Data Analysis and Visualization of Terrorist Attacks in the Philippines

Abraham Magpantay 1,2, Jay arr Buhain 2, Hadji Tejuco 1
Jennifer Contreras 1, Elisa Malasaga 1, Beau Gray Habal 1

1 FEU Institute of Technology, Manila, Philippines.
2 Ateneo de Manila University, Manila, Philippines.

Email: atmagpantay@feutech.edu.ph; buhainjararr.official@gmail.com; hjtejuco@feutech.edu.ph;
jocontreras@feutech.edu.ph; evmalasaga@feutech.edu.ph; bmhabal@feutech.edu.ph

Abstract - Nowadays, terrorism has inflicted chaos on humanity that causes fear to the people. The early detection of possible terrorist attacks is useful and important to avoid the serious damage it may cause. The Philippines was recently affected by terrorist attack situated in Marawi that caused a lot of destruction in the city and lost thousands of lives. This study focuses on the visualization and analysis of Global Terrorism Attacks (GTA) that may be helpful to the different organizations working in this area. We used the data of terrorist attacks from Kaggle website which is an open-source database named Global Terrorism Database, which comprises information of terrorist attacks around the world since 1970 until 2016 except for 1993. The Sharded Clustering algorithm was used in the study in the process of storing data records across multiple machine to meet the demands of data growth through MongoDB. The visualization focuses on the identification of terrorist attacks in the Philippines every year from 1970 and 2016, targeted Philippine provinces by terrorist attacks and the type of attack that is mostly used by the terrorists.

Keywords - Terrorism, Global Terrorism, Visualization, Analysis, MongoDB, Replica Sets, Sharding, Config Server, MapReduce, Aggregate Functions

I. INTRODUCTION

Terrorism is considered as a frightening phenomenon created by man that affects every facet of society. Nowadays, it emerged as a human threat no matter what nationality, culture or beliefs you have. Even though, the initiatives to start the international binding definition of terrorism was not successful, still we can say that terrorism causes human suffering and widespread fear. Modern attacks proved the disrupting potential of terrorism on social life and business [15].

According to Global Terrorism Index (2015), the universal economic costs of terrorism are its peak since the year 2001. In 2014, it was estimated that the cost of global economic terrorism reached 52.9 billion USD with a total economic impact of 105.8 billion USD [15].

Nowadays, terrorist attacks can be considered as one of the biggest problems in the world due to the constant threat of a well-planned, sophisticated and well-coordinated terrorist operations. Therefore, many countries are paying attention on counter-terrorism acts where they are continuously monitoring terrorist activities though some countries do not have enough trained officers to work on processing bulk data in a short time for decision-making regarding the terrorist attacks [2].

U.S. Government encourages different countries to make advantage of using advanced technologies to fight terrorism is succeeding [16]. In this study, we have analyzed a historical data using MongoDB, check on the terrorist group activities and their involvement different attacks in the Philippines.

A. Research Objectives

This study aims to present significant information in terms of terrorism attacks in the Philippines.

B. Research Questions

This study aims to answer the question: How were terrorists able to create successful attacks in the Philippines? In pursuit of this, the following will be addressed:

1. What year has the greatest number of terrorist attacks?
2. What is the most targeted province in the Philippines in terms of terrorist attacks?
3. What type of attack is the most used by terrorists?

C. Scope and Limitations

The focus of this study is to provide data analysis on the affected areas in terms of terrorism in the Philippines. This study utilizes the use of the Clustering algorithm that provides clear identification of the groups and types of attacks made in the Philippines. The researchers use MongoDB as its engine in extracting data and providing information. In our study, the database includes historical data of terrorism in all continents of the world from 1970 to 2016 except 1993. The simplicity and effectiveness of the
existing techniques in MongoDB will be utilized all throughout the study including the application of Replication Sets, Sharding and MapReduce algorithms.

