

**DATABASE**

**SPECIFICATIONS**

*Group Presentation and Project,*

*- Greg Eure, Muhammed Zahid, Calvin Ton, Junyan Liu, Estefania Lopez*

**ITMD 321: Data Modeling and Applications**

**Dr. Maurice Dawson**

November, 2020

**Revision Sheet**

|  |  |  |
| --- | --- | --- |
| **Release No.** | **Date** | **Revision Description** |
| Rev. 1 | 11/27/20 | Initial publication of specifications |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

|  |  |
| --- | --- |
| Illinois Institute of Technology Logo | **Database Specifications Authorization Memorandum** |

I have carefully assessed the Database Specifications for the Global Terrorism Database (GTD). This document has been completed in accordance with the requirements of the HUD System Development Methodology.

MANAGEMENT CERTIFICATION - Please check the appropriate statement.

\_\_\_X\_\_ The document is accepted.

\_\_\_\_\_\_ The document is not accepted.

We fully accept the changes as needed improvements and authorize initiation of work to proceed. Based on our authority and judgment, the continued operation of this system is authorized.

\_Greg Eure\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_11/27/2020\_\_\_\_\_\_\_\_\_\_\_

NAME DATE

Project Leader

\_Muhammed Zahid\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_11/27/2020\_\_\_\_\_\_\_\_\_\_\_

NAME DATE

Operations Division Director

\_Calvin Ton\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_11/27/2020\_\_\_\_\_\_\_\_\_\_\_

NAME DATE

Program Area/Sponsor Representative

\_Junyan Liu\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_11/27/2020\_\_\_\_\_\_\_\_\_\_\_

NAME DATE

Program Area/Sponsor Director

\_Estefania Lopez\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_11/27/2020\_\_\_\_\_\_\_\_\_\_\_

NAME DATE

Program Area/Sponsor Controller

**DATABASE SPECIFICATIONS**

**TABLE OF CONTENTS**

Page #

1.0 GENERAL INFORMATION

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 Database Specifications is to provide database architecture, design, configuration, controls, and administrative information for primary stakeholders, IT development, and database support personnel.

## 1.2 Scope

The scope of the Database Specifications as it relates to the ITMD 321 Group Presentation and Project are:

* Describe database architecture & design
* Describe database administrative information
* Describe database security controls

## 1.3 System Overview

The GTD is a database downloaded from the [National Consortium for the Study of Terrorism and Responses to Terrorism](https://www.start.umd.edu/data-tools/global-terrorism-database-gtd). It serves as a sample database for the ITMD 321 group project to demonstrate effective database documentation and security control practices.

* Responsible organization: ITMD 321 group members Greg Eure, Muhammed Zahid, Calvin Ton, Junyan Liu, Estefania Lopez
* System name: GroupProj
* System code: Github link: <https://github.com/SQLGROUPPROJ/SQLPROJ>
* System Category: general support system - provides general database functionality to demonstrate documentation and security controls
* Operational status: Operational
* System environment: any operating system running MySQL

## 1.4 Project References

The below references were used in preparation for this document:

* Group Presentation and Project Assignment Guidance document
* Database Security Project Guidance sample document
* HUD Database Specifications 15145 document

## 1.5 Acronyms and Abbreviations

* GTD: Global Terrorism Database
* ITMD: Information Technology & Management & Development
* POC: Points of Organizational Contact
* HUD: Housing and Urban Development
* DBMS: Database Management System
* QA: Quality Assurance
* SQL: Structured Query Language

## 1.6 Points of Contact

### 1.6.1 Information

List of POCs for informational and troubleshooting purposes:

* Architecture POC: Greg Eure, email: geure@hawk.iit.edu
* Infrastructure POC: Muhammed Zahid, email: [mzahid3@hawk.iit.edu](mailto:mzahid3@hawk.iit.edu)
* Helpdesk POC: Calvin Ton, email: [cton@hawk.iit.edu](mailto:cton@hawk.iit.edu)
* Development POC: Junyan Liu, email: [jliu192@hawk.iit.edu](mailto:jliu192@hawk.iit.edu)
* Operations POC: Estefania Lopez, email: [elopez17@hawk.iit.edu](mailto:elopez17@hawk.iit.edu)

### 1.6.2 Coordination

List of organizations that require coordination between the project and its specific support function:

* (Hardware Team) Server Procurement (on-premise or cloud provisioning, includes server and data storage): Schedule: week 1
* (Server Team) Operating System Installation and Configuration: Schedule: week 2
* (DBMS Team) DBMS Installation and Configuration: Schedule: week 3
* (DBMS Team) Database GTD Installation and Configuration: Schedule: week 4
* (Development Team) Database testing: Schedule: week 5
* (Security Team) Database security checking and system audit: week 6
* (QA Team) Rigorous system testing and sign-off: week 7

### 1.6.3 Additional Points of Contact

Additional points of contact are included in section 3.1.

