Extract (E)

Azure Data Lake Gen2 (Raw)

Databricks Auto Loader

Parsowanie JSON
+ schemat z schema_config.yaml

Transformacje (select, withColumn)

Walidacja (review_text IS NOT NULL)

Usuwanie duplikatów

Zapis do Table lub pliku Delta