

**DATABASE**

**SPECIFICATIONS**

*Global Terrorism Database*

**Guadalupe Gamero,**

**Griselda Pasillas, Erika Hernandez, Joseph Griffith, Julaam Diop, Vedant Vashishtha**

November, 2020

**Revision Sheet**

|  |  |  |
| --- | --- | --- |
| **Release No.** | **Date** | **Revision Description** |
| Rev. 0 | 11/16/20 | Database Specifications Template and Checklist |
| Rev. 1 | 11/30/20 | Additions and Revisions |
| Rev. 2 | 11/2/2020 | Additions and Revisions |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**DATABASE SPECIFICATIONS**

**TABLE OF CONTENTS**

Page #

1.0 GENERAL INFORMATION 1-1

1.1 Purpose

1.2 Scope

1.3 System Overview

1.4 Project References

1.5 Acronyms and Abbreviations

1.6 Points of Contact

1.6.1 Information

1.6.2 Coordination

1.6.3 Additional Points of Contact

1.6.4 Data Owners

2.0 DATABASE IDENTIFICATION AND DESCRIPTION

2.1 Naming Conventions

2.2 Database Identification

2.3 Systems Using the Database

2.4 Relationship to Other Databases

2.5 Schema Information

2.5.1 Description

2.5.2 Physical Design

2.5.3 Physical Structure

2.6 Data Dictionary

2.7 Special Instructions

3.0 DATABASE ADMINISTRATIVE INFORMATION

3.1 Responsibility

3.2 System Information

3.2.1 Database Management System (DBMS) Configuration

3.2.2 Hardware Configuration

3.2.3 Database Software Utilities

3.2.4 Support Software Available for Maintaining Database

3.2.5 Security

3.3 Storage Requirements

3.4 Recovery

3.5 Partition/File Information

3.5.1 Content

3.5.2 Description

3.5.3 Partition/File Interdependencies

3.6 Database Interfaces

3.6.1 Description of Operational Implications

3.6.2 Description of Data Transfer Requirements

3.6.3 Description of Formats of Data

3.7 Error Handling

**1.0 GENERAL INFORMATION**

# **GENERAL INFORMATION**

## **1.1 Purpose**

The purpose of the Global Terrorism Database (GTD) is to provide accurate information on terrorist events and activities worlwide. This collection of data can be traced back to 1970 and is relevant up to the year 2018.

Furthermore, we have taken data from the GTD, data only from 2016 through 2018, and have utilized this data to create our own database. Our database was created with the sole purpose of implementing security methods. It was not intended to be an improvement, change, update or anything of that matter to the GTD, that is provided by the National Consortium for the Study of Terrorism and Responses to Terrorism.

## **1.2 Scope**

Before implementing security protocols, a database must be designed beforehand. The design process included but was not limited to mapping entities to tables, attributes to columns, unique identifiers to unique keys and relationships to foreign keys. Once the database was created, security measures were taken into consideration. For example, accounts were created in order to comply with NIST AC-2. The end of this documentation will include a more detailed description of the controls implemented to the database.

## **1.3 System Overview -STILL MISSING**

<https://portal.helcom.fi/meetings/HELCOM%20PLUS%205-2014-135/MeetingDocuments/4-1%20Database%20functional%20specifications%20document.pdf>

Provide a brief system overview description as a point of reference for the remainder of the document. In addition, include the following:

* Responsible organization
* System name or title
* System code
* System category
* *Major application*: performs clearly defined functions for which there is a readily identifiable security consideration and need
* *General support system*: provides general ADP or network support for a variety of users and applications
* Operational status
* Operational
* Under development
* Undergoing a major modification
* System environment and special conditions

**EXAMPLE**

System Overview Details

System name HELCOM PLUS

System type Client

Server Application

Operational status In development

Database Name PLC Database

## **1.4 Project References**

List of the references that were used in preparation of this document

* National Vulnerability Database (NIST SP 800-53): <https://nvd.nist.gov/800-53/Rev4/impact/low>
* Global Terrorism Database (National Consortium for the Study of Terrorism and Responses to Terrorism) : <https://www.start.umd.edu/data-tools/global-terrorism-database-gtd>
* MySQL Workbench Manual: <https://dev.mysql.com/doc/workbench/en/>

## **1.5 Acronyms and Abbreviations**

|  |  |
| --- | --- |
| **Acronyms/Abbreviation** | **Meaning** |
| GTD | Global Terrorism Databse |
| START | National Consortium for the Study of Terrorism and Responses to Terrorism |
| DB | Database |

