

Oncology Encounter Star Schema – Explanation

A practical guide to the fact and dimensions in the uploaded star schema diagram.

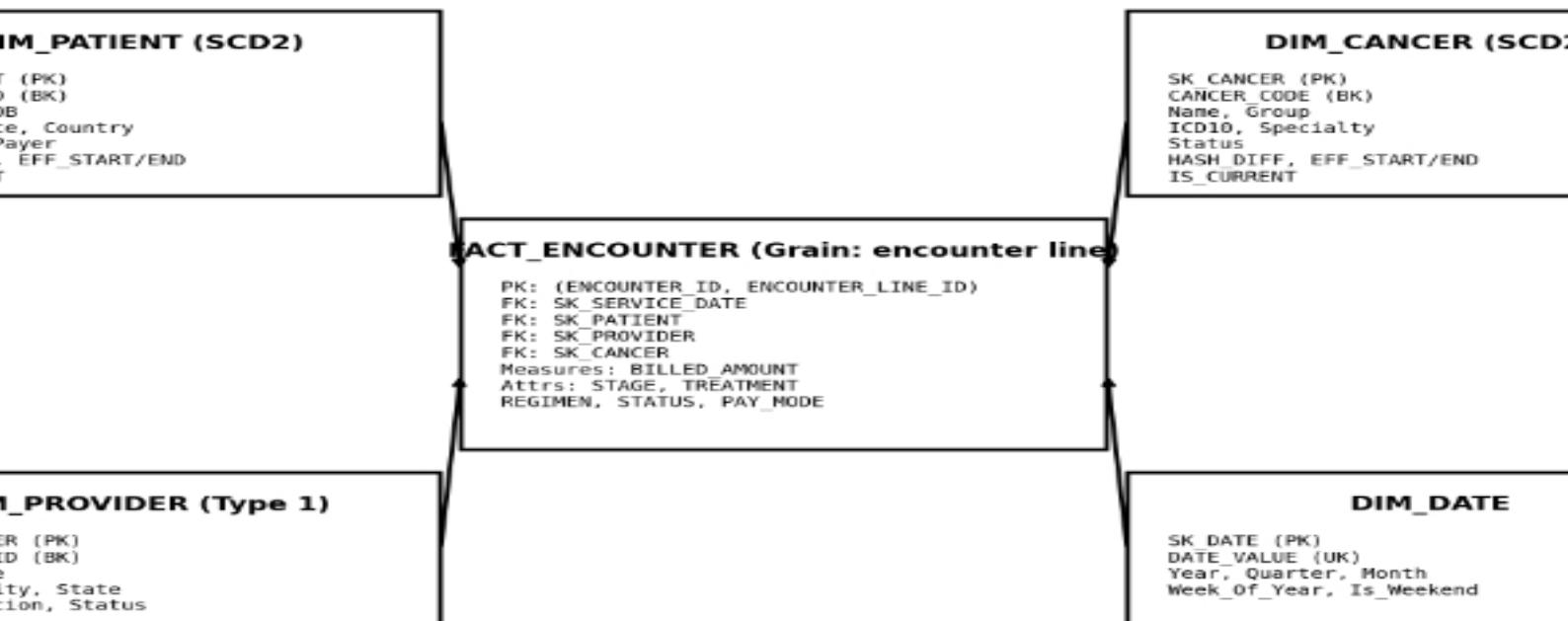


Figure: Cancer Encounter Star Schema (as provided).

1) What this model is about (business meaning)

The model stores oncology service activity. Each row in the central fact table represents one line within a patient encounter (i.e., a billed service line for a visit). You can analyze billed money and operational attributes by Patient, Cancer classification, Provider, and Date.

2) Center of the star — FACT_ENCOUNTER

Grain (level of detail): encounter line. That means each record is a specific service line within an encounter (visit), not a patient or a whole encounter summary.

- PK: (ENCOUNTER_ID, ENCOUNTER_LINE_ID) – unique line within an encounter.
- FKs: SK_SERVICE_DATE, SK_PATIENT, SK_PROVIDER, SK_CANCER – point to the dimension tables for consistent slicing.

Measure:

- BILLED_AMOUNT – the main numeric fact to aggregate (SUM for revenue-style reports).

Fact-level attributes:

- STAGE, TREATMENT_REGIMEN, STATUS, PAY_MODE – present on the fact because they describe the service line at that time.
- If cardinality stays low and you frequently filter by them, keep as degenerate attributes, move to a small Junk Dimension, or promote to separate conformed dimensions if reused across facts.

