**Exercise 6.1**Put the steps of the ETL process in the proper order:

C.

A.

B.

D.

**Exercise 6.2**Using OHDSI resources of your choice, spot four issues with the PERSON record show in Table [6.2](https://ohdsi.github.io/TheBookOfOhdsi/ExtractTransformLoad.html#tab:exercisePersonTable) (table abbreviated for space):

1. PERSON\_ID is not an integer
2. Race concept\_id is 0 indicating race cannot be determined, but race\_source\_value is white (there is a concept\_id for white, it is 8527)
3. 8527 is not a valid value for ethnicity\_concept\_id (it is the race\_concept\_id for white)
4. Year of birth is not provided, meaning that all this person’s data should be dropped from the CDM instance.

**Exercise 6.3**Let us try to generate VISIT\_OCCURRENCE records. Here is some example logic written for Synthea: Sort data in ascending order by PATIENT, START, END. Then by PERSON\_ID, collapse lines of claim as long as the time between the END of one line and the START of the next is <=1 day. Each consolidated inpatient claim is then considered as one inpatient visit, set:

* MIN(START) as VISIT\_START\_DATE
* MAX(END) as VISIT\_END\_DATE
* “IP” as PLACE\_OF\_SERVICE\_SOURCE\_VALUE

If you see a set of visits as shown in Figure [6.10](https://ohdsi.github.io/TheBookOfOhdsi/ExtractTransformLoad.html#fig:exerciseSourceData) in your source data, how would you expect the resulting VISIT\_OCCURRENCE record(s) to look in the CDM?

(see image in textbook)

VISIT\_OCCURENCE\_ID: 12

PERSON\_ID: 11

VISIT\_CONCEPT\_ID: 9201

VISIT\_START\_DATE: 2004-09-26

VISIT\_START\_DATETIME: 2004-09-26

00:00:00

VISIT\_END\_DATE: 2004-09-30

VISIT\_END\_DATETIME: 2004-09-30:

00:00:00

PLACE\_OF\_SERVICE\_SOURCE\_VALUE: IP