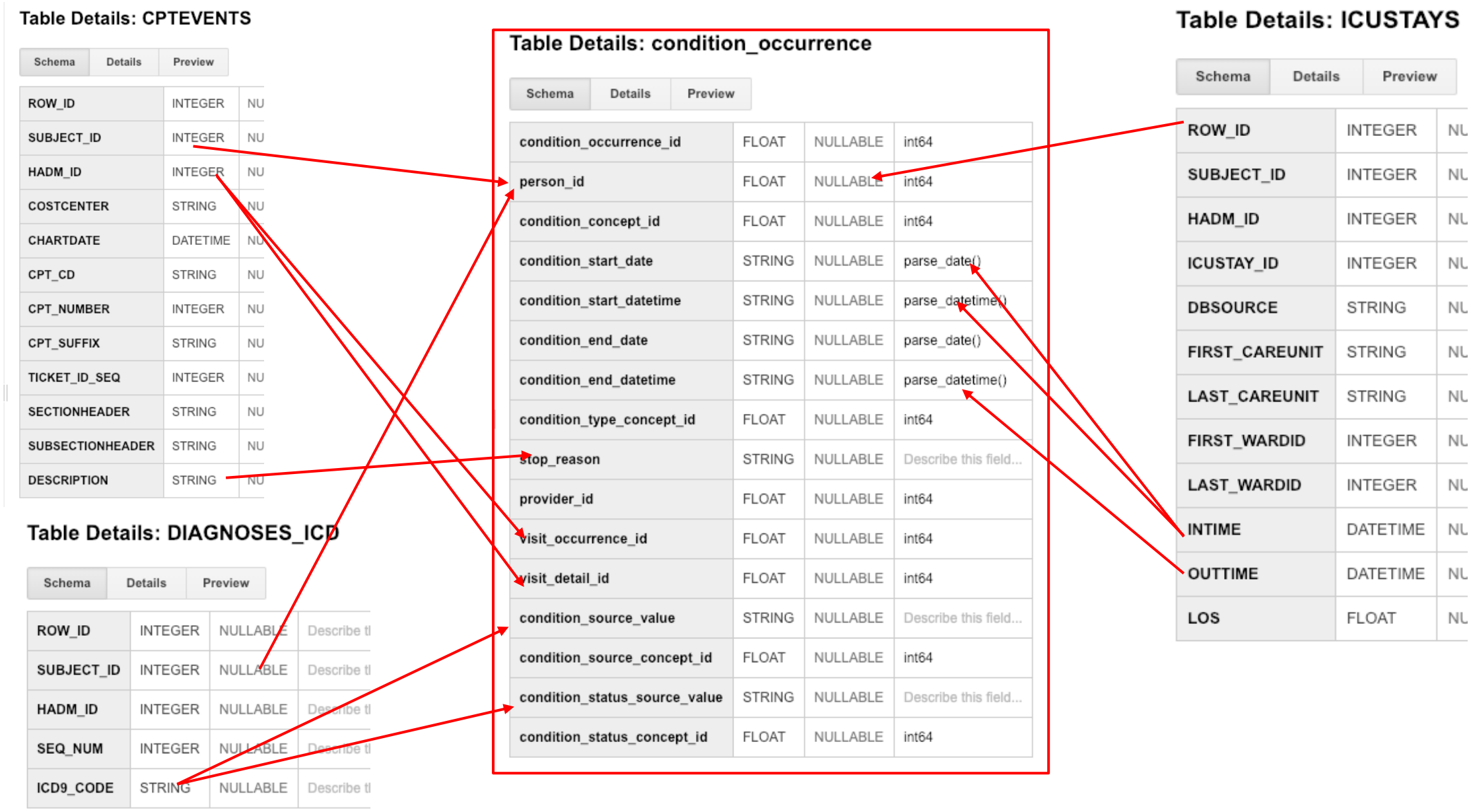
Course 2 Module 5
Programming Assignment

