Parallellprogrammering och multithreading i Python

En *process* inom programmering är ett program som körs. Datan som processen använder laddas in i datorns internminne, och programmet exekveras av datorns processor. Den utför alla beräkningar och lagrar och läser data i datorns internminne och hårddisk. Varje process har sitt eget allokerade minne.

En *tråd* (eng. thread) är en beräkningsenhet i en given process, och en process kan även köras på flera trådar samtidigt. Trådarna har då en del av processens minne, och delar det med dess andra trådar. Så kallad *multithreading* ämnar att öka effektiviteten av användningen av processorns kärnor genom att låta flera trådar skapas av en process och tilldelas olika uppgifter.

Multithreading är mindre användbart när större beräkningar ska utföras, och brukar istället användas mer för processer som involverar operativsystemet. Det handlar oftast om processer som läser från eller skriver till datorns hårddisk, eller som skickar eller tar emot data via datorns nätverk. Det är då vanligt att program sitter väntandes på nästa instruktion utan att göra något under tiden. Då går det att skynda på arbetet genom att låta ett program utföra andra uppgifter i väntan på något. Exempelvis som att utföra en beräkning i väntan på att en nedladdning ska bli klar, istället för att vänta med beräkningen tills efter att nedladdningen är klar.

Parallellprogrammering är bättre för program som enbart förlitar på processorn istället. Det rör sig oftast om program som utför stora beräkningar. Sådana program fördelas mellan processorns kärnor så att den kan utföra fler beräkningar samtidigt.

Parallellprogrammering

Vi kommer att implementera en enkel metod som härmar en beräkningsintensiv funktion. Vi kommer att köra programmet flera gånger och mäta exekveringstiden för programmet. Börja med att importera följande metoder i ett tomt Python-script `example.py`,

```
1 from time import perf_counter as pc
2 from time import sleep as pause
```

Vår metod kommer att använda `time`-modulens `sleep()`-method för att likna en process som tar en viss tid att slutföra.

```
1 def runner():
2         print("Performing a costly function")
3         pause(1)
4         print("Function complete")
```

Vi kör metoden och tar tid före och efter processen har körts. Vi skriver sedan ut hur lång tid körningen tog som en så kallad f-sträng, som helt enkelt är en enklare version av `"".format()` där vi kan sätta in variablerna direkt in i strängen.

```python
if __name__ == "__main__":
        start = pc()
        runner()
        end = pc()
        print(f"Process took {round(end-start, 2)} seconds")
```

```
$ python3 example.py
Performing a costly function
Function complete
Process took 1.0 seconds
```

Vi kör nu metoden tio gånger för att se om det tar ungefär tio sekunder att köra den som förväntat.

```python
if __name__ == "__main__":
        start = pc()
        for _ in range(10):
                runner()
        end = pc()
        print(f"Process took {round(end-start, 2)} seconds")
```

```
$ python3 example.py
Performing a costly function
Function complete
Performing a costly function
Function complete
Performing a costly function
Function complete
Performing a costly function
Function complete
Performing a costly function
Function complete
Performing a costly function
Function complete
Performing a costly function
Function complete
Performing a costly function
Function complete
Performing a costly function
Function complete
Performing a costly function
Function complete
Performing a costly function
Function complete
Process took 10.02 seconds
```

Då metoden anropades och exekverades tio gånger sekventiellt tog det ungefär 10.02 sekunder att slutföra programmet, vilket verkar stämma väl överens med våra förväntningar.

