a key identifiable need to ensure vehicle monitoring and safety of lives and properties of the pilgrims.

Big Data Management: This is the next phase; it is where the identified sources are managed. This stage involves the collection, storage and processing of data. The data are explored and processed. Data cleaning, transformation and packaging operation is performed. Big data management in the Hajj design consists of the data access, tools, operations

and security. The data management deals with the entities (objects) in the data model, the relations and granularity of data and the time window available of data (Felemban et al., 2020).

Thus, the mapping of data objects with one another in the existing business process and the supporting infrastructure is achieved. In data management, it is required to address the issue of data requirement and the capacity requirement of the data. The features of the data that must be considered for effective management, which include data availability and the timing and longevity of the data. Data manipulation operations are derived by joining data sources, filtering rows and fields of data sources, aggregating, combining and deriving new features, and appending data sources.

Big Data Analysis: After the processing stage of the data using all the required tools for data management, evaluation of the results for reporting, prediction and decision-making is carried out. The analytic process provides data access for informed decisions, improved and higher security and smooth pilgrimage. A data quality report is necessary; it describes the features of the analytic base table, using standard statistical measures such as central tendency and measures of dispersion. This process is followed by data visualisation using bar charts, graphs, histograms and other convenient data plots methods (Xuefeng et al., 2020).

Consequently, identifying and handling the data quality issues can be addressed here: missing values, outliers and cardinality. Thus, data report is examined here for all the use cases for the Hajj and Umrah events. There are two types of data examinations during the visualisation process: categorical and continuous data. In the categorical data examination, the modes are examined to observe level of domination in the dataset, whereas, in the continuous data examination, the mean and standard deviation are examined to make sense of the central tendency and variations within the dataset.

Big data is the new keyword today that is filled with large pool of data. The need for the use of data is to explore and identify pattern, prediction for informed guesses and infer to quantify what we know. The model for crowd control and management using big data analytics is to first know what is expected and also focus on possible harmful conditions and behaviours that might arise during the pilgrimage. Enhancing visitors experience can be obtained by ensuring public safety, explain changes in the rules, give proper direction and provide guidelines that must be enforced and followed during the Hajj and Umrah pilgrimages. Big data analytics enable the management to target and intervene quickly on danger situations and scenario of disturbance and violence. There must be rules and guidelines that reduce the potentials of misinterpretation of rules.

Additionally, the management and organisers must ensure consistent training and intervention prior and during the pilgrimage. The big data framework for Hajj is designed with three phases: big data uses case, the data management and data analysis. The use cases are the big data production; they are the identifiable sources of data that will be

critically needed and involved during the Hajj and Umrah pilgrimages. As elaborated earlier, Hajj and Umrah involve the largest mass gathering activities in the world (Ledhem et al., 2020).

Thus, there is need to effectively make informed decisions, predictions and innovations in the process to ensure reduction of hazards, risk factors, death and other vices that may meet the visitors. Therefore, big data deployment in Hajj and Umrah will help in minimising all the risk factors and ensure smooth operations and satisfaction of the pilgrimage, thus fostering effective realisation of the goals and objectives of the pilgrimage (Shambour & Gutub, 2021).

This research has also shown that big data implementation starts with feasibility study with formal report or document, providing the business understanding with the risk factors that will affect the potential development and implementation the design. This feasibility study is a major checkpoint with critical information on whether it is possible to develop a system given the project goals and constraints. This is followed by in-house big data development or purchasing the big data software applications and packages. Thus, packaged solutions sometimes may be generalised and may not provide some specialisation need as required by the development team in building the analytical and decision support system for predictive analysis in big data applications. Just as mentioned earlier on big data approaches, it is necessary to consider the nature of the approach, whether it requires stream processing or batch processing. A big data deployment requires a development process which includes methodologies that will be used in the big data development.

CONCLUSION

This paper designed big data framework for Hajj and Umrah Islamic pilgrimage, as it is now very essential in this Islamic pilgrimage that conglomerates an estimated 3 million people in a location. Therefore, because of the growing rate of people participating in the program, there is need for crowd control and management, crime detection and monitoring, congestion and transportation control, market transactions and health care services to deploy big data for the Hajj and Umrah pilgrimage. The framework for Hajj will solve and bring to the barest minimum the existing challenges faced in the pilgrimage. This will improve the efficiency and effective management of the pilgrimage. It will foster reliable surveillance of the process, with increased revenue generation, opportunities and satisfaction for the pilgrims.

REFERENCES

- Akerkar, R. (2014). *Big data computing*. CRC Press, Taylor & Francis Group.
- Algadhi, S.; Mufti, R.; Malick, D. (2002). Estimating the Total Number of Vehicles Active on the Road in Saudi Arabia. *J. King Abdulaziz Univ. Eng. Sci.* 14, 3–28
- Al-Nuaim, H., & Al-Masry, M. (2012). The use of mobile technology applications for crisis management during Hajj.

