CS 5200

PROJECT PROPOSAL

Prepared by :
Andrew Panzone,
Ishan Aggarwal,
Maalolan Bharaniraj

PROJECT DESCRIPTION

The "**CancerStatisticExplorer**" initiative aims to gather, archive, and make study of cancer incidence rates in all U.S. states easier. Researchers, healthcare providers, and public health representatives will be able to obtain current cancer statistics using the program's tools, examine historical patterns, and compare data from various states.

TOP INTEREST

The vital importance of cancer monitoring to public health is the driving force behind **CancerStatisticExplorer**. Determining at-risk groups, making informed policy decisions, and effectively allocating resources can all be aided by an understanding of cancer patterns. It is envisioned that this application will help in the battle against cancer by making comprehensive statistics easily accessible.

SQL OR NOSQL

We are using SQL in this project.

PROJECT RESOURCES

For this project, we intend to utilise our personal laptops. Two of us have M1 Pro models and one is Intel-based Systems.
We are using Java and varying IDEs, as well as git for source control. We will be using the Splash library for our graphical user Interface, as well as the Spring Framework for our application's framework. For database tools, we are using MySQLWorkbench.

APPLICATION

Top Level Functionality

APPLICATION OVERVIEW

1. **Data Ingestion:** Automate the process of gathering statistics on cancer rates from a range of reliable sources, like state health departments and the CDC.
2. **User Querying:** Enable users the ability to run intricate queries to retrieve information about particular cancer types, demographic groups, or time frames.
3. **Data Visualization:** To make the analysis of patterns and indicators better to comprehend, we provide graphical representations of the data using charts and maps.
4. **User Management:** Manage user accounts, Including authentication and authorization to access or modify data.
5. **Reporting:** Allow users to generate custom reports summarizing key statistics and trends.

USER-APP INTERACTION

1. **User Registration and Login:**
    a. Users must able to create an account and should be able to log In using valid credentials to access the data.
    b. The system will distinguish between various user roles and assign distinct permissions according to those roles (e.g., Viewer, Analyst, Administrator).
2. **Selecting Data Parameters:**
    a. Once logged In, users can filter data with various parameters such as state, cancer type, year, and demographic groups to tailer the data they wish to view.
3. **Querying the Database:**
    a. User queries are submitted, and the application's backend handles processing them.
    b. The database returns the relevant data, which the application the presents to the user.

USER-APP INTERACTION CONT.

4. Data Visualization:

a. Users have the option to view the data in a variety of ways, including maps and charts.

b. The application will dynamically generate these visualizations based on the query results.

5. Downloading Reports:

a. Based on their inquiries, users may generate and download reports.

i. These reports may Include narrative summaries, tables, and charts.

6. **Updating Database:**

a. As more data becomes available, authorised users can update the database.

b. To ensure data integrity and audit trails, all modifications are recorded in the ActivityLog.

7. **Logging Out:**

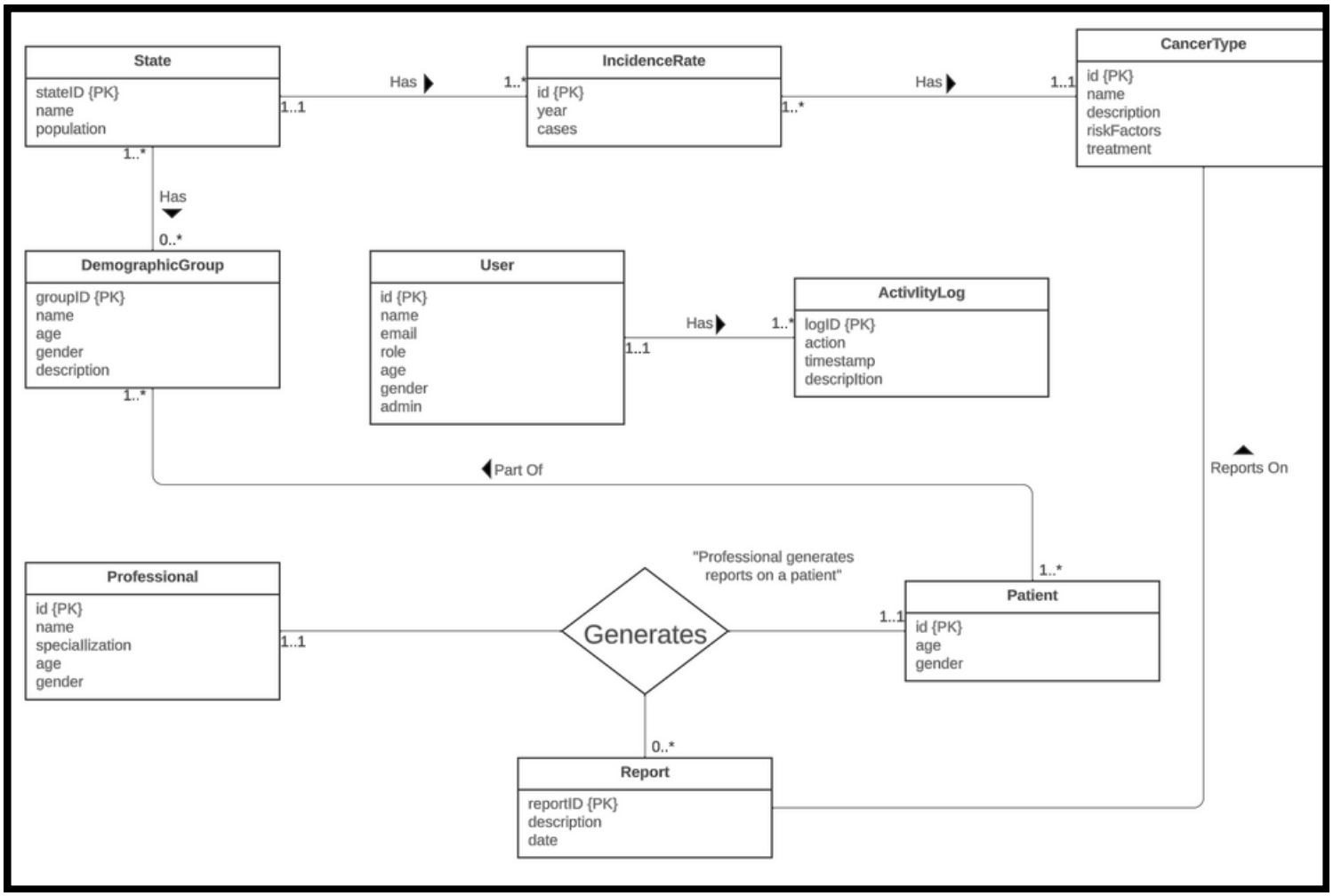
a. Users are required to log out of the set up after their session is over.

