

**DATABASE**

**SPECIFICATIONS**

*USCrimes - Final Group Project*

*Thomas Fracassi, Maria Soto, Tai Harrell, Erica Hernandez & Joshua Alaniz*

**Illinois Institute of Technology - Database Concepts with SQL ITM821**

**Dr. Maurice Dawson**

November 2020

|  |  |
| --- | --- |
|  | **Database Specifications Authorization Memorandum** |

I have carefully assessed the Database Specifications for the USCRIMES. This document has been completed in accordance with the requirements of the HUD System Development Methodology.

MANAGEMENT CERTIFICATION - Please check the appropriate statement.

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

\_\_\_\_\_\_ The document is accepted pending the changes noted.

\_\_\_\_\_\_ 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.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

NAME DATE

Database Administrator

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

NAME DATE

Tester

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

NAME DATE

Client

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

NAME DATE

User

**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

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**

This project has been developed for creating a central network for gathering data pertaining to crime occurring in the United States. The specificity of the crime being analyzed is hate crime due to its high frequency from the last few decades to today. The current state of the information consists of vast numbers of information, and some not so relevant information. With this design, all the relevant information will be readily available to all users who wish to view and compare the trend of hate crime over the past few decades.

## **1.2 Scope**

The database design for the database USCrimes is composed of database objects derived from crime databases. This includes tables, columns, identifiers, keys, and relationships with foreign keys.

For specific guidelines that the tables selected and controls that were picked can be found as listed in the [class specification.](https://learn-us-east-1-prod-fleet01-xythos.content.blackboardcdn.com/blackboard.learn.xythos.prod/585304babb2ec/7385084?X-Blackboard-Expiration=1606856400000&X-Blackboard-Signature=mHJBHS3ETVOWzOLZ5MpiHGePZzMV3HEfuc6OsnP6seI%3D&X-Blackboard-Client-Id=100773&response-cache-control=private%2C%20max-age%3D21600&response-content-disposition=inline%3B%20filename%2A%3DUTF-8%27%27%255BA%255D%2520Group%2520Presentation%2520and%2520Project.pdf&response-content-type=application%2Fpdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20201201T150000Z&X-Amz-SignedHeaders=host&X-Amz-Expires=21600&X-Amz-Credential=AKIAYDKQORRYTKBSBE4S%2F20201201%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Signature=4cde95c68784354969d46d00a77a4f92cc345fa07a00d659a0741a89eb501604)

## **1.3 System Overview**

The authors for the database are Erica Hernandez, Thomas Fracassi, Maria Soto, and Joshua Alaniz. This information system was deployed on MySql Workbench, hosted on a virtual Linux machine. The nature of the data amassed is relating to hate crime in the United States. The first set originating from the Global Terrorism database and is holding data from 1970 to the present. The second set originates from the Federal Bureau of Investigation, Hate Crime Collecting. The variables analyzed were: the target type, the weapon used., location, time, and more.

## **1.4 Project References**

Provide a list of the references that were used in preparation of this document.

* Global Terrorism Database
  + <https://start.umd.edu/gtd/>
* Federal Bureau of Investigation Hate Crime Statics
  + <https://www.fbi.gov/services/cjis/ucr/publications#Hate-Crime%20Statistics>

**2.0 DATABASE IDENTIFICATION AND DESCRIPTION**

# **DATABASE IDENTIFICATION AND DESCRIPTION**

## **2.1 Naming Conventions**

All datasets in the USCrimes database are identified by beginning with USCrimes then have an appreciated name or shorten down to one to two words. This accommodates users’ access to locate all datasets and locate specific datasets based on their naming convention.

## **2.2 Database Identification**

As stated above all database tables are uniquely identified by having USCrimes at the beginning to confirm it is for the USCrimes database. Since this is being utilized for academic purposes, datasets are SQL files for an effortless integration with MySQL Workbench on a Linux system.

## **2.3 Systems Using the Database**

The database is being utilized for academic purposes hence the host system for the database will be an Ubuntu (64-bit) Virtual Machine.

