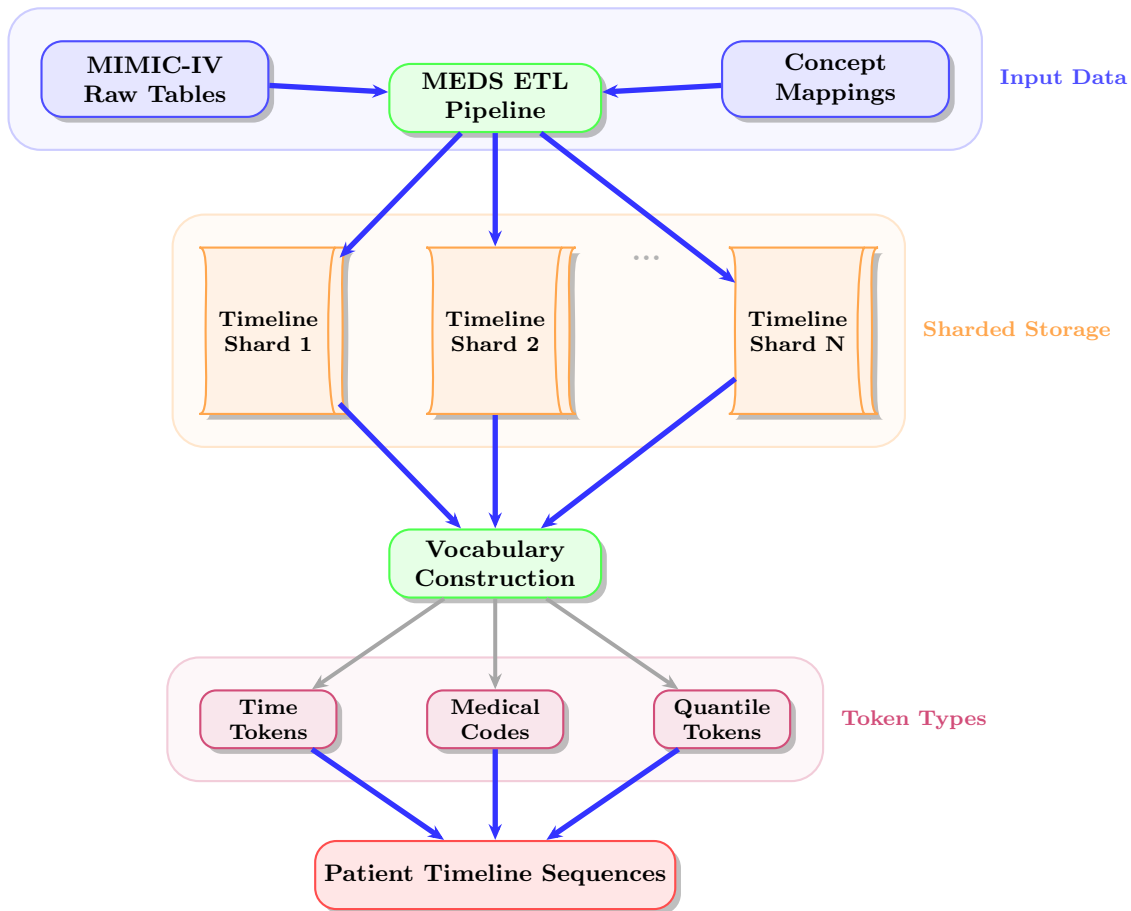# MIMIC-IV to MEDS Pipeline

Extract & Transform
- Longitudinal events
- LOINC/RxNorm mapping
- MEDS standardization

Input Data

MIMIC-IV Raw Tables → MEDS ETL Pipeline ← Concept Mappings

Sharded Storage

Timeline Shard 1 ⋯ Timeline Shard 2 ⋯ Timeline Shard N

Tokenization
- Code frequencies
- Time intervals
- Statistical binning

Vocabulary Construction

Token Types

Time Tokens | Medical Codes | Quantile Tokens

Patient Timeline Sequences