

ANSWER KEY

Welcome to the SD212 12-Week Exam for Spring 2023.

No calculators or other electronic devices may be used.

This is a multi-section exam that will be taken at different times by different Midshipmen. You may not discuss its contents with anyone until after everyone has taken the exam.

You may use a single double-sided (or two single-sided) handwritten study sheet. Write your name clearly on the top of the sheet(s) and turn it in along with your exam.

Do your best to write clearly and neatly. An unreadable answer is an incorrect answer.

Try to keep your answers inside the provided boxes and spaces. If you need more space, you may work on one of the blank pages at the end but must *indicate clearly that you have done so*.

Page	Points	Score
2	11	
3	9	
4	6	
5	6	
6	10	
7	10	
8	8	
Total:	60	

ANSWER KEY

1. Which one of these is closest to the number of processor instructions your laptop's CPU could execute in **1 hour**? [2 pts]
- a. 10 million
 - b. 10 billion
 - ! c. ! 10 trillion**
 - d. 10 googols
2. Which of these is a plausible amount of main memory (RAM) that you might find on a new laptop computer? [2 pts]
- a. 32 MB
 - b. 100 MB
 - c. 1 GB
 - ! d. ! 16 GB**
 - e. 1 TB
 - f. 1 PB
3. Where are the values of *variables* typically stored in a running computer program? [2 pts]
(Select the single best answer.)
- a. Registers
 - b. Disk
 - c. Operating System
 - ! d. ! Memory**
 - e. Cloud
 - f. CPU
4. Suppose we run a program on the command line using `time` which reports a “real (wall) time” of 53.2s and a “user (CPU) time” of 53.1s. [3 pts]
What conclusions can you make about this program? (Circle all that apply.)
- ! a. ! It is most likely CPU-bound**
 - b. It is most likely memory-bound
 - c. It is most likely IO-bound
 - d. It may benefit significantly from multi-threading
 - ! e. ! It may benefit significantly from multi-processing**
5. Select the best choice and then explain briefly what that means in your own words. [2 pts]
- ! a. ! Python is an interpreted language**
 - b. Python is a compiled language

Solution: To run a program, we feed it into a pre-compiled interpreter program like `python3` which translates source code and executes it on the fly.

6. Who or what is responsible for sharing hardware resources among multiple running processes? [2 pts]
- a. CPU
 - b. Bus
 - c. Power supply
 - ! d. ! Operating System**
 - e. VS Code
 - f. Programmer
 - g. ITSD

7. Which kinds of processor instructions would be needed to execute the following lines of Python code? [3 pts]

```
1 x = input('What is your favorite food? ')
2 if x.lower() == 'huaraches':
3     print('bienvenidos')
```

(Select all that apply.)

- a. Floating point instructions
- ! b. ! Memory (Load/Store) instructions**
- ! c. ! Control flow (branching) instructions**
- d. Library instructions
- ! e. ! Arithmetic instructions**
- f. Compiler (parsing) instructions

8. The following bash script downloads three files from the web. [4 pts]
Modify it so that the files are downloaded simultaneously
and the script exits when all downloads are finished.

```
1
2 # SOLUTION: add & after commands to run in background and wait
3
4 curl 'http://french-food.com/saucisson_sec.jpg' -o yum.jpg &
5
6 curl 'http://india.gov/saag_gosht.mp4' -o hungry.mp4 &
7
8 curl 'http://oh-canada.ca/poutine.png' -o heartattack.png &
9
10 wait
11
12
```

9. The file `midstore.txt` contains information about some recent purchases at the Midstore:

[6 pts]

```
midstore.txt
1 name;item;quantity
2 Ian Evans;toothpicks;200
3 Deborah Rosario;candy hearts;15
4 Renee Taylor;lucky charms;
5 CIV;t-shirt;5
6 Eric Lopez;hokas;
7 CIV;sword;1
8 Sarah Robinson;cup-o-noodles;12
```

Write a program **either in Python or in bash** that removes the CIV entries, fills in a quantity of 1 when none is specified, and changes semicolons to commas, saving the result to a new file `purchases.csv` which should look like:

```
purchases.csv
1 name,item,quantity
2 Ian Evans,toothpicks,200
3 Deborah Rosario,candy hearts,15
4 Renee Taylor,lucky charms,1
5 Eric Lopez,hokas,1
6 Sarah Robinson,cup-o-noodles,12
```

Solution:

```
1 grep -v 'CIV' midstore.txt | sed 's/;$;/1/' | tr ';' ',' >purchases.csv
```

```
1 import pandas as pd
2
3 df = pd.read_csv('midstore.txt', delimiter=';')
4 df = df[df['name'] != 'CIV']
5 df['quantity'].fillna(1, inplace=True)
6 df.to_csv('purchases.csv', index=False)
```

10. Assume you have successfully completed the previous question to create the file `purchases.csv` as shown on the previous page.

