

1. There are three solutions: add new data to rectify the situation, remove data causing the problem, or remove the foreign key. The necessary SQL for the first two is just to add or remove data in insertion, and for the third one it is to remove foreign keys.
2. The CSV file should be in the /var/lib/mysql-files/ on the ubuntu virtual machine. Using the insert statements, it takes 9 seconds. However, with the data loading it only takes 2.65 seconds.
3. The following questions have been solved in SQL, located in the sql file, and using RAs which will be inserted into this document.

- a. There are 449 players with unknown birthdates. The RA is as follows:

3.A) $\sigma_{\text{birth_day} = \emptyset \vee \text{birth_month} = \emptyset \vee \text{birth_year} = \emptyset}(\text{master})$

- b. There are 317 players inducted into the hall of fame. 250/449 are players, 23/449 are managers, 10/449 are umpires, and 34/449 are pioneers/executives. More people are dead than alive because 74 are dead and 243 are dead. The breakdown is described by the following graph:

	Alive	Dead
Player	65	185
Manager	5	18
Umpire	1	9
Pioneer/Executive	3	31

The RA is as follows:

3.B) $\sigma_{\text{inducted} = 'Y'}(\text{Hall Of Fame})$
 $\sigma_{\text{inducted} = 'Y' \wedge \text{category} = \text{"Player"}}(\text{Hall Of Fame})$
 $\sigma_{\text{inducted} = 'Y' \wedge \text{category} = \text{"manager"}}(\text{Hall Of Fame})$
 $\sigma_{\text{inducted} = 'Y' \wedge \text{category} = \text{"umpire"}}(\text{Hall Of Fame})$
 $\sigma_{\text{inducted} = 'Y' \wedge \text{category} = \text{"Pioneer/Executive"}}(\text{Hall Of Fame})$
 $\sigma_{\text{inducted} = 'Y' \wedge \text{death_year} = '' \wedge \text{death_month} = '' \wedge \text{death_day} = '' \wedge \text{death_country} = '' \wedge \text{death_state} = '' \wedge \text{death_city} = ''}(\text{Hall Of Fame})$
 $\sigma_{\text{inducted} = 'Y' \wedge (\text{death_year} \neq '' \vee \text{death_month} \neq '' \vee \text{death_day} \neq '' \vee \text{death_country} \neq '' \vee \text{death_state} \neq '' \vee \text{death_city} \neq '')}(\text{Hall Of Fame})$
 to add a category, add a "category = ..." to the end of the previous two statements

- c. The three people with the highest salaries are described by the following graph:

Alexander Enmanuel Rodriguez	398416252
Derek Sanderson Jeter	264618093
Mark Charles Teixeira	214275000

All three of the highest paid people are players.

By category, the graph looks like:

Player, 1st	Alexander Enmanuel Rodriguez	398416252
Player, 2nd	Derek Sanderson Jeter	264618093
Player, 3rd	Mark Charles Teixeira	214275000
Manager, 1st	Matthew Derrick Williams	78860000
Manager, 2nd	Robin Mark Ventura	67135000
Manager, 3rd	Paul Leo Molitor	38116666
Other, 1st	sabatic.01	25000000
Other, 2nd	rosajo01	12500000
Other, 3rd	dicker.01	12000000

For the other category, the player ID was used because the name did not exist in the database.

The RA is as follows:

3.C) $MAX_{salary}(TT_{playerID}, SUM(salary)(master)) \cup$
 $MAX_{salary}(TT_{playerID}, SUM(salary)(master)$
 $- MAX_{salary}(TT_{playerID}, SUM(salary)(master))) \cup$
 $MAX_{salary}(TT_{playerID}, SUM(salary)(master)$
 $- MAX_{salary}(TT_{playerID}, SUM(salary)(master)$
 $- MAX_{salary}(TT_{playerID}, SUM(salary)(master)))$

for doing it based on categories, use
 $category = "Player"(master)$ instead of $master$ or any of
the other categories instead of "Player"

- d. The average homeruns a player has is 15.3419 and the RA is:

$$3.D) \frac{\text{SUM}_{HR}(\text{Batting} \cup \text{Batting Post})}{\text{COUNT}(\pi_{\text{playerID}}(\text{Batting} \cup \text{Batting Post}))}$$

- e. The average is 37.4830 and the RA is:

$$3.E) \frac{\text{SUM}_{HR}(\sigma_{HR > 0}(\text{Batting} \cup \text{Batting Post}))}{\text{COUNT}(\pi_{\text{playerID}}(\sigma_{HR > 0}(\text{Batting} \cup \text{Batting Post})))}$$

- f. From 3d, we already know the average homeruns is 15.3419 and to find the average shutout games it is very similar, and it ends up being 2.1835. Using these values, we see that 39 players are both good batters and good pitchers.

The RA:

$$3.F) \text{COUNT}(\pi_{\text{playerID}}(\text{Batting} \cup \text{Batting Post}))$$

- $\pi_{\text{playerID}}(\sigma_{HR \geq \text{average of homeruns}}(\text{Batting} \cup \text{Batting Post}))$
- $\pi_{\text{playerID}}(\sigma_{SHO \geq \text{average of shutouts}}(\text{Batting} \cup \text{Batting Post}))$