**SDG 6 - Clean Water and Sanitation**

**Part 1: SDG Selection and Problem Definition**

**SDG Selection**:  
SDG 6: "Ensure availability and sustainable management of water and sanitation for all."

**Problem Definition**:  
One of the key challenges within SDG 6 is the **lack of access to clean drinking water** in rural areas. Using data, we can develop a solution to identify regions with poor water access, analyze the factors contributing to water scarcity, and propose improvements based on available data.

**Part 2: Database Design**

**ERD (Entity-Relationship Diagram):**

Key entities and relationships for the project:

1. **Regions**: Represents geographic areas (region ID, name, population, water sources).
2. **Water Sources**: Represents available water sources (source ID, type, capacity, contamination level).
3. **Households**: Represents individual households within a region (household ID, region ID, access to clean water, consumption).
4. **Sanitation Facilities**: Represents sanitation access (facility ID, region ID, type, condition).

**Relationships:**

* Regions can have multiple water sources and sanitation facilities.
* Households are located within regions and depend on water sources.

**SQL Schema:**

Here’s a basic SQL schema based on the ERD:

CREATE TABLE Regions (

region\_id INT PRIMARY KEY,

region\_name VARCHAR(255),

population INT

);

CREATE TABLE Water\_Sources (

source\_id INT PRIMARY KEY,

region\_id INT,

source\_type VARCHAR(50),

capacity INT,

contamination\_level DECIMAL(5,2),

FOREIGN KEY (region\_id) REFERENCES Regions(region\_id)

);

CREATE TABLE Households (

household\_id INT PRIMARY KEY,

region\_id INT,

has\_clean\_water BOOLEAN,

daily\_water\_consumption INT,

FOREIGN KEY (region\_id) REFERENCES Regions(region\_id)

);

CREATE TABLE Sanitation\_Facilities (

facility\_id INT PRIMARY KEY,

region\_id INT,

facility\_type VARCHAR(50),

condition VARCHAR(50),

FOREIGN KEY (region\_id) REFERENCES Regions(region\_id)

);

**Sample Data:**

You can populate these tables with data that mimics real-world scenarios:

-- Regions

INSERT INTO Regions (region\_id, region\_name, population) VALUES

(1, 'Region A', 15000),

(2, 'Region B', 23000),

(3, 'Region C', 8000);

-- Water\_Sources

INSERT INTO WaterSources (source\_id, region\_id, source\_type, capacity, contamination\_level) VALUES

(1, 1, 'River', 100000, 2.5),

(2, 2, 'Well', 50000, 1.2),

(3, 3, 'Rainwater Harvesting', 20000, 0.8);

-- Households

INSERT INTO Households (household\_id, region\_id, has\_clean\_water, daily\_water\_consumption) VALUES

(1, 1, FALSE, 15),

(2, 1, TRUE, 30),

(3, 2, TRUE, 50);

-- Sanitation\_Facilities

INSERT INTO SanitationFacilities (facility\_id, region\_id, facility\_type, condition) VALUES

(1, 1, 'Public Toilet', 'Poor'),

(2, 2, 'Sewage System', 'Good');

**Part 3: SQL Programming**

**Data Retrieval SQL Queries:**

1. **Identify regions with limited access to clean water**:

SELECT region\_name, COUNT(\*) AS households\_without\_clean\_water

FROM Households h

JOIN Regions r ON h.region\_id = r.region\_id

WHERE h.has\_clean\_water = FALSE

GROUP BY r.region\_name;

1. **Analyze water source contamination levels**:

SELECT region\_name, source\_type, contamination\_level

FROM WaterSources ws

JOIN Regions r ON ws.region\_id = r.region\_id

ORDER BY contamination\_level DESC;

1. **Compare daily water consumption by households across regions**:

SELECT region\_name, AVG(daily\_water\_consumption) AS avg\_water\_consumption

FROM Households h

JOIN Regions r ON h.region\_id = r.region\_id

GROUP BY r.region\_name;

**Part 5: Integration and Testing**

1. **Integration**:
   * I ensured that the SQL queries export data correctly into Excel using ODBC or another data connection.
   * I documented the process to show how the data flows from SQL to Excel, ensuring that any updates in the database are reflected in Excel.
2. **Testing**:
   * I verified the consistency of the data in Excel by checking that the analysis results match the SQL outputs.
   * I tested the dashboard by filtering data across different regions and making sure all graphs and tables update dynamically.