### 1.6.4 Data Owners

See section 1.6.1.

**2.0 DATABASE IDENTIFICATION AND DESCRIPTION**

# DATABASE IDENTIFICATION AND DESCRIPTION

## 2.1 Naming Conventions

Discuss the logical and physical naming standards and conventions.

## 2.2 Database Identification

Identify the names or labels by which the database may be uniquely identified. Specify the code name, tag, or label by which each database table or file may be uniquely identified.

Database name:

* GroupProj

Tables:

* Afghanistan
* Albania
* Algeria
* Angola
* Argentina
* Armenia
* Australia
* Austria
* Azerbaijan
* Bahrain
* Bangladesh
* Belgium
* Bolivia
* Brazil
* Bulgaria
* BurkinaFaso
* Burundi
* Cameroon
* Canada
* Chili
* Incidents

## 2.3 Systems Using the Database

No other systems will use the GroupProj database. This is a demonstration database only, as documented in this guide.

## 2.4 Relationship to Other Databases

The GroupProj database will not supersede or interface with other databases.

## 2.5 Schema Information

Describe the overall structure in the schema or other global definition of the database.

The GropProj is a schema which holds the information of terror attacks that happened between 2016 to 2018. There are a total of 21 tables in the schema, 20 of those tables correspond to a different country and contain 33 attributes listing information such as motive, attack type, weapons used and other pieces of useful information regarding terror attacks. The remaining table is a collection of all incidents and contains 2 attributes, incident\_id and test\_id.

### 2.5.1 Description

Describe the schema and each sub-schema of the system including name, file type and name, data description language, access control keys, concurrence locking, data name mapping, overall partition/file limitations and controls, redefinition and access path restrictions and any other limitations or restrictions.

The GroupProj schema was used to design our system. Our system connects all tables using the ‘incidence’ tables on incident\_id. In order to create our system we used DDL commands such as CREATE, DROP and ALTER. Our system has 4 distinct roles: administrator, end user, client and testing.

The schema GroupProj doesn’t have any sub-schema

### 

### 

### 

### 2.5.2 Physical Design

Graphically depict the physical design of the database.