Assignment is to ETL MIMIC data into the OMOP CONDITION_OCCURRENCE table

Detailed instructions with Slide Notes

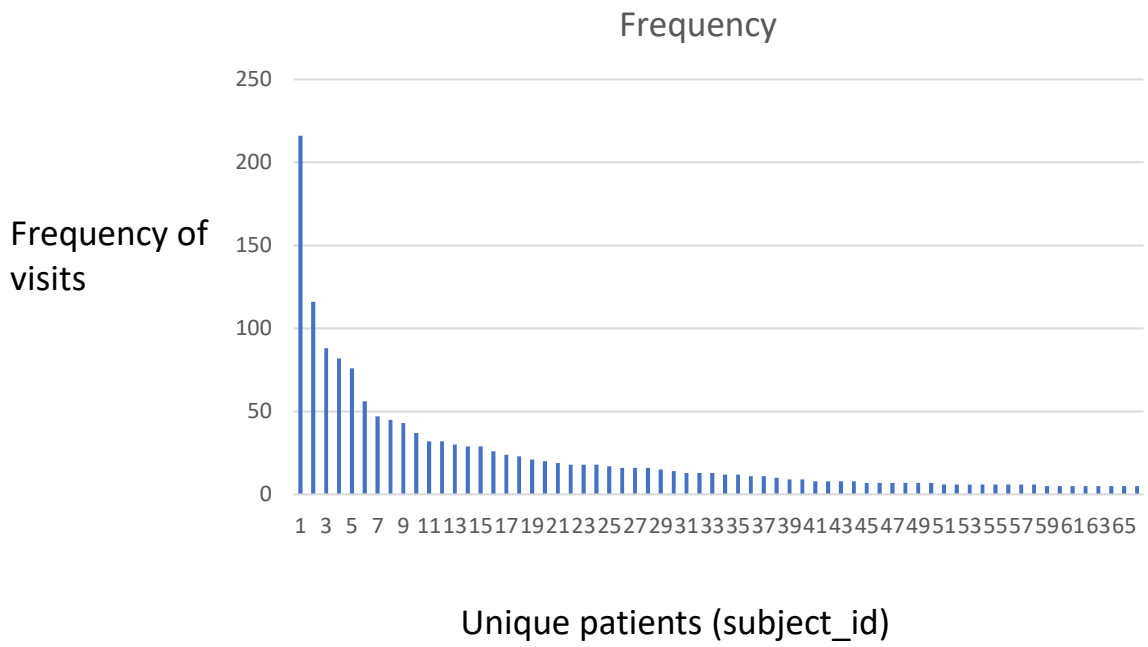
Step 1: Understand source/target data models



Step 2: Profile source table or tables

MIMIC Table: CPTEvents

Observation: A few patients account for a majority of ICU visits.



Step 3: Create ETL mappings

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

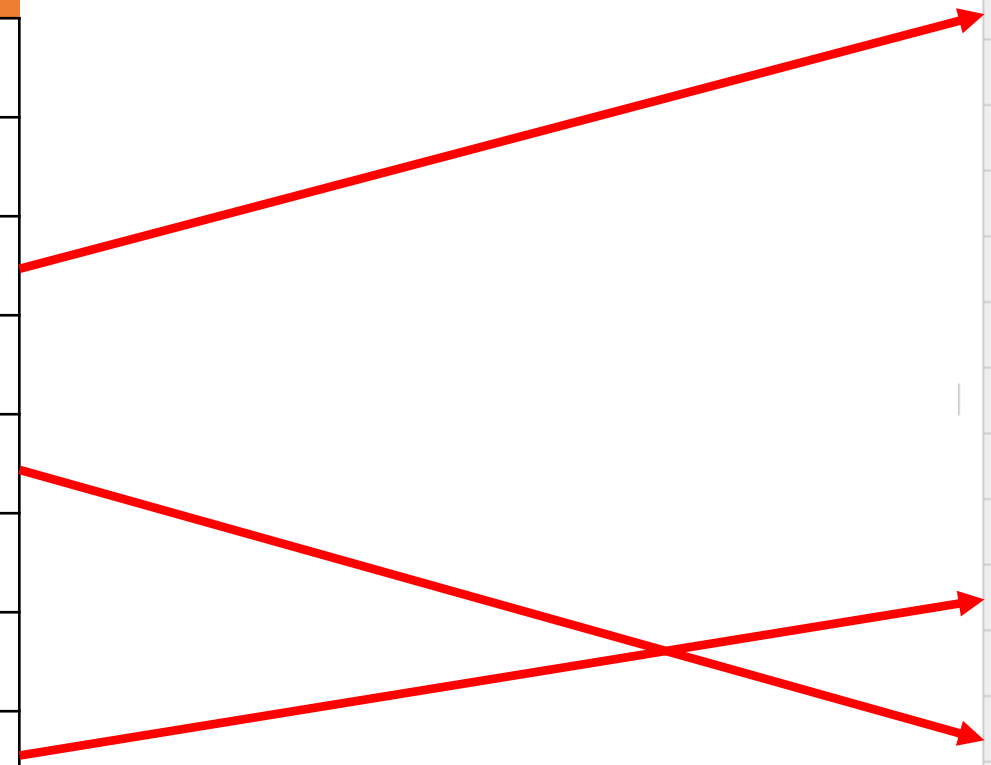


Table Details: condition_occurrence

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

Write your explanation here.

