PART – 1

Dataset

Dataset Selection:
Dataset: Healthcare Dataset (https://www.kaggle.com/datasets/prasad22/healthcare-dataset/data)

Format: Available in CSV format, suitable for importing into PostgreSQL.

Description: This dataset is a synthetic healthcare dataset designed for data science, machine learning, and data analysis enthusiasts. It serves as a valuable resource to practice and showcase data manipulation and analysis skills within the context of a healthcare system.

Here's a breakdown of the key points about the dataset:
Content: It consists of 10,000 records, each representing a synthetic patient healthcare record.
Data Elements: The records include various attributes related to patient demographics, medical conditions, admission details, diagnoses, medications, and more. This provides a comprehensive overview of simulated patient healthcare encounters.
Synthetic Nature: It's important to note that the data is synthetic, meaning it's artificially generated to resemble real-world healthcare data. While it offers valuable practice opportunities, it may not perfectly mirror real patient information.

PART – 2

Use case description

This section details 15 Use Cases, highlighting their connections:

Use Case 1: Register a New Patient (One patient - One record)
Description: Captures basic patient information for new registrations.
Relevance: Establishes the foundation for patient data within the system.
Interconnection: Links to future admissions, diagnoses, and medication records.

Use Case 2: Update Patient Information (One patient - Many updates)
Description: Allows modifying existing patient details (e.g., address change).
Relevance: Ensures patient information remains accurate for better care.
Interconnection: Updates patient information across all linked records (admissions, medications).

Use Case 3: Search for Patients (Many search criteria - Many patients)
Description: Finds patients based on specific criteria (name, date of birth).
Relevance: Expedites locating patients for treatment or record retrieval.

Interconnection: Provides access points to specific patient data based on search parameters, leveraging information from Patient, Insurance Provider (if applicable), and potentially Admissions (for recent interactions).

Use Case 4: Admit a Patient (One patient - One admission)
Description: Records details when a patient enters the healthcare facility, including assigning a doctor and admission type.
Relevance: Initiates a new episode of care and links to subsequent actions.
Interconnection: Creates an admission record tied to the patient, doctor (from Use Case 13: Assign Doctor to Admission), and admission type. Allows for further data collection (diagnoses, medications, billing).

Use Case 5: Record Discharge Information (One admission - One discharge record)
Description: Captures details when a patient is discharged, finalizes the billing process, and closes the admission record.
Relevance: Finalizes the admission cycle, tracks patient flow, and generates billing information.
Interconnection: Completes the admission cycle, triggers billing record creation and allows for analysis of length of stay (Use Case 8: Analyze Length of Stay).

Use Case 6: Diagnose a Patient (One patient - Many diagnoses)
Description: Links patients to diagnosed conditions during admissions.
Relevance: Documents health issues for treatment planning and future reference.
Interconnection: Connects diagnoses to specific admissions and facilitates tracking patient health history.

Use Case 7: Track Vital Signs (One admission - Many vital sign records)
Description: Records vital signs (temperature, blood pressure) measured at intervals.
Relevance: Monitors patient health status during admission.
Interconnection: Provides detailed data points for assessing patient progress within an admission.

Use Case 8: Analyze Length of Stay (Admission - Additional data for grouping)
Description: Studies how long patients from various categories stay admitted.
Relevance: Analyzes healthcare resource utilization and potential areas for improvement.
Interconnection: Leverages admission data along with patient group information for broader analyses.

Use Case 9: View Patient Medical History (One patient - Many diagnoses)
Description: Retrieves past diagnoses for a patient.
Relevance: Provides a comprehensive view of a patient's health history for informed decision-making.
Interconnection: Connects past diagnoses to a specific patient for continuity of care.

Use Case 10: Track Patient Allergies (One patient - Many allergies)
Description: Records allergies a patient may have.

Relevance: Ensures patient safety by preventing allergic reactions during treatment.
Interconnection: Flags potential risks associated with medications or procedures for a specific patient.

Use Case 11: Prescribe Medications (One admission - Many medications)
Description: Records medications prescribed for a patient during admission.
Relevance: Documents treatment plans and tracks medication use.
Interconnection: Links medications to a specific admission and patient for medication history and potential interaction checks.

Use Case 12: Check for Medication Interactions (Many medications - Analysis result)
Description: Analyzes potential interactions between prescribed medications.
Relevance: Ensures patient safety by identifying and flagging potentially harmful interactions.
Interconnection: Leverages medication data from admissions to analyze potential risks.

Use Case 13: Assign Doctor to Admission (One admission - One doctor)

Description: Links a doctor to a specific patient admission.
Relevance: Assigns responsibility for care and tracks doctor involvement.
Interconnection: Connects admissions to doctors for managing patient care and tracking doctor workload.

