Dimensional Data Model - Retail Sales Data Warehouse

# 1. Purpose & Goals

Build a small, analytic Retail Sales Data Warehouse to support reporting and analysis of retail transactions (sales), customer behavior, product performance, and location & payment trends. The DWH will:  
- Store cleansed, conformed dimensional data derived from CSVs in sales\_db.  
- Provide aggregated and drillable KPIs (daily/weekly/monthly).  
- Maintain customer history (SCD pattern).  
- Be performant for BI queries (sales by date, product, category, city/state, payment type, customer cohorting).

# 2. Source Data (summary)

Databases:  
- sales\_db: operational/source (CSV dumps).  
- sales\_stg: staging area.  
- sales\_dwh: curated data warehouse.  
  
custs: custid, fname, lname, age, profession  
txns: txnid, txndate, custid, amount, productname, categoryname, city, state, paymenttype

# 3. Business Questions / Key Use Cases

- Daily / monthly revenue and transaction counts  
- Top selling products and categories  
- Sales by city/state  
- Payment method analysis  
- Customer behavior (repeat, LTV, basket size)  
- Cohort analysis  
  
KPIs: Total Sales, Transaction Count, Avg Transaction Value, Unique Customers, Repeat Rate, Sales by Product/Category/City/State/PaymentType/Date

# 4. Dimensional Modeling — Design Decisions

Grain: One row per transaction.  
  
Dimensions:  
- dim\_date  
- dim\_customer (SCD Type 2)  
- dim\_product (SCD Type 1)  
- dim\_location  
- dim\_payment  
  
Fact:  
- fact\_sales with FKs to all dims and measure 'amount'.

# 5. ETL / ELT Flow

1. Load raw CSVs into sales\_stg.  
2. Cleanse & standardize data.  
3. Lookup/insert into dimensions.  
4. Load fact\_sales.  
5. Perform reconciliation & DQ checks.

# 6. Data Quality Rules

- txnid not null, unique  
- txndate parseable  
- custid exists or map to unknown  
- amount numeric > 0  
- productname/categoryname not null

# 7. ER Diagram