Step 3: Create ETL mappings

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: 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

Step 3: Create ETL mappings

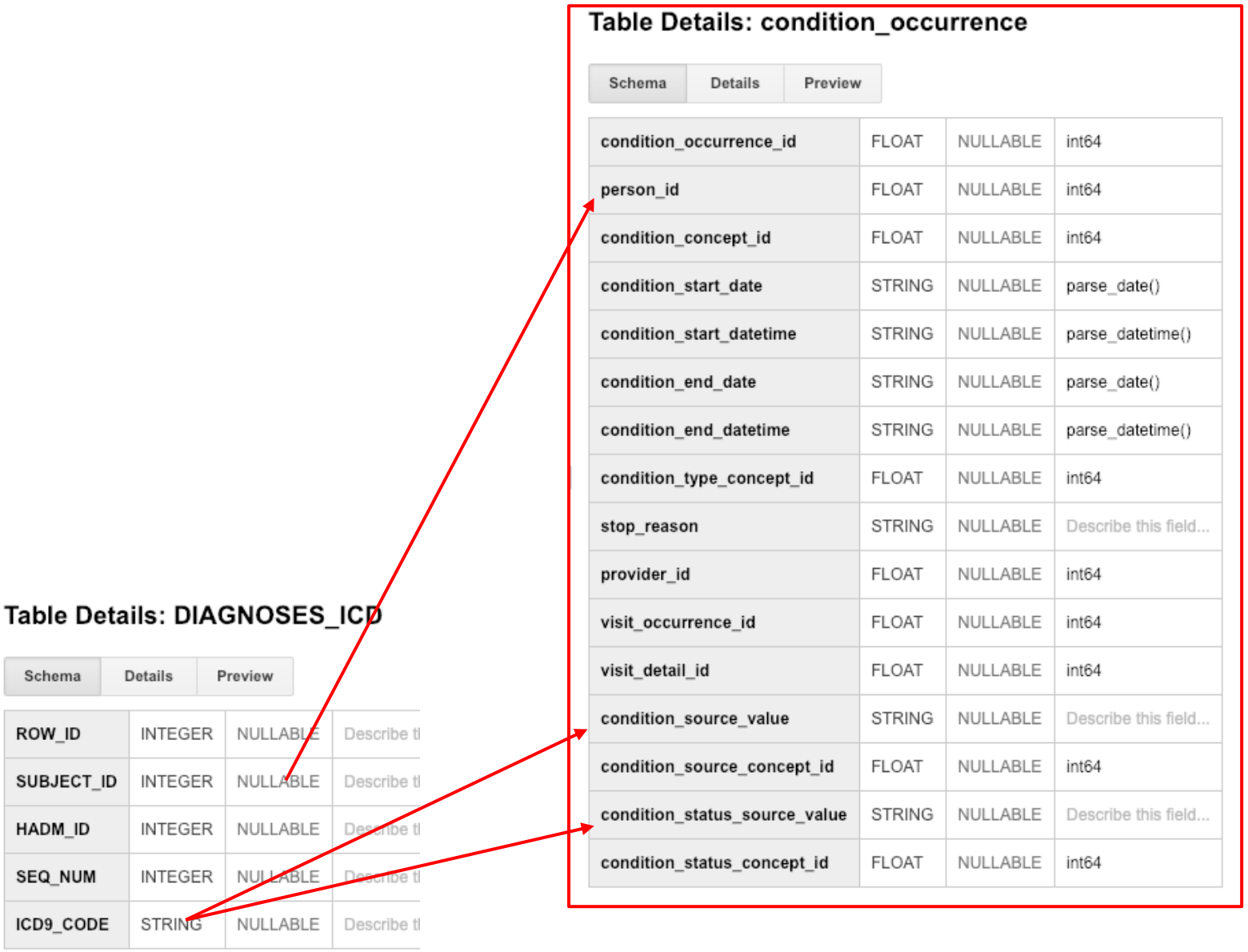


Table Details: DIAGNOSES_ICD

Schema | Details | Preview

ROW_ID	INTEGER	NULLABLE	Describe tl
SUBJECT_ID	INTEGER	NULLABLE	Describe tl
HADM_ID	INTEGER	NULLABLE	Describe tl
SEQ_NUM	INTEGER	NULLABLE	Describe tl
ICD9_CODE	STRING	NULLABLE	Describe tl

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

Step 3: Create ETL mappings

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: ICUSTAYS

Schema | Details | Preview

ROW_ID	INTEGER	NU
SUBJECT_ID	INTEGER	NU
HADM_ID	INTEGER	NU
ICUSTAY_ID	INTEGER	NU
DBSOURCE	STRING	NU
FIRST_CAREUNIT	STRING	NU
LAST_CAREUNIT	STRING	NU
FIRST_WARDID	INTEGER	NU
LAST_WARDID	INTEGER	NU
INTIME	DATETIME	NU
OUTTIME	DATETIME	NU
LOS	FLOAT	NU

