SDM - Final Project

Name: Disease Data Analysis and Visualization

Team:

- Prarthana Bahuriya
- Jetendra Mulinti
- Goutham Vemula
- Prajeeth Nakka

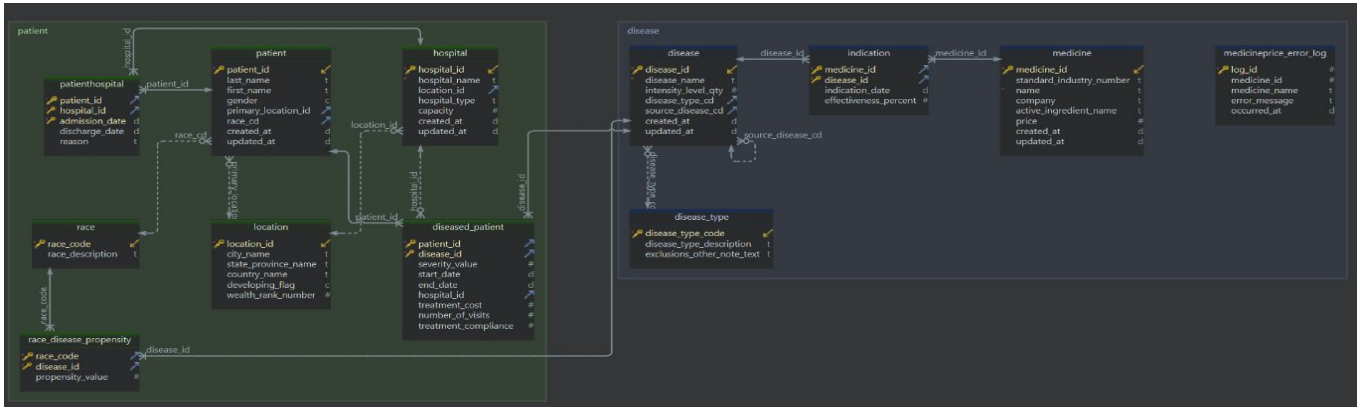
Agenda

- Introduction to the Project
- Business Problem Overview
- ER Diagram and Data Dictionary
- Operational Database
- Dimensional Database and ETL Process
- Dashboard and Business Analytics
- NO SQL vs. Relational Database Structure
- AWS Architecture and Implementation
- Snowflake in Our Project
- Conclusion

Business Problem :

- **Problem Statement:** Identifying patterns and insights in disease data to understand healthcare business.
- **Project Objective:** Utilize disease data to create a robust data model, perform analytical queries, and provide actionable insights through visualizations.

ER Diagram and Data Dictionary



- [Data Dictionary]

Operational Database

Create Statements for Base tables

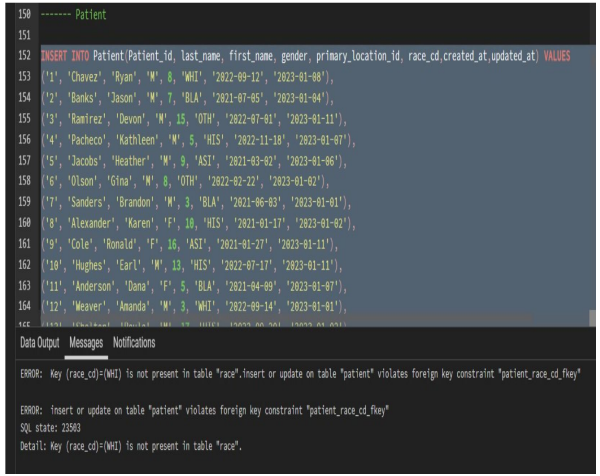
```sql
DROP TABLE IF EXISTS PatientHospital;
DROP TABLE IF EXISTS Diseased_Patient;
DROP TABLE IF EXISTS Indication;
DROP TABLE IF EXISTS Race_Disease_Propensity;
DROP TABLE IF EXISTS Medicine;
DROP TABLE IF EXISTS Patient;
DROP TABLE IF EXISTS Disease;
DROP TABLE IF EXISTS Hospital;
DROP TABLE IF EXISTS Disease_Type;
DROP TABLE IF EXISTS Race;
DROP TABLE IF EXISTS Location;


-- Creating Location Table
CREATE TABLE Location (
    Location_ID SERIAL PRIMARY KEY,
    City_Name VARCHAR(100),
    State_Province_Name VARCHAR(100),
    Country_Name VARCHAR(100),
    Developing_Flag CHAR(1),
    Wealth_Rank_Number INT
);

-- Creating Disease Type Table
CREATE TABLE Disease_Type (
    Disease_Type_Code CHAR(5) PRIMARY KEY,
    Disease_Type_Description VARCHAR(255),
    Exclusions_Other_Note_Text VARCHAR(255)
);

-- Creating Race Table
CREATE TABLE Race (
```

