Healthcare Insurance Claim Data Warehouse Project Documentation

Project Title: Healthcare Insurance Claim Processing Data Warehouse

Objective: To design and implement a healthcare-specific data warehouse schema focused on insurance claim processing. This solution enables efficient data analysis and reporting of claim approvals, processing times, treatment costs, and hospital-wise revenue, aiding decision-making across healthcare operations.

Use Case: Insurance Claim Processing Insurance claim processing is critical for hospitals and patients to manage treatment reimbursements efficiently. The warehouse tracks treatments, doctors, hospitals, insurance providers, and processing performance metrics.

Key Business Questions Answered:

1. What is the total revenue generated by each hospital through insurance claims?
2. What is the average time taken by different insurance providers to process claims?
3. What is the distribution of claim statuses (Approved, Rejected, Pending)?
4. Which treatments are most frequently claimed?
5. How do claims differ by patient city or demographics?

Schema Type: Star Schema was chosen for its simplicity and suitability for analytical queries. The fact table is surrounded by six dimension tables.

Schema Design:

Fact Table: `fact_insurance_claims`

- claim_ref_id (PK)
- claim_patient_id (FK)
- claim_doctor_id (FK)
- claim_treatment_id (FK)
- claim_hospital_id (FK)
- provider_id (FK)
- claim_date_id (FK)
- claim_amount
- treatment_duration_min
- claim_decision_status

- approval_days

Dimension Tables:

1. `dim_claim_patient` — patient demographic info
2. `dim_claim_doctor` — doctor specialty & department
3. `dim_claim_treatment` — treatment type & category
4. `dim_claim_hospital` — hospital & location
5. `dim_insurance_provider` — insurance company & plan type
6. `dim_claim_date` — calendar details

Sample Data Fields:

- 5 patients, 5 doctors, 5 treatments, 5 hospitals, 5 insurance providers, 5 dates
- 5 claim records inserted into the fact table.

Sample Queries & Business Insights:

1. Total Revenue per Hospital

   SELECT h.institution_name, SUM(f.claim_amount) AS total_revenue

   FROM fact_insurance_claims f

   JOIN dim_claim_hospital h ON f.claim_hospital_id = h.claim_hospital_id

   GROUP BY h.institution_name; (Insight: Identify hospitals generating the most revenue from claims.)

2. Claim Status Breakdown

   SELECT claim_decision_status, COUNT(*) AS total_claims

   FROM fact_insurance_claims

   GROUP BY claim_decision_status; (Insight: Track how many claims are approved, pending, or rejected.)

3. Average Processing Time by Insurance

   SELECT i.company_name, AVG(approval_days) AS avg_days

   FROM fact_insurance_claims f

JOIN dim_insurance_provider i ON f.provider_id = i.provider_id

GROUP BY i.company_name; (Insight: Determine which insurance providers process claims faster.)