## **2.4 Relationship to Other Databases**

The datasets of the USCrime database originate from the Global Terrorism Database and Federal Bureau of Investigation Hate Crime Statics. The datasets that were utilized were in reference to hate crimes in the United States as a starting off point for the USCrimes database.

## **2.5 Schema Information**

The structure of the USCrime schema is 20 tables related to hate crimes in the United States and descriptions of hate crimes. Users and clients are allowed to view and access the database however they have no authorization to change any datasets. Administration and testing have full access to the database allowing them to edit the database as they see fit.

### **2.5.1 Description**

The 20 tables are sql files and the following is all the table names:

USCrimes\_attack\_type; USCrimes\_disability; USCrimes\_fbi\_hate\_crime; USCrimes\_gender; USCrimes\_gender\_identity; USCrimes\_group\_name; USCrimes\_gtd; USCrimes\_kill\_wounded; USCrimes\_lattitude\_longitude; USCrimes\_location; USCrimes\_motive; USCrimes\_multi\_bias; USCrimes\_property; USCrimes\_property\_damage; USCrimes\_race\_ethnicity\_ancestry; USCrimes\_religion; USCrimes\_sexual\_orientation; USCrimes\_single\_bias; USCrimes\_target\_type; USCrimes\_weapon\_type.

All tables are viewable to users, clients, administration, and testing. However, administration and testing have full accessibility which allows them to edit the database and make any changes they see fit.

### **2.5.2 Physical Design**

The database is being used for academic purposes, hence a simple physical design to the database. The database is made up of 20 tables, but 18 tables are connected to either the USCrimes\_fbi\_hate\_crime dataset or USCrimes\_gtd dataset.

### **2.5.3 Physical Structure**

USCrimes database was created for academic purposes so the physical structure does not require any criteria to achieve operating efficiency.

## **2.6 Data Dictionary**

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

## **2.7 Special Instructions**

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.

* DBA is the only authorized user to input data into the database
* To submit data for entry please contact DBA via email with name, date, contact info, data sample, and reason for entry
* The USCrimes database was created using MySQL workbench. For instructions to navigate this program visit:<https://dev.mysql.com/doc/workbench/en/>
* Any changes made to this database are to be included in the Data Specifications document with a description of the changes made, the date the changes were made, and information on the individual who made these changes.

**3.0 DATABASE ADMINISTRATIVE INFORMATION**

# **DATABASE ADMINISTRATIVE INFORMATION**

## **3.1 Responsibility**

There are four roles for this database: Database Administrator, User, Client and Testing.

