COVID-19 Analytics System (CAMS)

The COVID-19 pandemic has been one of the most significant global health challenges in recent history. It has affected millions of people around the world, causing severe morbidity and mortality, as well as social and economic disruptions. To understand the dynamics and impacts of the pandemic, it is essential to have reliable and timely data on the cases, deaths, and vaccinations across different countries and regions. However, collecting, storing, and analyzing such data can be challenging, especially when dealing with large and diverse datasets from multiple sources.

1. Project Scope (Introduction)

The **COVID-19 Analytics Management System (CAMS)** is a SQL project, serving as a Database Management System (DBMS) that aims to create a database that can store, manage and analyze data related to the COVID-19 pandemic. The database will allow users to query and retrieve relevant information about the pandemic, such as the number of confirmed cases, deaths, and vaccinations, the testing and hospitalization rates. The project will focus on leveraging SQL to create a well-made and scalable database system that facilitates in-depth analytics and visualization of COVID-19 trends at various levels.

2. Project Objectives

The main objectives of the **CAMS** project that would be the Expected Outcomes of this project, are:

a. Data Integration:

- Collect and integrate COVID-19 data from reliable sources, including daily case reports, vaccination statistics, testing data, and demographic information.

b. Database Design:

- Design an efficient and normalized database schema, with tables for countries, states, daily case reports, vaccination data, testing data, and demographics.
- To ensure the quality and integrity of the data, by applying appropriate data validation and normalization techniques.
- To provide an interface that can allow users to query and retrieve the data, using SQL commands.
- To demonstrate the functionality and usability of the database, by performing various data analyses and visualizations, using SQL queries.

c. Geographic Visualization:

- Implement SQL queries and views to visualize the geographic spread of COVID-19 cases, allowing users to explore trends globally, nationally, and regionally.

d. Trend Analysis:

- Develop SQL queries to analyze trends in COVID-19 cases, deaths, recoveries, and vaccination rates over time, enabling comparisons between different countries or regions.

e. User-Friendly Queries:

- Develop a set of user-friendly SQL queries that enable users to retrieve specific information, such as daily case counts, vaccination rates, or trends for a particular region.

3. Technical Scope

- **Database Engine:** Microsoft SQL Server.
- **Development Tools:** SQL Server Management Studio (SSMS)
- **Data Sources:** Publicly available COVID-19 datasets (e.g., Johns Hopkins CSSE COVID-19 Data, WHO COVID-19 Dashboard)
- **Programming Language:** SQL
  Following SQL concepts will be brought to inclusion to this project:
  - DML commands (Insert, Update and Delete).

- DQL commands (Select).
- DDL commands (Create, Drop, Alter, Rename and truncate).
- Constraints (Primary Key, Foreign Key, Not Null, Unique).
- Aggregate functions.
- Other concepts (Where clause, Having clause, Order, Group by, Aliases, In and Between operator).
- Join and Sub queries.
- Procedures and more.

4. Methodology: Project Timeline

The Methodology of the CAMS project consists of following steps:

a. Step 1: Data Collection:

The data will be collected from various sources, such as the WHO, JHU, and OWID, using their official websites. The data will be downloaded and stored in a local directory (local PC).

b. Step 2: Data Modelling and Normalization:

The data will be modeled and structured, by defining the entities, attributes, and relationships that represent the data. The data will also be normalized, by applying the rules of normalization, such as eliminating redundancies, dependencies, and anomalies. The data model will be represented by an entity-relationship diagram (ERD), using a standard notation, such as Crow's Foot.

The inclusion of the following tables will significantly contribute to the design of our project:

- Countries
- States/Regions
- COVID-19 Reports (Cases, Deaths, Recoveries)
- Vaccination Data
- Testing Data
- Demographics
- Public Health Measures

c. Step 3: Database creation:

The database will be created and implemented, by using SQL as the programming language. The database will consist of tables, columns, keys, indexes, and constraints, that correspond to the data model. The database will also include views, functions, and procedures, that facilitate the data access and manipulation. The database will be hosted on a local server, using a SQL database management system (DBMS), such as MS SQL Server.

d. Step 4: Data querying:

The data will be queried and retrieved, by using SQL commands or tools, such as select, join, group by, or order by. The data will be filtered, aggregated, sorted, or calculated, according to the user's needs and preferences.

5. Conclusion

In conclusion, the COVID-19 Analytics Management System project serves as a pivotal learning opportunity, allowing me to deepen my SQL and database management skills. Crafting a robust database schema, implementing advanced queries, and addressing real-world challenges underscore my commitment to continuous learning in the dynamic field of data management. This project not only enhances my technical capabilities but also fosters a mindset of responsible and ethical data handling.