ER Diagram: The ER diagram shows a star schema with one central fact table and six connected dimension tables. All relationships are 1:M from dimensions to the fact table.
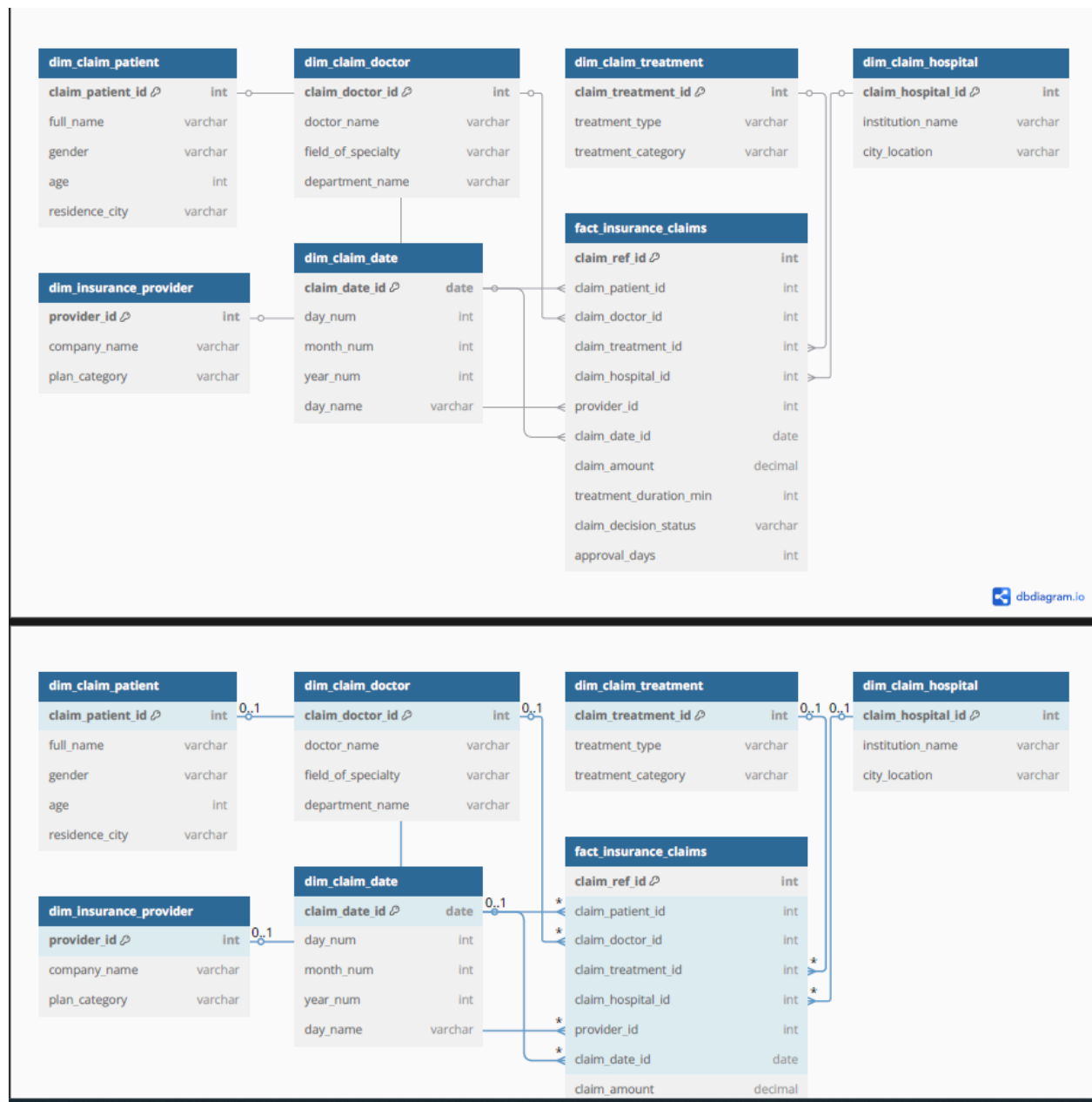
Scalability & Optimization:

- Star schema supports fast analytical queries
- Dimension tables are normalized to avoid data duplication
- Can be extended with new dimensions (diagnosis, billing agent, etc.)
- Suitable for Power BI, Tableau, or Excel dashboards

Conclusion: This healthcare data warehouse efficiently supports claim analytics by providing insights into hospital billing, claim processing delays, insurance efficiency, and patient treatment patterns. It is scalable, well-structured, and optimized for real-time business intelligence needs.

