## Telecom Customer Churn Analytics Project

**Tools:** Spark · PostgreSQL

### **1. Raw Data Ingestion**

* **Dataset:** [IBM Telco Customer Churn](https://www.kaggle.com/datasets/blastchar/telco-customer-churn/data)
* **Source File:** WA\_Fn-UseC\_-Telco-Customer-Churn.csv
* **Initial Load:** Data ingested into PostgreSQL as telco\_customer\_churn (raw layer).

### **2. Silver Layer (Cleaned & Standardized Data)**

**Source table:** silver\_telco\_customer\_churn  
**Goal:** Prepare clean, consistent, and analytics-ready data for modeling.

**Steps Performed:**

* Removed invalid or malformed rows (e.g., blank TotalCharges).
* Dropped duplicate records based on customerid.
* Converted numeric fields (TotalCharges, MonthlyCharges) to FLOAT.
* Cast SeniorCitizen to BOOLEAN.
* Standardized Churn to boolean (Yes → TRUE, No → FALSE).
* Trimmed whitespace from string fields.
* Handled missing values via null replacement or filtering.
* **Slowly Changing Dimensions (SCD) Handling:**
  + Identified SCD attributes (e.g., Contract, PaymentMethod).
  + Introduced surrogate keys using snapshots.
  + Added is\_current, effective\_from, effective\_to metadata.
  + Preserved full change history.

**Output:** Clean, enriched table feeding into dimensions and fact tables.

### **3. Dimensional Modeling (Star Schema with Date Dimension)**

#### **Fact Table**

**fact\_churn\_events**  
Columns:

* customerid
* contract
* internetservice
* monthlycharges
* totalcharges
* tenure
* churn
* date\_id

#### **Dimensions**

* **dim\_customer**
  + customerid (PK), gender, seniorcitizen, partner, dependents
* **dim\_service**
  + customerid (FK), phoneservice, multiplelines, internetservice, onlinesecurity, onlinebackup, deviceprotection, techsupport, streamingtv, streamingmovies
* **dim\_billing**
  + customerid (FK), contract, paperlessbilling, paymentmethod, monthlycharges, totalcharges
* **dim\_date**
  + date\_id (PK), date, year, month, quarter

### **4. Business Questions Answered**

From this star schema, we can answer key business questions:

1. **Which contract type has the highest churn rate?**
2. **Does churn rate differ between customers with and without tech support?**
3. **What is the churn rate trend over the past months or quarters?**
4. **Do customers with paperless billing churn more than those without?**
5. **Which internet service type generates the highest average monthly revenue from churned customers?**
6. **What is the average tenure of churned customers by contract type?**