Healthcare EHR Semantic Database Layer

Information Schema Documentation

This document describes the semantic database layer schema designed for diabetic patient care cost analysis. The schema follows a star/snowflake architecture with fact and dimension tables optimized for analytical queries across the patient care journey.

Schema Property	Value
Database Type	Semantic Layer (Star/Snowflake Schema)
Primary Use Case	Diabetic Patient Care Cost Analysis
Source Systems	EHR, Claims, Pharmacy, Lab Systems
Refresh Frequency	Daily ETL with Real-time Updates
Data Retention	7 Years (HIPAA Compliant)

DIMENSION TABLES

DIM_PATIENT

Central patient dimension containing demographic and clinical classification data for diabetic patients.

Column Name	Data Type	Nullable	Description
patient_key	BIGINT	NO	Surrogate key (PK)
patient_id	VARCHAR(20)	NO	Natural key from source EHR
mrn	VARCHAR(30)	NO	Medical Record Number
first_name	VARCHAR(100)	YES	Patient first name (encrypted)
last_name	VARCHAR(100)	YES	Patient last name (encrypted)
date_of_birth	DATE	NO	Date of birth
gender	CHAR(1)	NO	M/F/O/U
diabetes_type	VARCHAR(20)	NO	Type1/Type2/Gestational/MODY/Other
diagnosis_date	DATE	YES	Initial diabetes diagnosis date
hba1c_baseline	DECIMAL(4,2)	YES	Baseline HbA1c at diagnosis
risk_score	DECIMAL(5,2)	YES	Calculated risk score (0-100)
pcp_provider_key	BIGINT	YES	FK to DIM_PROVIDER
insurance_key	BIGINT	YES	FK to DIM_INSURANCE
address_key	BIGINT	YES	FK to DIM_GEOGRAPHY
effective_date	DATE	NO	SCD2 effective date
expiration_date	DATE	YES	SCD2 expiration date
is_current	BOOLEAN	NO	Current record flag

DIM_PROVIDER

Healthcare provider dimension including physicians, specialists, and care team members.

Column Name	Data Type	Nullable	Description
provider_key	BIGINT	NO	Surrogate key (PK)
provider_id	VARCHAR(20)	NO	NPI or internal provider ID
npi	VARCHAR(10)	YES	National Provider Identifier
provider_name	VARCHAR(200)	NO	Full provider name
specialty	VARCHAR(100)	NO	Medical specialty
provider_type	VARCHAR(50)	NO	MD/DO/NP/PA/RN/etc.
facility_key	BIGINT	YES	FK to DIM_FACILITY
department	VARCHAR(100)	YES	Department or practice area
is_diabetes_specialist	BOOLEAN	NO	Endocrinology/diabetes certified
accepting_patients	BOOLEAN	NO	Currently accepting new patients

DIM_FACILITY

Healthcare facility dimension for hospitals, clinics, pharmacies, and labs.

Column Name	Data Type	Nullable	Description
facility_key	BIGINT	NO	Surrogate key (PK)
facility_id	VARCHAR(20)	NO	Facility code (maps to Excel)
facility_name	VARCHAR(200)	NO	Full facility name
facility_type	VARCHAR(50)	NO	Hospital/Clinic/Pharmacy/Lab/DME
address_line1	VARCHAR(200)	YES	Street address
city	VARCHAR(100)	YES	City
state	CHAR(2)	YES	State code
zip_code	VARCHAR(10)	YES	ZIP/Postal code
geography_key	BIGINT	YES	FK to DIM_GEOGRAPHY
bed_count	INT	YES	Number of beds (hospitals)
is_teaching	BOOLEAN	NO	Teaching hospital flag
cms_certification	VARCHAR(20)	YES	CMS certification number

DIMENSION TABLES (Continued)

DIM_DATE

Date dimension for time-based analysis with fiscal and calendar hierarchies.

Column Name	Data Type	Nullable	Description
date_key	INT	NO	Surrogate key (YYYYMMDD format)
full_date	DATE	NO	Actual date value
day_of_week	TINYINT	NO	1-7 (Sunday=1)
day_name	VARCHAR(10)	NO	Monday, Tuesday, etc.
day_of_month	TINYINT	NO	1-31
day_of_year	SMALLINT	NO	1-366
week_of_year	TINYINT	NO	1-53
month_number	TINYINT	NO	1-12
month_name	VARCHAR(10)	NO	January, February, etc.
quarter	TINYINT	NO	1-4
year	SMALLINT	NO	4-digit year
fiscal_quarter	TINYINT	NO	Fiscal quarter 1-4
fiscal_year	SMALLINT	NO	Fiscal year
is_weekend	BOOLEAN	NO	Weekend flag
is_holiday	BOOLEAN	NO	US federal holiday flag

DIM_DIAGNOSIS

Diagnosis dimension with ICD-10 codes and clinical groupings for diabetes-related conditions.