[6 pts]

There is another file `prices.csv` which looks like this:

```
prices.csv
1  object,price
2  lucky charms,10.00
3  hokas,150.00
4  sword,100.00
5  cup-o-noodles,1.50
6  toothpicks,0.01
```

Write a Python program that determines the total price each person paid at the Midstore, ignoring any purchases which don't appear on the prices list.

Your program should **print the names sorted in increasing order of the total amount they spent**. For the provided example files, the output would be:

```
1  Ian Evans
2  Renee Taylor
3  Sarah Robinson
4  Eric Lopez
```

Solution:

```
1  import pandas as pd
2
3  purchases = pd.read_csv('purchases.csv')
4  prices = pd.read_csv('prices.csv')
5  prices.rename(columns={'object': 'item'}, inplace=True)
6  combined = pd.merge(purchases, prices, on=['item'])
7  combined['total'] = combined['price'] * combined['quantity']
8  combined.sort_values(by=['total'], inplace=True)
9  for name in combined['name']:
10     print(name)
```

11. Suppose a restaurant has **three csv files** `breakfast.csv`, `lunch.csv`, and `dinner.csv`, each of which is structured like this:

[10 pts]

```
_____ breakfast.csv _____  
1 customer,food,price,tip  
2 Albert Zamora,eggs,7.99,1.00  
3 Laura Smith,latte,3.50,0.00  
4 Anna Wittrock,scrapple platter,12.50,4.00  
5 James Deitz,waffles,8.00,1.50
```

Write a Python program that **prints a single number** for the restaurant's total revenue for the day, including the price AND the tip of every order for breakfast, lunch, and dinner, **all added up**.

Your program should process the three files **in parallel**.

Solution:

```
1 from threading import Thread  
2 import pandas as pd  
3  
4 subtotals = []  
5 def add_revenues_from(fname):  
6     global subtotals  
7     df = pd.read_csv(fname)  
8     price_tot = df['price'].sum()  
9     tip_tot = df['tip'].sum()  
10    together = price_tot + tip_tot  
11    subtotals.append(together)  
12  
13 children = []  
14 for fname in ['breakfast', 'lunch', 'dinner']:  
15     c = Thread(target=add_revenues_from, args=[f'{fname}.csv'])  
16     c.start()  
17     children.append(c)  
18 for c in children:  
19     c.join()  
20 print(sum(subtotals))
```

12. The following Python program reads in a list of names (one per line) from a file `names.txt` and prints out any name that repeats within 20 positions in the file: [10 pts]

```
1 names = []
2 for line in open('names.txt'):
3     names.append(line.strip())
4
5 for i in range(len(names)):
6     for j in range(i+1, i+21):
7         if j < len(names) and names[i] == names[j]:
8             print(names[i])
9             break
```

Write a modified version of this program that does the same thing (perhaps printing in a different order), but operates **in parallel** to run more efficiently.

You may assume that `names.txt` has exactly 1,000,000 lines.

Solution:

```
1 from multiprocessing import Process
2
3 names = []
4 for line in open('names.txt'):
5     names.append(line.strip())
6
7 def search_range(start, end):
8     for i in range(start, end):
9         for j in range(i+1, i+21):
10            if j < len(names) and names[i] == names[j]:
11                print(names[i])
12                break
13
14 if __name__ == '__main__':
15     nprocs = 10
16     breaks = [round(i*len(names)/nprocs) for i in range(nprocs+1)]
17     children = []
18     for i in range(nprocs):
19         p = Process(target=search_range, args=[breaks[i], breaks[i+1]])
20         p.start()
21         children.append(p)
22     for p in children:
23         p.join()
```

13. This is a two-part question, based on the following scenario.

Imagine you work for a company that does DNA sequencing. People voluntarily submit their own DNA and your company looks for generic markers of certain diseases and sends back the results.

Recently, a pharmaceutical company has developed a new data-based method to develop vaccines for some common forms of cancer. They need access to a large variety of human DNA sequences in order to minimize side effects for a large portion of the population.

The pharma company has asked to collaborate with the DNA sequencing company in order to use their database of DNA sequences to develop new cancer drugs.

(A) Identify and briefly describe **two ethical issues** raised by this potential collaboration.

[4 pts]

(B) Assume the proposed collaboration with the pharmaceutical company is going to happen and make **two concrete recommendations** for specific steps that could be taken to minimize any ethical hazards you identified.

[4 pts]

Blank page

Blank page