I

Infsci 2725 | Assignment 1: Relational Database SQL | Due week 4

———Goal:

Create relations to store data

normalize relations

employ foreign key constraints on necessary columns

load data into relations

perform queries to answer questions

——-Deliverables:

describe the schemas of your relations(CREATE statements)

explain your data import strategy (python library/csv importer/shell command/etc)

queries / answers to my questions (answer first query after)

D:\Program Files\Microsoft SQL Server\MSSQL11.SQLEXPRESS\MSSQL\DATA

Remember you may not need all the columns so you may choose to build a model that will be the minimal to answer the questions.

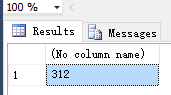
Automation is always the goal.

There is a deadline ;)

————————————questions——————————————————————

1

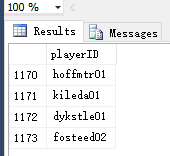
1. how many players are inducted into the hall of fame?



SELECT COUNT(playerID) FROM HallOfFame

WHERE inducted='Y'

b)how many distinct players in the batting data set are in the hall of fame?



SELECT DISTINCT Batting.playerID FROM Batting

INTERSECT

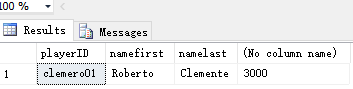
SELECT HallOfFame.playerID FROM HallOfFame

SELECT COUNT(distinct Batting.playerID) from Batting

SELECT COUNT(distinct HallOfFame.playerID) from HallOfFame

2

a)what player has exactly 3000 hits total for his career?



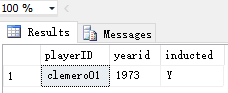
SELECT Batting.playerID,namefirst,namelast,SUM(Hits) FROM Batting,Players

WHERE Batting.playerID=Players.playerid

GROUP BY Batting.playerID,namefirst,namelast

Having SUM(Hits)=3000

b)what year was he inducted into the hall of fame?

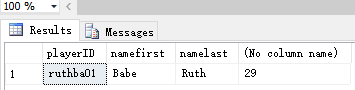


SELECT playerID,yearid,inducted FROM HallOfFame

WHERE playerID='clemero01'

3

what player had the highest number of homeruns in 1919?



SELECT TOP 1 Batting.playerID,namefirst,namelast,SUM(Homeruns) FROM Batting,Players

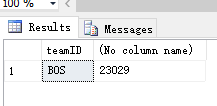
WHERE yearid=1919 AND Batting.playerID=Players.playerid

GROUP BY Batting.playerID,namefirst,namelast

ORDER BY SUM(Homeruns) DESC

4

what team has had the most hits after the year 2000, how many?



SELECT TOP 1 teamID,SUM(Hits) FROM Batting

WHERE yearID>2000

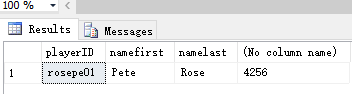
GROUP BY teamID

ORDER BY SUM(Hits) DESC

5

a)what player has the most cumulative hits?

b)how many does he have?



SELECT TOP 1 Batting.playerID,namefirst,namelast,SUM(Hits) FROM Batting,Players

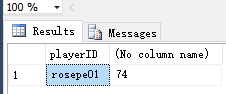
WHERE Batting.playerID=Players.playerid

Group by Batting.playerID,namefirst,namelast

Order by SUM(Hits) DESC

c)is he in the hall of fame?

d)how many votes has he received total?



SELECT playerID,SUM(votes) FROM HallOfFame

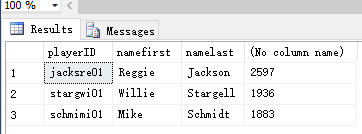
WHERE playerID='rosepe01'

Group by playerID

6

a)what 3 players have the highest number of cumulative strikeouts over their career that have been inducted into the hall of fame?

b)how many total strikeouts do they each have?



SELECT TOP 3 HallOfFame.playerID,namefirst,namelast,SUM(strikeOuts) FROM Batting,HallOfFame,Players

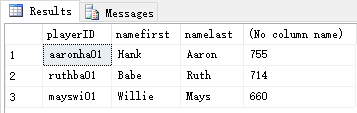
WHERE HallOfFame.playerID=Batting.playerID AND HallOfFame.playerID=Players.playerid AND inducted='Y'

Group by HallOfFame.playerID,namefirst,namelast

ORDER BY SUM(strikeOuts) DESC

7

a)what 3 players have the highest cumulative total number of homeruns and have been inducted into the hall of fame?



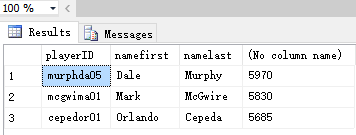
SELECT TOP 3 HallOfFame.playerID,namefirst,namelast,SUM(Homeruns) FROM Batting,HallOfFame,Players

WHERE HallOfFame.playerID=Batting.playerID AND HallOfFame.playerID=Players.playerid AND inducted='Y'

Group by HallOfFame.playerID,namefirst,namelast

ORDER BY SUM(Homeruns) DESC

b)what 3 players have the highest cumulative total number of homeruns and have NOT been inducted into the hall of fame?