*Table 1. Acronyms/Abbreviations and their meanings*

## **1.6 Points of Contact**

### **1.6.1 Information**

Table of the points of contact (POCs) that may be needed by the user for informational and troubleshooting purposes. All listed POCs are team members of the project.

|  |  |
| --- | --- |
| **Point of Contact Name** | **Email Address** |
| Guad Gamero | [ggamero@hawk.iit.edu](mailto:ggamero@hawk.iit.edu) |
| Erika Hernandez | [ehernandez14@hawk.iit.edu](mailto:ehernandez14@hawk.iit.edu) |
| Griselda Pasillas | [gpasillas@hawk.iit.edu](mailto:gpasillas@hawk.iit.edu) |
| Joseph Griffith | [jgriffith@hawk.iit.edu](mailto:jgriffith@hawk.iit.edu) |
| Julaam Diop | [jdiop@hawk.iit.edu](mailto:jdiop@hawk.iit.edu) |
| Vedant Vashishtha | [vvashishtha@hawk.iit.edu](mailto:vvashishtha@hawk.iit.edu) |

### **1.6.2 Coordination**

No additional coordination is needed for this database.

### **1.6.3 Additional Points of Contact**

There are no additional points of contact.

### **1.6.4 Data Owners**

The data for this database derived from the START. Consequently, since what we have done is only extract the data and we have not manipulated, edited, deleted or altered the data, this means that we are not the owners of the data. Meaning that the owners of this data within our database is START and they are responsible for data accuracy integrity and relevancy.

**2.0 DATABASE IDENTIFICATION AND DESCRIPTION**

# **DATABASE IDENTIFICATION AND DESCRIPTION**

## **2.1 Naming Conventions**

To avoid any confusion the database has been named GTD. This is due to the fact that the data for this particular database derived from the START provided GTD. Since this is just a narrowed down version of that same database, the name has remained the same. Additionally, the tables within this database have all been named relatively the same, meaning that the begin with the word event followed by a name of a column, this column identifies what the data in the table means.

## **2.2 Database Identification**

This database may be identified as the GTD database.

## **2.3 Systems Using the Database**

|  |  |  |  |
| --- | --- | --- | --- |
| **Vendor** | **Product** | **Version** | **Comments** |
| Oracle Corporation | MySQL Workbench | 8.0.21 | The version utilized for the creation of this database was MySQL Workbench Community. |
|  |  |  |  |

*Table 2. Database Software Utilities*

## **2.4 Relationship to Other Databases**

This database that has been created is not intended to supersede the existing database by START. On the contrary, this database is simply intended to interface with the existing database, as it makes it less complex by having a limited time frame, recall that START runs from 1970 to the present.