Column Name	Data Type	Nullable	Description
diagnosis_key	BIGINT	NO	Surrogate key (PK)
icd10_code	VARCHAR(10)	NO	ICD-10-CM code
icd10_description	VARCHAR(500)	NO	Full ICD-10 description
diagnosis_category	VARCHAR(100)	NO	Clinical category
is_diabetes_related	BOOLEAN	NO	Diabetes-related flag
is_complication	BOOLEAN	NO	Diabetes complication flag
complication_type	VARCHAR(50)	YES	Retinopathy/Nephropathy/Neuropathy/etc.
severity_level	VARCHAR(20)	YES	Mild/Moderate/Severe
hcc_code	VARCHAR(10)	YES	HCC risk adjustment code
hcc_weight	DECIMAL(5,3)	YES	HCC risk weight

DIM_PROCEDURE

Procedure dimension with CPT/HCPCS codes for medical services and treatments.

Column Name	Data Type	Nullable	Description
procedure_key	BIGINT	NO	Surrogate key (PK)
cpt_code	VARCHAR(10)	NO	CPT or HCPCS code
procedure_description	VARCHAR(500)	NO	Full procedure description
procedure_category	VARCHAR(100)	NO	E&M/Lab/Imaging/Surgery/DME
is_diabetes_specific	BOOLEAN	NO	Diabetes-specific procedure
typical_duration_min	INT	YES	Typical duration in minutes
rvu_work	DECIMAL(6,2)	YES	Work RVU
rvu_practice	DECIMAL(6,2)	YES	Practice expense RVU
rvu_malpractice	DECIMAL(6,2)	YES	Malpractice RVU

DIM_MEDICATION

Medication dimension covering diabetes therapies, supplies, and related pharmaceuticals.

Column Name	Data Type	Nullable	Description
medication_key	BIGINT	NO	Surrogate key (PK)
ndc_code	VARCHAR(15)	NO	National Drug Code
medication_name	VARCHAR(200)	NO	Brand or generic name
generic_name	VARCHAR(200)	YES	Generic equivalent name
drug_class	VARCHAR(100)	NO	Therapeutic class
is_insulin	BOOLEAN	NO	Insulin product flag
insulin_type	VARCHAR(50)	YES	Rapid/Short/Intermediate/Long
is_diabetes_supply	BOOLEAN	NO	Diabetes supply (strips, lancets)
requires_prior_auth	BOOLEAN	NO	Prior authorization required
awp_unit_price	DECIMAL(10,4)	YES	Average wholesale price
route_of_admin	VARCHAR(50)	YES	Oral/Injection/Inhalation

DIM_INSURANCE

Insurance/payer dimension for coverage and reimbursement analysis.

Column Name	Data Type	Nullable	Description
insurance_key	BIGINT	NO	Surrogate key (PK)
payer_id	VARCHAR(20)	NO	Payer identifier
payer_name	VARCHAR(200)	NO	Insurance company name
plan_type	VARCHAR(50)	NO	HMO/PPO/EPO/POS/HDHP/Medicare/Medicaid
plan_name	VARCHAR(200)	YES	Specific plan name
coverage_tier	VARCHAR(50)	YES	Bronze/Silver/Gold/Platinum
is_medicare	BOOLEAN	NO	Medicare plan flag
is_medicaid	BOOLEAN	NO	Medicaid plan flag
is_commercial	BOOLEAN	NO	Commercial insurance flag
diabetes_coverage_level	VARCHAR(20)	YES	Basic/Enhanced/Comprehensive

FACT TABLES

FACT_ENCOUNTER

Primary encounter fact table capturing all patient visits and associated costs.

Column Name	Data Type	Nullable	Description
encounter_key	BIGINT	NO	Surrogate key (PK)
encounter_id	VARCHAR(20)	NO	Source encounter ID
patient_key	BIGINT	NO	FK to DIM_PATIENT
provider_key	BIGINT	NO	FK to DIM_PROVIDER
facility_key	BIGINT	NO	FK to DIM_FACILITY
service_date_key	INT	NO	FK to DIM_DATE (service)
admission_date_key	INT	YES	FK to DIM_DATE (admission)
discharge_date_key	INT	YES	FK to DIM_DATE (discharge)
encounter_type	VARCHAR(20)	NO	Inpatient/Outpatient/ED/Telehealth
primary_diagnosis_key	BIGINT	NO	FK to DIM_DIAGNOSIS
drg_code	VARCHAR(10)	YES	DRG for inpatient
length_of_stay	INT	YES	Days (inpatient only)
total_charges	DECIMAL(12,2)	NO	Total billed charges
allowed_amount	DECIMAL(12,2)	YES	Payer allowed amount
paid_amount	DECIMAL(12,2)	YES	Amount paid by payer
patient_responsibility	DECIMAL(12,2)	YES	Patient copay/coinsurance
adjustment_amount	DECIMAL(12,2)	YES	Write-offs and adjustments