- Journal of Occupational Safety and Health, 9(2).
- Amro, A., & Nijem, Q. (2012). Pilgrims Hajj tracking system (e-Mutawwif). *Contemporary Engineering Sciences*, 5(9), 437–446.
- Archenaa J, Anita EM. A. (2015). Survey of big data analytics in healthcare and government. *Procedia Computer Sci* 2015; 50:408–13.
- Apollos, E. C., Adeshina, S. A., & Nnanna, N. A. (2019). Memristor-based CiM architecture for big data era .15th International Conference on Electronics, Computer and Computation (ICECCO) 1–6. <https://doi.org/10.1109/ICECCO48375.2019.9043218>.
- Bihani, P., & Patil, S. T. (2014). A comparative study of data analysis techniques. *International Journal of Emerging Trends & Technology in Computer Science*, 3(2), 95–101.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662–679.
- Brian F. Kingshott (2014). Crowd Management: Understanding Attitudes and Behaviours, *Journal of Applied Security Research*. 9:3, 273-289, DOI: 10.1080/19361610.2014.913229 from: <http://www.cdsi.gov.sa/english/index.php>
- Central Department of Statistics and Information. (2014). Annual statistical report. <http://www.cdsi.gov.sa/english/index.php>
- Chen H., Chiang R.H.L., Storey V.C., (2012). Business intelligence and analytics: From big data to big impact *MIS Quarterly*, 36 (4), 1165-1188.
- Crawford, K. (2013). The hidden biases of big data. *Harvard Business Review Blog*. <http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/>
- Creagh, B., (2016). Crown Melbourne capitalizes on big data to enhance FM. *FM online Magazine*. <https://www.fmmagazine.com.au/sectors/crown-melbourne-capitalises-on-big-data-to-enhance-fm/>
- Davenport, T. H., & Dyché, J. (2013). Big data in big companies. SAS Institute Inc. Available at <http://www.sas.com/resources/asset/Big-Data-in-Big-Companies.pdf>
- Felemban, E. A., Rehman, F. U., Biabani, S. A. A., Ahmad, A., Naseer, A., Majid, A. R. M. A., ... & Zanjir, F. (2020). Digital revolution for Hajj crowd management: a technology survey. *IEEE Access*, 8, 208583-208609]
- Fruin, J., (1993). The causes and prevention of crowd disasters. In S. R. Dickie J (Ed.), *Engineering for crowd safety*, 99–108.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144.
- Halevi, G., & Moed, H. (2012). The evolution of big data as a research and scientific topic: overview of the literature.
- Research Trends, Special Issue on Big Data, 30, 3–6.
- Hargittai, E. (2015). Is bigger always better? Potential biases of big data derived from social network sites. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 63–76.
- Hinton, G., & Salakhutdinov, R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 504–507.
- IBM Student Guide. (2018). *Introduction to Big Data Ecosystem*. Skills Academy: Big Data Engineer 2018.
- Joseph, R. C., & Johnson, N. A. (2013). Big data and transformational government. *IT Professional*, 15(6), 43–48.
- Kankanhalli, A., Hahn, J., Tan, S., & Gao, G. (2016). Big data and analytics in healthcare: Introduction to the special section. *Information Systems Frontiers*, 18, 233–235. <https://doi.org/10.1007/s10796-016-9641-2>
- Keshavarz, H.; Mahdzir, A.M.; Talebian, H.; Jalaliyoon, N.; Ohshima, N. (2021). The Value of Big Data Analytics Pillars in Telecommunication Industry. *Sustainability*, 13, 7160. <https://doi.org/10.3390/su13137160>.
- Khalid, Z. M., & Zebaree, S. R. (2021). Big data analysis for data visualization: A review. *International Journal of Science and Business*, 5(2), 64-75.
- Khozium, M. O., Abuarafah, A. G., & AbdRabou, E. (2012). A Proposed Computer-Based System Architecture for Crowd Management of Pilgrims using Thermography. *Life Science Journal* 9(2): 377- 383. <http://www.lifesciencesite.com>.
- Klievink, B., Romijn, B. J., Cunningham, S., & de Bruijn, H. (2017). Big data in the public sector: Uncertainties and readiness. *Information systems frontiers*, 19(2), 267–283.
- Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, 5(12), 2032–2033.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Baraba'si, A., Brewer, D., & Van Alstyne, M. (2009). Computational social science. *Science*, 323(5915), 721–723.
- Ledhem, M. A., & Moussaoui, W. (2020). The impact of accommodation entrepreneurship activities on Islamic tourism (Umrah) development: An empirical evidence from Saudi Arabia. *International Journal of Hospitality and Tourism Studies*, 1(2), 111-118.]
- Luz, N. (2020). Pilgrimage and religious tourism in Islam. *Annals of Tourism Research*, 82, 102915.]
- Maciej Szczodrak, Jozef Kotus, Krzysztof Kopaczewski, Kuba Lopatka, Andrzej Czyzewski, Henryk Krawczyk, (2011). Behaviour analysis and dynamic crowd management in video surveillance system, 22nd International Workshop on Database and Expert Systems Applications.
- Mahmassani H. and Sheffi Y. (1981). Using gap sequences to estimate gap acceptance functions. *Transportation*

