CSC343H Summer 2003.
Assignment 2.
Deadline: Monday July 7th 2003 9:00 pm

1 Rules

For this assignment you have to submit one report and one archive.

1.1 Report

The report should be put in the course drop box, located next to SF/PT bridge on the SF side (second floor), by Monday July 7th 2003 9:00 pm.

The report should contain a front page where you clearly identify your last name, first name, student id, login name, section (L5101 or L5201). It is also strongly advised to put your report in an envelope (letter format) and repeat the front page information on it.

1.2 Archive

You should submit your archive file by the deadline using the cdf facility. For the Unix command submit, the name of this assignment is: "Assignment2" (submit -a Assignment2 ...). The name of the archive should be: “a2_<studentid>.tar.gz”.

1.3 Late submission

The late policy is very simple: 20% for each late day regardless whether or not you partially submitted the assignment. The penalty will be calculated on the whole assignment.

Late assignment should be handed in person to an instructor (not to a TA) and the archive file must be sent via email to your intructor.

1.4 What happens if I do not follow the rules?

Penalty! We do not have time to find the name for a report or student id for an archive.

2 Queries (100 points)

This assignment consists of two parts. In the first part you should write a report containing the transcription of the following 18 queries in relational algebra when it is possible (otherwise state it and explain why) and SQL. For each query give also the first three tuples answered by the query.

The second part consists of an archive file where you will insert 18 files containing the query for each question (named query01.sql to query18.sql). These files must be written is such way that one can run them on db2 in shell mode (`db2 -f queryxx.sql`)

Question	R.A	SQL	First three tuples	Electronic
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1
5	1	1	1	1
6	1	1	1	1
7	2	2	1	1
8	2	2	1	1
9	2	2	1	1
10	1	3	1	1
11	1	2	1	1
12	2	2	1	1
13	2	3	1	1
14	2	3	1	1
15	2	3	1	1
16	2	3	1	1
17	2	3	1	1
18	1	3	1	1
Total	27	37	18	18

1. Give the names of the franchises with their id.

Franchise

```
SELECT *
FROM Franchise

FRANCHID FRANCHNAME
-------- ----------------------
ANA      Anaheim Angels
ARI      Arizona Diamondbacks
ATL      Atlanta Braves
BAL      Baltimore Orioles
BOS      Boston Red Sox
CHC      Chicago Cubs
CHW      Chicago White Sox
CIN      Cincinnati Reds
CLE      Cleveland Indians
COL      Colorado Rockies
DET      Detroit Tigers
FLA      Florida Marlins
HOU      Houston Astros
KCR      Kansas City Royals
LAD      Los Angeles Dodgers
MIL      Milwaukee Brewers
MIN      Minnesota Twins
MON      Montreal Expos
NYM      New York Mets
NYY      New York Yankees
OAK      Oakland Athletics
PHI      Philadelphia Phillies
PIT      Pittsburgh Pirates
SDP      San Diego Padres
SEA      Seattle Mariners
SFG      San Francisco Giants
STL      St. Louis Cardinals
TBD      Tampa Bay Devil Rays
TEX      Texas Rangers
TOR      Toronto Blue Jays
```

2. Give the names of the franchise for the team playing in 'AL' league

$$\pi_{franchname}(\sigma_{lgid='AL}(Teams) \bowtie Franchise)$$

```
SELECT distinct f.franchname
FROM Teams t, Franchise f
WHERE t.franchId = f.franchId and t.lgID = 'AL'

FRANCHNAME
------------------------
Anaheim Angels
Baltimore Orioles
Boston Red Sox
Chicago White Sox
Cleveland Indians
Detroit Tigers
Kansas City Royals
Milwaukee Brewers
Minnesota Twins
New York Yankees
Oakland Athletics
Seattle Mariners
Tampa Bay Devil Rays
Texas Rangers
Toronto Blue Jays
```

3. Give the number of players born in Canada.

$$Impossible$$

```
SELECT count(*) AS Number
FROM Players
WHERE birthCountry = 'Canada'

Number
-----------
         29
```

4. Give the names of the players who have also been a manager.

$$\pi_{nameFirst,nameLast}(\sigma_{managerId!=''}(Players))$$

```
SELECT nameFirst, nameLast
FROM Players
WHERE managerId != ''
```

NAMEFIRST	NAMELAST
Dusty	Baker
Don	Baylor
Buddy	Bell
Bruce	Bochy
Bob	Boone
Larry	Bowa
Bob	Brenly
Terry	Francona
Ron	Gardenhire
Phil	Garner
Mike	Hargrove
Toby	Harrah
Glenn	Hoffman
Clint	Hurdle
Mike	Jorgensen
Ray	Knight
Davey	Lopes
Buck	Martinez
Lloyd	McClendon
Hal	McRae
John	Mizerock
Jerry	Narron
Larry	Parrish
Tony	Pena
Tony	Perez
Pete	Rose
Jerry	Royster
Bill	Russell
Mike	Scioscia
Joel	Skinner
John	Wathan