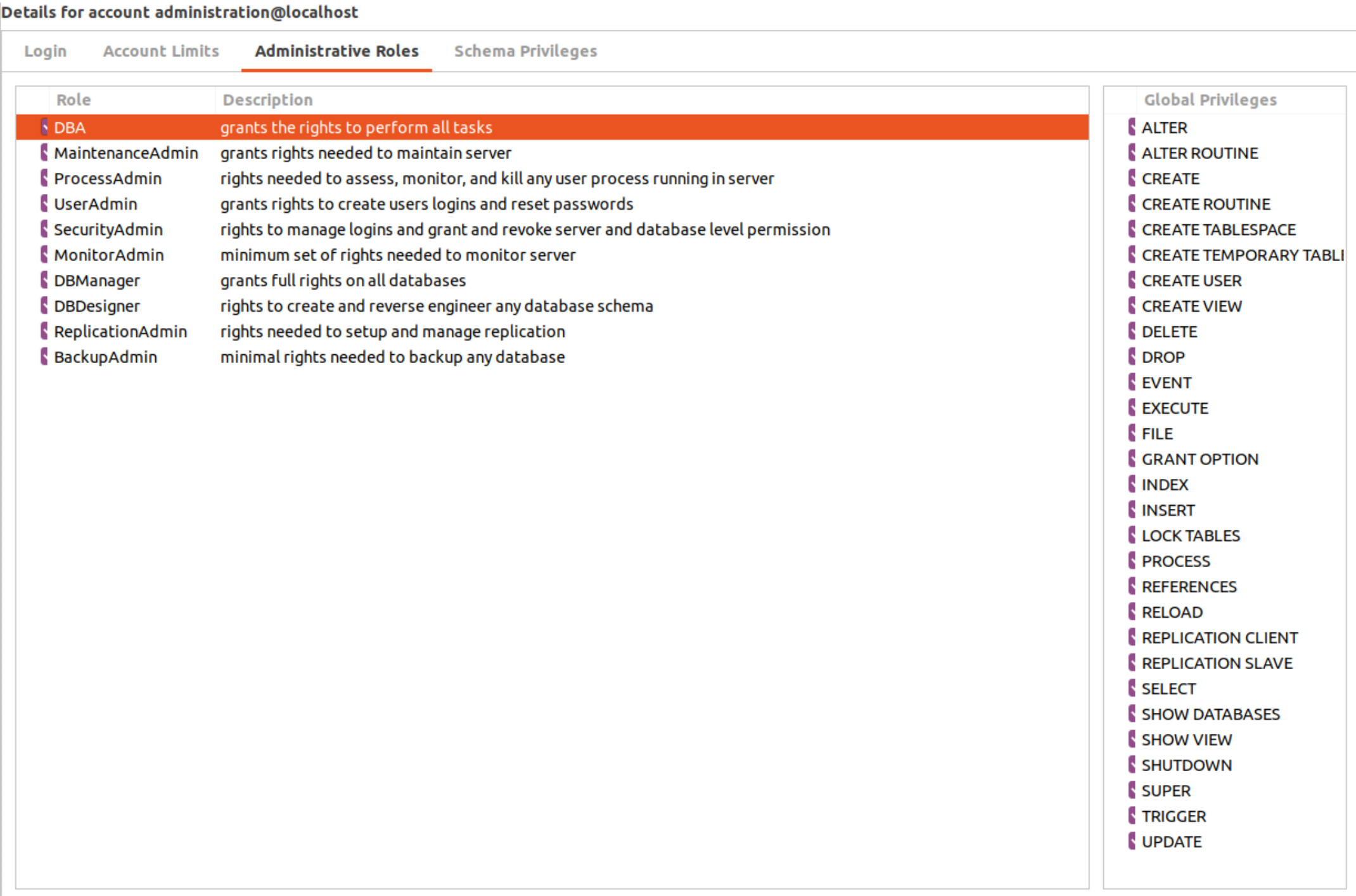
**AC-2 Account Management**

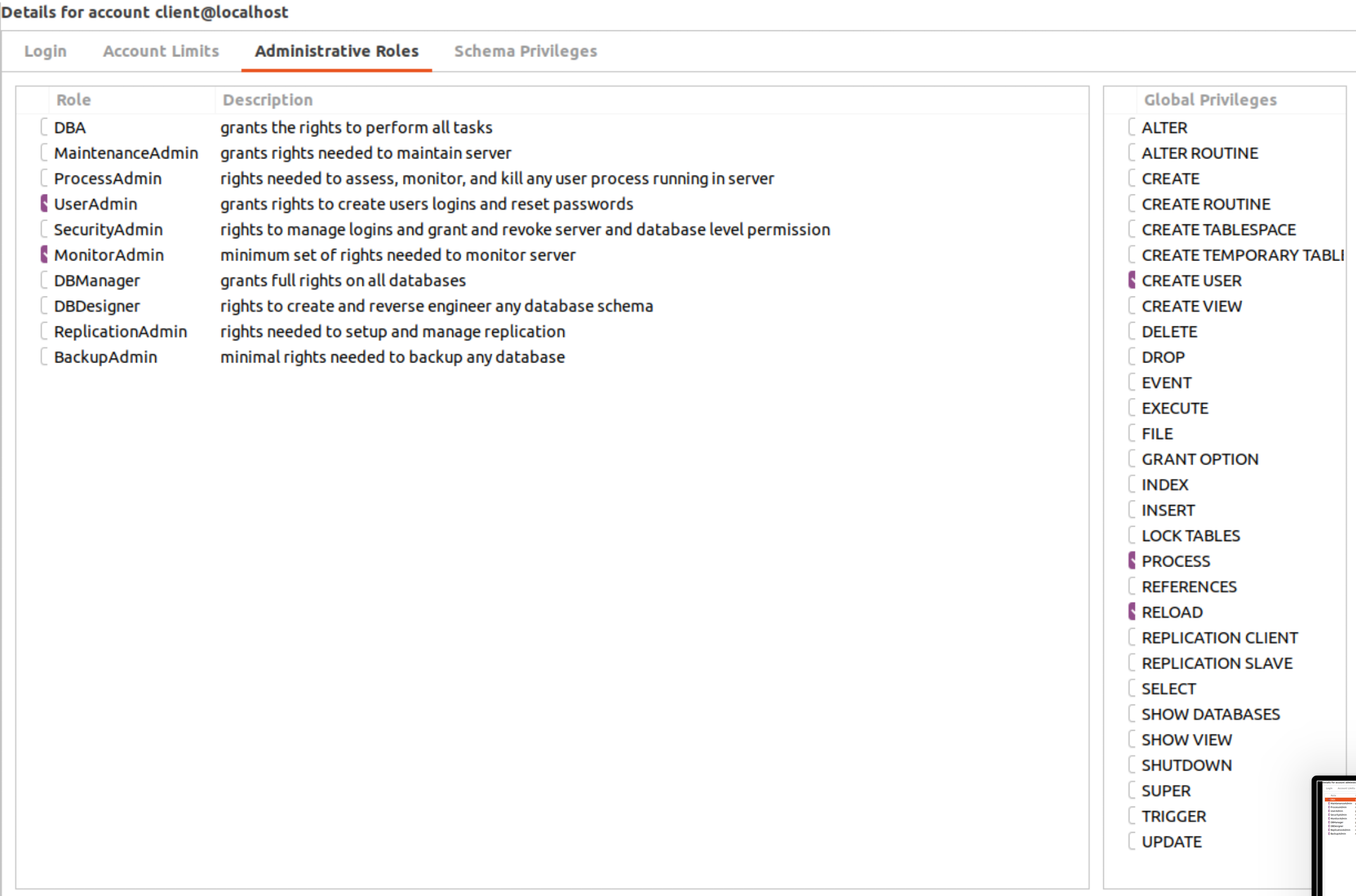