Power BI DashBoard
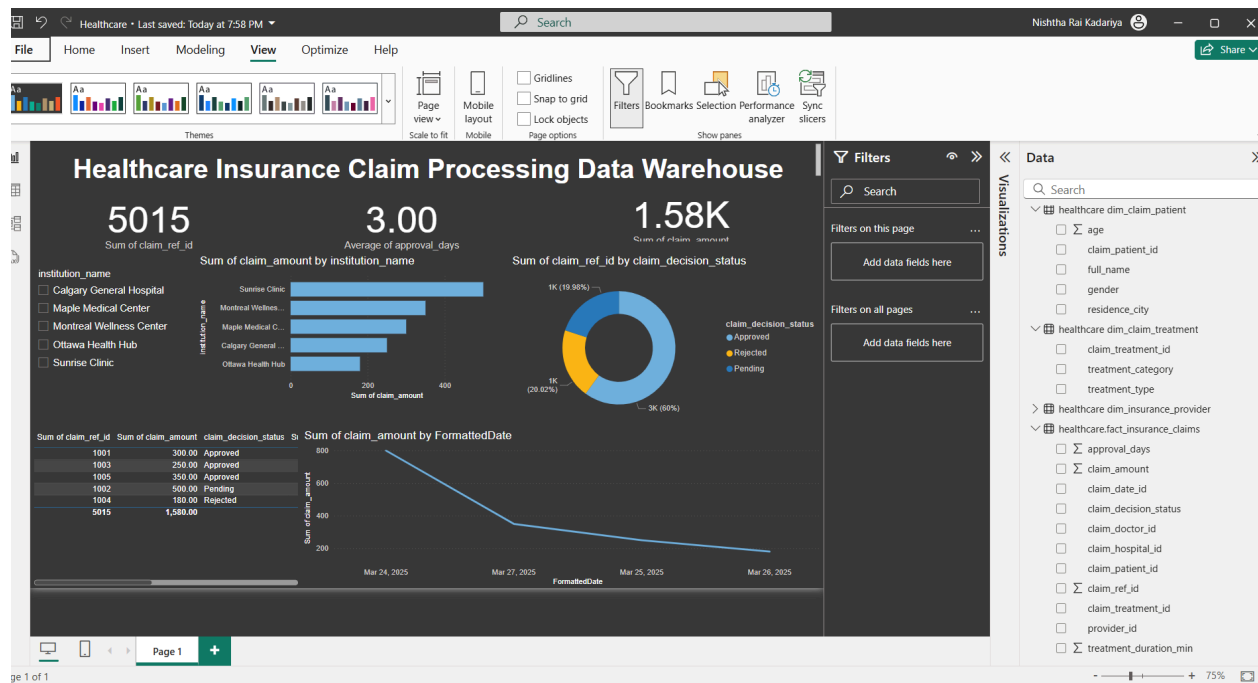
SQL Queries :

step 1

```sql
USE healthcare;

SHOW TABLES;




## create patient Dimension

CREATE TABLE dim_claim_patient (

  claim_patient_id INT PRIMARY KEY,

  full_name VARCHAR(100),

  gender VARCHAR(10),

  age INT,

  residence_city VARCHAR(50)

);


#3 doctor dimension

CREATE TABLE dim_claim_doctor (

  claim_doctor_id INT PRIMARY KEY,

  doctor_name VARCHAR(100),

  field_of_specialty VARCHAR(50),

  department_name VARCHAR(50)

);



CREATE TABLE dim_claim_treatment (
```

```sql
    claim_treatment_id INT PRIMARY KEY,

    treatment_type VARCHAR(100),

    treatment_category VARCHAR(50)

);


CREATE TABLE dim_claim_hospital (

    claim_hospital_id INT PRIMARY KEY,

    institution_name VARCHAR(100),

    city_location VARCHAR(100)

);


CREATE TABLE dim_insurance_provider (

    provider_id INT PRIMARY KEY,

    company_name VARCHAR(100),

    plan_category VARCHAR(50)

);


CREATE TABLE dim_claim_date (

    claim_date_id DATE PRIMARY KEY,

    day_num INT,

    month_num INT,

    year_num INT,

    day_name VARCHAR(10)

);
```

create a fact table

```sql
CREATE TABLE fact_insurance_claims (

  claim_ref_id INT PRIMARY KEY,

  claim_patient_id INT,

  claim_doctor_id INT,

  claim_treatment_id INT,

  claim_hospital_id INT,

  provider_id INT,

  claim_date_id DATE,

  claim_amount DECIMAL(10,2),

  treatment_duration_min INT,

  claim_decision_status VARCHAR(20),

  approval_days INT,

  FOREIGN KEY (claim_patient_id) REFERENCES dim_claim_patient(claim_patient_id),

  FOREIGN KEY (claim_doctor_id) REFERENCES dim_claim_doctor(claim_doctor_id),

  FOREIGN KEY (claim_treatment_id) REFERENCES dim_claim_treatment(claim_treatment_id),

  FOREIGN KEY (claim_hospital_id) REFERENCES dim_claim_hospital(claim_hospital_id),

  FOREIGN KEY (provider_id) REFERENCES dim_insurance_provider(provider_id),

  FOREIGN KEY (claim_date_id) REFERENCES dim_claim_date(claim_date_id)

);
```

