

Course 2 Module 5 Programming Assignment

Assignment is to ETL MIMIC data into the OMOP CONDITION_OCCURRENCE table

Detailed instructions with Slide Notes

Assignment is to ETL MIMIC data into the OMOP CONDITION_OCCURRENCE table

ETL Steps

- 1. Understand source/target data models
- 2. Profile source File display
- 3. Create ETL mappings
- 4. Write transformation code
- 5. Execute transformation
- 6. Perform data quality assessment
- 7. Package documentation

Sten 1: Understand source/target data models

CONDITION_OCCURRENCE is the TARGET OMOP table.

Read the OMOP documentation about the type of data stored in CONDITION_OCCURRENCE and for three fields below that are in that table:

- person id
- visit_occurrence_id
- · condition source value

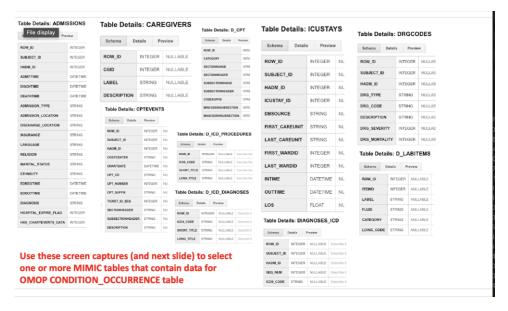


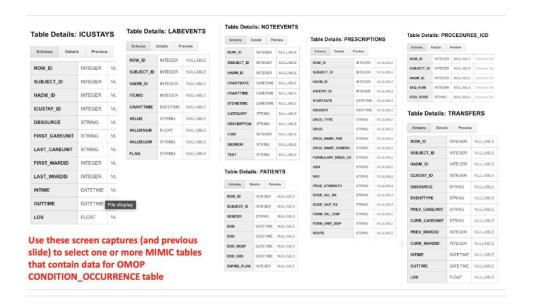
File displayed 1: Understand source/target data models

CONDITION_OCCURRENCE is the TARGET OMOP table.

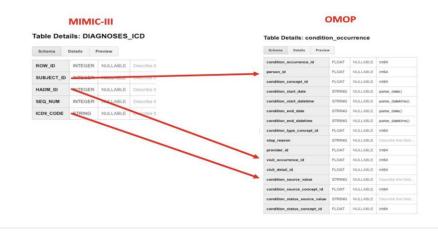
Select one or more MIMIC tables from the table screen shots on the next slides that you feel are most related to the three fields in CONDITION_OCCURRENCE.







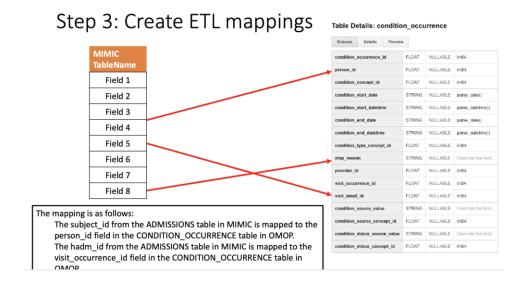
Step 1: Understand source/target data models



Step 2: Profile source table or tables

Using the White Rabbit profiling data from the 100 patient MIMIC database provided in the Assessment to comment on the distribution of the SUBJECT_ID field from one of the MIMIC tables selected in Step 1

- DIAGNOSES_ICD
 - SUBJECT_ID is a heavily right-skewed distribution, with just SUBJECT_ID 41976 accounting for about 15% of the available data in the report. A similar amount of records can be obtained with the next 6 most common patients, on which there exist 2–3% of records each. The rest of the patients are more uniformly represented in the table.



Step 4: Write transformation code

WITH CONDITION_OCCURRENCE1 as (select adm.subject_id as person_id, adm.hadm_id as visit_occurrence_id from mimic3_demo.ADMISSIONS adm),

CONDITION_OCCURRENCE as (select

oc1.person_id, oc1.visit_occurrence_id, diag.icd9_code as condition_source_value from CONDITION_OCCURRENCE1 as oc1 join mimic3_demo.DIAGNOSES_ICD as diag

on oc1.visit_occurrence_id = diag.hadm_id)

select * from CONDITION_OCCURRENCE as oc order by oc.person_id, oc.visit_occurrence_id;

Paste the SQL statements that transform data from one or more MIMIC tables into the three OMOP CONDITION_OCCURRENCE fields (patient-id, visit_occurrence_id, condition_source_value) into the Coursera Submission Site

Transformation code shown here is from the Course 2 videos showing transformation of MIMIC PATIENTS to OMOP PERSON