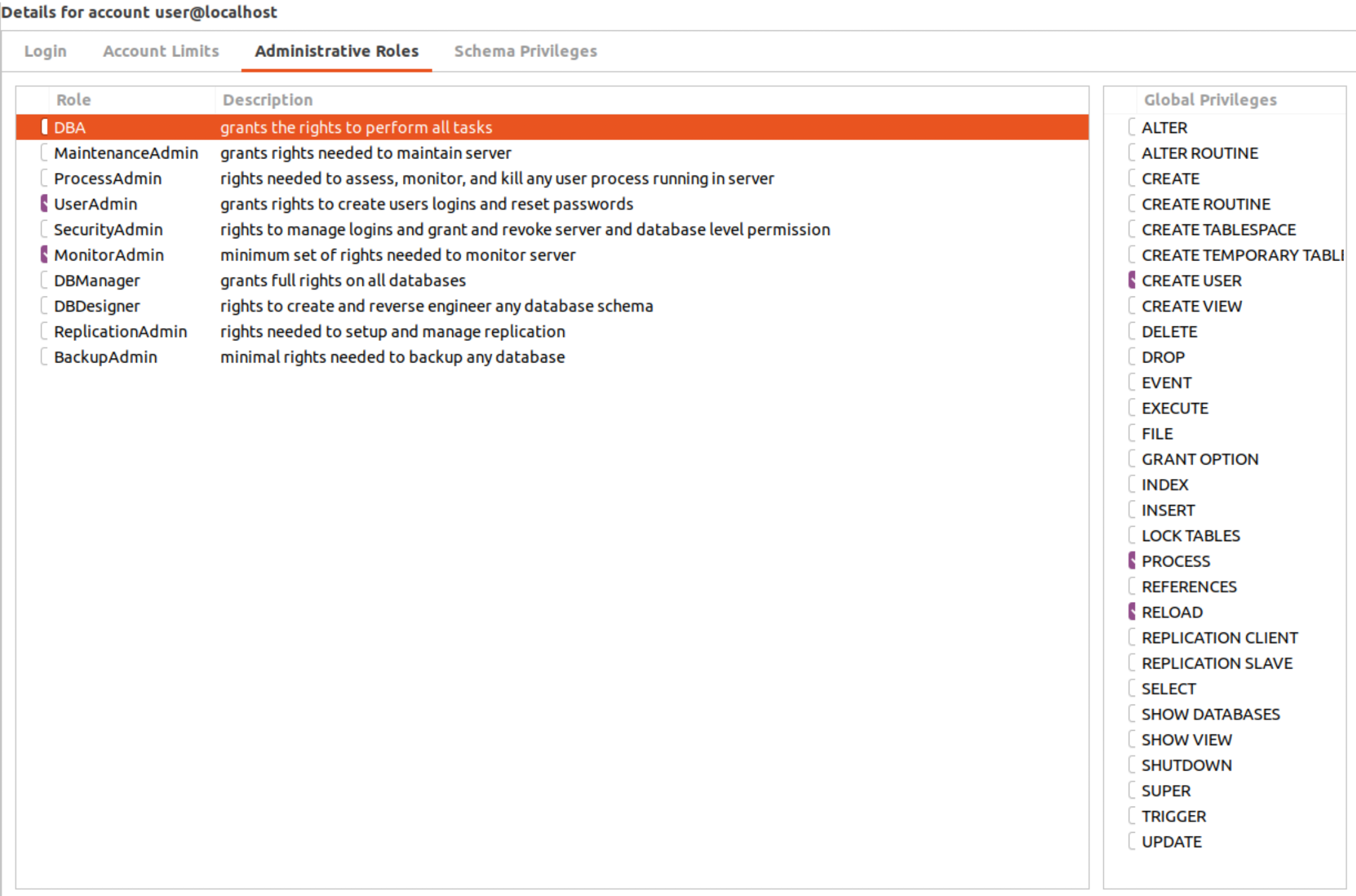
**For all information systems:**