Table

Description automatically generated

### 

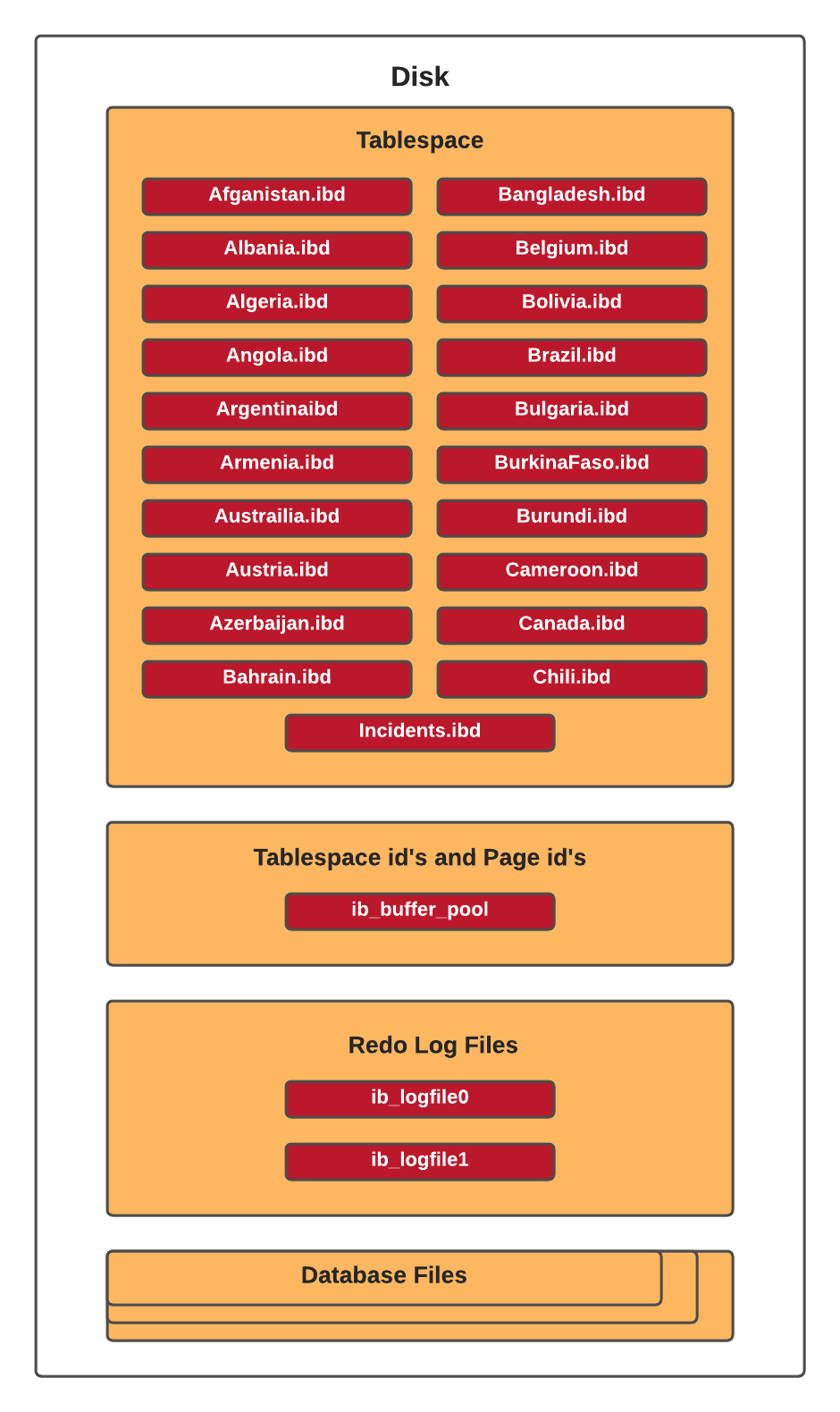
### 

### 

### 2.5.3 Physical Structure

Describe and depict in a graphic representation the physical structure (partitions, files, indexes, pointers) and the logical components of the database. Identify the criteria required to achieve operating efficiency.

The DBMS resides on the filesystem in the /var/lib/mysql directory. The database resides on the filesystem in the /var/lib/mysql/GroupProj directory. There are no partitions or pointers. There are at least 3 indexes per table, consisting of the primary key index, the foreign key index, and an eventid index.



## 2.6 Data Dictionary

Reference the data dictionary and attach it as an appendix to this document.



## 2.7 Special Instructions

Identify instructions to be followed by personnel who will contribute to the generation of the database and who will use it for testing and operational purposes. Such instructions may include:

There are no specific instructions

**3.0 DATABASE ADMINISTRATIVE INFORMATION**

# DATABASE ADMINISTRATIVE INFORMATION

## 3.1 Responsibility

Identify the organizations and personnel responsible for the following database administrative functions: database administrator, system administrator, and security administrator. Describe specific administration skill requirements.

Database administrator organization: DBMS Services

DBMS Services responsible personnel: Muhammed Zahid

Database administrator required skills:

* Knowledge of database queries
* Knowledge of database theory
* Knowledge of database design
* Knowledge about the RDBMS itself, e.g. Microsoft SQL Server or MySQL
* Knowledge of structured query language (SQL), e.g. SQL/PSM or Transact-SQL
* General understanding of distributed computing architectures, e.g. Client–server model
* General understanding of operating system, e.g. Windows or Linux
* General understanding of storage technologies and networking
* General understanding of routine maintenance, recovery, and handling failover of a database

System Administrator organization: Operations Services

Operations Services responsible personnel: Estefania Lopez

System Administrator required skills:

* Problem-Solving and Administration
* Networking
* Cloud
* Automation and Scripting, including HTML, JavaScript, Go, Bash, Python, and Node.js
* Security and Monitoring
* Account Access Management
* IoT/Mobile Device Management
* Hardware Management
* SQL

Security Administrator organization: Security Management

Security Management responsible personnel: Calvin Ton

Security Administrator required skills:

* Defending systems against unauthorized access, modification and/or destruction
* Scanning and assessing network for vulnerabilities
* Monitoring network traffic for unusual activity
* Configuring and supporting security tools such as firewalls, anti-virus software and patch management systems
* Implementing network security policies, application security, access control and corporate data safeguards
* Training fellow employees in security awareness and procedures
* Developing and updating business continuity and disaster recovery protocols

## 3.2 System Information

Document the Database Management System configuration, hardware configuration, database software utilities, and any support software used:

DMBS configuration: MySQL Server installation on test environment with mysql\_secure\_installation script execution

Hardware configuration: Test environment running on virtual machine with 2 cpu’s, 4Gb RAM, and 32Gb disk, with Ubuntu linux installed

Database software utilities: MySQL Workbench installed on test environment

### 3.2.1 Database Management System (DBMS) Configuration

Identify the vendor, version or release date and targeted hardware for the DBMS. Describe any restrictions on the initialization and use of the DBMS to support any intended distributed processing.

The targeted hardware can be any x86-based commodity hardware, including dedicated or virtualized environments. No restrictions are identified.

### 3.2.2 Hardware Configuration

Identify the hardware configurations on which the database will reside.

No strict hardware configurations are required, but the test system is running 2 vcpu’s, 4 Gb memory, and 32 Gb disk.

### 3.2.3 Database Software Utilities

List and reference the documentation of any DBMS utility software available to support the use or maintenance of the database.

DBMS support: https://dev.mysql.com/doc/

MySQL Workbench support: https://dev.mysql.com/doc/workbench/en/wb-intro.html

### 3.2.4 Support Software Available for Maintaining Database

Describe all support software, including the operating system, directly related to the database, including name, version, function, and major operating characteristics. Cite documentation by title, number, and appropriate sections. Examples of such software include database management systems, query language, report writers, storage allocation software, database loading software programs, and file processing programs, and data cleaning software.

Operating System: Ubuntu, version 20.04.01 LTS, function: database server

DBMS System: MySQL Community Server, version 8.0.22, function: DBMS server

Tool: MySQL Workbench, version 8.0.22, function: MySQL support tool

Query Language: structured query language (SQL)

### 3.2.5 Security

Describe the use and management of integrity and access controls that apply to all database components such as schema, sub-schema, partitions or physical files, records or tables, sets or relations, and data elements.

The database shall be secured initially by executing the mysql\_secure\_installation script provided by MySQL. The database instance shall be secured in a manner consistent and compliant with NIST 800-53. All relevant controls shall be documented in separate control documents (defined below), along with respective test case results.

|  |  |  |
| --- | --- | --- |
| **NIST 800-53 Control Description** | | |
| **Control Number** | **Control Description** | **Control Enforcement Procedure** |
| AC-7 | Unsuccessful Login Attempts | NIST AC-7 Control Document |
| AU-2 | Audit Events | NIST AU-2 Control Document |
| CM-10 | Software Usage Restrictions | NIST CM-10 Control Document |
| IA-6 | Authenticator Feedback | NIST IA-6 Control Document |
| SA-2 | Allocation of Resources | NIST SA-2 Control Document |
| SC-5 | Denial of Service Protection | NIST SC-5 Control Document |
| SI-3 | Malicious Code Protection | NIST SI-3 Control Document |
| AC-2 | Account Management | NIST AC-2 Control Document |
| AC-5 | Separation of Duties | NIST AC-5 Control Document |
| AT-3 | Role-Based Security Training | NIST AT-3 Control Document |
| AU-3 | Content of Audit Records | NIST AU-3 Control Document |
| CM-6 | Configuration Settings | NIST CM-6 Control Document |
| AU-8 | Session Audit | NIST AU-8 Control Document |
| AC-8 | System Use Notification | NIST AC-8 Control Document |
| PE-15 | Water Damage Protection | NIST PE-15 Control Document |
| CA-5 | Plan of Action and Milestones | NIST CA-5 Control Document |
| CP-9 | Information System Backup | NIST CP-9 Control Document |
| CP-10 | Information System Recover & Reconstitution | NIST CP-10 Control Document |
| CM-2 | Baseline Configuration | NIST CM-2 Control Document |
| MP-2 | Media Access | NIST MP-2 Control Document |
| PE-8 | Visitor Access Records | NIST PE-8 Control Document |
| AC-14 | Permitted Actions Without Identification or Authentication | NIST AC-14 Control Document |
| AC-17 | Remote Access | NIST AC-17 Control Document |
| AC-22 | Publicly Accessible Content | NIST AC-22 Control Document |
| SA-3 | System Development Life Cycle | NIST SA-3 Control Document |
| IR-4 | Incident Handling | NIST IR-4 Control Document |
| SC-7 | Boundary Protection | NIST SC-7 Control Document |
| AC-3 | Access Enforcement | NIST AC-3 Control Document |
| AC-20 | Use of External Information System | NIST AC-20 Control Document |
| AT-4 | Security Training Record | NIST AT-4 Control Document |
| CM-11 | User-Installed Software | NIST CM-11 Control Document |
| IR-5 | Incident Monitoring | NIST IR-5 Control Document |
| PS-7 | Third-Party Personnel Security | NIST PS-7 Control Document |
| PS-6 | Access Agreements | NIST PS-6 Control Document |
| PL-4 | Rules of Behavior | NIST PL-4 Control Document |
| PE-13 | Fire Protection | NIST PE-13 Control Document |