Genom att använda Pythons `multiprocessing`-modul kan vi köra metoden flera gånger parallellt. Dagens datorer är oftast flerkärniga, och kan därför köra flera program samtidigt. Den här modulen kommer att hjälpa oss med att skicka arbetsuppgifter till de olika kärnorna för att köra delar av programmet parallellt. Importera modulen längst upp i din kod.

```
import multiprocessing as mp
```

För att skapa en process måste vi skapa ett processobjekt med metoden som vi vill köra som parametern `target`. (Det finns även en parameter vid namn `args` som kan användas för att skicka in en lista till metoden.) Vi skapar nu flera processobjekt genom att skriva följande kod.

```
p1 = mp.Process(target=runner)
p2 = mp.Process(target=runner)
```

För att starta processerna behöver vi anropa metoden `start` som är associerad med processens objekt.

```
p1.start()
p2.start()
```

Låt oss nu beräkna hur länge programmet kommer att köras.

```
if __name__ == "__main__":
        start = pc()
        p1 = mp.Process(target=runner)
        p2 = mp.Process(target=runner)
        p1.start()
        p2.start()
        end = pc()
        print(f"Process took {round(end-start, 2)} seconds")
```

```
$ python3 example.py
Process took 0.02 seconds
Performing a costly function
Performing a costly function
Function complete
Function complete
```

Det som vi ser nu är att processerna verkar ha tagit 0.02 sekunder. Det är så klart fel då att vi kan se att det finns kvar utskrifter efter den sista utskriften i vår metod. Det som hände här var att processerna startade samtidigt, men att programmet fortsatte att köras trots att processerna inte var klara än. Problemet här är att vi inte vet exakt hur lång tid programmet tog att utföra processerna eftersom att metoderna inte längre körs sekventiellt.

Vi löser det här problemet genom att använda metoden `join` för att vänta in processerna tills dess att alla är klara.

```
p1.join()
p2.join()
```

Vi mäter nu hur lång tid programmet egentligen tog att köra.

```
if __name__ == "__main__":
        start = pc()
        p1 = mp.Process(target=runner)
        p2 = mp.Process(target=runner)
        p1.start()
        p2.start()
        p1.join()
        p2.join()
        end = pc()
        print(f"Process took {round(end-start, 2)} seconds")
```

```
$ python3 example.py
Performing a costly function
Performing a costly function
Function complete
Function complete
Process took 1.1 seconds
```

Nu ser vi hur programmet enbart tog ungefär en sekund att köra, trots att metoden kördes två gånger och stannades en sekund varje gång. Vi halverade alltså exekveringstiden av programmet.

Om vi vill köra många processer samtidigt så kan vi starta och köra dem i en loop. Om vi vill samla in dem så måste vi däremot göra det utanför loopen. Då måste vi lagra våra processer i en lista så att vi kan iterera över listan av processer och samla in dem då de är klara. Om vi samlar in dem direkt i loopen så kommer de att köras sekventiellt igen.

Här är generisk kod för att köra 10 parallella körningar `some_method` med argument `some_var`.

```
processes = []
for _ in range(10):
        p = mp.Process(target=some_method, args=[some_var])
        processes.append(p)
for p in processes:
        p.start()
for p in processes:
        p.join()
```

Futures

Om vi även vill få tillbaka resultaten av våra metoder så måste vi använda klassen `Manager` för att använda oss av några arbetsmetoder som lagrar värden åt oss. Den här klassen kan vara lite svårhanterlig, så vi kommer istället att betrakta en annan Python-modul som kan köra parallellprogram och hantera insamlingen och returerna åt oss, som på så sätt är mer lätthanterlig. Den här modulen är bara en av många sätt att visa hur de här processerna fungerar innan vi använder den mer användarvänliga modulen `concurrent.futures`.

Ta bort importen av `multiprocessing`, och importera istället `concurrent.futures`.

```
import concurrent.futures as future
```

Modulen Futures har en gruppexekverare som är lättast att köra i en så kallad context manager (ett `with`-statement i Python). Vi kallar då på metoden `submit` genom att skicka in metoden som vi vill köra tillsammans med eventuella inparametrar som vi vill ska följa med vår metod då den körs. För att sedan komma åt våra returvärden kallar vi på metoden `result` för att komma åt processobjekten. Den här modulen låter oss samla processerna genom att automatiskt vänta på processer tills dess att de är klara innan vi lämnar `with`-delen av koden.

```
with future.ProcessPoolExecutor() as ex:
        p1 = ex.submit(some_method, some_arg, some_other_arg, ...) # Starts first
        ↪  process
        p2 = ex.submit(some_method, some_arg, some_other_arg, ...) # Starts second
        ↪  process

        r1 = p1.result() # Program waits until p1 is complete before assigning r2
        r2 = p2.result()

print("all done") # Will be printed once all processes are completed
```