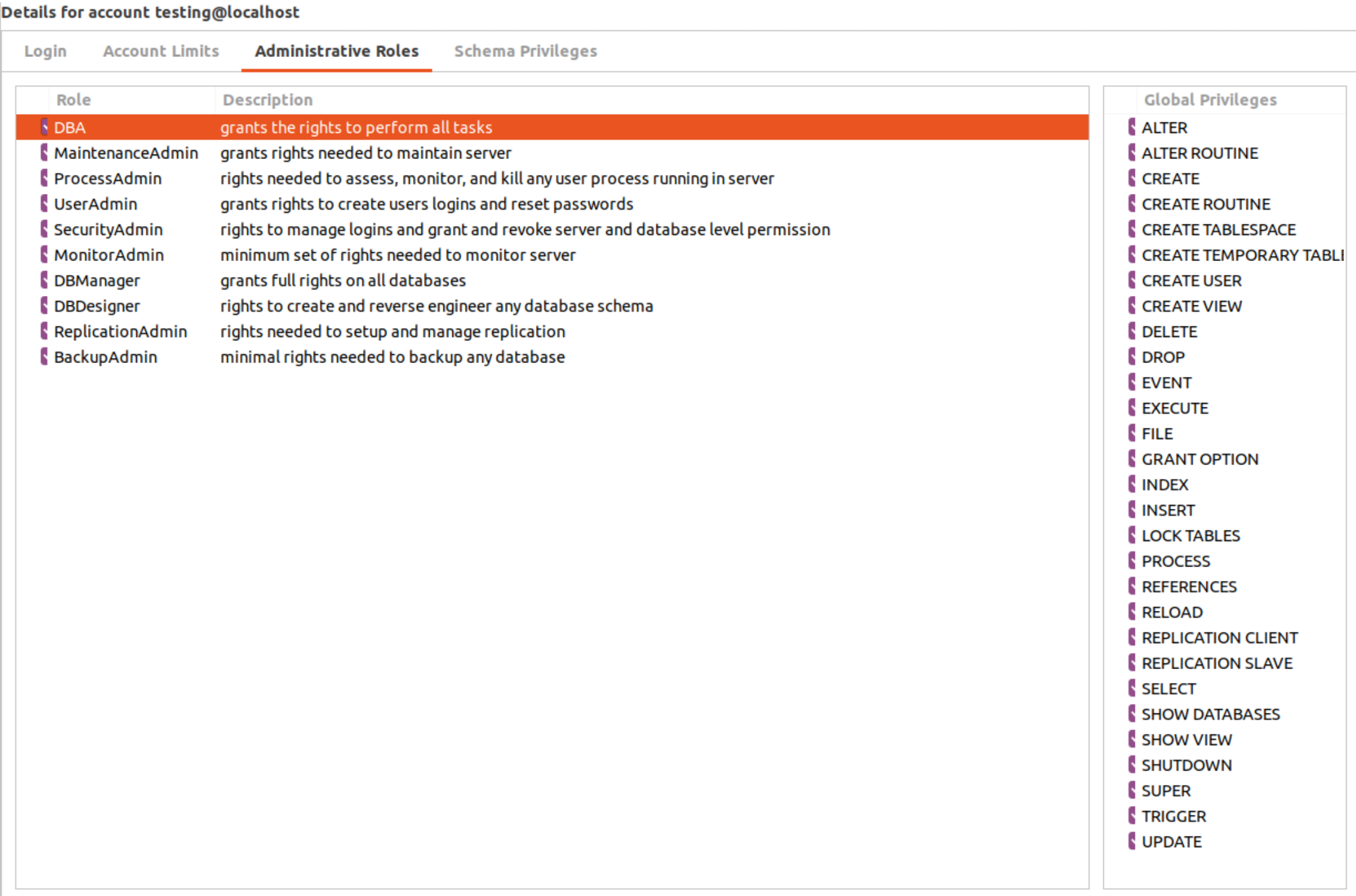
The database administrator (DBA), has created this database (DB) for academic purposes only.

