

File display


## 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 
3. Create ETL mappings
4. Write transformation code
5. Execute transformation
6. Perform data quality assessment
7. Package documentation

### Step 1: Understand source/target data models

File display

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

Table Details: condition\_occurrence

Schema	Details	Preview
condition_occurrence_id	Float	Nullable
person_id	Float	Nullable
condition_concept_id	Float	Nullable
condition_start_date	String	Nullable
condition_start_datetime	String	Nullable
condition_end_date	String	Nullable
condition_end_datetime	String	Nullable
condition_type_concept_id	Float	Nullable
stop_reason	String	Nullable
provider_id	Float	Nullable
visit_occurrence_id	Float	Nullable
visit_detail_id	Float	Nullable
condition_source_value	String	Nullable
condition_source_concept_id	Float	Nullable
condition_status_source_value	String	Nullable
condition_status_concept_id	Float	Nullable

File display

## Step 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**.

Table Details: condition\_occurrence

Schema	Details	Preview
condition_occurrence_id	FLOAT	NULLABLE int64
person_id	FLOAT	NULLABLE int64
condition_concept_id	FLOAT	NULLABLE int64
condition_start_date	STRING	NULLABLE parse_date()
condition_start_datetime	STRING	NULLABLE parse_datetime()
condition_end_date	STRING	NULLABLE parse_date()
condition_end_datetime	STRING	NULLABLE parse_datetime()
condition_type_concept_id	FLOAT	NULLABLE int64
stop_reason	STRING	NULLABLE Describe this field...
provider_id	FLOAT	NULLABLE int64
visit_occurrence_id	FLOAT	NULLABLE int64
visit_detail_id	FLOAT	NULLABLE int64
condition_source_value	STRING	NULLABLE Describe this field...
condition_source_concept_id	FLOAT	NULLABLE int64
condition_status_source_value	STRING	NULLABLE Describe this field...
condition_status_concept_id	FLOAT	NULLABLE int64

Table Details: ADMISSIONS

Schema	Details	Preview
ROW_ID	INTEGER	
SUBJECT_ID	INTEGER	
HADM_ID	INTEGER	
ADMITTIME	DATETIME	
DISCHTIME	DATETIME	
DEATHTIME	DATETIME	
ADMISSION_TYPE	STRING	
ADMISSION_LOCATION	STRING	
DISCHARGE_LOCATION	STRING	
INSURANCE	STRING	
LANGUAGE	STRING	
RELIGION	STRING	
MARITAL_STATUS	STRING	
ETHNICITY	STRING	
EDREGTIME	DATETIME	
EDOUTTIME	DATETIME	
DIAGNOSIS	STRING	
HOSPITAL_EXPIRE_FLAG	INTEGER	
HAS_CHARGEVENTS_DATA	INTEGER	

Table Details: CAREGIVERS

Schema	Details	Preview
ROW_ID	INTEGER	NULLABLE
CGID	INTEGER	NULLABLE
LABEL	STRING	NULLABLE
DESCRIPTION	STRING	NULLABLE

Table Details: CPTEVENTS

Schema	Details	Preview
ROW_ID	INTEGER	NU
SUBJECT_ID	INTEGER	NU
HADM_ID	INTEGER	NU
COSTCENTER	STRING	NU
CHARTDATE	DATETIME	NU
CPT_CD	STRING	NU
CPT_NUMBER	INTEGER	NU
CPT_SUFFIX	STRING	NU
TICKET_ID_SEQ	INTEGER	NU
SECTIONHEADER	STRING	NU
SUBSECTIONHEADER	STRING	NU
DESCRIPTION	STRING	NU

Table Details: D\_CPT

Schema	Details	Preview
ROW_ID	INT64	
CATEGORY	INT64	
SECTIONRANGE	STRN	
SECTIONHEADER	STRN	
SUBSECTIONRANGE	STRN	
SUBSECTIONHEADER	STRN	
CODEPREFIX	STRN	
MIMICCODESUBSECTION	INT64	
MAXCODESUBSECTION	INT64	

Table Details: D\_ICD\_PROCEDURES

Schema	Details	Preview
ROW_ID	INTEGER	NULLABLE Describe this field...
ICD9_CODE	STRING	NULLABLE Describe this field...
SHORT_TITLE	STRING	NULLABLE Describe this field...
LONG_TITLE	STRING	NULLABLE Describe this field...

Table Details: D\_ICD\_DIAGNOSES

Schema	Details	Preview
ROW_ID	INTEGER	NULLABLE Describe this field...
ICD9_CODE	STRING	NULLABLE Describe this field...
SHORT_TITLE	STRING	NULLABLE Describe this field...
LONG_TITLE	STRING	NULLABLE Describe this field...