Vi börjar med att uppdatera vår metod `runner` genom att ge den en inparameter och en returnsträng.

```
def runner(n):
        print(f"Performing costly function {n}")
        pause(n)
        return f"Function {n} has completed"
```

Istället för att köra processerna i en loop kan vi använda exekveringsmetodens metod `map`. Med *exekveringsmetod* menar vi den metod som exekverar metoden `map`. Python har en inbyggd metod map som tar en metod och en lista av variabler som sedan sedan itereras över för att exekvera metoden varje element i listan. Användningen av den inbyggda metoden visas i följande exempel.

```
>>> def sq(x):
        return x**2

>>> for x in map(sq, [2,3,4,5]):
        print(x)
4
9
16
25
```

Låt oss nu deklarera en lista heltal som vi tar som inparameterar till vår metod `runner`. Vi skapar sedan en variabel vid namn `results` för att lagra våra returvärden som exekveringsmetodens metod `map` returnerar. Vi skickar sedan vår metod tillsammans med listan av heltal in i anropet. Slutligen så itererar vi över det returnerade objektet `map` och skriver ut resultaten som det returnerar.

```python
if __name__ == "__main__":
        start = pc()

        with future.ProcessPoolExecutor() as ex:
                p = [5, 4, 3, 2, 1]
                results = ex.map(runner, p)

                for r in results:
                        print(r)

        end = pc()
        print(f"Process took {round(end-start, 2)} seconds")
```

```
$ python3 example.py
Performing costly function 5
Performing costly function 4
Performing costly function 3
Performing costly function 2
Performing costly function 1
Function 3 has completed
Function 2 has completed
Function 1 has completed
Function 5 has completed
Function 4 has completed
Process took 5.21 seconds
```

Som vi ser så exekverades metoderna i en annan ordning än i den som de anropades. Det sker eftersom att det inte finns något sätt att veta hur upptagen en given processorkärna är då programmet skickar uppgiften till den. Så metoderna kommer att startas i olika ordning varje gång du kör programmet. Den process som kör processerna håller däremot koll på deras returvärden och försäkrar oss om att de returneras i den ordning som de exekverades i. Modulen `futures` har även en metod vid namn `as_completed` som returnerar resultaten i den ordning som de kom in. (Den kan inte användas tillsammans med `map`, men det går att använda den tillsammans med metoden submit som tidigare användes i en loop).

Multithreading

Användningen av modulen `concurrent.futures` genom att köra flera trådar är inte svårare än att ändra `ProcessPoolExecutor` i en klass till `ThreadPoolExecutor`. I övrigt är proceduren precis densamma.

```python
if __name__ == "__main__":
        start = pc()

        with future.ThreadPoolExecutor() as ex:
                p = [5, 4, 3, 2, 1]
                results = ex.map(runner, p)

                for r in results:
                        print(r)

        end = pc()
        print(f"Process took {round(end-start, 2)} seconds")
```

```
$ python3 example.py
Performing costly function 5
Performing costly function 4
Performing costly function 3
Performing costly function 2
Performing costly function 1
Function 5 has completed
Function 4 has completed
Function 3 has completed
Function 2 has completed
Function 1 has completed
Process took 5.01 seconds
```

Som ni ser så är utdatan från programmet lite annorlunda. Metoderna exekveras här i den ordning som de anropades. Det beror på att vi inte längre använder flera kärnor för att utföra uppgiften. Syftet med den här koden är så klart bara att illustrera proceduren, men det är inte beräkningsintensivt eller knutet till operativsystemet, så trådproceduren fungerar precis lika bra i det här fallet, och processerna slutfördes återigen efter cirka fem sekunder.

Länkar till den som vill lära sig mer, och se andra exempel:

https://www.machinelearningplus.com/python/parallel-processing-python/

https://www.geeksforgeeks.org/parallel-processing-in-python/