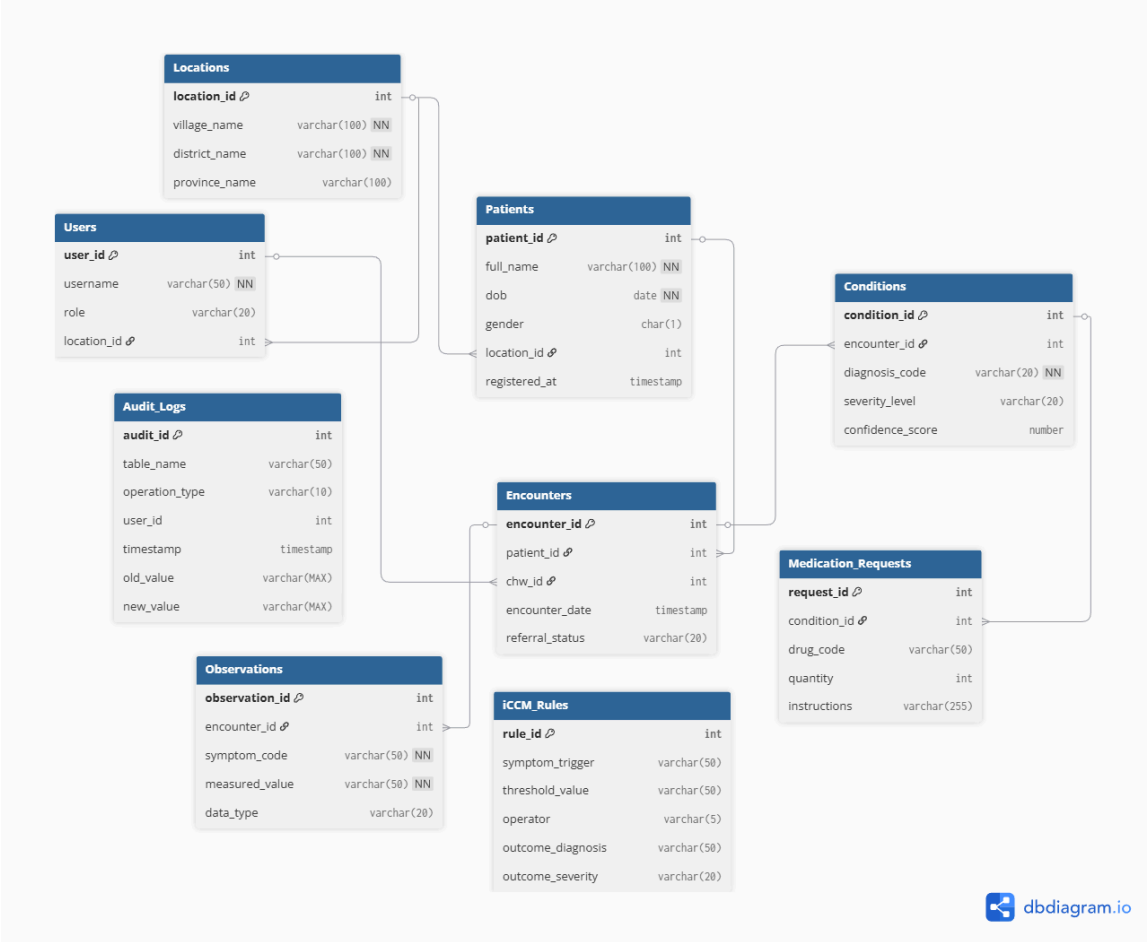
Phase III: Logical Model Design & BI Strategy

Student: Keza Shania | ID: 25793 | Database Development with PL/SQL (INSY 8311) | Group C

1. Entity-Relationship Model Overview

The logical model is designed to support high-volume transactional data (OLTP) while remaining structured enough for analytical queries (OLAP). It follows a strict hierarchical structure:

- **Dimensions (Reference Data):** Locations, Users, Patients, iCCM_Rules.
- **Facts (Transactional Data):** Encounters, Observations, Conditions, Medication_Requests.