Table Details: ICUSTAYS

Schema	Details	Preview
ROW_ID	INTEGER	NL
SUBJECT_ID	INTEGER	NL
HADM_ID	INTEGER	NL
ICUSTAY_ID	INTEGER	NL
DBSOURCE	STRING	NL
FIRST_CAREUNIT	STRING	NL
LAST_CAREUNIT	STRING	NL
FIRST_WARDID	INTEGER	NL
LAST_WARDID	INTEGER	NL
INTIME	DATETIME	NL
OUTTIME	DATETIME	NL
LOS	FLOAT	NL

Table Details: DIAGNOSES\_ICD

Schema	Details	Preview
ROW_ID	INTEGER	NULLABLE Describe this field...
SUBJECT_ID	INTEGER	NULLABLE Describe this field...
HADM_ID	INTEGER	NULLABLE Describe this field...
SEQ_NUM	INTEGER	NULLABLE Describe this field...
ICD9_CODE	STRING	NULLABLE Describe this field...

Table Details: DRGCODES

Schema	Details	Preview
ROW_ID	INTEGER	NULLAB
SUBJECT_ID	INTEGER	NULLAB
HADM_ID	INTEGER	NULLAB
DRG_TYPE	STRING	NULLAB
DRG_CODE	STRING	NULLAB
DESCRIPTION	STRING	NULLAB
DRG_SEVERITY	INTEGER	NULLAB
DRG_MORTALITY	INTEGER	NULLAB

Table Details: D\_LABITEMS

Schema	Details	Preview
ROW_ID	INTEGER	NULLABLE
ITEMID	INTEGER	NULLABLE
LABEL	STRING	NULLABLE
FLUID	STRING	NULLABLE
CATEGORY	STRING	NULLABLE
LOINC_CODE	STRING	NULLABLE

Use these screen captures (and next slide) to select one or more MIMIC tables that contain data for OMOP CONDITION\_OCCURRENCE table

Table Details: ICUSTAYS

Schema	Details	Preview
ROW_ID	INTEGER	NL
SUBJECT_ID	INTEGER	NL
HADM_ID	INTEGER	NL
ICUSTAY_ID	INTEGER	NL
DBSOURCE	STRING	NL
FIRST_CAREUNIT	STRING	NL
LAST_CAREUNIT	STRING	NL
FIRST_WARDID	INTEGER	NL
LAST_WARDID	INTEGER	NL
INTIME	DATETIME	NL
OUTTIME	DATETIME	File display
LOS	FLOAT	NL

Table Details: LABEVENTS

Schema	Details	Preview
ROW_ID	INTEGER	NULLABLE
SUBJECT_ID	INTEGER	NULLABLE
HADM_ID	INTEGER	NULLABLE
ITEMID	INTEGER	NULLABLE
CHARTTIME	DATETIME	NULLABLE
VALUE	STRING	NULLABLE
VALUENUM	FLOAT	NULLABLE
VALUEUOM	STRING	NULLABLE
FLAG	STRING	NULLABLE

Table Details: NOTEEVENTS

Schema	Details	Preview
ROW_ID	INTEGER	NULLABLE
SUBJECT_ID	INTEGER	NULLABLE
HADM_ID	INTEGER	NULLABLE
CHARTDATE	DATETIME	NULLABLE
CHARTTIME	DATETIME	NULLABLE
STORETIME	DATETIME	NULLABLE
CATEGORY	STRING	NULLABLE
DESCRIPTION	STRING	NULLABLE
CMD	INTEGER	NULLABLE
ISERROR	STRING	NULLABLE
TEXT	STRING	NULLABLE

Table Details: PRESCRIPTIONS

Schema	Details	Preview
ROW_ID	INTEGER	NULLABLE
SUBJECT_ID	INTEGER	NULLABLE
HADM_ID	INTEGER	NULLABLE
ICUSTAY_ID	INTEGER	NULLABLE
STARTDATE	DATETIME	NULLABLE
ENDDATE	DATETIME	NULLABLE
DRUG_TYPE	STRING	NULLABLE
DRUG	STRING	NULLABLE
DRUG_NAME_POE	STRING	NULLABLE
DRUG_NAME_GENERIC	STRING	NULLABLE
FORMULARY_DRUG_CD	STRING	NULLABLE
QDN	STRING	NULLABLE
NDC	STRING	NULLABLE
PROD_STRENGTH	STRING	NULLABLE
DOSE_VAL_RX	STRING	NULLABLE
DOSE_UNIT_RX	STRING	NULLABLE
FORM_VAL_DISP	STRING	NULLABLE
FORM_UNIT_DISP	STRING	NULLABLE
ROUTE	STRING	NULLABLE