## **2.5 Schema Information**

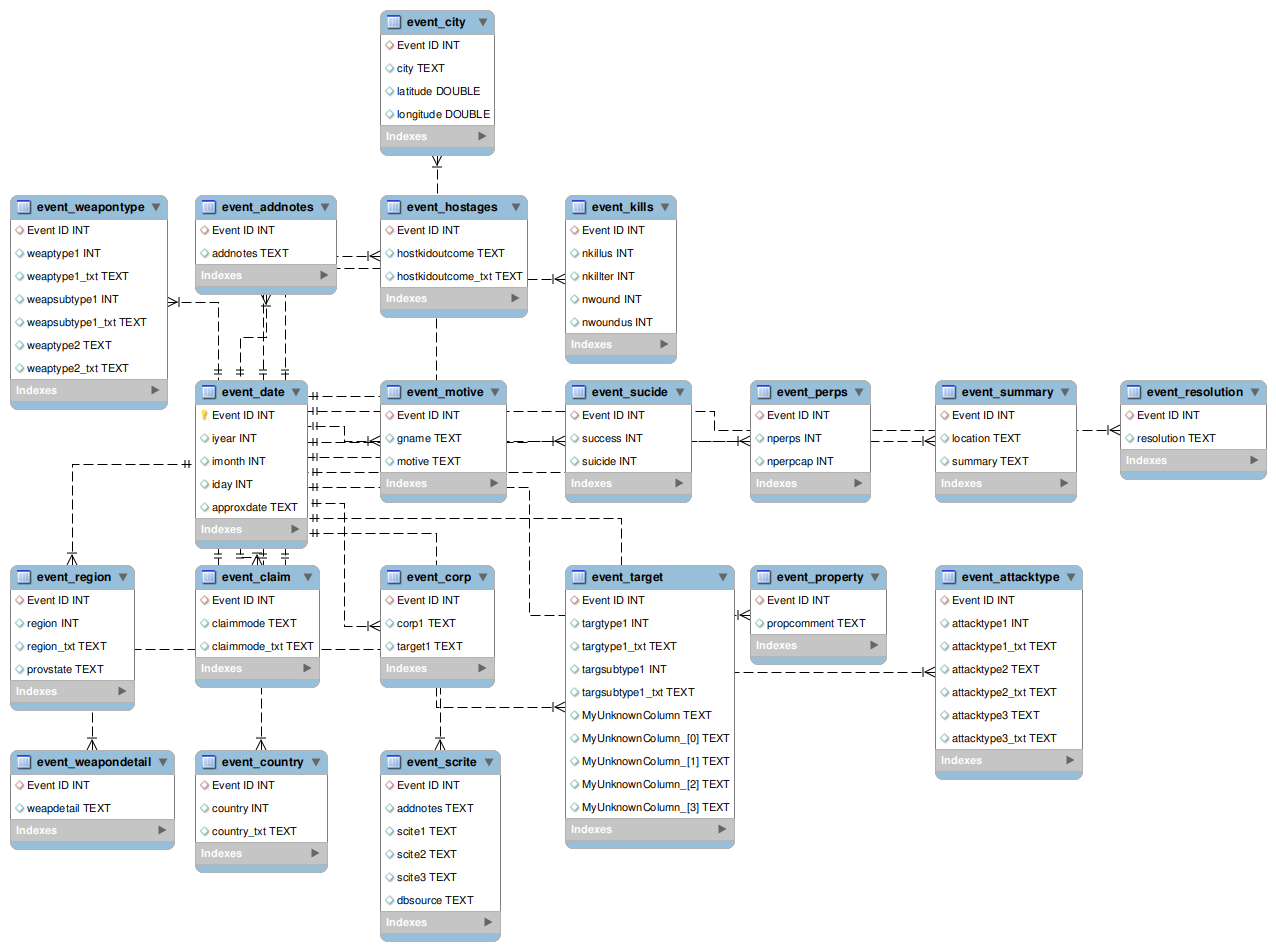
The schema consists of 20 different tables derived from the GTD by START. The data was then cleaned and shortened to a smaller time period, since the START database is traced back to 1970. The data used was from 2016 through 2018.

### **2.5.1 Description**

The 20 tables in the schema are all named to include the word “event”. This is due to the fact that the correlation between all of the tables is the Event ID, the Event ID is an integer assigned to each instance of terrorism beginning in the year 2016 though 2018. Therefore, in order to emphasize the relationship the tables are all named “event\_\*blank\*”.

### **2.5.2 Physical Design**

* 1. Below is the entity relationship diagram, which shows the physical design of the database.



*Fig. 1 Global Terrorism Database Entity Relationship Diagram*

### **2.5.3 Physical Structure**

The database structure revolves around a specific data entry. This data entry is that of Event ID, the Event ID is the identification number for every instance of terrorism commited. By centering the structure of the database to this information the design results in the end user, or any role in particular, being able to identify other corresponding information to the instance.

