RELATIONAL ALGEBRA ASSIGNMENT

NAME: RIDHIMA SHINDE   UCID: rs949

1. Select yearid, lgid, teamid, playerid and the number of home runs (HR) from the BATTING table.

    -   $\prod$ YEARID, LGID, TEAMID, PLAYERID, HR   (BATTING)

2. Modify the query in #1 so it also shows the number of hits that were for "extra bases" and the batting average (H/AB). The Extra Bases column must be calculated by adding the following columns together (2B, 3B and HR). Rename the derived column Extra Base Hits.

    -   $\prod$ YEARID, LGID, HR
        TEAMID, PLAYERID, σ (ρ (2B+3B+HR) | EXTRA BASES,   (BATTING)
        EXTRA BASES,   (H/AB) | BATTING AVG
        BATTING AVG

3. Select the yearid, lgid, Teamid, PlayerID and HR from the Batting table for all players who hit 20 or more home runs (HR) in 2015 or 2016 and played on the New York Yankees. Hint: Use teamid NYA in your where statement

    -   $\prod$ YEARID, LGID, HR σ HR >= 20 $\wedge$ (BATTING)
        TEAMID, PLAYERID   (YEARID = 2015 $\vee$ 2016) $\wedge$
                                         TEAMID = 'NYA'
                                         IN (2015,2016)

4. Write a query that lists the playerid, birthcity, birthstate, salary and batting average for all players born in New Jersey sorted by last name and year in ascending order.

    -   A $\leftarrow$ $\prod$ PLAYERID, BIRTHCITY σ BIRTHSTATE = NJ   (PEOPLE)
                 BIRTHSTATE,
                 LASTNAME

        B $\leftarrow$ $\prod$ PLAYERID, BIRTHCITY σ SALARY.PLAYERID=A.PLAYERID   (SALARY $\bowtie$ A)
                 BIRTHSTATE, YEARID,
                 SALARY, LASTNAME

        C $\leftarrow$ $\prod$ PLAYERID, BIRTHCITY σ B.YEARID=BAT.YEARID $\wedge$ (BATTING $\bowtie$ B)
                 BIRTHSTATE, YEARID,   B.LGID=BAT.LGID $\wedge$
                 SALARY, LASTNAME   B.TEAMID=BAT.TEAMID $\wedge$
                                               B.LASTNAME=BAT.LASTNAME $\wedge$
             ρ (H/AB) | BATTING AVG   B.PLAYERID=BAT.PLAYERID

∏　　σ BIRTHSTATE=NJ ∧　　　　　　　　　　　　　　　　　　(PEOPLE ⋈ SALARY ⋈ BATTING)
　　　　　PEOPLE.PLAYERID=SALARY.PLAYERID ∧
　　　　　SALARY.PLAYERID=BAT.PLAYERID

　　⊤ LASTNAME, YEARID (C)

5. Write the same query as #2 but use a LEFT JOIN.

- A ← ∏ PLAYERID, BIRTHCITY　　σ BIRTHSTATE = NJ　　　　　　　　　　　　(PEOPLE)
　　　　　　BIRTHSTATE, LASTNAME

　B ← ∏ PLAYERID, BIRTHCITY　　σ SALARY.PLAYERID=A.PLAYERID ∧　　(SALARY ⋈ A)
　　　　　　BIRTHSTATE, YEARID,　　　A.PLAYERID NOT NULL)
　　　　　　SALARY, LASTNAME

　C ← ∏ PLAYERID, BIRTHCITY　　σ (B.PLAYERID=BAT.PLAYERID ∧　　(BATTING ⋈ B)
　　　　　　BIRTHSTATE, YEARID,　　　　B.YEARID=BAT.YEARID ∧),
　　　　　　SALARY, LASTNAME　　　　(BAT.TEAMID=B.TEAMID ∧
　　　　　　ρ (H/AB) | BATTING AVG　　　BAT.LGID=B.LGID
　　　　　　　　　　　　　　　　　　　B.PLAYERID NOT NULL)

