Database Model Description
Course Project: Heart Attack Risk Prediction Database

1. Project Overview

This database model is designed to store and analyze patient data for predicting heart attack risks. The schema adheres to **third normal form (3NF)** and includes 10 tables to ensure scalability, clarity, and data integrity.

2. Entities and Relationships

2.1 Core Tables

1. **Patient**: Stores patient demographics (e.g., `Age`, `Sex`, `Income`).
2. **Country**: Maps countries to their continents and hemispheres.
3. **Continent/Hemisphere**: Reference tables for geographic categorization.
4. **Diet**: Defines dietary categories (Healthy, Average, Unhealthy).
5. **RiskFactor**: Lists risk factors (e.g., Diabetes, Smoking, Obesity).

2.2 Specialized Tables

- **HealthMetrics**: Clinical measurements (Cholesterol, Blood Pressure).
- **Lifestyle**: Exercise, sedentary hours, and sleep patterns.
- **MedicalHistory**: Heart-related medical records.

2.3 Many-to-Many Relationship

- **PatientRiskFactor**: Links patients to multiple risk factors (e.g., a patient can have both Diabetes and Obesity).

3. Normalization Process

3.1 First Normal Form (1NF)

- Atomic values enforced (e.g.,
  splitting `BloodPressure` into `SystolicBP` and `DiastolicBP`).
- Eliminated repeating groups by creating separate tables (e.g., `RiskFactor`).

3.2 Second Normal Form (2NF)

- Removed partial dependencies:
  - `Country` depends on `CountryName`, not `PatientID`.
  - `Continent` and `Hemisphere` stored independently.

3.3 Third Normal Form (3NF)

- Eliminated transitive dependencies:

    - Geographic data (Continent, Hemisphere) moved to reference tables.
    - Health metrics split into HealthMetrics, Lifestyle, and MedicalHistory to avoid redundancy.

4. Relationships and Constraints

4.1 Primary Keys

- All tables have primary keys (e.g., PatientID, CountryName).

4.2 Foreign Keys

- Patient references Country and Diet.
- Country references Continent and Hemisphere.

4.3 Constraints

- **Data Validation**:
    - CHECK (Cholesterol > 0) in HealthMetrics.
    - CHECK (Sex IN ('Male', 'Female')) in Patient.
- **Referential Integrity**: Foreign keys ensure valid relationships (e.g., CountryName in Patient must exist in Country).

5. Data Types

- **Numeric**: INT for age, DECIMAL for BMI and exercise hours.
- **Boolean**: Flags like HeartAttackRisk use BOOLEAN.
- **Text**: VARCHAR for names, ENUM for categorical data (e.g., Sex).

6. Schema Clarity

- **Naming Conventions**:
    - Tables: Descriptive names (e.g., MedicalHistory, Lifestyle).
    - Columns: Self-explanatory (e.g., ExerciseHoursPerWeek, SleepHoursPerDay).
- **Consistency**:
    - All foreign keys use the format [TableName]ID or [TableName]Name.

7. Key Design Decisions

1. **Modularity**: Separated clinical, lifestyle, and geographic data for flexibility.
2. **Scalability**: Adding new risk factors or regions requires no schema changes.
3. **Performance**: Indexed primary/foreign keys for efficient querying.

8. Conclusion

This schema achieves:

- **3NF Compliance**: No redundancy or dependency issues.
- **Business Problem Alignment**: Supports heart attack risk analysis through structured patient data.
- **Usability**: Clear naming and constraints make the database easy to maintain.