D. Significance of Study

This study may contribute to the improvement of ways of the Philippine government to prevent terrorist attacks. It will allow the government agencies to further speed up and maintain accuracy in their work, eliminating problems in growing number of terrorist activities all over the world that leads to injuries and deaths of innocent people. This study may also contribute in the field of computer science, from the results, it might show that there’s a database like MongoDB that could produce visualizations on the terrorist activities that help towards combating terrorism.

Data Visualization (DV) is useful in understanding data about terrorist attacks and can represent data in a more organized way. It is used to let people know the importance of data by placing visual context. In fact, an example of DV is a simple table that can be used to elaborate “a map of geographic data depicting an additional layer in Google Earth” or represents the complex Facebook’s social relationships data.

The research is designed as follows: 1) Introduction, 2) The review of literature where we give an overview of important concepts for this study. 3) The third section discusses the methodology of the study that includes the selection of the Dataset, Setting up the MongoDB and the application of the different MongoDB functions. 4) the fourth section is the results and analysis and the last section is 5) the conclusion and the recommendation.

II. REVIEW OF RELATED LITERATURE

From the wide range of techniques and disciplines covered by the study, this literature review intends to focus on 3 main topics: terrorism, threats of terrorism, Data Visualization and MongoDB Each of these topics will be divided into relevant subtopics to give further discussion in the study

A. Terrorism

The word terrorism can be integrated with the deprivation of human rights and denial of democracy. It is factual, universal and multifaceted, though it must be assessed cautiously to ensure that it is not exaggerated or understated. Each state is responsible to prevent terrorism and protect the rights of mankind. However, the state needs to draw on the support of society in general, including civil societies and businesses to successfully fight this phenomenon [12]. In the Philippines, ‘terrorism’ is considered as a major social issue and usually linked to rebellions as a terrorism act.

a) Regional Drivers of Terrorism: Greed, Grievance and Geography: The most common reasons why some people or group of people are doing terrorism acts like rebellions are greed, grievance and the geographic aspect. In the Philippines, the terrorism’s frequent drivers are the manifestation of the armed forces, political exclusion and group grievances remain the vital factors. The social, economic and political grievances, the economic variables are found to have the most explanatory power in addressing causes of internal armed conflict [6].

B. Threats of Terrorism

Studying about terrorist organizations and terrorism is a complex job since its main purpose is to understand the concept of terrorism acts. In fact, the word terrorism itself is terrifying and contentious since it has no agreed explanation of the word ‘terrorism’ in the international community [12]. Different terrorist organizations participate in several activities like bombings, hostage takings, assassinations, kidnappings, armed assault and hijackings [8] can cause death and destruction to mankind.

a) External and Internal Threats: Terrorism is the act where one group attacks the others thru killings, placing fires to houses or offices. On the other hand, suicide bombing is the common scenario in different places where the terrorists carry bombs and blow themselves up that can cause death and damage to the surroundings. Furthermore, another threat is the hijacking of planes where terrorists group will perform the airplane hijacking as they ask to release political prisoners. These attacks typically happen in a place where you can find a lot of people. The event of 911 is the best example of a terrorism act.

The acts stated are considered as terrorist attacks and these are being heard by people every day in the news. Unfortunately, some people are taking advantage of the terrorism issues in the country where they put the blame to the terrorist’s group though it is a simply an act of crime like shooting and killing someone. This in a way is considered as terrorism, but these acts are more difficult to detect and identified [12]. Moreover, the threats inside the state are threats from people in an organization attacking others around them but they are not using bombs nor airplanes. Some just simply used a sinister mechanism instead. A sample insider threats includes an agent in an intelligence agency committing espionage or it could be a person from a specific company that gives information to a competitor of proprietary products. Insider threats can occur at all levels that can be dangerous since it is hard to know who these terrorists are.

C. Data Visualization

Challenges of working with “Big Data” lies in capturing, storing, analyzing, sharing, searching, and visualization [5].
Visualization is considered as the “front end” of “Big Data” and data visualization (DV) is an organized data representation like charts or diagrams[10]. One other hand, advanced analytics is incorporated in various approaches to assist the development of interactive and animated graphics on desktop pc, laptop, or mobile devices such as tablets and smartphones [9]. Table 1.0 [14] illustrates the benefits of visualization of data corresponding to the respondent percentages of a survey.