　∏　　σ BIRTHSTATE=NJ ∧　　　　　　　　　　　　　　　(PEOPLE ⟞⋈ BATTING)
　　　　　PEOPLE.PLAYERID=SALARY.PLAYERID ∧
　　　　　SALARY.PLAYERID=BAT.PLAYERID

　　⊤ LASTNAME, YEARID (C)

6. You get into a debate regarding the level of school that professional sports players attend. Your stance is that there are plenty of baseball players who attended Ivy League schools and were good batters in addition to being scholars. Write a query to support your argument using the People, CollegePlaying and Batting tables. You must use an IN clause in the WHERE clause to identify the Ivy League schools. You have also decided that a batting average less than .4 indicates a good batter.

- A ← ∏ PLAYERID,　　　　　σ (C.PLAYERID=P.PLAYERID) ∧　　(COLLEGE PLAYING⋈PEOPLE)
　　　　　NAME,　　　　　　　C.SCHOOLID IN ('HARVARD',
　　　　　SCHOOLID　　　　　'COLUMBIA','DARTMOUTH',
　　　　　　　　　　　　　　'CORNELL','YALE',
　　　　　　　　　　　　　　'PRINCETON','UNIVOFPENN',
　　　　　　　　　　　　　　'BROWN')

　∏ PLAYERID, SCHOOLID　　σ (BATTING AVG > 0.4) ∧　　　　　　(BATTING ⋈ A)
　　　NAME　　　　　　　　　(AB > 0)
　　　ρ (H/AB) | BATTING AVG

7. Using the Appearances table and the appropriate JOIN, find the players that played for the same teams in 2007 and 2010. Your query only needs to return the playerid and teamids.

   - A ← Π PLAYERID, TEAMID, YEARID σ YEARID=2007   (APPEARANCES)

   B ← Π PLAYERID, TEAMID, YEARID σ YEARID=2010   (APPEARANCES)

   A $\cap$ B

8. Using the Appearances table and the appropriate JOIN, find the players that played for the different teams in 2007 and 2010. Your query only needs to return the playerids and the 2007 teamid.

   - A ← Π PLAYERID, TEAMID, YEARID σ YEARID=2007   (APPEARANCES)

   B ← Π PLAYERID, TEAMID, YEARID σ YEARID=2010   (APPEARANCES)

   A - B

9. Using the Salaries table, calculate the average and total salary for each player. Make sure the amounts are properly formatted.

   - Π PLAYERID, AVG SAL, TOTAL SAL   PLAYERID ρ ($\mathcal{G}$ AVG(SALARY) | AVG SAL),   (SALARY)

   PLAYERID ρ ($\mathcal{G}$ SUM(SALARY) | TOTAL SAL)

10. Write a query that identifies all the playerids, the players full name and the team names who in 2010 that were playing on teams that existed in 1910. You should use the appearances table to identify the players years and the TEAMS table to identify the team name.

    - A ← Π PLAYERID, TEAMID, ρ (FIRSTNAME+LASTNAME | FULL NAME σ (APPEARANCES, YEARID=2010) $\wedge$ (P.PLAYERID= APPEARANCES.PLAYERID)   (APPEARANCES $\bowtie$ PEOPLE)

$\prod$ PLAYERID, TEAMID, FULLNAME $\quad\quad$ σ (A.TEAMID=TEAMS.TEAMID) $\wedge$ (TEAMS.YEARID=1910) $\quad\quad$ (APPEARANCES)

11. Using the Salaries table, find the players full name, average salary and the last year they played for each team they played for during their career. Also find the difference between the players salary and the average team salary.

- A $\leftarrow$ $\prod$ TEAMID, ρ (TEAMID ^{G} AVG(SALARY) | TEAM SAL) $\quad\quad$ (SALARY)

  B $\leftarrow$ $\prod$ PLAYERID, TEAMID, SALARY, ρ (P.SALARY - TEAM SAL | Diff) $\quad\quad$ σ S.TEAMID=A.TEAMID $\quad\quad$ (SALARY $\bowtie$ A)