5. How many players bat with both hands and throw with the left hand?

$$Impossible$$

```
SELECT count(*) AS Number
FROM Players
WHERE bath = 'B' and throws = 'L'

Number
-----------
         35
```

6. Give the average attendance per teamid ordered from the largest number to the smallest.

Impossible

```
SELECT teamid, avg(attendance) AS AVG
FROM Teams
GROUP BY teamid
ORDER BY 2 DESC

TEAMID AVG
------ -----------
COL        3540790
ARI        3101524
LAN        2662986
TOR        2406698
ANA        2152196
MIL        1973440
NYN        1907631
TEX        1790606
BAL        1760134
CAL        1754941
ATL        1727236
FLO        1720640
SEA        1709919
KCA        1690780
HOU        1599424
TBA        1576580
SDN        1533108
SFN        1482489
ML4        1436526
NYA        1398250
OAK        1389180
MON        1381598
MIN        1355228
BOS        1158017
DET        1112193
SLN        1090608
CHN        1031804
CHA         984118
CLE         977931
CIN         932093
PHI         906084
PIT         805250
```

7. Give the first and last name of the best payed player ever with the salary that he earned the year he made this record amount.

$$Sal = \pi_{salary}(Salaries)$$

$$Max = Sal - \pi_1(\sigma_{2>1}(Sal \times Sal)$$

$$Result = \pi_{nameFirst,nameLast,salary,yearId}((Salaries \bowtie Max) \bowtie Players)$$

```
SELECT nameFirst, nameLast, salary, yearId
FROM Salaries s, Players p
WHERE p.playerid = s.playerId
   AND s.salary = (SELECT max(salary)
                   FROM Salaries
                  )
```

```
NAMEFIRST   NAMELAST    SALARY      YEARID
----------  ---------   -------     -------
Alex        Rodriguez   22000000    2001
Alex        Rodriguez   22000000    2002
```

8. Give the first name and last name of the players who have been division winner, league champion, world series champion batting with both hand throwing the ball with his left hand with the name of the team which they played for when the won these titles and the year.

$$PS = \sigma_{bath='B' \wedge throws='L'}(Players) \bowtie Salary$$

$$\pi_{yearId,name,nameFirst,nameLast}(\sigma_{WSWIN='Y' \wedge DivWin='Y' \wedge lgWin='Y'}(Teams) \bowtie PS)$$

```
SELECT s.yearId, t.name AS TeamName, p.nameFirst, p.nameLast
FROM Players p, Teams t, Salaries s
WHERE bath = 'B' and throws = 'L'
   AND p.playerId = s.playerID
   AND s.teamId = t.teamId
   AND s.yearId = t.yearId
   AND t.WSWIN = 'Y'
   AND DivWin = 'Y'
   AND lgWin = 'Y'
```

```
YEARID      TEAMNAME         NAMEFIRST    NAMELAST
-------     ---------        ----------   ----------
1990        Cincinnati Reds  Norm         Charlton
```

9. Give the name and the year of creation of the oldest team in the database.

$$T = \pi_{yearId}(Teams)$$

$$Min = T - \pi_2(\sigma_{2>1}(T \times T))$$

$$\Pi_{teamId,name,yearId}(Teams \bowtie Min)$$

```
SELECT t1.teamid, t1.yearId, t1.name
FROM Teams t1
WHERE NOT EXISTS ( SELECT t2.yearId
                   FROM Teams t2
                   WHERE t2.yearId < t1.yearID
                 )

TEAMID YEARID      NAME
------ ----------- ------
CHN    1876        Chicago White Stockings
```

10. Give the first and last names with the total amount earned by players who made more than 70,000,000.

$$Impossible$$

```
SELECT nameFirst, nameLast, SUM(salary) AS TotalAmount
FROM Salaries s, Players p
WHERE p.playerid = s.playerid
GROUP BY p.playerID, nameFirst, nameLast
HAVING SUM(salary) > 70000000
ORDER BY 3 DESC

NAMEFIRST    NAMELAST    TOTALAMOUNT
----------   ---------   ------------
Barry        Bonds       97879882
Greg         Maddux      95095000
Roger        Clemens     87901000
Albert       Belle       84376294
Kevin        Brown       83747644
Ken          Griffey Jr. 82659654
Gary         Sheffield   81593335
Randy        Johnson     79295000
Mark         McGwire     74688354
Sammy        Sosa        74568000
Rafael       Palmeiro    73295996
Larry        Walker      72363430
Tom          Glavine     70614437
```