insert the sample data in dim_claim_patient

```sql
INSERT INTO dim_claim_patient VALUES

(1, 'Alice Johnson', 'Female', 32, 'Toronto'),

(2, 'Bob Smith', 'Male', 45, 'Vancouver'),

(3, 'Clara Davis', 'Female', 28, 'Calgary'),

(4, 'David Lee', 'Male', 52, 'Ottawa'),

(5, 'Ella Brown', 'Female', 39, 'Montreal');


INSERT INTO dim_claim_doctor VALUES

(1, 'Dr. Susan Lee', 'Cardiology', 'Heart Dept'),

(2, 'Dr. Mark Adams', 'Neurology', 'Neuro Dept'),

(3, 'Dr. Nina Patel', 'Radiology', 'Imaging Dept'),

(4, 'Dr. James Wu', 'Orthopedics', 'Bone Clinic'),

(5, 'Dr. Rachel Kim', 'General Medicine', 'Outpatient');


INSERT INTO dim_claim_treatment VALUES

(1, 'ECG', 'Diagnostics'),

(2, 'MRI Scan', 'Imaging'),

(3, 'X-Ray', 'Radiology'),

(4, 'Blood Test', 'Laboratory'),

(5, 'Physiotherapy', 'Rehabilitation');


INSERT INTO dim_claim_hospital VALUES

(1, 'Maple Medical Center', 'Toronto'),

(2, 'Sunrise Clinic', 'Vancouver'),
```

```sql
(3, 'Calgary General Hospital', 'Calgary'),

(4, 'Ottawa Health Hub', 'Ottawa'),

(5, 'Montreal Wellness Center', 'Montreal');


INSERT INTO dim_insurance_provider VALUES

(1, 'SunLife', 'Premium'),

(2, 'Manulife', 'Standard'),

(3, 'GreatWest', 'Basic'),

(4, 'BlueCross', 'Gold'),

(5, 'GreenShield', 'Essential');


INSERT INTO dim_claim_date VALUES

('2025-03-24', 24, 3, 2025, 'Monday'),

('2025-03-25', 25, 3, 2025, 'Tuesday'),

('2025-03-26', 26, 3, 2025, 'Wednesday'),

('2025-03-27', 27, 3, 2025, 'Thursday'),

('2025-03-28', 28, 3, 2025, 'Friday');


INSERT INTO fact_insurance_claims VALUES

(1001, 1, 1, 1, 1, 1, '2025-03-24', 300.00, 30, 'Approved', 2),

(1002, 2, 2, 2, 2, 2, '2025-03-24', 500.00, 45, 'Pending', 1),

(1003, 3, 3, 3, 3, 3, '2025-03-25', 250.00, 20, 'Approved', 3),

(1004, 4, 4, 4, 4, 4, '2025-03-26', 180.00, 25, 'Rejected', 5),

(1005, 5, 5, 5, 5, 5, '2025-03-27', 350.00, 60, 'Approved', 4);
```

Total claim revenue per hospital

```sql
SELECT h.institution_name, SUM(f.claim_amount) AS total_revenue
FROM fact_insurance_claims f
JOIN dim_claim_hospital h ON f.claim_hospital_id = h.claim_hospital_id
GROUP BY h.institution_name;
```

claim status breakdown

```sql
SELECT claim_decision_status, COUNT(*) AS total_claims
FROM fact_insurance_claims
GROUP BY claim_decision_status;
```

average claim processing time per insurance

```sql
SELECT i.company_name, AVG(approval_days) AS avg_days
FROM fact_insurance_claims f
JOIN dim_insurance_provider i ON f.provider_id = i.provider_id
GROUP BY i.company_name;
```