FACT_CLAIM

Claims fact table linking to expenditure transactions for cost analysis. The transaction_id column maps to the Excel spreadsheet Transaction_ID.

Column Name	Data Type	Nullable	Description
claim_key	BIGINT	NO	Surrogate key (PK)
claim_id	VARCHAR(30)	NO	Source claim number
transaction_id	VARCHAR(20)	NO	Maps to Excel Transaction_ID
encounter_key	BIGINT	YES	FK to FACT_ENCOUNTER
patient_key	BIGINT	NO	FK to DIM_PATIENT
provider_key	BIGINT	NO	FK to DIM_PROVIDER
facility_key	BIGINT	NO	FK to DIM_FACILITY
insurance_key	BIGINT	NO	FK to DIM_INSURANCE
service_date_key	INT	NO	FK to DIM_DATE

claim_type	VARCHAR(20)	NO	Professional/Institutional/Pharmacy
claim_status	VARCHAR(20)	NO	Paid/Pending/Denied/Appealed
billed_amount	DECIMAL(12,2)	NO	Total billed
allowed_amount	DECIMAL(12,2)	YES	Contracted rate
paid_amount	DECIMAL(12,2)	YES	Paid to provider
copay_amount	DECIMAL(10,2)	YES	Patient copay
coinsurance_amount	DECIMAL(10,2)	YES	Patient coinsurance
deductible_amount	DECIMAL(10,2)	YES	Applied to deductible

FACT_PRESCRIPTION

Pharmacy claims and prescription activity for medication cost tracking.

Column Name	Data Type	Nullable	Description
prescription_key	BIGINT	NO	Surrogate key (PK)
rx_number	VARCHAR(20)	NO	Prescription number
patient_key	BIGINT	NO	FK to DIM_PATIENT
provider_key	BIGINT	NO	FK to prescribing provider
pharmacy_key	BIGINT	NO	FK to DIM_FACILITY (pharmacy)
medication_key	BIGINT	NO	FK to DIM_MEDICATION
fill_date_key	INT	NO	FK to DIM_DATE
quantity_dispensed	DECIMAL(10,2)	NO	Units dispensed
days_supply	INT	NO	Days supply
refill_number	INT	NO	Refill count (0=new)
ingredient_cost	DECIMAL(10,2)	NO	Drug ingredient cost
dispensing_fee	DECIMAL(8,2)	NO	Pharmacy dispensing fee
patient_pay	DECIMAL(10,2)	YES	Patient out-of-pocket
plan_pay	DECIMAL(10,2)	YES	Insurance payment
is_specialty	BOOLEAN	NO	Specialty pharmacy flag
is_mail_order	BOOLEAN	NO	Mail order flag

FACT_LAB_RESULT

Laboratory results fact table for clinical outcomes and quality metrics.

Column Name	Data Type	Nullable	Description
lab_result_key	BIGINT	NO	Surrogate key (PK)
order_id	VARCHAR(30)	NO	Lab order number
patient_key	BIGINT	NO	FK to DIM_PATIENT
ordering_provider_key	BIGINT	NO	FK to DIM_PROVIDER
lab_facility_key	BIGINT	NO	FK to DIM_FACILITY
collection_date_key	INT	NO	FK to DIM_DATE
result_date_key	INT	NO	FK to DIM_DATE
loinc_code	VARCHAR(20)	NO	LOINC test code
test_name	VARCHAR(200)	NO	Test description
result_value	DECIMAL(12,4)	YES	Numeric result
result_text	VARCHAR(500)	YES	Text/qualitative result
result_unit	VARCHAR(50)	YES	Unit of measure
reference_low	DECIMAL(12,4)	YES	Normal range low

reference_high	DECIMAL(12,4)	YES	Normal range high
abnormal_flag	CHAR(1)	YES	H/L/A/N (High/Low/Abnormal/Normal)
is_hba1c	BOOLEAN	NO	HbA1c test flag
is_glucose	BOOLEAN	NO	Glucose test flag

BRIDGE & MAPPING TABLES

BRIDGE_PATIENT_TRANSACTION

Critical mapping table that links Patient_ID from the semantic layer to Transaction_ID in the Excel expenditure data. This enables joining EHR data with financial transactions.