The information found is this database is gathered from the Global Terrorism Database (GTD) made by The National Consortium for the Study of Terrorism and Responses to Terrorism (START) and the Federal Bureau of Investigation (FBI) Annual Hate Crime Statistics publication. There are four roles in relation to this database: DBA, End User, Client, and Testing. The following presents the privileges assigned to each role by the DBA.

Database Administrator System Privileges:



Client PrivilegesEnd User System Privileges: 

System Testing Privileges: 

## **3.2 System Information**

This section will document the Database Management System, hardware, database software utilities, software, support and security for USCrimes database.

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

The Database Management System (DBMS) that was utilized was MySQL Workbench Community version 1.15.10. The GUI interface made it easier to work with datasets and still bring restrictions to unauthorized users.

### 3.2.2 Hardware Configuration

MySQL Workbench Community (GPL) for Linux/Unix version 6.3.8 CE build 1228 (64 bit)

Configuration Directory: /home/student/.mysql/workbench

Data Directory: /usr/share/mysql-workbench

Cairo Version: 1.15.10

OS: Linux 5.4.0-52-generic

CPU: Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz (1991.996MHz) - 981.31MiB RAM

Distribution: elementary OS 5.1.7 Hera

### **3.2.3 Database Software Utilities**

The USCrimes database was developed using MySQL RDBMS which supports the use and maintenance of the database. Not much else is needed for the database for it to be used for academic purposes.

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

MySQL Workbench software has the capability to manage systems, utilize query language, report writes, file programs and clean data. The software functions are enough for there not to be a need for any other software for academic purposes.