## **2.6 Data Dictionary**

The following table is the Data Dictionary for this database. Please refer to it when necessary.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Element or Value Display Name** | **Description** | **Data Type** | **Character Maximum Length** | **Acceptable Values (Example)** | **Required (Y/N)** | **Accepts Null Value (Y/N)** |
| addnotes | additional notes/comments about the event | TEXT | 65,535 | Casualty numbers for this incident were below 50. | N | Y |
| approxdate | approximate date of event | TEXT | 65,535 | 1-Jan-16 | N | Y |
| attacktype1 | category number of the attack | INT | 4294967295 | 3 | Y | N |
| attacktype1\_txt | attack description | TEXT | 65,535 | Bombing | Y | N |
| attacktype2 | category number of the attack | INT | 4294967295 | 2 | N | Y |
| attacktype2\_txt | attack description | TEXT | 65,535 | Armed Assault | N | Y |
| attacktype3 | category number of the attack | INT |  | 1 | N | Y |
| attacktype3\_txt | attack description | TEXT | 65,535 | Assasination | N | Y |
| city | City where attack took place | TEXT | 65,535 | Kabul | Y | N |
| claimmode | a number associated with the claim | INT | 4294967295 | 8 | N | Y |
| claimmode\_txt | claim in text | TEXT | 65,535 | Personal Claim | N | Y |
| corp1 | corporation involved | TEXT | 65,535 | Le Jardin | Y | N |
| country | number associated with the country of the event | INT | 4294967295 | 95 | Y | N |
| country\_txt | name of the country | TEXT | 65,535 | Iraq | Y | N |
| dbsource | Database Source, will always be START | TEXT | 5 | START | Y | N |
| eventID | unique number to identify the event | INT | 4294967295 | 345 | Y | N |
| gname | name of the group behind attack | TEXT | 65,535 | Taliban | N | Y |
| hostkidoutcome | number of hostages | INT | 4294967295 | 2 | N | Y |
| hostkidoutcome\_txt | outcome of the hostages | TEXT | 65,535 | hostages released by perpetrators | N | Y |
| iday | day of event | INT | 31 | 1 | Y | N |
| imonth | month of event | INT | 31 | 1 | Y | N |
| iyear | year of event | INT | 4 | 2016 | Y | N |
| latitude | latitude coordinations | DOUBLE | 8 bytes | 33.303567 | N | Y |
| longitude | longitude coordinations | DOUBLE | 8 bytes | 44.371771 | N | Y |
| motive | motive behind attack | TEXT | 65,535 | revenge | N | Y |
| nkillter | number of kills | INT | 4294967295 | 0 | N | N |
| nkillus | number of kills | INT | 4294967295 | 0 | N | N |
| nperpcap | perpetrators captured | INT | 4294967295 | 1 | N | N |
| nperps | perpetrators involved | INT | 4294967295 | 5 | N | N |
| nwound | number of wounded amongst attackers | INT | 4294967295 | 0 | N | N |
| nwoundus | number of wounded | INT | 4294967295 | 0 | N | N |
| propcomment | description of property involved with attack | TEXT | 65,535 | A vehicle was stolen | N | Y |
| provstate | province or state | TEXT | 65,535 | Baghdad | N | Y |
| region | region number | INT | 4294967295 | 6 | Y | N |
| region\_txt | region name | TEXT | 65,535 | South Asia | Y | N |
| scite1 | news title | TEXT | 65,535 | 1 dead, 12 injured in latest attack by... | N | Y |
| scite2 | news title | TEXT | 65,535 | 1 dead, 12 injured in latest attack by... | N | Y |
| scite3 | news title | TEXT | 65,535 | 1 dead, 12 injured in latest attack by... | N | Y |
| suicide | people who commited suicide | INT | 4294967295 | 1 | N | Y |
| target1 | target of the attack | TEXT | 65,535 | Civilians | Y | N |
| targsubtype1 | target subtype number | INT | 4294967295 | 68 | N | Y |
| targsubtype1\_txt | target subtype | TEXT | 65,535 | Unnamed Civilian | N | Y |
| targtype1 | target type number | INT | 4294967295 | 3 | N | Y |
| targtype1\_txt | target type | TEXT | 65,535 | Police | N | Y |
| weapdetail | detail of weapons | TEXT | 65,535 | Rockets were used. | N | Y |
| weapsubtype1 | subtype weapon number | INT | 4294967295 | 6 | N | Y |
| weapsubtype1\_txt | subtype weapon | TEXT | 65,535 | Explosives | N | Y |
| weaptype1 | weapon type number | INT | 4294967295 | 16 | N | Y |
| weaptype1\_txt | weapon text | TEXT | 65,535 | Unknown Explosives | N | Y |
| weaptype2 | weapon type 2 | INT | 4294967295 | 6 | N | Y |
| weaptype2\_txt | weapon text | TEXT | 65,535 | Firearms | N | Y |

## **2.7 Special Instructions**

Since this data is provided by a third party and the records are from previous years, it is not expected that new data will be imported or configured into the database. However, if needed previous data can be altered or updated if it is deemed necessary. Refer to the Data Dictionary in section 2.6 for information on the data type, examples, and other information.

Otherwise, the data within this database is not expected to change. The purpose of this database is to be utilized rather than modified. To utilize this database MySQL Workbench Community software is preferred. Please refer to MySQL Workbench Reference Manual for information on the software and more. Refer to section 1.6.1 to contact team members for inquiries.

**3.0 DATABASE ADMINISTRATIVE INFORMATION**

# **DATABASE ADMINISTRATIVE INFORMATION**

## **3.1 Responsibility**

START is responsible for fulfilling an administrator position, the team was able to thoroughly test databases for troubleshooting, patient, dedicated to keeping databases up to date, able to design databases. Other duties can be having a backup route in case the original database does not execute correctly, implementing security measures, and more. Aside from that, our group was able to allow further testing of these already set responsibilities. START may be contacted via phone at301.405.6600 or via email at infostart@start.umd.edu

## **3.2 System Information**

For SQL servers TSQL scripts can be used to control the embedded database. SQL servers are almost always needed to be in control of source also known as a VCS. VCS is a Version control system repository. Databases would consist of physical layers or logical objects such as columns, rows, legends and other identifying information. Processes like ETL will help with the test building. Defining and source control can rely on Schema Baseline which is a simple process for a simple database. This would further allow for the security and system control.