Use Case 14: Schedule Appointments for Admissions (Admission - Doctor - Appointment schedule)
Description: Assign doctors with available time slots to upcoming admissions based on their schedule.
Relevance: Ensures continuity of care by assigning qualified doctors with available time for upcoming admissions.
Interconnection: Connects admissions to doctors with appropriate schedules, leveraging Use Case 13 ("Assign Doctor to Admission") for doctor assignment during scheduling.

Use Case 15: Generate Patient Reports (Patient - Report)
Description: This use case allows healthcare providers to generate comprehensive reports summarizing patient information.
Relevance: Provides a consolidated view of patient data for informed decision-making, treatment planning, and recordkeeping.
Interconnection: Leverages data from various entities (admissions, diagnoses, medications, allergies) to generate reports specific to a patient. Reports can be customized to include relevant details from linked Use Cases.

PART – 3

ERD

Entities:

1. Patient:
    a. Attributes:
        i. PatientID (Primary Key)
        ii. BloodTypeID (Foreign Key references Patient)
        iii. Name
        iv. Age
        v. Gender
2. BloodGroup:
    a. Attributes:
        i. BloodTypeID (Primary Key)
        ii. BloodType
3. InsuranceProvider:
    a. Attributes:
        i. InsuranceProviderID (Primary Key)
        ii. PatientID (Foreign Key references Patient)
        iii. InsuranceNumber (Candidate Key)
        iv. InsuranceProviderName
4. Admission:
    a. Attributes:
        i. AdmissionID (Primary Key)
        ii. PatientID (Foreign Key references Patient)
        iii. DoctorID (Foreign Key references Doctor)
        iv. HospitalID (Foreign Key references Hospital)
        v. MedicationID (Foreign Key references Medication)
        vi. AdmissionDate
        vii. DischargeDate
5. AdmissionType:
    a. Attributes:
        i. AdmissionTypeID (Primary Key)
        ii. AdmissionID (Foreign Key references Admission)
        iii. AdmissionType (Emergency, Elective, Urgent)
6. Billing:
    a. Attributes:
        i. BillingID (Primary Key)
        ii. AdmissionID (Foreign Key references Admission)
        iii. BillingAmount
7. Test:
    a. Attributes:
        i. TestID (Primary Key)

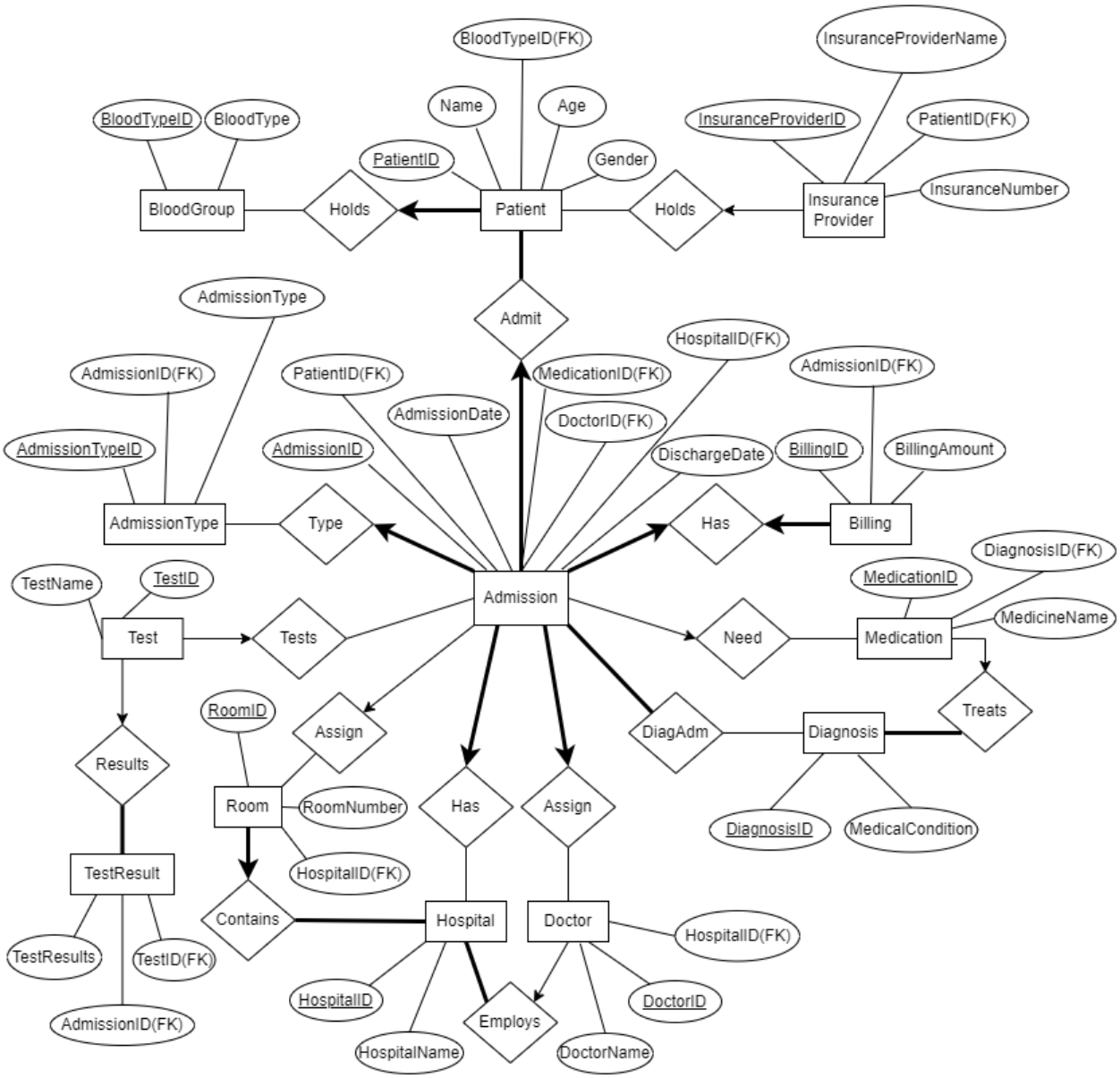
        ii. TestName
8. TestResult:
    a. Attributes:
        i. AdmissionID (Foreign Key references Admission)
        ii. TestID (Foreign Key references Test)
        iii. TestResults
9. Hospital:
    a. Attributes:
        i. HospitalID (Primary Key)
        ii. HospitalName
10. Room:
    a. Attributes:
        i. RoomID (Primary Key)
        ii. HospitalID (Foreign Key references Hospital)
        iii. RoomNumber
11. Doctor:
    a. Attributes:
        i. DoctorID (Primary Key)
        ii. HospitalID (Foreign Key references Hospital)
        iii. DoctorName
12. Medication:
    a. Attributes:
        i. MedicationID (Primary Key)
        ii. DiagnosisID (Foreign Key references Diagnosis)
        iii. MedicineName
13. Diagnosis:
    a. Attributes:
        i. DiagnosisID (Primary Key)
        ii. MedicalCondition
14. DiagAdm (Associative Entity):
    a. Attributes:
        i. AdmissionID (Foreign Key references Admission) – Primary Key part 1
        ii. DiagnosisID (Foreign Key references Diagnosis) – Primary Key part 2
    b. Participation: Links one Admission to many Diagnoses, and one Diagnosis to many admissions (Many-to-Many)

