

# Star Schema Activity and Examination of Administrative Claims Data

UC Davis Continuing and Professional Education

July 10, 2024

## Part 1: The Star Schema Activity

### Step 1: Review the SQL Used to Create the Date Dimension Table (DATE\_DIM)

The provided DATE\_DIM table SQL creation script creates a table with one row for every day from January 3, 2000, to December 31, 2015. This table contains multiple columns that describe each date in various ways, such as day of the week, month, and year, which allows for efficient querying and analysis.

#### DATE\_DIM Table Creation Script

```
CREATE TABLE Date_Dim(  
    Date_Key number(18,0) NOT NULL,  
    Date_Value Date NOT NULL,  
    my_Day Char(10),  
    Day_Of_Week number(18,0),  
    Day_Of_Month number(18,0),  
    Day_Of_Year number(18,0),  
    Week_Of_Year number(18,0),  
    my_Month Char(10),  
    Month_Of_Year number(18,0),  
    Quarter_Of_Year number(18,0),  
    my_Year number(18,0)  
);
```

The table is populated using the following SQL script:

```
INSERT INTO Date_Dim  
SELECT  
    to_number(to_char(CurrDate, 'YYYYMMDD')) as Date_Key,  
    CurrDate AS Date_Value,  
    TO_CHAR(CurrDate, 'Day') as Day,  
    to_number(TO_CHAR(CurrDate, 'D')) AS Day_Of_Week,  
    to_number(TO_CHAR(CurrDate, 'DD')) AS Day_Of_Month,  
    to_number(TO_CHAR(CurrDate, 'DDD')) AS Day_Of_Year,  
    to_number(TO_CHAR(CurrDate+1, 'IW')) AS Week_Of_Year,  
    TO_CHAR(CurrDate, 'Month') AS my_Month,  
    to_number(TO_CHAR(CurrDate, 'MM')) AS Month_of_Year,  
    to_number((TO_CHAR(CurrDate, 'Q')) AS Quarter_Of_Year,  
    to_number(TO_CHAR(CurrDate, 'YYYY')) AS my_Year  
FROM  
    (select level n, TO_DATE('01-02-2000', 'MM-DD-YYYY') + NUMTODSINTERVAL(level, 'DAY') CurrDate  
    from dual
```

```
connect by level <= 5842);  
COMMIT;
```

The purpose of the DATE\_DIM table is to provide a single interface to query date-related information efficiently. It reduces the need for repetitive date calculations across multiple fact tables by storing all relevant date attributes in one place.

## Step 2: Compare Standard Queries vs. Those Written for a Star Schema

Standard Query:

```
SELECT *  
FROM encounters_fact  
WHERE trunc(enc_start_datetime) >= to_date('01-01-2008','MM-DD-YYYY')  
AND trunc(enc_start_datetime) < to_date('01-01-2009','MM-DD-YYYY');
```

Star Schema Query:

```
SELECT *  
FROM encounters_fact enc  
JOIN date_dim ddim ON enc.date_key = ddim.date_key  
WHERE ddim.my_year = 2008;
```

The standard query directly references the date field in the `encounters_fact` table. The star schema query uses the DATE\_DIM table to achieve the same result. The star schema query is generally faster because it joins with the DATE\_DIM table, which has fewer rows to scan for date attributes compared to scanning all rows in the `encounters_fact` table directly.

## Comparison and Contrast: Standard Relational Model vs. Star Schema

- **Performance:** Star schema queries can be faster due to the separation of dimension data into smaller, indexed tables.
- **Complexity:** Star schema design may be more complex to set up but simplifies querying.
- **Use Case:** Star schema is ideal for analytical processing in data warehouses, whereas the standard relational model is more suited for transactional systems.

## Application to Different Scenarios

- **Data Warehousing:** Star schemas are particularly useful for data warehousing applications where large volumes of data need to be aggregated and analyzed efficiently.
- **Transactional Systems:** Standard relational models are better suited for transactional systems where data integrity and normalization are paramount.

## Part 2: Examining Administrative Claims Data

### Accessing and Examining CMS 2008-2010 Data Entrepreneurs Synthetic Public Use File (DE-SynPUF)

The DE-SynPUF dataset contains synthetic Medicare claims data, including:

- Beneficiary Summary
- Inpatient Claims
- Outpatient Claims

- Carrier Claims
- Prescription Drug Events

This data provides a realistic set of claims data for querying and exploring healthcare data. Although synthetic, it retains many features of original Medicare claims data, offering an opportunity to work with large datasets and derive meaningful insights.

## Conclusion

This exercise demonstrates the importance of star schema design in data warehouses and its advantages over traditional relational models for analytical queries. By examining both approaches and understanding their applications, we can better design data systems to meet the needs of different healthcare data scenarios.