2. Normalization Strategy (3NF Justification)

The database is normalized to the **Third Normal Form (3NF)** to ensure data integrity and reduce redundancy.

First Normal Form (1NF)

- **Requirement:** Atomic values, no repeating groups.
- **Implementation:** No table contains lists of values (e.g., Symptoms is not a comma-separated column in Encounters). Instead, symptoms are split into individual rows in the Observations table.

Second Normal Form (2NF)

- **Requirement:** No partial dependencies (all attributes depend on the *entire* Primary Key).
- **Implementation:** All tables use single-column Surrogate Primary Keys (e.g., patient_id, encounter_id). No composite keys are used for main entities, ensuring that attributes like Patient_Name depend entirely on Patient_ID.

Third Normal Form (3NF)

- **Requirement:** No transitive dependencies (Attributes must depend *only* on the Key, not on other non-key attributes).
- **Implementation:**
  - **Violation Avoided:** We did not store Village_Name and District_Name in the Patients table. If we did, District would depend on Village, not Patient.
  - **Resolution:** We created a separate Locations table. The Patients table only stores Location_ID. This allows us to update a District name in one place without scanning the entire Patient list.

3. Data Dictionary

A. Core Dimensions

TABLE: PATIENTS

Stores demographic data. Acts as a Type 2 Slowly Changing Dimension (SCD) candidate.

Column	Data Type	Key	Constraints	Description
patient_id	NUMBER(10)	PK	NOT NULL	Surrogate Key
full_name	VARCHAR2(100)		NOT NULL	
dob	DATE		NOT NULL	Used for age

				calculation
gender	CHAR(1)		CHECK ('M', 'F')	
location_id	NUMBER(10)	FK	Ref(Locations)	Link to Village

TABLE: LOCATIONS

Normalized geographical hierarchy.

Column	Data Type	Key	Constraints	Description
location_id	NUMBER(10)	PK	NOT NULL	
village_name	VARCHAR2(100)		NOT NULL	
district_name	VARCHAR2(100)		NOT NULL	

B. Transactional Facts

TABLE: ENCOUNTERS

The central fact table linking a patient to a provider and time.

Column	Data Type	Key	Constraints	Description
encounter_id	NUMBER(15)	PK	NOT NULL	
patient_id	NUMBER(10)	FK	Ref(Patients)	
chw_id	NUMBER(10)	FK	Ref(Users)	
encounter_date	TIMESTAMP		DEFAULT SYSDATE	Key for Time Dimension

TABLE: OBSERVATIONS

Vertical storage of symptoms (EAV Model specific to clinical data).

Column	Data Type	Key	Constraints	Description
observation_id	NUMBER(15)	PK	NOT NULL	
encounter_id	NUMBER(15)	FK	Ref(Encounters)	
symptom_code	VARCHAR2(50)		NOT NULL	Matches iCCM Rules
measured_val	VARCHAR2(50)		NOT NULL	The raw input value

4. Business Intelligence (BI) Considerations

Fact vs. Dimension Identification

To support the BI Phase, the schema is designed to be easily transformed into a **Star Schema**:

- **Fact Tables:** Encounters (Count of visits), Conditions (Count of diseases), Medication_Requests (Volume of stock used).
- **Dimensions:** Dim_Time (derived from encounter_date), Dim_Location (from Locations), Dim_Patient (from Patients).

Audit Trails

- **Requirement:** Strictly track data changes for accountability.
- **Strategy:** A dedicated AUDIT_LOGS table will be populated via **Database Triggers** (Phase VII). It will capture the OLD_VALUE and NEW_VALUE whenever a diagnosis is updated or deleted.

Assumptions

1. **Identity:** A Patient is uniquely identified by Patient_ID. We assume biometric or ID card verification happens at the application level.
2. **Location:** Villages do not move between Districts often (Slowly Changing Dimension Type 1).
3. **Connectivity:** The database acts as the master sync point; mobile devices may cache data but the Oracle DB is the "Source of Truth."