Table Details: PROCEDURES\_ICD

Schema	Details	Preview
ROW_ID	INTEGER	NULLABLE
SUBJECT_ID	INTEGER	NULLABLE
HADM_ID	INTEGER	NULLABLE
SEQ_NUM	INTEGER	NULLABLE
ICD9_CODE	STRING	NULLABLE

Table Details: TRANSFERS

Schema	Details	Preview
ROW_ID	INTEGER	NULLABLE
SUBJECT_ID	INTEGER	NULLABLE
HADM_ID	INTEGER	NULLABLE
ICUSTAY_ID	INTEGER	NULLABLE
DBSOURCE	STRING	NULLABLE
EVENTTYPE	STRING	NULLABLE
PREV_CAREUNIT	STRING	NULLABLE
CURR_CAREUNIT	STRING	NULLABLE
PREV_WARDID	INTEGER	NULLABLE
CURR_WARDID	INTEGER	NULLABLE
INTIME	DATETIME	NULLABLE
OUTTIME	DATETIME	NULLABLE
LOS	FLOAT	NULLABLE

Use these screen captures (and previous slide) to select one or more MIMIC tables that contain data for OMOP CONDITION\_OCCURRENCE table

## Step 1: Understand source/target data models

### MIMIC-III

Table Details: DIAGNOSES\_ICD

Schema	Details	Preview
ROW_ID	INTEGER	NULLABLE
SUBJECT_ID	INTEGER	NULLABLE
HADM_ID	INTEGER	NULLABLE
SEQ_NUM	INTEGER	NULLABLE
ICD9_CODE	STRING	NULLABLE

### OMOP

Table Details: condition\_occurrence

Schema	Details	Preview
condition_occurrence_id	FLOAT	NULLABLE
person_id	FLOAT	NULLABLE
condition_concept_id	FLOAT	NULLABLE
condition_start_date	STRING	NULLABLE
condition_start_datetime	STRING	NULLABLE
condition_end_date	STRING	NULLABLE
condition_end_datetime	STRING	NULLABLE
condition_type_concept_id	FLOAT	NULLABLE
stop_reason	STRING	NULLABLE
provider_id	FLOAT	NULLABLE
visit_occurrence_id	FLOAT	NULLABLE
visit_detail_id	FLOAT	NULLABLE
condition_source_value	STRING	NULLABLE
condition_source_concept_id	FLOAT	NULLABLE
condition_status_source_value	STRING	NULLABLE
condition_status_concept_id	FLOAT	NULLABLE

## 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 3: Create ETL mappings

MIMIC TableName
Field 1
Field 2
Field 3
Field 4
Field 5
Field 6
Field 7
Field 8

Table Details: condition\_occurrence

Schema	Details	Preview
condition_occurrence_id	FLOAT	NULLABLE int64
person_id	FLOAT	NULLABLE int64
condition_concept_id	FLOAT	NULLABLE int64
condition_start_date	STRING	NULLABLE parse_date()
condition_start_datetime	STRING	NULLABLE parse_datetime()
condition_end_date	STRING	NULLABLE parse_date()
condition_end_datetime	STRING	NULLABLE parse_datetime()
condition_type_concept_id	FLOAT	NULLABLE int64
stop_reason	STRING	NULLABLE Describe this field...
provider_id	FLOAT	NULLABLE int64
visit_occurrence_id	FLOAT	NULLABLE int64
visit_detail_id	FLOAT	NULLABLE int64
condition_source_value	STRING	NULLABLE Describe this field...
condition_source_concept_id	FLOAT	NULLABLE int64
condition_status_source_value	STRING	NULLABLE Describe this field...
condition_status_concept_id	FLOAT	NULLABLE int64

The mapping is as follows:

The subject\_id from the ADMISSIONS table in MIMIC is mapped to the person\_id field in the CONDITION\_OCCURRENCE table in OMOP.  
The hadm\_id from the ADMISSIONS table in MIMIC is mapped to the visit\_occurrence\_id field in the CONDITION\_OCCURRENCE table in OMOP.

## Step 4: Write transformation code

```
WITH CONDITION_OCCURRENCE1 as (select
  adm.subject_id as person_id, adm.hadm_id as
  visit_occurrence_id
  from mimick3_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 mimick3_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