

Database Level

defaults:

subset
strategy: enum (sample, all, fk_closure)
max_rows: int
time_window_days: int
masking:
deterministic_salt: string (e.g.- "CHANGE-ME")
email_domain: string (e.g.- "dev.local")
preserve_length: bool

Table Level

tables:

[table_name]:
subset:
strategy: enum (sample, all, fk_closure)
time_window_days: int
root: bool
order_by: string "created_at DESC"
max_rows: int
fk: (only if strategy = fk_closure)
references:
- column: string [column name from this table]
table: string [parent table name]
column_ref: string [parent column name]
processing_order: int
post_load:
sequence_offset: int (e.g.- 1000000)

Column Level

columns:

[column_name]:
references: string (parent_table_name.parent_column_name)
sensitivity: enum (public, internal, pii, secret)
strategy: enum (keep, mask, synthesize)
primary_key: bool
mask: (only if strategy = mask)
type: enum (username, category_map, numeric_noise)
deterministic_key: id
percent: int (only if type = numeric_noise)
preserve_domain: bool (only if type = category_map)
synth: (only if strategy = synthesize)
type: enum (email, password_hash, credit_card, address)
deterministic_key: id
domain: dev.local (only if type = email)
format: enum (last4-only) (only if type = credit_card)
unique: bool