Insert Statements for Base tables

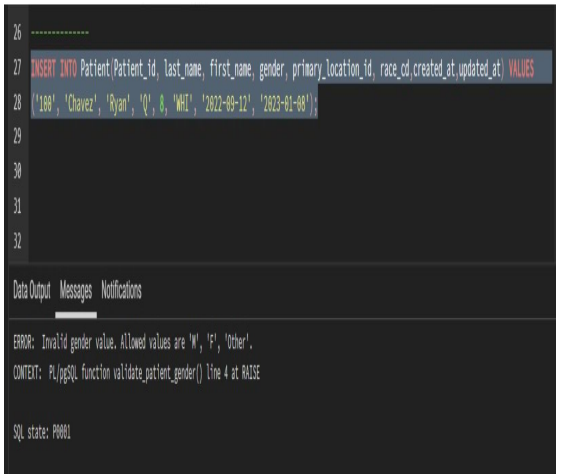
```sql
---- Location
INSERT INTO Location (Location_ID, City_Name, State_Province_Name, Country_Name, Developing_Flag, Wealth_Rank_Number) VALUES
(1, 'New York', 'New York', 'USA', 'N', 1),
(2, 'Los Angeles', 'California', 'USA', 'N', 2),
(3, 'Chicago', 'Illinois', 'USA', 'N', 3),
(4, 'Houston', 'Texas', 'USA', 'N', 4),
(5, 'Phoenix', 'Arizona', 'USA', 'N', 5),
(6, 'Philadelphia', 'Pennsylvania', 'USA', 'N', 6),
(7, 'San Antonio', 'Texas', 'USA', 'N', 7),
(8, 'San Diego', 'California', 'USA', 'N', 8),
(9, 'Dallas', 'Texas', 'USA', 'N', 9),
(10, 'San Jose', 'California', 'USA', 'N', 10),
(11, 'Austin', 'Texas', 'USA', 'N', 11),
(12, 'Jacksonville', 'Florida', 'USA', 'N', 12),
(13, 'Fort Worth', 'Texas', 'USA', 'N', 13),
(14, 'Columbus', 'Ohio', 'USA', 'N', 14),
(15, 'Charlotte', 'North Carolina', 'USA', 'N', 15),
(16, 'San Francisco', 'California', 'USA', 'N', 16),
(17, 'Indianapolis', 'Indiana', 'USA', 'N', 17),
(18, 'Seattle', 'Washington', 'USA', 'N', 18),
(19, 'Denver', 'Colorado', 'USA', 'N', 19),
(20, 'Washington', 'D.C.', 'USA', 'N', 20);


select * from Location;


--- Disease_Type
INSERT INTO Disease_Type (Disease_Type_Code, Disease_Type_Description, Exclusions_Other_Note_Text) VALUES
('DTC01', 'Infectious', 'Excludes chronic infections'),
('DTC02', 'Metabolic', 'Excludes inherited metabolic disorders'),
```