Step 4: Write transformation code

```
WITH person1 as (select distinct mp.subject_id as person_id,
                        mp.subject_id as person_source_value from mimic3_demo.PATIENTS mp),
    person2 as (select distinct p1.person_id, p1.person_source_value
                ,mp.GENDER as gender_source_value
                ,CASE mp.GENDER
                 WHEN 'F' then 8532
                 WHEN 'M' then 8507
                 ELSE 0 END  as gender_concept_id
              from person1 p1 join mimic3_demo.PATIENTS mp on p1.person_id = mp.subject_id),
    person3 as (select distinct p2.person_id, p2.person_source_value
                        ,p2.gender_source_value, p2.gender_concept_id
                        ,extract(year from mp.dob) as year_of_birth
                        ,extract(month from mp.dob) as month_of_birth
                        ,extract(day from mp.dob) as day_of_birth
                        ,dob as birth_datetime
              from person2 p2 join mimic3_demo.PATIENTS mp on p2.person_id = mp.subject_id),
    person4 as (select distinct p3.person_id, p3.person_source_value
                        ,p3.gender_source_value,p3.gender_concept_id
                    ,p3.year_of_birth, p3.month_of_birth, p3.day_of_birth, p3.birth_datetime
                    ,ma.ethnicity as race_source_value
                    ,case ma.ethnicity
                                                            when 'WHITE' then 8527
                                                            when 'BLACK/AFRICAN AMERICAN' then 8516
                                                            when 'ASIAN' then 8515
                                                            when 'HISPANIC/LATINO-PUERTO RICAN' then 44814653
                                                            when 'HISPANIC OR LATINO' then 44814653
                                                            when 'UNKNOWN/NOT SPECIFIED' then 44814653
                                                            when 'OTHER' then 44814653
                                                            when 'AMERICAN INDIAN/ALASKA NATIVE FEDERALLY RECOGNIZED TRIBE' then 8657
                                                            when 'UNABLE TO OBTAIN' then 44814650
                                                            else 0 end as race_concept_id
                from person3 p3 join mimic3_demo.ADMISSIONS ma on p3.person_id = ma.subject_id),
    person as (select distinct p4.person_id, p4.person_source_value
                    ,p4.gender_source_value,p4.gender_concept_id
                    ,p4.year_of_birth, p4.month_of_birth, p4.day_of_birth, p4.birth_datetime
                    ,p4.race_source_value, p4.race_concept_id
                    ,ethnicity as ethnicity_source_value
                    ,case ma.ethnicity
                                                            when 'WHITE' then 38003564
                                                            when 'BLACK/AFRICAN AMERICAN' then 38003564
                                                            when 'ASIAN' then 38003564
                                                            when 'HISPANIC/LATINO-PUERTO RICAN' then 38003563
                                                            when 'HISPANIC OR LATINO' then 38003563
                                                            when 'UNKNOWN/NOT SPECIFIED' then 44814653
                                                            when 'OTHER' then 38003564
                                                            when 'AMERICAN INDIAN/ALASKA NATIVE FEDERALLY RECOGNIZED TRIBE' then 38003564
                                                            when 'UNABLE TO OBTAIN' then 44814650
                                                            else 0 end as ethnicity_concept_id
                from person4 p4 join mimic3_demo.ADMISSIONS ma on p4.person_id = ma.subject_id)
select * from person
```

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

Step 5: Execute transformation code

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	person_id	visit_occurrence_id	condition_source_value
1	40286	109698	78603
2	40286	109698	42731
3	40286	109698	45829
4	40286	109698	E9352
5	40286	109698	2851
6	40286	109698	4019
7	40286	109698	2768
8	40286	109698	71615
9	40286	109698	3488
10	40286	109698	2449
11	40286	109698	2930

Step 6: Perform data quality assessment

Define, implement, execute one or more data quality measures. Submit final DQ measure and an explanation why you created your measure(s).

DQ Measure:
Maximum number of Visit Occurrences by any one patient

15

```
WITH person1 as (select distinct mcpte.subject_id as person_id
                       , mcpte.hadm_id as visit_occurrence_id
   from mimic3_demo.CPTEVENTS mcpte),

    person as (select distinct p1.person_id, p1.visit_occurrence_id
                      , md.ICD9_CODE as condition_source_value
       from person1 p1 join mimic3_demo.DIAGNOSES_ICD md on p1.person_id = md.subject_id)


   SELECT person_id, count(distinct person.visit_occurrence_id) as visits_count
   from person
   group by person_id
   order by visits_count desc
```

Query results

JOB INFORMATION | RESULTS | JSON

Row	person_id	visits_count
1	41976	15
2	44083	3
3	10088	3
4	42346	2
5	41795	2
6	10059	2
7	43881	2
8	10094	2
9	40310	2
10	10124	2
11	42135	2

Step 7: Package documentation

- Congratulations! The materials in the previous slides constitute a complete ETL package.

There is no submission for this Step.