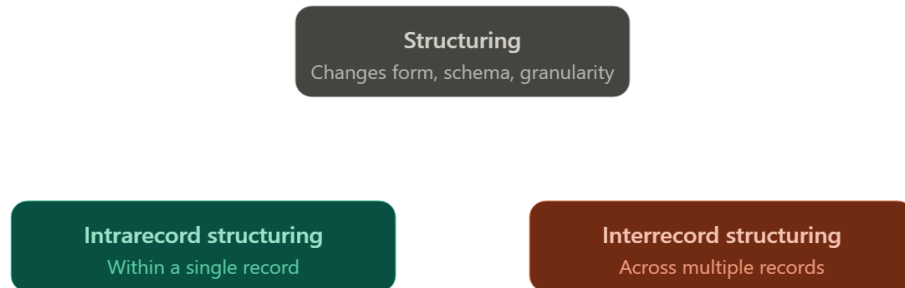


# **TRANSFORMING: STRUCTURING**

UDSC-402

Rajesh Kumar Sahoo, III B.Sc.(Mathematics),  
24010107007



It outlines the fundamental processes of data structuring, a transformation phase used to modify the form or granularity of a dataset. It is categorized into intrarecord structuring, which focuses on manipulating individual fields through techniques like extracting substrings or merging columns, and interrecord structuring, which involves broader shifts in data organization. Specific methods such as positional and pattern-based extraction are explained for isolating specific data values, while filtering and aggregation are introduced as ways to refine dataset focus. Finally, it details how pivots can reshape information by converting rows to columns or vice versa to prepare data for analytical tools.

## 1. Intrarecord *(within one record)*

- **Reorder fields** — rearrange columns, values unchanged
- **Extract values** — pull a substring into a new field
  - Simple positional (fixed start/end, e.g. day from a date)
  - Complex positional (delimiter-based, e.g. split "LAST, FIRST")
  - Pattern-based (matching a general pattern, e.g. order id which start with ORD-some number)
  - Complex structure (pull elements out of JSON maps/arrays)
- **Combine fields** — merge fields into one (reverse of extraction), e.g. City + State → "MCIPHERSON, KS"

## 2. Interrecord *(across many records)*

- **Filter** — remove records (row-based) or columns (field-based) to change granularity
- **Aggregate & pivot** — reshape granularity
  - Simple aggregation: many records → one (sum, mean, count)
  - Column-to-row pivot ("unpivot"): one record → many
  - Row-to-column pivot: many-to-many, new column per category