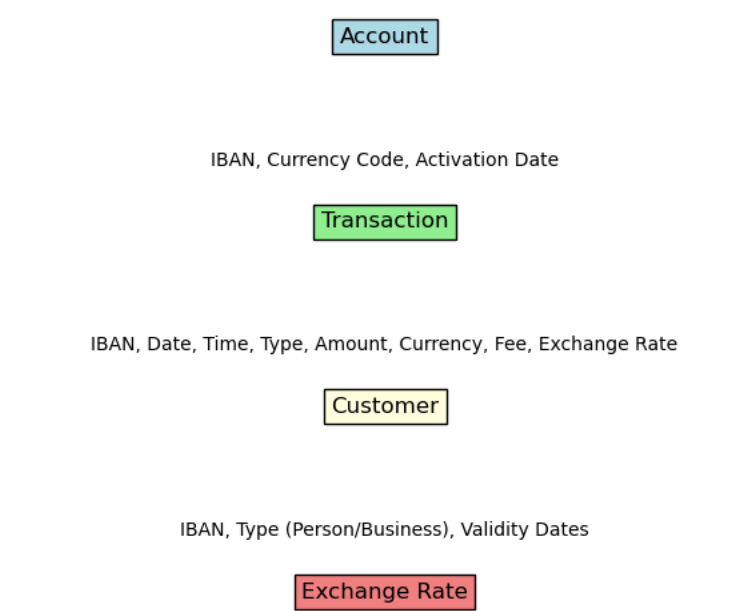# SOF Data Pipeline Report

1. UML-Style Data Model

This diagram shows the key entities and their relationships in the SOF data pipeline:

Account

IBAN, Currency Code, Activation Date

Transaction

IBAN, Date, Time, Type, Amount, Currency, Fee, Exchange Rate

Customer

IBAN, Type (Person/Business), Validity Dates

Exchange Rate

2. Source-to-Target Mapping

This table maps the source datasets and fields to the final output fields used in the fee statement:

Target Field	Source Dataset	Source Field(s)
iban	dc_trx_clrg_issng_df, bsc_pymt_agrm_df, rel_x_ar_hist_df	CARDHDR_IBAN, IBAN, AR_AC_IBAN
sactiongroup	Derived	Based on CARD_TP_CODE, CARD_TXN_ATM_POS_
sactiondate	dc_trx_clrg_issng_df	CARD_TXN_LCL_DT
sactioncount	Derived	Count of transactions per group
ansactioncount	Derived	Based on fee applicability
rinterestrate	dc_trx_clrg_issng_df	CLRG_MU_TRFF_PCT
stcurrencysuamount	Derived	Based on fee calculations
restcurrencycode	Static or derived	Usually EUR

erestbasecode	Static	1 or 2 depending on fee type

3. Architecture Diagram

This layered diagram illustrates the architecture of the SOF pipeline, from ingestion to orchestration:

Ingestion Layer

Sources: ART, COK, POD, Timescape, Siebel, Cross Border
Mounts: Azure Data Lake containers mounted via DBFS

Data Preparation Layer

Filtering: Based on date ranges (April 1 – March 31)
Column Selection: Only required fields are selected
Data Cleaning: Handling nulls, formatting dates, etc.

Transformation Layer

Business Logic: SQL queries for each transaction group (9101, 9107, etc.)
Joins: With customer and account data
Calculations: Fees, exchange rate surcharges, etc.

Aggregation & Output Layer