11. Give the number of players per year for the team whose name is: “Toronto Blue Jays”.

$$Impossible$$

```
SELECT s.yearID, count(*) As Number
FROM Salaries s, Teams t
WHERE t.name = 'Toronto Blue Jays'
   AND s.teamid = t.teamid
   AND s.yearId = t.yearId
GROUP BY s.yearId
```

YEARID	Number
1985	20
1986	27
1987	23
1988	26
1989	29
1990	27
1991	24
1992	36
1993	33
1994	30
1995	33
1996	34
1997	33
1998	32
1999	32
2000	25
2001	28
2002	29

12. Give the id of the players who won the world series in 1992 with the franchise named: “Toronto Blue Jays” and played for this team in 2002.

$$TBJ = \pi_{franchid}(\sigma_{franchName='TorontoBlueJays'}(Franchise))$$

$$1992 = \pi_{playerId}(\sigma_{yearId=1992 \wedge WSWIN='Y'}(Teams) \bowtie TBJ \bowtie Salary)$$

$$2002 = \pi_{playerId}(\sigma_{yearId=2002}(Teams) \bowtie TBJ \bowtie Salary)$$

$$Result = 1992 \cap 2002$$

```
SELECT DISTINCT s.playerID
FROM Teams t1, teams t2, salaries s1, salaries s2, Franchise f
WHERE f.franchName = 'Toronto Blue Jays'
   AND f.franchid = t1.franchid
   AND f.franchid = t2.franchid
   AND t1.yearid = 1992
   AND t1.WSWIN = 'Y'
   AND t2.yearId = 2002
   AND t1.teamID = s1.teamId
   AND t1.yearid = s1.yearid
   AND t2.teamID = s2.teamId
   AND t2.yearid = s2.yearid
   AND s1.playerID = s2.playerId

PLAYERID
---------------------------------------------------------------

  0 record(s) selected.
```

13. Name of the player who never earned more than 100,000 per year.

$$PlayerMore100K = \pi_{playerId}(\sigma_{salary>100000}(Salaries)$$

$$PlayerAlwaysLess100K = \pi playerId(Salaries) - PlayerMore100K$$

$$Result = \pi_{nameFirst,nameLast}(Players \bowtie PlayerAlwaysLess100K)$$

```
SELECT nameFirst, nameLast
FROM Players
WHERE playerId in (SELECT playerId
                   FROM Salaries s1
                   WHERE NOT EXISTS (SELECT s2.playerId
                                     FROM Salaries s2
                                     WHERE s2.salary > 100000
                                        AND s1.playerId = s2.playerId
                                    )
                  )


SELECT DISTINCT p.nameFirst, p.nameLast
FROM Players p , Salaries s1
WHERE p.playerId = s1.playerId
   AND NOT EXISTS (SELECT s2.playerId
                   FROM Salaries s2
                   WHERE s2.salary > 100000
                      AND s1.playerId = s2.playerId
                  )


SELECT nameFirst, nameLast
FROM Players
WHERE playerId IN (SELECT playerId
                   FROM Salaries
                   GROUP BY playerId
                   HAVING max(salary) < 100001
                  )


159 answers .....
```

14. The name of the team with the year and the number of wins for the team who won the world championship and the most wins during this year.

$$WY = \pi_{yearId,W}(Teams)$$

$$MAXW = WY - \pi_{1,2}(\sigma_{3>1}(WY))$$

$$Result = \pi_{teamId,W,yearId}(Teams \bowtie MAXW)$$

```
SELECT teamId, W, yearId
FROM Teams
WHERE wswin = 'Y'
   AND (yearId,w) in (SELECT yearId, MAX(W)
                      FROM Teams
                      GROUP by yearId
                     )
```

```
TEAMID W           YEARID
------ ----------- -----------
BOS             91        1903
CHN            107        1907
CHN             99        1908
PIT            110        1909
BOS            105        1912
BOS            101        1915
BOS             91        1916
CHA            100        1917
CIN             96        1919
CLE             98        1920
NYA             98        1923
PIT             95        1925
NYA            110        1927
NYA            101        1928
SLN            101        1931
NYA            107        1932
NYA            102        1936
NYA            102        1937
NYA             99        1938
NYA            106        1939
CIN            100        1940
NYA            101        1941
SLN            106        1942
SLN            105        1944
NYA             97        1947
CLE             97        1948
NYA             97        1949
NYA             98        1950
NYA             98        1951
```

```
NYA             95        1952
NYA             99        1953
NYA             97        1956
NYA             92        1958
NYA            109        1961
BAL             97        1966
SLN            101        1967
DET            103        1968
BAL            108        1970
CIN            108        1975
CIN            102        1976
NYA            100        1978
DET            104        1984
NYN            108        1986
OAK             99        1989
NYA            114        1998
```