b. To secure user information and access credentials, the system will make sure that the session is closed securely.

Throughout the user Interaction, security measures will be In place to protect sensitive data and user Information, ensuring that **CancerStatisticExplorer** Is not only a powerful tool for understanding cancer rates but also a secure repository of sensitive health data.

DB SCHEMA

Conceptual Design



DB SCHEMA

Textual Description

ENTITIES

1. **State:**
   a. **Attributes**: StateID {PK}**,** Name, Population
   b. **Description:** Represents each U.S. state with demographic information**.**
2. **Demographic Group:**
   a. **Attributes**: GroupID {PK}**,** Name, Age, Gender, Description.
   b. **Description**: Represents different demographic segments for more detailed analysis.
3. **Incidence Rate:**
   a. **Attributes**: ID {PK}**,** Year, Cases.
   b. **Description**: Captures the incidence rate of a particular cancer type in a given state and year.
4. **Cancer Type**:
   a. **Attributes**: ID {PK}, Name, Descriptions, riskFactors, Treatment
   b. **Description**: Details different types of cancer tracked in the database.
5. **Professional**:
   a. **Attributes**: ID {PK}, Name, Specialization, Age, Gender
   b. **Description**: An expert that is treating and creating reports on patients.
6. **Report**:
   a. **Attributes**: ReportId {PK}, Description, Date
   b. **Description**: Report on patient that tracks their cancer types and appointments.
7. **Patient:**
   a. **Attributes**: ID {PK}, Age, Gender, GroupId
   b. **Description**: Part of a demographic group who is being treated by a professional.
8. **User:**
   a. **Attributes**: ID {PK}, Name, Email, Role, Age, Gender, Admin
   b. **Description**: Information about users who can access the database.
9. **ActivityLog:**
   a. **Attributes**: LogId {PK}, Action, TimeStamp, Description
   b. **Description**: Tracks user interactions with the database for auditing and security purposes.

RELATIONSHIPS

1. **State to Incidence Rate:**
    a. **Multiplicity:** One to many, one to one
    b. **Description:** State has Incidence rate, one incidence rate can be a part of one state.
2. **State to Demographic Group:**
    a. **Multiplicity:** Zero to many, one to many
    b. **Description:** A state can have one to many demographic groups, a demographic group can be a part of 0 to many states.
3. **Incidence Rate to Cancer Type:**
    a. **Multiplicity:** One to one, one to many
    b. **Description:** Incidence rate will have one to many cancer type but a cancer type can be a part of only one incidence rate.
4. **Demographic Group to Patients:**
    a. **Multiplicity:** One to many, one to many
    b. **Description:** Patients are a part of one to many demographic group, and a demographic group can have one to many patients
5. **Reports to Cancer Type:**
    a. **Multiplicity:** One to many, one to many
    b. **Description:** reports report on 1 to many cancer types, cancer type will be reported by 1 to many reports
6. **User to Activity Log:**
    a. **Multiplicity:** One to many, one to one
    b. **Description:** User has 1 to many activity logs, an activity log is part of one patient.
7. **Professional generates reports on patient:**
    a. **Multiplicity:** One to one, zero to many, one to one
    b. **Description:** Ternary relationship represents the instance of a medical professional filing any amount of reports on a patient.