- Research 15B, (143-148).
- Manyika J., Chui M., Brown B., Bughin J., Dobbs R., & Roxburgh, C. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey and Company.
- Mawed, M., & Aal-Hajj, A. (2017). Using big data to improve the performance management: a case study from the UAE FM industry. Facilities, 35(13-14), 746 765. <https://doi.org/10.1108/F-01-2016-0006>
- McPadden, J., Durant T. J., Bunch, D. R., Coppi, A., Price N., Rodgerson, K., & Schuls W. L. (2018). A scalable Data Science Platform for Healthcare and Precision Medicine Research. ArXiv preprint arXiv:1808.04849.
- Perrot, A., Bourqui, R., Hanusse, N., Lalanne, F., & Auber, D. (2015). Large interactive visualization of density functions on big data infrastructure. In Large data analysis and visualization (LDAV) (99–106) [Symposium]. 2015 IEEE 5th Symposium, Chicago, IL, USA.
- Pin, S., Fazilah, H., Siamak, S., Abdullah, Z., & Ahamad, K. (2011). Applying TRIZ principles in crowd management. Safety Science 49(2011), 286–291.
- Preko, A., Allaberganov, A., Mohammed, I., Albert, M., & Amponsah, R. (2020). Understanding spiritual journey to hajj: Ghana and Uzbekistan perspectives. Journal of Islamic Marketing.
- Qi, G. J., & Luo, J. (2020). Small data challenges in big data era: A survey of recent progress on unsupervised and semi-supervised methods. IEEE Transactions on Pattern Analysis and Machine Intelligence.
- Reffat, R. (2012). An Intelligent Computational Real-time Virtual Environment Model for Efficient Crowd Management, International Journal of Transportation Science and Technology, 1, (4), 365-378.
- Root, C., & Mostak, T. (2016). A GPU-powered big data analytics and visualization platform. Proceedings of the ACM SIGGRAPH 2016 Talks (pp. 1–2), Anaheim, CA, USA.
- Shambour, M. K., & Gutub, A. (2021). Progress of IoT research technologies and applications serving Hajj and Umrah. Arabian Journal for Science and Engineering, 1-21.]
- Sivarajah, U., Kamal, M., Irani, Z., and Weerakkody, V. (2016). Critical analysis of Big Data challenges and analytical methods. Brunel University London, Brunel Business School United Kingdom
- Sonia, H., & Mohamed, A. (2015). Efficient wireless sensor network rings overlay for crowd management in Arafat area of Makkah. Signal Processing, Informatics, Communication and Energy Systems (SPICES), 2015 IEEE International Conference.
- Stefania, B., Mizar, F., & Giuseppe V. (2007). Situated cellular agents approach to crowd modeling and simulation. Cybernetics and Systems, 38(7).
- Still, K., Papalexis, M., Fan, Y. and Bamford, D., (2020). Place crowd safety, crowd science? Case studies and application. Journal of Place Management and Development.
- Szongott, C., Henne, B., & von Voigt, G. (2012). Big data privacy issues in public social media . 6th IEEE International Conference on Digital Ecosystems Technologies (DEST) 1–6.
- Tole, A. A. (2013). Big data challenges. Database Systems Journal, 4(3), 31–40.
- Vanani, I., & Majidian, S. (2019). Literature review on big data analytics methods. Intechopen. <https://doi.org/10.5772/intechopen.86843>
- Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. Journal of Business Logistics, 34(2), 77–84.
- Wu X., Zhu X., Wu G.-Q. and Ding W., (2014). Data mining with big data, IEEE Trans. Knowl. Data Eng., 26 (1), 97-107.
- Xuefeng, G., Chong, X., Linxu, H., Yumei, Z., Dannam, S., & Weiran, X. (2020). MAP-Vis: A distributed spatio-temporal big data visualization framework based on a multi-dimensional aggregation pyramid model. Appl. Sci., 10(2), 598. <https://doi.org/10.3390/app10020598>
- Zicari, R. V. (2014). Big data: Challenges and opportunities. In R. (Ed.), Big data computing 103–128. CRC Press, Taylor & Francis Group.