<table>
<thead>
<tr>
<th>Benefits</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved Decision Making</td>
<td>77</td>
</tr>
<tr>
<td>Better ad-hoc data analysis</td>
<td>43</td>
</tr>
<tr>
<td>Improved collaboration/information sharing</td>
<td>41</td>
</tr>
<tr>
<td>Provide self-service capabilities to end users</td>
<td>36</td>
</tr>
<tr>
<td>Increased return on investment (ROI)</td>
<td>34</td>
</tr>
<tr>
<td>Time savings</td>
<td>20</td>
</tr>
<tr>
<td>Reduced burden on IT</td>
<td>15</td>
</tr>
</tbody>
</table>

The advantages of using Data Visualization tools is shown in Table 1 that can be a useful aide to improve decision-making, improved collaboration or sharing of information. Understanding information in a visual manner is easier than having a huge table size with lots of rows and columns.

There is a tendency that the person creating visual representation may be formulating the visualization results incorrectly since there are a lot of options that are present. In fact, in some scenarios, there already pre-defined visuals you can select for a specific data. Showing an alternative visual like DV can convey information and clearly and can provide just what’s needed to easily grasp the most relevant factors affecting important decisions.

D. MongoDB

MongoDB [11][13] can be identified “as an open source NoSQL document store database and commercially reinforced by 10gen [1]. Though it offers different features or manipulating databases like secondary indexing, range queries and sorting, It’s not organizing data in tables using rows and columns but using ‘documents’ to store data where each data is associative with lists and scalar or associative nested arrays.

“MongoDB documents are serialized” naturally as “JavaScript Object Notation (JSON) objects and are in fact stored internally using a binary encoding of JSON” called BSON [3].

a) Sharding: To scale MongoDB’s performance thru a server clustering, MongoDB practices a method named Sharding, that pertains to the process of evenly splitting data throughout the cluster to parallelize the use. It is applied thru dividing the MongoDB server front-end routing servers (mongos) set which routes operations to a group of back-end data servers (mongod).

III. METHODOLOGY

To clearly layout the outputs expected from this study and ensure that they are properly in line with the questions it aims to answer, this methodology was divided into several stages to achieve the goal of this study: Data Set, Setting up MongoDB, MapReduce and Aggregate Function and Data Representation.

A. Dataset

The data set used in this study was taken from Kaggle, it is a website wherein you can discover big sets of open data and apply analysis on it depending on your need. In this study, we used an open-source database named Global Terrorism Database (GTD). GTD comprises information of terrorist attacks around the world since 1970 until 2016 except for 1993. It also includes domestic and international terrorist data about the terrorism acts occurred in 1970s until today. Currently, GTD have greater than 170,000 terrorist attacks cases. The maintenance of data id done daily at the Maryland University by the researchers of National
Consortium for the Study of Terrorism and Responses to Terrorism (START) and is updated on a regular basis.

B. Setting up MongoDB

Installer for MongoDB was downloaded from the MongoDB Download Center website and was installed to the machine. By default, MongoDB will be installed inside the program files folder under your drive C.

a) Deploying Replica Sets: Replica sets in MongoDB are used to provide redundancy on the collections in case of system failure. In this study, there were three replica sets that were used and deployed, each replica set contains three mongod instances. The first replica set is used to be the config server and the other two is for the shard servers. By default, each mongod instance runs on the port 27017, if you’re running a several mongod instances in one machine, you may change the port number to avoid conflict. Replica set for config server was named ‘configReplSet’ then ‘rs0’ and ‘rs1’ for the two shards.