Checks, Triggers & Views

Referential Integrity Check

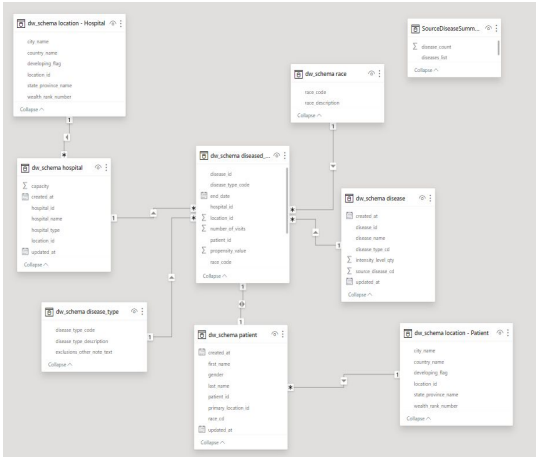


Gender Trigger Check



PriceTrigger Check

Dimensional Database and ETL Process



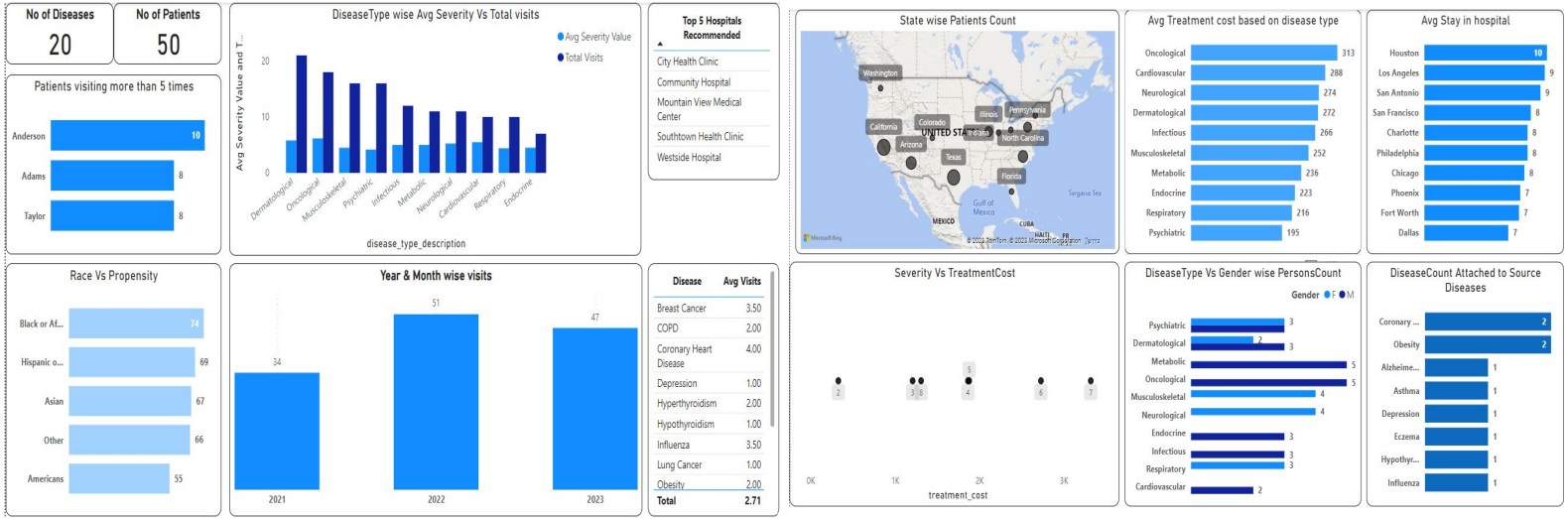
Create Statements for OLAP

```sql
--- CREATE SCHEMA IF NOT EXISTS dw_schema;

DROP TABLE IF EXISTS dw_schema.location;
DROP TABLE IF EXISTS dw_schema.race;
DROP TABLE IF EXISTS dw_schema.disease_type;
DROP TABLE IF EXISTS dw_schema.hospital;
DROP TABLE IF EXISTS dw_schema.Patient;
DROP TABLE IF EXISTS dw_schema.disease;
DROP TABLE IF EXISTS dw_schema.diseased_patient;


-- DW Table: Location
CREATE TABLE dw_schema.location (
    location_id INT PRIMARY KEY,
    city_name VARCHAR(100),
    state_province_name VARCHAR(100),
    country_name VARCHAR(100),
    developing_flag CHAR(1),
    wealth_rank_number INT
);

-- DW Table: Race
CREATE TABLE dw_schema.race (
    race_code CHAR(3) PRIMARY KEY,
    race_description VARCHAR(255)
);

-- DW Table: Disease_Type
CREATE TABLE dw_schema.disease_type (
    disease_type_code CHAR(5) PRIMARY KEY,
    disease_type_description VARCHAR(255),
    exclusions_other_note_text VARCHAR(255)
```