## 3.3 Storage Requirements

Describe the storage device. Provide sizing formulas for determining the storage required to support the database content and associated software. Estimate the internal and peripheral storage requirements. Identify multiple storage requirements for distributed processing.

The storage device can be any internal or external storage that is recognized and mounted by the operating system. Minimum recommended storage for the operating system is 10 Gb. Minimum required storage for installation of MySQL is 300 Mb (minimum recommended is 1 Gb). Minimum required storage for the GroupProj database is 15Mb. For distributed processing, multiply these requirements by the number of servers to be added to the resource cluster.

## 3.4 Recovery

Describe the methodology for reestablishment or recreation of the necessary data schema and system support files.

Regular system back-ups are recommended, in which case system restore will be an option for complete recovery. Optionally, database exports should be performed as a secondary back-up option. Full database exports should be performed within MySQL Server or MySQL Workbench, and kept on a separate and independent system. In case recreation of the database is necessary, a fresh installation of MySQL Community Server and MySQL Workbench can be performed, followed by importing the latest export of the GroupProj database.

## 3.5 Partition/File Information

### 3.5.1 Content

Describe the content of each partition/file, listing the records it contains and explaining the purpose.

The GroupProj database is not partitioned. The database is contained under the file directory /var/lib/mysql/GroupProj. Files in this directory are associated with the defined tables in the database:

* Afghanistan.ibd
* Albania.ibd
* Algeria.ibd
* Angola.ibd
* Argentina.ibd
* Armenia.ibd
* Australia.ibd
* Austria.ibd
* Azerbaijan.ibd
* Bahrain.ibd
* Bangladesh.ibd
* Belgium.ibd
* Bolivia.ibd
* Brazil.ibd
* Bulgaria.ibd
* BurkinaFaso.ibd
* Burundi.ibd
* Cameroon.ibd
* Canada.ibd
* Chili.ibd
* Incidents.ibd

### 3.5.2 Description

Describe the design and format of each partition/file, including name, type, code, mapping, limitations and controls, access procedures, and mechanisms.

The file names are listed in section 3.5.1. There is no unique type, code, mapping, limitation, or mechanism. All controls and access procedures are handled by the operating system (for file/directory access) or MySQL (for database, table, data access)

### 3.5.3 Partition/File Interdependencies

Identify the interdependencies of each partition/file in the database.

All identified files (as listed in section 3.5.1) are dependent on file incidents.ibd, as all other tables have foreign key relationships with the primary key in this file. See section 2 for detailed database architecture.

## 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:

The GlobalProj database doesn’t have any interfaces with other application software not defined in this document.

### 3.6.1 Description of Operational Implications

Describe operational implications of data transfer, including security considerations.

N/A

### 3.6.2 Description of Data Transfer Requirements

Describe data transfer requirements to and from the software, including data content, format, sequence, and any conversion issues.

N/A

### 3.6.3 Description of Formats of Data

Describe formats of data for both the sending and receiving systems, including the data item names, codes, or abbreviations that are to be interchanged, as well as any units of measure/conversion issues.

N/A

## 3.7 Error Handling

Describe those system error handling routines and procedures that are available during execution of database software.

Error handling is performed inherently by the MySQL DBMS software. No custom or proprietary error handling routines or procedures are implemented.