### **3.2.5 Security**

The USCrimes Project was developed using MySQL RDBMS to create the USCrimes database and was designed to take advantage of the built-in security features within the application software.

In addition, a number of controls were identified to be met from the NIST Special Publication 800-53 (Rev. 4). The following table contains a list of 35 NIST controls which were a part of the USCrimes

|  |  |  |
| --- | --- | --- |
| **NIST 800-53 (Rev. 4) Control Family** | **Completed** | **Partially Completed** |
| Access Control (AC) | 5 | 0 |
| Awareness & Training (AT) | 1 | 0 |
| Audit & Accountability (AU) | 1 | 0 |
| Certification, Accreditation, & Security Assessments (CA) | 7 | 0 |
| Configuration Management (CM) | 1 | 0 |
| Contingency Planning (CP) | 2 | 0 |
| Identification and Authentication (IA) | 1 | 0 |
| Incident Response (IR) | 2 | 0 |
| Maintenance (MA) | 1 | 0 |
| Media Protection (MP) | 3 | 0 |
| Physical and Environmental Protection (PE) | 2 | 0 |
| Planning (PL) | 0 | 0 |
| Personnel Security (PS) | 3 | 0 |
| Risk Assessment (RA) | 4 | 0 |
| System and Services Acquisition (SA) | 0 | 0 |
| System and Communications Protection (SC) | 1 | 0 |
| System and Information Integrity (SI) | 1 | 0 |
| **Control Totals** | 35 | 0 |

## **3.3 Storage Requirements**

The USCrimes database was developed using the MySQL RDBMS with the default schema. All tables are created as InnoDB tables, with a row\_format of Dynamic. In order to monitor storage requirements for the database, a query of the INFORMATION\_SCHEMA grouped by database, to calculate and retrieve the total size (data and index) for each database can be used. Refer to the MySQL Reference Manual for further information about database storage requirements.

## **3.4 Recovery**

Backup and Recovery for the USCrimes database will be supported using the methodology found in the MySQL Reference Manual – Example Backup and Recovery Strategy. This strategy includes Establishing a Backup Policy, Using Backups for Recovery and Backup Strategy Summary.

## **3.5 Partition/File Information**

### **3.5.1 Content**

The USCrimes database was created with the MySQL RDBMS using the default schema. The current database content is made up of 20 tables, in addition to the standard schema data used to support the database.

### **3.5.2 Description**

Based on the size of the USCrimes database, and the fact that it has been created with the default schema, there is no partitioning present in the initial USCrimes database. Generally, unless dealing with very large datasets or millions of rows, partitioning may not provide much benefit.

### **3.5.3 Partition/File Interdependencies**

None.

## **3.6 Database Interfaces**

MYSQl Workbench was the primary GUI used in the creation of this database. Other software applications were also used hand in hand to bring all of the data together. Microsoft excel and libri office were used to convert datasets into .csv files. Google drive was used to share the data sets with the members of this project. Additionally google drive was used to bring over files to the virtual machine which MYSQL workbench was operating off of.

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

Some security considerations taken for this database were renaming database, minimizing the value, and monitoring the database activity

### **3.6.2 Description of Data Transfer Requirements**

For data transfer to the software the Comma separated value (.CSV) format was used. Our members first gathered data from databases then opened them in a spreadsheet software. The data was stripped from the spreadsheet and converted into a .CSV file. One member from our party would then enter those .CSV files into the database using an import wizard. It was imperative that the data was properly labeled to make sure it was easy to identify the data and enter it into the database. No conversion issues occurred.

### **3.6.3 Description of Formats of Data**

All of the data within this database was converted to the comma separated value format (.CSV). This format works seamlessly with MYSQL Workbench so issues and errors were easily avoided.

## **3.7 Error Handling**

When encountering an error the first step was to locate the error code. The error code is essential because it can show exactly what you're doing wrong. Upon locating the error code the group would then look on the MYSQL workbench forms page and type in the error code that we received. After completing that a list of many other individuals with the same problem would appear with solutions to the problem.