Insert Statements for OLAP

```sql
-- Insert data into Location
INSERT INTO dw_schema.location (location_id,city_name, state_province_name, country_name, developing_flag, wealth_rank_number)
SELECT location_id,city_name, state_province_name, country_name, developing_flag, wealth_rank_number
FROM public.location;

-- Insert data into Race
INSERT INTO dw_schema.race (race_code, race_description)
SELECT race_code, race_description
FROM public.race;

-- Insert data into Disease_Type
INSERT INTO dw_schema.disease_type (disease_type_code, disease_type_description, exclusions_other_note_text)
SELECT disease_type_code, disease_type_description, exclusions_other_note_text
FROM public.disease_type;


-- Insert data into Hospital
INSERT INTO dw_schema.hospital (hospital_id,hospital_name, location_id, hospital_type, capacity, created_at, updated_at)
SELECT hospital_id,hospital_name, location_id, hospital_type, capacity, created_at, updated_at
FROM public.hospital;


-- Insert data into Patient
INSERT INTO dw_schema.Patient (Patient_id,last_name, first_name, gender, primary_location_id, race_cd, created_at, updated_at)
SELECT Patient_id,last_name, first_name, gender, primary_location_id, race_cd, created_at, updated_at
FROM public.Patient;


-- Insert data into Disease
INSERT INTO dw_schema.disease (disease_id, disease_name, intensity_level_qty, disease_type_cd, source_disease_cd, created_at,
SELECT disease_id, disease_name, intensity_level_qty, disease_type_cd, source_disease_cd, created_at, updated_at
```

Dashboard and Business Analytics



NO SQL vs. Relational Database Structure

NoSQL Database:

- Schema Flexibility
- Horizontal Scaling
- Data Model Options
- Denormalization

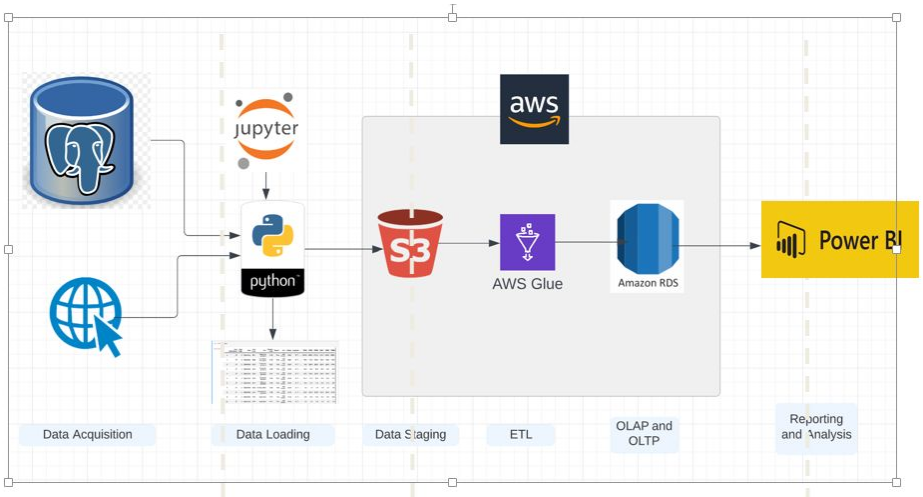
Relational Database:

- Structured Schema
- ACID Transactions
- Complex Query Support
- Normalization

```
+-----------+-----------+------------+--------+--------------------+----------+
| Person_ID | Last_Name | First_Name | Gender | Primary_Location_ID| Race_CD  |
+-----------+-----------+------------+--------+--------------------+----------+
|         1 | Doe       | Jane       | F      |                 123| WHT      |
|         2 | Smith     | John       | M      |                 456| BLK      |
+-----------+-----------+------------+--------+--------------------+----------+
```

```
{
  "Person_ID": 1,
  "Last_Name": "Doe",
  "First_Name": "Jane",
  "Gender": "F",
  "Primary_Location": {
    "Location_ID": 123,
    "City": "Springfield",
    "State": "IL"
  },
  "Race": "White"
},
{
  "Person_ID": 2,
  "Last_Name": "Smith",
  "First_Name": "John",
  "Gender": "M",
  "Primary_Location": {
    "Location_ID": 456,
```

AWS Workflow



Snowflake Commentary

- **Scalability:** Effortlessly scales to meet analytical demands with on-the-fly compute resource adjustment.
- **Performance:** Delivers rapid data processing for real-time insights, vital for healthcare decision-making.
- **Time Travel:** Allows users to access historical data, providing the ability to query past states of the database for auditing, recovery, or analytical purposes.
- **Concurrent Access:** Multiple users can analyze data simultaneously without performance lag.
- **Data Sharing:** Facilitates seamless and secure sharing with stakeholders, promoting collaborative research.
- **Security:** Offers robust encryption and compliance features, ensuring patient data is protected.
- **Ecosystem Integration:** Connects with a vast array of tools and platforms for streamlined workflows.

Thank you……