

Steps for Preparing Environment

1. Download Python 2.7 from (https://www.continuum.io/downloads#_macosx)
2. After downloading ensure anaconda is your default python environment
3. If you are not using Anaconda ensure you have Python 2.7.10 and following python libraries (latest version) installed
 - pandas
 - glob
 - urllib
 - urllib2

Steps for Executing

1. Create a folder (you can give it any name)
2. Download and save the "baseball_solution.py" in folder created above
3. Open terminal and cd to directory created in step1
4. Execute command "python baseball_solution.py"

(* The script was executed/tested in Mac environment. This will not work in windows machine)

Total Execution time = 65 seconds

Problem 1:

input folder has all the downloaded files from 1871 to 2014 as YYYY.csv

Problem 2:

There are 11 triple pairs satisfying the condition mentioned:

```
('BOS', 'CHA', 'CLE')
('NYA', 'CHA', 'CLE')
('PIT', 'CHN', 'PHI')
('PIT', 'CHN', 'CIN')
('PIT', 'PHI', 'CIN')
('PIT', 'PHI', 'SLN')
('PIT', 'CIN', 'SLN')
('CHN', 'PHI', 'CIN')
('CHN', 'PHI', 'SLN')
('CHN', 'CIN', 'SLN')
('PHI', 'CIN', 'SLN')
```

Procedure create_triples has a run time of 1.38 seconds

Time & Space complexity of create_triples:

(See comments in "baseball_solution.py" for Time & Space complexity for each step)

Let n -- Total Size of concatenated data

Let m -- Total teams

Let p -- Total players

Time Complexity = $O(n + m^3p)$
Space Complexity = $O(n + m^3)$

Problem 3:

Here is the list for top 100 players by number of years played

(Player Name, Years Played)
('Cap Anson', 27)
('Nolan Ryan', 27)
('Deacon McGuire', 26)
('Tommy John', 26)
('Bobby Wallace', 25)
('Charlie Hough', 25)
('Eddie Collins', 25)
('Jamie Moyer', 25)
('Jim Kaat', 25)
('Rickey Henderson', 25)
('Carlton Fisk', 24)
('Dennis Eckersley', 24)
('Jesse Orosco', 24)
('Omar Vizquel', 24)
('Pete Rose', 24)
('Phil Niekro', 24)
('Rick Dempsey', 24)
('Roger Clemens', 24)
('Steve Carlton', 24)
('Ty Cobb', 24)
('Brooks Robinson', 23)
('Carl Yastrzemski', 23)
('Dennis Martinez', 23)
('Don Sutton', 23)
('Early Wynn', 23)
('Greg Maddux', 23)
('Hank Aaron', 23)
('Jack Quinn', 23)
("Jim O'Rourke", 23)
('Julio Franco', 23)
('Rabbit Maranville', 23)
('Rogers Hornsby', 23)
('Rusty Staub', 23)
('Tim Lincecum', 23)
('Tony Perez', 23)
('Al Kaline', 22)
('Babe Ruth', 22)
('Barry Bonds', 22)
('Bert Blyleven', 22)
('Bill Buckner', 22)
('Cy Young', 22)
('Dave Winfield', 22)
('Gary Sheffield', 22)
('Gaylord Perry', 22)
('Graig Nettles', 22)
('Harmon Killebrew', 22)
('Harold Baines', 22)

('Harry Davis', 22)
('Herb Pennock', 22)
('Jerry Reuss', 22)
('Jim Thome', 22)
('Jimmy Dykes', 22)
('Joe Morgan', 22)
('Joe Niekro', 22)
('Ken Griffey', 22)
('Kid Gleason', 22)
('Mel Ott', 22)
('Mike Morgan', 22)
('Phil Cavarretta', 22)
('Randy Johnson', 22)
('Red Ruffing', 22)
('Rich Gossage', 22)
('Sad Sam Jones', 22)
('Stan Musial', 22)
('Tom Glavine', 22)
('Tris Speaker', 22)
('Willie Mays', 22)
('Willie McCovey', 22)
('Andre Dawson', 21)
('Bill Dahlen', 21)
('Bob O'Farrell', 21)
('Cal Ripken', 21)
('Clark Griffith', 21)
('Danny Darwin', 21)
('Darrell Evans', 21)
('David Wells', 21)
('Eddie Murray', 21)
('Eppa Rixey', 21)
('Frank Robinson', 21)
('Frank Tanana', 21)
('Fred Clarke', 21)
('George Brett', 21)
('Honus Wagner', 21)
('Hoyt Wilhelm', 21)
('Ivan Rodriguez', 21)
('Jack O'Connor', 21)
('John Franco', 21)
('John Smoltz', 21)
('Lave Cross', 21)
('Lindy McDaniel', 21)
('Nap Lajoie', 21)
('Paul Molitor', 21)
('Reggie Jackson', 21)
('Rick Honeycutt', 21)
('Ron Fairly', 21)
('Ted Lyons', 21)
('Ted Simmons', 21)
('Tim McCarver', 21)
('Tom Gordon', 21)
('Waite Hoyt', 21)