Participation:

1. Patient Can have many Admissions
    a. one-to-many
    b. Total Participation - Every patient must have <u>at least</u> one admission record
    c. Total Participation - Every admission <u>must be</u> associated with a patient
2. Patient can have many insurances providers
    a. one-to-many

    b. Partial Participation - A patient may or may not have an insurance provider
    c. Partial Participation - An insurance provider may or may not have associated with many patients
3. Patient can have only one blood type, but a single blood type can be associated with many patients
    a. many-to-one
    b. Total Participation – Every patient must be having exactly one blood type
    c. Partial Participation – The blood type may or may not be associated with the patient
4. Many admissions can have one medication
    a. many-to-one
    b. Partial Participation - An admission may or may not have a specific medication prescribed
    c. Partial Participation - One medication may or may not be prescribed in an admission
5. An admission can only assign one doctor, but one doctor can be assigned to multiple admissions
    a. many-to-one
    b. Total Participation - Every admission must be assigned to one doctor
    c. Partial Participation – A specific doctor may or may not be assigned to an admission
6. An admission can have many Diagnoses, and one Diagnosis to many admissions
    a. Many-to-Many - linked through a separate entity "DiagAdm" (Short of "DiagnosisAdmission")
    b. Total Participation - Every admission must have at least one diagnosis associated with it
    c. Partial Participation – One diagnosis (medical condition) may or may not be associated with an admission
7. An admission has one Billing record
    a. one-to-one
    b. Total Participation - Every admission must have a corresponding billing record
    c. Total Participation – Every billing record must be associated with an admission
8. An admission has one or more tests to be performed
    a. one-to-many
    b. Partial Participation - An admission may or may not have tests performed
    c. Partial Participation – A test may or may not be performed on the admission
9. An admission has one specific admission type but there can be many admissions with the same admission type
    a. many-to-one
    b. Total Participation - Every admission must have a specific admission type assigned
    c. Partial Participation – An admission type may or may not be associated with an admission
10. A hospital can have many admissions
    a. one-to-many
    b. Total Participation - Every admission must occur in a specific hospital
    c. Partial Participation – A hospital may or may not have admissions
11. A hospital can employ many doctors