3) The dimensions (the points of the star)

a) DIM_PATIENT (SCD2)

- Keys: SK_PATIENT (PK), PATIENT_ID (BK)
- Attributes: Gender, DOB, City/State/Country, Segment, Payer
- SCD2 columns: HASH_DIFF, EFF_START, EFF_END, IS_CURRENT — when patient details change (e.g., city, payer, segment), create a new row with a new SK to preserve history.

b) DIM_CANCER (SCD2)

- Keys: SK_CANCER (PK), CANCER_CODE (BK)
- Attributes: Name, Group, ICD10, Specialty, Status
- SCD2 columns: HASH_DIFF, EFF_START, EFF_END, IS_CURRENT — coding/grouping changes are versioned to keep historically accurate reporting.

c) DIM_PROVIDER (Type 1)

- Keys: SK_PROVIDER (PK), PROVIDER_ID (BK)
- Attributes: Name, Type, Region, City, State, Accreditation, Status
- Change type: Type 1 (overwrite) — provider attributes are updated in place (no history). Simpler, but loses historical provider attribute context.

d) DIM_DATE

- Keys: SK_DATE (PK), DATE_VALUE (UK)
- Attributes: Year, Quarter, Month, WeekOfYear, Is_Weekend — standard calendar dimension for grouping and time drill-down.

Note on keys:

BK (Business Key) is the natural identifier from the source (e.g., PATIENT_ID). SK (Surrogate Key) is a warehouse-generated integer used for joins and SCD history.

4) How SCD2 affects reporting

Because Patient and Cancer are SCD2, the ETL should resolve the correct surrogate key that was valid at the time of service (usually by SERVICE_DATE). This enables as-of analytics, e.g., reporting by the patient segment that was true when the service occurred.

5) Typical queries you can run

1) Revenue trend by cancer group

```
SELECT
    d.Year,
    d.Month,
    c."Group"      AS Cancer_Group,
    SUM(f.BILLED_AMOUNT) AS Total_Billed
FROM FACT_ENCOUNTER f
JOIN DIM_DATE      d ON f.SK_SERVICE_DATE = d.SK_DATE
JOIN DIM_CANCER    c ON f.SK_CANCER       = c.SK_CANCER
GROUP BY d.Year, d.Month, c."Group"
ORDER BY d.Year, d.Month, c."Group";
```

2) Provider performance by region & specialty

```

SELECT
    p.Region,
    c.Specialty,
    COUNT(DISTINCT f.ENCOUNTER_ID) AS Encounters,
    SUM(f.BILLED_AMOUNT)           AS Billed
FROM FACT_ENCOUNTER f
JOIN DIM_PROVIDER p ON f.SK_PROVIDER = p.SK_PROVIDER
JOIN DIM_CANCER   c ON f.SK_CANCER   = c.SK_CANCER
GROUP BY p.Region, c.Specialty;

```

3) Patient segment impact, preserving history

```

SELECT
    d.Year,
    pat.Segment,
    SUM(f.BILLED_AMOUNT) AS Billed
FROM FACT_ENCOUNTER f
JOIN DIM_DATE      d  ON f.SK_SERVICE_DATE = d.SK_DATE
JOIN DIM_PATIENT  pat ON f.SK_PATIENT     = pat.SK_PATIENT
GROUP BY d.Year, pat.Segment
ORDER BY d.Year, pat.Segment;

```

6) ETL & data modeling notes

Dimension loading

- For SCD2 (DIM_PATIENT, DIM_CANCER): detect changes via HASH_DIFF; end-date old row (EFF_END) and insert a new row with IS_CURRENT = 1.
- For Type 1 (DIM_PROVIDER): overwrite attributes; SK_PROVIDER remains constant.

Fact loading

- Assign SK_PATIENT, SK_CANCER, SK_PROVIDER, SK_SERVICE_DATE during ETL. For SCD2 dims, pick the version valid on SERVICE_DATE.
- Keep ENCOUNTER_ID and ENCOUNTER_LINE_ID as degenerate identifiers for drill-through and reconciliation.

Integrity checks

- All foreign keys resolve to existing dimension rows (or to a defined unknown member).
- SCD2 ranges do not overlap per business key (EFF_START/EFF_END).
- BILLED_AMOUNT is non-negative (or exceptions are flagged).
- Date keys map to valid calendar rows.

7) When you might extend/refine this model

- Multiple dates: If you need Admit Date, Discharge Date, Post Date, add multiple role-playing DIM_DATE keys.
- Payment collections: Create a separate FACT_PAYMENT linked to Claim/Encounter to analyze cash vs. billed.
- High-cardinality attributes: If STAGE or TREATMENT_REGIMEN are reused across facts, promote them to conformed dimensions.
- Many-to-many (e.g., patient with multiple concurrent cancers): use a bridge table to attribute measures accurately.
- Privacy & PHI: Exclude direct identifiers or tokenize them to comply with privacy policies.

8) How to read it quickly (cheat sheet)

- Star center = FACT_ENCOUNTER (money and line-level activity).
- Points = Dimensions: Patient (SCD2), Cancer (SCD2), Provider (Type 1), Date.
- Grain = encounter line → aggregate before mixing with encounter-level facts.
- SCD2 enables as of reporting; Type 1 always shows latest provider attributes.
- Degenerate attributes on the fact (stage, regimen, status, pay mode) can remain or be dimensionized later.