b) Config Server: Config server is used to store in the sharded cluster the metadata where it indicates the design for all data and elements within the ‘sharded’ cluster. It includes the chunks listing on every shard and the ranges that define the chunks. It is responsible for the grouping of data and its distribution to the different shards. In this study, config server was deployed as a replica set since “using a replica set for the config servers improves consistency across the config servers.”

c) Shard Cluster: Once the replica sets for config server were established, replica sets for each of the two shard are added using the mongos. In MongoDB, the mongos are the routing service shard configurations which handles queries from the application layer. It identifies the data position in the “sharded cluster”, to accomplish the procedures issued to them.

d) Construction of Database and its Collection: After the config server and shard were setup, it is time to construct the database and its collection. Database was created together with the collection to be used in it in the mongos prompt. Once the database and collection exist, we may now enable the “sharding” on the database level allowing it to chunk its data and distribute it to the available shards. Here, it is important to determine the shard key since it identifies how MongoDB distributes the documents between shards. The shard key was identified and indexed then “sharding” on the collection was enabled. Once “sharding” on the collection is ready, the balancer redistributes chunk of documents on the next run. Finally, once we insert more documents in the collection, the mongos will route the documents to the appropriate shard.

e) Loading of Datasets to MongoDB: The GTD was downloaded on the Kaggle’s website as csv file then it was converted to JSON using the available online tools/converter. It was then imported into the mongos using mongo import command.
C. MapReduce and Aggregate Function

To be able to address the research question, series of MapReduce function were used to extract data in the collection. First the Successful Terrorist Attacks and its Targets (STAT) were extracted using the MapReduce. To answer the research question 1, MapReduce was again applied on the previous result of the MapReduce (SAT) to get the year and its total successful terrorist attacks then aggregate function was used to determine the year with highest successful terrorist attacks. For the research question 2, The result of MapReduce (STAT) was again used to another MapReduce get the most attack targeted by the terrorist and aggregate function was again applied. For the last research question, the same process was done. The result from STAT was again used to provide the output of another MapReduce to obtain the most kind of attack used, and aggregate function was used again.

D. Data Representation

To visually represent the data produced by the MapReduce and aggregate function of MongoDB, this study has used PHP to be the programming language to perform these queries and display them to be visualize using HTML and bootstrap.

IV. RESULTS AND DISCUSSIONS

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted a n d  s t y l e d .  D o  n o t  u s e  h a r d  t a b s ,  a n d  l i m i t  u s e  o f  h a r d
returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

This section concludes the study by answering the research questions that were mentioned beforehand. To clearly layout the output obtained it was presented in such a way that it answers each of the research questions.

A. Answers to Research Questions

a) What year has the greatest number of terrorist attacks?
Based from the result using MapReduce and aggregate function applied by the researchers, it was found out that the year that has the greatest number of terrorist attacks in the Philippines was in the year 2015. In the figure 1, we can see the scatterplot of the terrorist attacks from 1970 up to 2016.

![Terrorist Attacks per Year](image)

Figure 3. Terrorist from 1970 to 2016

b) What is the most targeted province in the Philippines in terms of terrorist attacks? On the result obtained from the MapReduce and aggregate function applied, the top three most target by terrorist attacks are as follows: Maguindanao Metropolitan Manila, North Cotabato. In figure 2, it is shown that the most targeted province of terrorist attacks is Maguindanao.
c) What type of attack is the most used by terrorists? In figure 3, the results obtained using the MapReduce and aggregate functions used in MongoDB shows that using bombing or explosion is the most used attack by the terrorist in their successful attacks followed by armed assault and assassination.

V. CONCLUSIONS

This study has successfully demonstrated the use of MapReduce and aggregate function in querying useful information using MongoDB. It also has implemented the use of Replica Sets, setting up Config servers and perform Sharding to the cluster. This study was able to apply the concepts on MongoDB successfully to ensure the data consistency and backups in case of failure. Replica sets, config servers and shard clusters were used in the setup of the MongoDB environment.

REFERENCES


DOI 10.5013/IJSSST.a.20.S2.28 28.6 ISSN: 1473-804x online, 1473-8031 print