    a. one-to-many
    b. Partial Participation - A doctor may or may not be employed by a specific hospital
    c. Total Participation – Every hospital must have a doctor
12. A hospital contains many rooms
    a. one-to-many
    b. Total Participation - Every hospital must have at least one room
    c. Total Participation – Every room must be in a hospital
13. A room can be assigned to many admissions over time
    a. one-to-many
    b. Partial Participation - A room may or may not be assigned to an admission at a particular time
    c. Partial Participation – An admission may or may not be assigned a room
14. A medical condition can be treated with multiple medicine
    a. one-to-many
    b. Total Participation – Every diagnosis (medical condition) must be having associated medications
    c. Partial Participation – A medication may or may not be used for treating the medical condition
15. Many tests can have same test result, but one test result is associated with only one test
    a. one-to-many
    b. Partial Participation – A test may or may not have the test result at a particular time
    c. Total Participation - Every test result must be linked to the specific test performed


BloodTypeID(FK)
InsuranceProviderName
BloodTypeID
BloodType
Name
Age
InsuranceProviderID
PatientID(FK)
PatientID
Gender
InsuranceNumber
BloodGroup
Holds
Patient
Holds
Insurance Provider
Admit
AdmissionType
HospitalID(FK)
AdmissionID(FK)
PatientID(FK)
MedicationID(FK)
AdmissionID(FK)
AdmissionDate
DoctorID(FK)
AdmissionTypeID
AdmissionID
DischargeDate
BillingID
BillingAmount
AdmissionType
Type
Has
Billing
DiagnosisID(FK)
TestName
TestID
Admission
MedicationID
Test
Tests
Need
Medication
MedicineName
RoomID
Assign
Treats
Results
DiagAdm
Diagnosis
Room
RoomNumber
Has
Assign
DiagnosisID
MedicalCondition
TestResult
HospitalID(FK)
Contains
Hospital
Doctor
HospitalID(FK)
TestResults
TestID(FK)
HospitalID
DoctorID
Employs
AdmissionID(FK)
HospitalName
DoctorName

PART – 4

Write-Up: NewSQL Databases and the Evolving Landscape of DBMS

Issue: The Rise of Big Data and Limitations of Traditional DBMS

Traditional relational databases (RDBMS) have been the cornerstone of data management for decades. However, the ever-growing volume, variety, and velocity of data, especially with the rise of big data, challenges their capabilities. RDBMS can struggle with scaling efficiently and handling complex data structures like semi-structured or unstructured data.

Impact and Significance:

Data Explosion: Organizations across industries are generating massive datasets that RDBMS weren't built to handle effectively.
New Data Types: Big Data often includes unstructured data (e.g., sensor data, social media posts) that doesn't fit neatly into the rigid schema of RDBMS.
Performance Bottlenecks: Traditional RDBMS can struggle with real-time analytics and complex queries on large datasets, impacting decision-making agility.

Existing Approaches:

Vertical Scaling (Upsizing Hardware): Scaling RDBMS by adding more powerful hardware can be expensive and eventually reach a limit.
Sharding: Dividing a large database into smaller, more manageable chunks can improve scalability but introduces complexity in managing distributed data.
NoSQL Databases: These non-relational databases offer greater flexibility and scalability for handling diverse data types and structures. However, they often lack the strong data consistency guarantees of RDBMS.

Chosen Approach: NewSQL Databases

NewSQL databases emerge as a promising solution by combining the scalability and flexibility of NoSQL databases with the ACID (Atomicity, Consistency, Isolation, Durability) transaction guarantees traditionally associated with RDBMS.

Justification:

Bridges the Gap: NewSQL offers the best of both worlds, handling diverse data structures while maintaining strong data consistency for reliable transactions.
Scalability and Performance: NewSQL databases can scale horizontally, adding more nodes to handle increasing data volumes and query workloads efficiently.
Real-Time Analytics: NewSQL can handle complex queries on large datasets faster, facilitating real-time data analysis and decision support.

Challenges and Future Directions:

Complexity: Implementing and managing NewSQL databases can be more complex than traditional RDBMS.

Maturity: NewSQL is a relatively new technology, and ongoing development is needed to optimize performance and explore advanced features.

Conclusion:

The data landscape is constantly evolving, and traditional RDBMS may not always suffice for big data needs. NewSQL databases offer a promising solution that combines scalability, flexibility, and data consistency. As NewSQL technology matures and offers more user-friendly tools, it will likely play a significant role in the future of data management for large-scale, complex data environments.

Reference:

NewSQL Databases: https://en.wikipedia.org/wiki/SQL This Wikipedia article provides a comprehensive overview of NewSQL databases, including their functionalities, limitations, and a list of popular NewSQL database products.

What is NewSQL? https://www.g2.com/compare/dremio-vs-sql-server-2019 This resource offers a clear explanation of NewSQL databases, highlighting their key features and how they address the limitations of traditional RDBMS in the context of big data.