15. The name of the players who won at least once the world series as a player and as a manager (not necessarily the same year).

$$Player = \sigma_{wswin='Y'}(Players \bowtie Salaries \bowtie Teams)$$

$$Manager = \sigma_{wswin='Y'}(Players \bowtie Salaries \bowtie Teams)$$

$$PlayerManager = \pi_{nameFirst,nameLast}(\sigma_{player.playerId=Manager.managerId}(Player \bowtie Manager))$$

```
SELECT p1.nameFirst, p1.nameLast
FROM Players p1, Players p2, Teams t1, Teams t2, salaries s1, salaries s2
WHERE p1.playerId = s1.playerId
   AND p2.playerId = s2.playerId
   AND s1.teamid = t1.teamid
   AND s1.yearid = t1.yearid
   AND s2.yearid = t2.yearid
   AND s2.teamid = t2.teamid
   AND t1.wswin = 'Y'
   AND t2.WSWin = 'Y'
   AND p1.playerid = p2.managerId

0 answer
```

16. Name of the franchise that were always with the same teamId.

$$T1 = \pi_{teamId,franchId}(Teams)$$

$$T2 = \pi_{teamId,franchId}(Teams)$$

$$F = \pi_{franchId}(Franchise) \bowtie \pi_{franchId}(\sigma_{T1.teamIdneqT2.teamId \wedge T1.franchId=T2.franchId}(T1 \times T2))$$

$$Result = \pi_{franchName}(Franchise \bowtie F)$$

```
SELECT franchName
FROM Franchise f
WHERE f.franchId IN (SELECT t1.franchId
                     FROM Teams t1
                     WHERE NOT EXISTS (SELECT *
                                       FROM Teams t2
                                       WHERE t1.teamId != t2.teamId
                                          AND t1.franchId = t2.franchID
                                       )
                    )
```

```
FRANCHNAME
------------------------
Arizona Diamondbacks
Atlanta Braves
Baltimore Orioles
Boston Red Sox
Chicago Cubs
Chicago White Sox
Cincinnati Reds
Cleveland Indians
Colorado Rockies
Detroit Tigers
Florida Marlins
Houston Astros
Kansas City Royals
Los Angeles Dodgers
Minnesota Twins
Montreal Expos
New York Mets
New York Yankees
Oakland Athletics
Philadelphia Phillies
```

Pittsburgh Pirates
San Diego Padres
Seattle Mariners
San Francisco Giants
St. Louis Cardinals
Tampa Bay Devil Rays
Texas Rangers
Toronto Blue Jays

17. Name of the players who won at least once the world series with a team of the American league (AL) and with a team of the national league (NL).

$$ALW = \pi_{playerId}\sigma_{lgId='AL' \wedge wswin='Y'}(Teams \bowtie Salaries))$$

$$NLW = \pi_{playerId}(\sigma_{lgId='NL' \wedge wswin='Y'}(Teams \bowtie Salaries))$$

$$Result = \pi_{nameFirst,nameLast}(Player \bowtie (ALW \cap NLW))$$

```
SELECT p.nameFirst, p.nameLast
FROM Players p
WHERE p.playerId IN (SELECT DISTINCT s.playerId
                     FROM Salaries s, Teams t
                     WHERE s.lgid = 'AL'
                        AND s.teamId = t.teamId
                        AND s.yearId = t.yearId
                        AND t.WSWin = 'Y'
                    )
   AND p.playerID IN (SELECT DISTINCT s.playerId
                      FROM Salaries s, Teams t
                      WHERE s.lgid = 'NL'
                         AND s.teamId = t.teamId
                         AND s.yearId = t.yearId
                         AND t.WSWin = 'Y'
                     )
```

```
NAMEFIRST   NAMELAST
----------  ----------
Rick        Aguilera
Mark        Hutton
Mariano     Duncan
Paul        O'Neill
Devon       White
Dennis      Cook
Al          Leiter
Dwight      Gooden
Danny       Jackson
Todd        Stottlemyre
Mike        Stanton
Steve       Bedrosian
Darryl      Strawberry
Alfredo      Griffin
```

18. The name of the player who won most world series.

Impossible

```
 1) Impossible without creating a view or,
select p.playerid, count(*) from Players p, teams t, Salaries s where
p.playerId = s.playerId and s.teamId = t.teamId and s.yearId = t.yearId
and t.WSWin = 'Y' group by p.playerId order by 2 asc fetch 1

NAMEFIRST  NAMELAST  WON
---------- --------  ----
Paul       O'Neill   5
```