SELECT TOP 3 HallOfFame.playerID,namefirst,namelast,SUM(Homeruns) FROM Batting,HallOfFame,Players

WHERE HallOfFame.playerID=Batting.playerID AND HallOfFame.playerID=Players.playerid AND inducted='N'

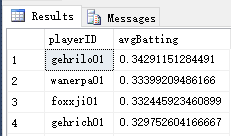
Group by HallOfFame.playerID,namefirst,namelast

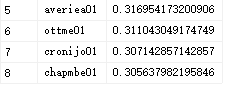
ORDER BY SUM(Homeruns) DESC

8)

what player had the highest batting average from 1930 to 1940 with a cumulative total of hits over 1800 and

has NOT been elected into the hall of fame? batting average = hits/atbats





SELECT Batting.playerID,SUM(Hits) as sh,SUM(Atbats) as sab into Answ81

FROM Batting

WHERE yearID between 1930 AND 1940

GROUP BY playerID

SELECT Answ81.playerID,cast(Answ81.sh as float)/Answ81.sab as avgBatting

FROM Answ81

WHERE Answ81.sab>0 AND Answ81.sh>1800 AND Answ81.playerID IN

(SELECT HallOfFame.playerID

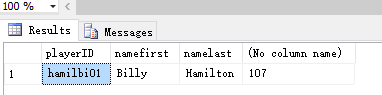
FROM HallOfFame

WHERE HallOfFame.inducted='N')

ORDER BY avgBatting DESC

9)

what player averaged the highest number of walks from 1900-1920 with over 500 atbats?



SELECT TOP 1 Batting.playerID,namefirst,namelast,AVG(Walks) FROM Batting,Players

WHERE Batting.playerID=Players.playerid AND Atbats>500 AND yearID BETWEEN '1900' AND '1920'

Group by Batting.playerID,namefirst,namelast

Order by AVG(Walks) DESC

10)

write a query that uses a non equi join and explain what you are doing.

11)

write a query that uses an OUTER Join and explain what you are doing.

Create Database

CREATE DATABASE BatPlayers

ON

(NAME=' BatPlayers\_Data',

FILENAME= D:\Program Files\Microsoft SQL Server\MSSQL11.SQLEXPRESS\MSSQL\DATA\ BatPlayers.mdf',

SIZE=5MB,

MAXSIZE=50MB,

FILEGROWTH=1MB

)

LOG ON

(NAME=' BatPlayers \_Log',

FILENAME= D:\Program Files\Microsoft SQL Server\MSSQL11.SQLEXPRESS\MSSQL\DATA\ BatPlayers \_Log.ldf',

SIZE=2MB,

FILEGROWTH=10%

)

GO

CREATE TABLES

CREATE TABLE Batting

(

playerID VARCHAR(20) NOT NULL,

yearID INT NOT NULL,

teamID VARCHAR(10) NOT NULL,

Games INT NULL,

Atbats INT NULL,

RunsScored INT NULL,

Hits INT NULL,

Homeruns INT NULL,

RunsBattedIn INT NULL,

Walks INT NULL,

strikeOuts INT NULL

)

USE BatPlayers

CREATE TABLE HallOfFame

(

playerID VARCHAR(20) NOT NULL,

yearid INT NOT NULL,

votedBy VARCHAR(25) NULL,

ballots INT NULL,

needed INT NULL,

votes INT NULL,

inducted CHAR(2) NULL,

category CHAR(20) NULL

)

USE BatPlayers

CREATE TABLE Players

(

playerid CHAR(20) NOT NULL,

namefirst CHAR(15) NULL,

namelast CHAR(20) NULL

)

INSERT CSV FILE

BULK INSERT Batting

FROM 'F:\GraduateStudy\2016Fall\DatAna\A1\_questions\_handout\Batting.csv'

WITH

(

FIRSTROW=2,

FIELDTERMINATOR=',',

ROWTERMINATOR='\n',

ERRORFILE='F:\GraduateStudy\2016Fall\DatAna\A1\_questions\_handout\BatErroe.csv'

)

BULK INSERT HallOfFame

FROM 'F:\GraduateStudy\2016Fall\DatAna\A1\_questions\_handout\HallOfFame.csv'

WITH

(

FIRSTROW=2,

FIELDTERMINATOR=',',

ROWTERMINATOR='\n',

ERRORFILE='F:\GraduateStudy\2016Fall\DatAna\A1\_questions\_handout\HallErroe.csv'

)

BULK INSERT Players

FROM 'F:\GraduateStudy\2016Fall\DatAna\A1\_questions\_handout\Players.csv'

WITH

(

FIRSTROW=2,

FIELDTERMINATOR=',',

ROWTERMINATOR='\n',

ERRORFILE='F:\GraduateStudy\2016Fall\DatAna\A1\_questions\_handout\PlayerError.csv'

)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Batting Table Structure | | | | | |
| column | Data type | length | null | Defaults | Instruction |
| playerID | VARCHAR | 20 | N |  |  |
| yearID | INT |  | N |  |  |
| teamID | VARCHAR | 10 | N |  |  |
| Games | INT |  | Y |  |  |
| Atbats | INT |  | Y |  |  |
| RunsScored | INT |  | Y |  |  |
| Hits | INT |  | Y |  |  |
| Homeruns | INT |  | Y |  |  |
| RunsBattedIn | INT |  | Y |  |  |
| Walks | INT |  | Y |  |  |
| strikeOuts | INT |  | Y |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| HallOfFame Table Structure | | | | | |
| column | Data type | length | null | defaults | instruction |
| playerID | varchar | 20 | N |  |  |
| yearid | INT |  | N |  |  |
| votedBy | VARCHAR | 25 | Y |  |  |
| ballots | INT |  | Y |  |  |
| needed | INT |  | Y |  |  |
| votes | INT |  | Y |  |  |
| Inducted | CHAR | 2 | Y |  |  |
| category | CHAR | 15 | Y |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Players Table Structure | | | | | |
| column | Data type | length | null | defaults | instruction |
| playerid | CHAR | 20 | N |  |  |
| namefirst | CHAR | 15 | Y |  |  |
| namelast | CHAR | 20 | Y |  |  |