Column Name	Data Type	Nullable	Description
mapping_key	BIGINT	NO	Surrogate key (PK)
mapping_id	VARCHAR(20)	NO	Natural mapping key
patient_key	BIGINT	NO	FK to DIM_PATIENT
patient_id	VARCHAR(20)	NO	Source patient ID (maps to Excel)
transaction_id	VARCHAR(20)	NO	Excel Transaction_ID reference
encounter_id	VARCHAR(20)	YES	Source encounter ID
claim_key	BIGINT	YES	FK to FACT_CLAIM
mapping_date	DATE	NO	Date mapping was created
data_source	VARCHAR(50)	NO	EHR_Import/Claims_Feed/Manual
is_validated	BOOLEAN	NO	Validation status
last_updated	TIMESTAMP	NO	Last modification timestamp

BRIDGE_ENCOUNTER_DIAGNOSIS

Many-to-many bridge for encounters with multiple diagnoses.

Column Name	Data Type	Nullable	Description
encounter_key	BIGINT	NO	FK to FACT_ENCOUNTER (composite PK)
diagnosis_key	BIGINT	NO	FK to DIM_DIAGNOSIS (composite PK)
diagnosis_sequence	INT	NO	1=Primary, 2+=Secondary
present_on_admission	CHAR(1)	YES	Y/N/U/W POA indicator

BRIDGE_ENCOUNTER_PROCEDURE

Many-to-many bridge for encounters with multiple procedures.

Column Name	Data Type	Nullable	Description
encounter_key	BIGINT	NO	FK to FACT_ENCOUNTER (composite PK)
procedure_key	BIGINT	NO	FK to DIM_PROCEDURE (composite PK)
procedure_sequence	INT	NO	Procedure order
modifier_1	VARCHAR(5)	YES	CPT modifier 1
modifier_2	VARCHAR(5)	YES	CPT modifier 2
units	DECIMAL(8,2)	NO	Service units

line_charge	DECIMAL(12,2)	NO	Line item charge

ENTITY RELATIONSHIPS

The following describes the key relationships in the schema:

Relationship	Cardinality	Join Columns
DIM_PATIENT → FACT_ENCOUNTER	1:N	patient_key
DIM_PATIENT → FACT_CLAIM	1:N	patient_key
DIM_PATIENT → FACT_PRESCRIPTION	1:N	patient_key
DIM_PATIENT → FACT_LAB_RESULT	1:N	patient_key
DIM_PROVIDER → FACT_ENCOUNTER	1:N	provider_key
DIM_FACILITY → FACT_ENCOUNTER	1:N	facility_key
DIM_DATE → FACT_* tables	1:N	date_key variations
DIM_DIAGNOSIS ↔ FACT_ENCOUNTER	M:N	via BRIDGE_ENCOUNTER_DIAGNOSIS
DIM_PROCEDURE ↔ FACT_ENCOUNTER	M:N	via BRIDGE_ENCOUNTER_PROCEDURE
DIM_PATIENT ↔ Excel Transactions	M:N	via BRIDGE_PATIENT_TRANSACTION
DIM_INSURANCE → DIM_PATIENT	1:N	insurance_key
FACT_CLAIM → Excel (Transaction_ID)	1:1	transaction_id

DATA INTEGRATION NOTES

Linking EHR to Expenditure Data:

The semantic layer connects to the Excel expenditure spreadsheet through the BRIDGE_PATIENT_TRANSACTION table. The primary join path is:

1. **DIM_PATIENT.patient_id → BRIDGE_PATIENT_TRANSACTION.patient_id**
2. **BRIDGE_PATIENT_TRANSACTION.transaction_id → Excel.Transaction_ID**

This enables cost analysis queries that combine clinical data (diagnoses, procedures, lab results) with financial data (charges, payments, adjustments) for comprehensive patient journey cost modeling.

Sample Join Query:

```
SELECT p.patient_id, p.diabetes_type, e.encounter_type,
    SUM(excel.Amount_USD) as total_cost
FROM DIM_PATIENT p
JOIN BRIDGE_PATIENT_TRANSACTION bpt ON p.patient_id = bpt.patient_id
JOIN [Excel_Expenditure] excel ON bpt.transaction_id = excel.Transaction_ID
LEFT JOIN FACT_ENCOUNTER e ON p.patient_key = e.patient_key
GROUP BY p.patient_id, p.diabetes_type, e.encounter_type
```