### **3.2.1 Database Management System (DBMS) Configuration**

MySQL is considered a database management system for software databases. In the creation of this project's database, mysql was used to portray the information in an organized manner. Other examples of databases are Apache CouchDB, PostgreSQL and more. Mysql was downloaded using the 2020 version .

**3.2.2 Hardware Configuration**

Heavily depends on processors, ram and HDD equipped devices. The requirements would be, a dual core processor, 4gb ram minimum and 40 gb of hard drive disk memory as database creation, editing and downloading requires a significant amount of energy from a device. Something like a phone would not be ideal at all. Database servers rely on multiple applications that run in the background and on the program as data collection focuses on large amounts of data. These are ideal requirements for any web or database servers. When working with programs, you must not limit your capabilities on your machinery as that can interrupt the process of data collection, organization of data, and even the publication of data.

### **3.2.3 Database Software Utilities**

DBMS software can include System profilers, Network Managers, Application Launchers, disk repair, file manager, file compression and data security. These all apply to the database as data was transferred into the database and managed to properly handle the opening of all test tables. These software utilities help to open the program, mySQL, and allows for the programming of data. Although the efforts seem to be minimal, there are surfaces of development that are ignored. A consumer normally does not think about the efforts made and establish the level of necessity for all operations to run under simple utilities mentioned above. These all work together to keep the created database up to date and the software allows real time translation and projection.

### **3.2.4 Support Software Available for Maintaining Database**

Support software such as Sql, mySQL, oracle, informix, seqel, cloudera and many more are considered normal support software but for this project specifically, mySQL, oracle and SQL were considered essential as they were directly related to the database.

SQL, Oracle, and mySQL are the latest version for 2020 and were used to operate the transition of data and implement it onto the database.

## **3.3 Storage Requirements**

As long as the device being used has approximately the same requirements for downloading mySQL such as 8gb of RAM, a supporting system in this case we used a virtual machine to download Ubuntu. Running Ubuntu will allow for future downloads of databases with mySQL. Standard devices/ laptops will be compatible for the job as long as they have 128gb in storage and have available space. May need to download Oracle, Ubuntu, and then install mySQL workbench. This will open a database with easy access to query tables and more.

## **3.4 Recovery**

In order to recreate the database, you would need to look for data sets online that are relevant to the topic. You may choose to keep the same data as the one in the database or switch it out for any similar information. Once you find the data sets, you will download any csv files you wish to include and then proceed to clean out the sets to make sure only relevant information is included. From there , the new edited data sets will be imported into the workbench using the new schema feature.

## **3.6 Database Interfaces**

Provide a description of the interfaces with other application software including these of other operational capabilities and from other organizations. For each interface, specify the following information:

### **3.6.1 Description of Operational Implications**

Data transfer must always be handled with care as your system and device run a great risk when it comes to outside downloads. It is essential to keep the data integrity of your own devices as well as keeping your private information protected. Your data can be tampered with when not taking the right precautions. In order to prevent loss of data or tampering with data, the use of two data backups is suggested to guarantee the restoration of data. With computers, storage can be overwhelming so it is recommended to use an external hardrive to make all the data from your current state of device and you can also proceed to use cloud storage option that is most beneficial to you. Apple offers a time machine which lets you store everything from the device onto a device and transfer it to another device.

### **3.6.2 Description of Data Transfer Requirement**

* Must download csv files with file containing desired data
* You may open the file using spreadsheet in this case libreoffice calc
* Clean data for relevant information
* Open workbench
* Navigate to new schema creation
* Import libreoffice calc files onto workbench which will appear as tables
* Create eer diagrams
* Save csv files and database

Issues you may experience are the data downloaded was not downloaded properly or formatting changed, this can be fixed by editing the files on libreoffice calc when running ubuntu. Other problems can be that the workbench may lose connection to the local host and prevent the creation of the schema. For that you must establish a connection to the workbench.

### **3.7 Error Handling**

Exception handling is the process where you can switch between one device to another using servers. Error handling refers to the execution of code and data when being transferred. In particular, it would be the interruptions of information transfer. When referring to error handling, it would imply the correction and smoothing data so that it can potentially be transferred to another company.